

Diana Eva Krause *Hrsg.*

# Personalauswahl

Die wichtigsten diagnostischen Verfahren  
für das Human Resources Management

---

# Personalauswahl

---

Diana Eva Krause  
(Hrsg.)

# Personalauswahl

Die wichtigsten diagnostischen  
Verfahren für das Human  
Resources Management

*Herausgeber*  
Diana Eva Krause  
Berlin, Deutschland

ISBN 978-3-658-14566-8      ISBN 978-3-658-14567-5 (eBook)  
DOI 10.1007/978-3-658-14567-5

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer Gabler

© Springer Fachmedien Wiesbaden GmbH 2017

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Gabler ist Teil von Springer Nature

Die eingetragene Gesellschaft ist Springer Fachmedien Wiesbaden GmbH

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

---

## Geleitwort

Die Zahl der Buchveröffentlichungen zum Thema Personalauswahl ist nicht gering. In anderen Fachgebieten könnte man vielleicht sagen, der Markt sei schon gesättigt. Das gilt jedoch nicht für die Personalpsychologie und ganz besonders nicht für die Personalauswahl und ihre psychologische Grundlage, die Berufseignungsdiagnostik. In diesem Feld trifft eine Vielzahl innovativer Forschungsgebiete zusammen, um die praktische Nutzung fundierter und effektiver zu gestalten: Persönlichkeits- und Fähigkeitsforschung, Leistungsbemessung, Statistik und Psychometrie, Verfahrensentwicklung und Evaluation, um nur die wichtigsten zu nennen. Dies hat nicht nur Komplexität zur Folge, die der Übersichtsangebote bedarf, sondern auch ein Tempo der Innovation, das nach beständiger Beobachtung verlangt, wenn der Transfer gelingen und dafür sorgen soll, dass in der Personalpraxis gemäß aktuellen Standards verfahren werden kann.

Das Verhältnis von Wissenschaft und Praxis ist kein einfaches. Zwar lässt sich zeigen, dass wissenschaftlich fundierte Auswahlverfahren der „reinen“ praktischen Intuition überlegen sind; viel schwieriger ist es aber, forschungsbasierte Einsicht mit praktischer Intuition und Erfahrung so zu verbinden, dass sowohl den methodischen als auch den Verwenderperspektiven gerecht getan wird – dass also die Kombination dieser beiden Welten bestmöglich gelingt und dass zudem die Interessen der Bewerber angemessen berücksichtigt werden.

Diese anspruchsvolle und lohnende Aufgabe verdient unsere laufende Bemühung. Deshalb ist der vorliegende Band sehr zu begrüßen. Diana Eva Krause bringt für die Rolle der Herausgeberin die besten Voraussetzungen mit – vereint sie doch in ihrer Person wissenschaftliche Kompetenz mit praxisorientiertem Engagement und dem Geschick, bestens ausgewiesene Kollegen zu guten, aktuellen Beiträgen zu diesem Buch zu versammeln. Hier wird über die wichtigsten diagnostischen Verfahren der Personalauswahl berichtet, über Anforderungsanalyse als ihr Fundament, über Qualitäts- und Nutzenfragen der Eignungsdiagnostik, über die Perspektive der Bewerber, über spezielle Gesichtspunkte wie Innovationsorientierung und Personalauswahl im internationalen Vergleich; gleich zum Auftakt wird dazu eingeladen, über das Verhältnis von Praxis und Forschung nachzudenken. Damit bietet dieser Band eine abgewogene Gesamtschau dessen, was die psychologische Eignungsdiagnostik heute für eine erfolgreiche und verantwortungsbewusste Personalauswahl zur Verfügung stellt.

Nachdem mir die Arbeiten der Autoren dieses Bandes gut bekannt sind, weiß ich, dass sie alle sowohl viel von unserem Fach verstehen als auch gut schreiben können. Mit diesem Band zu arbeiten, wird deshalb für die Leserinnen und Leser nicht nur ein Informationsgewinn sein, sondern auch ein Lesevergnügen. Beides wünsche ich ihnen. Der Herausgeberin und den Autoren wünsche ich den verdienten Erfolg.

Hohenheim, Deutschland  
Frühjahr 2017

Heinz Schuler

---

# Inhaltsverzeichnis

<b>1</b>	<b>Das Verhältnis zwischen Praxis und Forschung in der Personalauswahl: Weiß die linke Hand, was die rechte tut? . . . . .</b>	<b>1</b>
	Neil Anderson	
1.1	Einleitung . . . . .	2
1.2	Fragmentierung oder Spezialisierung: Wege zu einem Modell der Psychologie der Personalauswahl . . . . .	3
1.3	Forschung-Praxis-Beziehungen in der Personalauswahl . . . . .	6
1.4	Beziehungen zwischen Wissenschaft und Praxis: Ein neuer Forschungsbereich in der Personal- und Organisationspsychologie . . . .	10
1.5	Synergie zwischen Forschung und Praxis: Vier historische Szenarien der psychologischen Personalauswahl . . . . .	11
1.6	Anregungen für die zukünftige Forschung im Bereich des Forschung-Praxis-Verhältnisses in der Personalauswahl . . . . .	23
	Literatur. . . . .	26
<b>2</b>	<b>Berufliche Eignung und ihre Diagnostik. . . . .</b>	<b>31</b>
	Carolin Palmer und Martin Kersting	
2.1	Berufliche Eignung. . . . .	32
2.2	Diagnostik der beruflichen Eignung. . . . .	35
2.3	Passung von Person und Organisation . . . . .	41
2.4	Methoden der Eignungsdiagnostik. . . . .	44
2.5	Balanced Scorecard eignungsdiagnostischer Verfahren. . . . .	48
2.6	Fazit . . . . .	53
	Literatur. . . . .	54
<b>3</b>	<b>Qualität in der Personalauswahl . . . . .</b>	<b>57</b>
	Klaus G. Melchers	
3.1	Einleitung . . . . .	58
3.2	Indikatoren für die diagnostische Güte von Auswahlverfahren . . . . .	58
3.3	Qualitätsstandards für die Personalauswahl. . . . .	60

3.4	Wie wirkt sich die Einhaltung zentraler Qualitätsaspekte auf die Qualität der Personalauswahl aus? . . . . .	62
3.5	Welchen Nutzen haben auswählende Organisationen von der Berücksichtigung der Qualitätsaspekte? . . . . .	66
3.6	Fazit . . . . .	68
	Literatur. . . . .	69
<b>4</b>	<b>Anforderungsanalysen</b> . . . . .	<b>71</b>
	Friedemann W. Nerdinger	
4.1	Anforderungen und Anforderungsanalysen . . . . .	72
4.2	Konzepte anforderungsrelevanter Personenmerkmale . . . . .	74
4.3	Methoden und Verfahren zur Bestimmung von Anforderungen . . . . .	76
4.4	Fazit . . . . .	85
	Literatur. . . . .	87
<b>5</b>	<b>Interviewverfahren und biografischer Fragebogen</b> . . . . .	<b>91</b>
	Anja Strobel und Luise Franke-Bartholdt	
5.1	Interviewverfahren . . . . .	92
5.2	Biografischer Fragebogen . . . . .	106
5.3	Fazit . . . . .	116
	Literatur. . . . .	117
<b>6</b>	<b>Testverfahren I: Kognitive Fähigkeiten</b> . . . . .	<b>127</b>
	Martin Kersting und Carolin Palmer	
6.1	Intelligenz: Bedeutung, Messung und Aussagekraft . . . . .	128
6.2	Zur Treffsicherheit intelligenztestbasierter Aussagen . . . . .	129
6.3	Intelligenzmodelle . . . . .	131
6.4	Ausgewählte Intelligenztests . . . . .	136
6.5	Zur Einsatzhäufigkeit und zur Akzeptanz von Intelligenztests . . . . .	139
6.6	Kriterien der Testauswahl. . . . .	141
6.7	Kreativität. . . . .	142
6.8	Kreativität und Intelligenz . . . . .	143
6.9	Kreativitätstests . . . . .	148
6.10	Erweiterungen des Intelligenzbegriffs . . . . .	149
6.11	Fazit . . . . .	150
	Literatur. . . . .	151
<b>7</b>	<b>Testverfahren II: Persönlichkeit und personenbezogene Attribute</b> . . . . .	<b>159</b>
	Rüdiger Hossiep und Sabine Weiß	
7.1	Zur Relevanz persönlichkeitsorientierter Verfahren im eignungsdiagnostischen Kontext. . . . .	160
7.2	Definitiorische Überlegungen zu persönlichkeitsorientierten Verfahren im berufsbezogenen Kontext. . . . .	161
7.3	Kriteriumsvalidität persönlichkeitsorientierter Testverfahren . . . . .	162
7.4	Arten persönlichkeitsorientierter Verfahren . . . . .	164



7.5	Einsatzhäufigkeit persönlichkeitsorientierter Verfahren in Deutschland. . .	168
7.6	Zur Verfälschbarkeit persönlichkeitsorientierter Testverfahren . . . . .	170
7.7	Rechtliche Rahmenbedingungen für den Einsatz persönlichkeitsorientierter Testverfahren. . . . .	171
7.8	Einbettung persönlichkeitsorientierter Verfahren in den Personalauswahlprozess und darauf aufbauende Personalentwicklungsprozesse. . . . .	175
	Literatur. . . . .	178
<b>8</b>	<b>Arbeitsprobe und Situational-Judgment-Test . . . . .</b>	<b>181</b>
	Arleta Franczukowska	
8.1	Arbeitsprobe. . . . .	182
8.2	Situational-Judgment-Test . . . . .	188
8.3	Fazit . . . . .	194
	Literatur. . . . .	195
<b>9</b>	<b>Assessment-Center . . . . .</b>	<b>197</b>
	Diana Eva Krause	
9.1	AC: Ziele, Definition, Beschreibung . . . . .	198
9.2	Anforderungsdimensionen. . . . .	199
9.3	Übungen, Simulationen und zusätzliche Diagnoseinstrumente. . . . .	203
9.4	Beobachter und Beobachtertraining . . . . .	206
9.5	Prädiktive und inkrementelle Validität . . . . .	209
9.6	Fazit . . . . .	211
	Literatur. . . . .	211
<b>10</b>	<b>„Online-Assessments“ und „Self-Assessments“ in der Eignungsdiagnostik . . . . .</b>	<b>215</b>
	Michael Ott, Anna-Sophie Ulfert und Martin Kersting	
10.1	Einleitung. . . . .	216
10.2	Unterscheidungsdimensionen . . . . .	219
10.3	Spezifika der Durchführung als Folge des Online-Charakters. . . . .	228
10.4	Bewertung der Qualität von „Online-Assessments“ und „Self-Assessments“. . . . .	231
10.5	Fazit . . . . .	234
	Literatur. . . . .	238
<b>11</b>	<b>Anwendung der Personalauswahlverfahren im internationalen Vergleich . . .</b>	<b>243</b>
	Diana Eva Krause	
11.1	Einleitung. . . . .	244
11.2	Vergleich der Anwendung der Verfahren der Personalauswahl im internationalen Kontext. . . . .	245
11.3	Fazit . . . . .	265
	Literatur. . . . .	266

---

<b>12</b>	<b>Fairness und Akzeptanz von Personalauswahlmethoden</b> .....	271
	Uwe Peter Kanning	
12.1	Einleitung .....	272
12.2	Fairness .....	273
12.3	Akzeptanz .....	285
12.4	Fazit .....	294
	Literatur .....	295
<b>13</b>	<b>Innovationsorientierte Personalauswahl</b> .....	301
	Kristina Potočnik und Neil Anderson	
13.1	Einleitung .....	302
13.2	Individuelle Unterschiede und innovative Leistung .....	304
13.3	Empfehlungen für die zukünftige Forschung .....	311
13.4	Empfehlungen für die Praxis .....	313
13.5	Fazit .....	315
	Literatur .....	316
	<b>Verzeichnis der Mitwirkenden</b> .....	321

---

# Das Verhältnis zwischen Praxis und Forschung in der Personalauswahl: Weiß die linke Hand, was die rechte tut?

1

Neil Anderson

---

## Zusammenfassung

In diesem Kapitel werden vier Hauptpunkte und -themen vertreten. Zum Ersten, dass eine natürliche Distanz zwischen den Berufsfeldern des wissenschaftlichen bzw. praktischen Bereichs der psychologischen Personalauswahl besteht. Die absolut wichtige, logische Folge davon ist, dass Vernetzungsmechanismen, die Informations- und Kommunikationskanäle und die strukturellen Mittel zur Gewährleistung einer gegenseitigen Beeinflussung beider Bereiche stark und gesund sein müssen. Zweitens wird argumentiert und illustriert, dass es vier Wissenschaftstypen in der Psychologie der Personalauswahl gibt (popularistische, pragmatische, pedantische und puerile bzw. infantile Wissenschaft), und dass unter Beibehaltung der ursprünglichen Formulierung eines Vier-Quadranten-Modells, die Zukunft beider Bereiche unserer Disziplin nur durch die pragmatische Wissenschaft gewährleistet werden kann. Drittens kann man im Verlauf der Geschichte der psychologischen Personalauswahl vier Szenarien in Bezug auf das Forschungs-Praxis-Verhältnis identifizieren: Szenario 1, in welchem fundierte Forschung die Praxis auf angemessene Weise beeinflusst, Szenario 2, in welchem unzuverlässige Forschung die Praxis glücklicherweise nicht beeinflussen kann, Szenario 3, in welchem Fortschritte in der Praxis neue Wege für die Forschung anregen, und Szenario 4, in dem fundierte Forschung die Praxis in der Personalauswahl bedauerlicherweise nicht beeinflussen kann. Es werden Beispiele angeführt, um das Bestehen dieser Szenarien zu veranschaulichen, jedoch mit dem Vorbehalt, dass in unserem Bereich das erste Szenario im Laufe der Zeit insgesamt am häufigsten vertreten ist.

---

Aus dem Englischen übersetzt von Katharina Rodgers, Julija Schellander-Obid und Diana E. Krause

---

N. Anderson (✉)  
Brunel University London, Uxbridge, UK  
E-Mail: Neil.Anderson@brunel.ac.uk

Ferner wird argumentiert, dass die Forschung zum Inhaltsbereich durch Studien zum Prozessbereich, welche Forschung-Praxis- und Praxis-Forschung-Verhältnisse beleuchten, ergänzt werden sollte. Viertens und abschließend spricht sich dieses Kapitel für zwei Hauptausrichtungen für die zukünftige Forschung zu diesen vernachlässigten Themen in der Personalpsychologie aus: Erstens die Anschauungen der Praktiker und Strategien, wie man Praktiker von Forschungserkenntnissen überzeugen kann, und zweitens Validitätsforschung zum derzeitigen Ansturm auf CPD-Veranstaltungen auf internationaler Ebene.

## Inhaltsverzeichnis

1.1	Einleitung.....	2
1.2	Fragmentierung oder Spezialisierung: Wege zu einem Modell der Psychologie der Personalauswahl.....	3
1.3	Forschung-Praxis-Beziehungen in der Personalauswahl.....	6
1.4	Beziehungen zwischen Wissenschaft und Praxis: Ein neuer Forschungsbereich in der Personal- und Organisationspsychologie.....	10
1.5	Synergie zwischen Forschung und Praxis: Vier historische Szenarien der psychologischen Personalauswahl.....	11
1.6	Anregungen für die zukünftige Forschung im Bereich des Forschung-Praxis-Verhältnisses in der Personalauswahl.....	23
	Literatur.....	26

### 1.1 Einleitung

In letzter Zeit haben zahlreiche Autoren der Arbeits-, Personal- und Organisationspsychologie Besorgnis über eine wachsende Kluft zwischen Forschung und Praxis in der Personalauswahl geäußert. Während viele die Ansicht vertreten, dass sich die Personalauswahl als forschungsbasiertes Feld entfalten konnte, gibt es auch eindeutige Anzeichen dafür, dass der praxisbezogene und der wissenschaftliche Bereich der Disziplin seit ca. 20 Jahren immer weiter auseinanderdriftet [5, 22, 32, 57]. Es wird die problematische Entwicklung einer „Praktiker-Forscher-Kluft“ konstatiert, deren allgemeine Auswirkungen der synergetischen Funktionsweise der Berufsseignungsdiagnostiker zweifellos abträglich sind [5]. Das vorliegende Kapitel verfolgt vier Ziele:

Die Schnittstelle zwischen Wissenschaft und Praxis der Personalauswahl soll als Begriffsfeld „Prozessbereich“ etabliert werden, der per se der Erforschung würdig ist. Es soll argumentiert werden, dass sich der pragmatische Weg des Fortschritts dort findet, wo eine „natürliche Distanz“ zwischen Forschung und Praxis existiert, die mit ausreichenden und angemessenen Kanälen für den Austausch zwischen den beiden Bereichen „kombiniert“ wird. Es soll ein typografisches Modell von vier aus der Personalpsychologie abgeleiteten Forschungsarten beschrieben werden. Es sollen vier historische

Beispiele für die Schnittstelle zwischen Forschung und Praxis in unserem Bereich dargestellt werden, die zur Illustration möglicher Zukunftsszenarien herangezogen werden.

Dieses Kapitel befasst sich mit diesen strittigen Themen der Personalauswahl und -rekrutierung und beruht auf historischen Beispielen (dys-)funktionaler Beziehungen zwischen Forschung und Praxis in der Personalauswahl, um mögliche Szenarien für den Austausch zwischen diesen beiden Bereichen zu illustrieren. Dabei zielt das Kapitel darauf ab, mögliche zukünftige Verbindungen zwischen Forschern und Praktikern der Disziplin zu generieren und so unterschiedliche Mechanismen aufzuzeigen, welche die Schnittstelle zwischen Praxis und Forschung optimieren können. Die zugrunde liegende Annahme lautet, dass Forschung die Personalauswahlpraxis in eine bestimmte Richtung lenken sollte und parallel Veränderungen in der Personalauswahlpraxis neue Richtungen in der Forschung stimulieren könnten. Die Psychologie der Personalauswahl sollte also von einem bidirektionalen und synergetischen Netzwerk von Verbindungen zwischen Forschung und Praxis profitieren, wobei jeder Bereich der Disziplin in ausreichend engem Kontakt zum jeweils anderen stehen sollte, um Isolation und Abspaltung zu vermeiden [42, 56]. Auch wenn diese Annahme axiomatisch erscheinen mag, plädiert dieses Kapitel – vielleicht eher strittig – für eine „natürliche Distanz“ zwischen Forschung und Praxis der Psychologie der Personalauswahl. Ferner wird argumentiert, dass eine solche Distanz nicht nur heilsam für den Zustand des Berufsstandes ist, sondern angesichts der komplexen Umstände in der gegenwärtigen Organisationswissenschaft geradezu notwendig ist. Schließlich wird – als Gegengewicht – ausdrücklich darauf hingewiesen, dass uns nicht diese Distanz oder „Kluft“ über alle Maßen beschäftigen sollte, sondern die Mechanismen und Verbindungen, die den gegenseitigen Wissenstransfer zwischen Praxis- und Forschungsbereich der Personalauswahl begünstigen. Ebenso wird argumentiert, dass die Vorteile des natürlichen Auseinanderdriftens teilweise von kompensatorischen Mechanismen des bidirektionalen Wissenstrfers abhängig sind. In einem ersten Schritt soll nun das Argument für die Spezialisierung der Funktionen als unverzichtbarer Weg in die Zukunft für die Personalpsychologie bzw. Berufseignungsdiagnostik dargelegt werden.

---

## **1.2 Fragmentierung oder Spezialisierung: Wege zu einem Modell der Psychologie der Personalauswahl**

Obwohl die psychologische Personalauswahl aufgrund ihrer wissenschaftsbasierten Praxis lange als prototypisches Beispiel für einen höchst erfolgreichen Bereich der Arbeits-, Personal- und Organisationspsychologie galt, wurden im Laufe der Jahre relativ wenige Modelle entwickelt, die diese Beziehungen illustrieren und zusammenfassen [63, 73]. Dies stellt ein bedauerliches Versäumnis in der Literatur dar – sowohl aus der Perspektive der Forscher als auch aus jener der Praktiker –, da wichtige Fragen zu den Beziehungen zwischen den beiden Interessensgruppen, zu Mechanismen zur Förderung des Austausches zwischen Forschung und Praxis, zu konkurrierenden Belohnungen und

Anreizen zwischen Forschern versus Praktikern und zum Transfer von zentralen Ergebnissen der Forschung in die organisationale Praxis der Personalauswahl auffallend gering erforscht geblieben sind. Im Laufe der Jahre hat sich eine Reihe von vergleichsweise naiven, präskriptiven, aber nicht empirisch überprüften Annahmen etabliert, die im Grunde davon ausgehen, dass die Beziehung zwischen Forschung und Praxis möglichst eng sein sollte [5]. Diese Annahmen suggerieren: „Je enger die Beziehung zwischen Forschung und Praxis, umso besser“. Dies mag zutreffend sein, dies mag nicht zutreffend sein. An dieser Stelle wird argumentiert, dass eine natürliche Distanz zwischen Forschung und Praxis notwendig ist, damit beide Bereiche unabhängig, aber gleichzeitig auch abhängig vom jeweils anderen, gedeihen können – ähnlich, wie dies bei Forschung und Praxis in der Medizin der Fall ist [51]. Zweifelsohne besteht heute im Bereich der Personalauswahl eine Spezialisierung der Arbeitsbereiche, wobei sich Berufseinsteiger in den ersten Jahren ihrer Karriere entscheiden, ob sie sich eher zu einer wissenschaftlichen Ausrichtung (Dissertation, Post-doc, Professur) oder einem praktischen Tätigkeitsfeld (Trainee, Junior Consultant oder Senior Consultant) berufen fühlen. Es ist zunehmend schwieriger geworden, den eingeschlagenen Weg zu einem späteren Zeitpunkt der Berufsausübung zu wechseln, und da empirische und statistische Forschungsmethoden immer komplexer werden, ist unter Nachwuchswissenschaftlern eine steigende Spezialisierung auf einen, höchstens zwei Teilbereiche des Fachs zu beobachten [37].

Zu fragen bleibt: „Ist diese Spezialisierung so schlecht?“ Mehrere Belege können angeführt werden, um das stichhaltige Argument zu begründen, dass ein gewisser Grad an Spezialisierung positiv und sogar absolut notwendig ist, da die psychologische Personalauswahl in ihren wissenschaftlichen Untersuchungen und Bestrebungen zunehmend komplexer wird. Betrachten wir als erstes die Parallele zur Medizin: Die Erwartung, ein Medizinstudent könnte später in seiner beruflichen Laufbahn als erfolgreicher praktischer Arzt tätig sein und sich gleichzeitig ernsthafter wissenschaftlicher Forschung widmen, wäre eindeutig unhaltbar. Die Spezialisierung auf einen der beiden Karrierewege würde hier in Anbetracht der Komplexität der medizinischen Forschung einerseits und der Anforderungen an praktizierende Ärzte, die Krankheiten ihrer Patienten richtig zu diagnostizieren und zu behandeln, andererseits als absolut notwendig erachtet. Wieso sollte das in der Psychologie der Personalauswahl anders sein? In der Tat ist die Forschung in der modernen Arbeits-, Personal- und Organisationspsychologie so komplex, dass eine Spezialisierung ähnlich notwendig erscheint. Zudem ist unser Arbeitsbereich in letzter Zeit so rapide gewachsen, dass eine Spezialisierung auf bestimmte Forschungsschwerpunkte und Interessen heutzutage geradezu essenziell ist, um mit einem bestimmten Forschungsbereich Schritt halten zu können [73]. Zweitens – und hier soll erneut ein etwas weiterer Blickwinkel eingenommen werden – weisen alle gegenwärtigen Berufe mehr oder weniger stark ausgeprägte Elemente der Trennung von Forschung und Praxis auf. Dies gilt für die Managementwissenschaften [31], die Rechtswissenschaften,

die Medizin und Gesundheitswissenschaften, die Versicherungsmathematik und das kommerzielle Versicherungswesen, den Bereich der Industrial Economics, und, hier besonders relevant, die klinische Psychologie und Psychotherapie [51], um nur einige Beispiele für mögliche vergleichbare Berufsfelder zu nennen. Noch einmal: Warum sollte sich die Psychologie der Personalauswahl davon unterscheiden? Vielmehr soll die Frage nach einem angemessenen Grad an Spezialisierung gestellt werden, die mit ausreichenden Mechanismen zur Einbindung von wissenschaftlich bzw. praktisch orientierten Teilbereichen gepaart ist, um sie vor einer irreversiblen Spaltung zu bewahren. Drittens kann argumentiert werden, dass die Wissenschaft ein gewisses Maß der Unabhängigkeit von kommerziellen Interessen in der Personalauswahl beibehalten sollte, und dass dieses Maß an Unabhängigkeit deshalb ganz zu Recht besteht [22]. Eine übermäßig pragmatische oder einzig von kommerziellen Interessen bestimmte Herangehensweise würde die Wissenschaft lähmen sowie die Bandbreite und Art der durchgeführten Studien und untersuchten Forschungsfragen in der Arbeits-, Personal- und Organisationspsychologie bedenklich einschränken. Außerdem könnten kontroverse Themen und Forschungsfragen, die aktuelle Geschäftspraktiken infrage stellen, Gefahr laufen, niemals behandelt zu werden, wenn Forscher darauf beschränkt wären, nur noch durch aktuelle kommerzielle Interessen und kurzlebige Launen der Beraterbranche geprägte Forschung zu betreiben. Viertens wäre es ähnlich restriktiv für Praktiker, in gewisser Hinsicht darauf beschränkt zu sein, nur in jenen Bereichen Beratung anbieten zu können, in denen die Validität bestimmter Methoden oder Zugänge in der Personalauswahl durch Forschung zweifelsfrei bestätigt wurde. Dies würde eine bedenkliche Einschränkung gewisser Fähigkeiten der Praktiker bedeuten, nämlich die Fähigkeit, neue Techniken und Methoden, deren Validität erst gezeigt werden muss, zu erkunden, um auf neu entstehende Marktnachfragen zu reagieren, sowie ihre Fähigkeit, Entwicklungen organisationaler Vorgehensweisen bereits vor längerfristigen strategischen Forschungsvorhaben zu erfassen [42]. Gleichermaßen existiert also eine natürliche Distanz, die von der Praxis abseits der wissenschaftlichen Forschung eingenommen werden kann, und dieser Zustand natürlicher Distanz ist für beide Bereiche der Personalpsychologie sowohl konstruktiv als auch funktional. Fünftens und abschließend können gelegentlich auch wertvolle Vorteile entstehen, wenn Wissenschaftler Untersuchungen, ganzen Themengebieten und kommerziell sensibler Forschung jenseits der kurzfristigen Forderungen nach sofort anwendbaren Forschungsergebnissen nachgehen. Frühe Forschungen zur Persönlichkeitsstruktur und die Entwicklung von Techniken explorativer Faktorenanalyse sind gute Beispiele dafür, wie im Frühstadium spekulative Forschung letztendlich wichtige praktische und kommerzielle Bedeutung erlangt hat (in diesem Fall Jahrzehnte später). Sogar heute würde ein Forschungsantrag auf ein Mehrjahresprojekt, für das Personalwissenschaftler tausende von Eigenschaftsdeskriptoren aus Wörterbüchern und anderen Quellen sortieren und mit umstrittenen und unausgereiften statistischen Methoden in Gruppen einordnen wollen,

keineswegs hoch eingeschätzt. In anderen Fällen ermöglicht eine gewisse natürliche Distanz den Forschern, Studien nachzugehen, die aktuellen Trends oder vorübergehenden kommerziellen Interessen zuwiderlaufen (z. B. kritische Forschung zur emotionalen Intelligenz), jedoch längerfristig Ergebnisse hervorbringen können, die den Zeitgeist umstürzen (z. B. weisen metaanalytische Ergebnisse nunmehr darauf hin, dass sogar nicht strukturierte Interviews eine gewisse Validität haben). Wiederum handelt es sich eher um die entscheidende Frage des Ausmaßes dieser Distanz und der Existenz von Mechanismen, um Wissenstransfer zwischen Forschung und Praxis längerfristig zu integrieren und zu ermöglichen, als um die naive Annahme, dass Forschung und Praxis in genau demselben professionellen Umfeld koexistieren und auf tagesaktueller Basis vollkommen voneinander abhängig sein müssen.

---

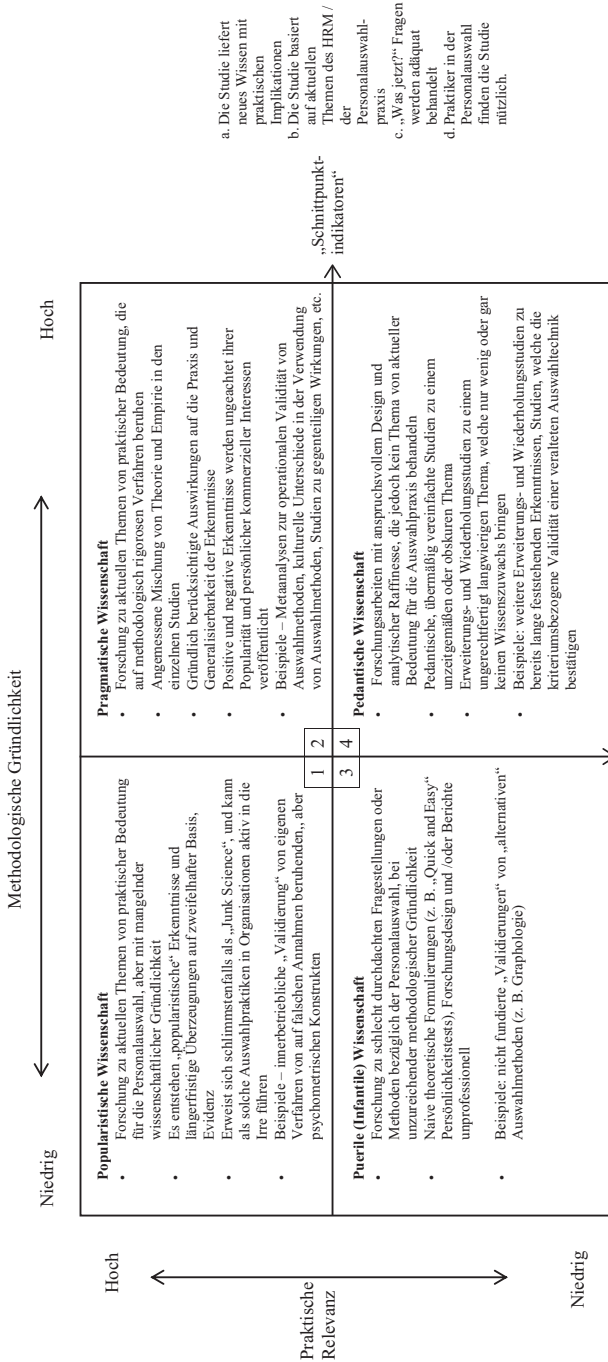
### 1.3 Forschung-Praxis-Beziehungen in der Personalauswahl

Nach der Argumentation zugunsten einer natürlichen Distanz zwischen dem wissenschaftlichen und dem praktischen Bereich der Personalauswahl als forschungsbasierte Disziplin ist es angebracht, sich wieder der Modellherleitung bezüglich der Beziehungen zwischen diesen beiden Bereichen zuzuwenden. Trotz Dunnettes [22] bahnbrechender Abhandlung, die eine breitere Diskussion des Themas innerhalb der Arbeits-, Personal- und Organisationspsychologie fordert, wurde diesen wichtigen Beziehungen bis jetzt auffallend wenig Beachtung zuteil. Dies führte dazu, dass in der Literatur nur wenige Modelle zu diesen Beziehungen generiert, geschweige denn durch empirische Studien und Feldforschung überprüft worden sind. Eine Ausnahme von diesem generellen Trend stellt ein von Anderson et al. [5] entwickeltes Modell dar.

In diesem Modell wird ein einfaches  $2 \times 2$  faktorielles Design mit den Dimensionen der wissenschaftlichen Gründlichkeit (rigor) und Relevanz für die berufliche Praxis erstellt. Dies ergab vier Quadranten der Forschungstypen: den popularistischen, pragmatischen, pedantischen und puerilen (bzw. infantilen) Wissenschaftszweig. In Erweiterung und Anwendung dieses Modells auf die Personalauswahl präsentieren Anderson, Lievens, van Dam und Ryan [3] eine Reihe von Forschungsbeispielen, wobei jeder der genannten Quadranten belegt wurde, zusammen mit Schnittpunktindikatoren zur Darstellung der Übergänge zwischen hohen und niedrigen Bedingungen der jeweiligen Dimensionen von methodologischer Gründlichkeit und praktischer Relevanz. Abb. 1.1 illustriert dieses Modell.

In Fällen von niedriger methodologischer Gründlichkeit und hoher praktischer Relevanz liegt popularistische Wissenschaft vor. Obwohl das Forschungsthema hierbei von aktueller Natur sein kann, untersuchen popularistische Studien die Forschungsfrage(n) jedoch nicht mit ausreichender methodologischer Gründlichkeit. Solche Studien wurden





**Abb. 1.1** Arten von Forschungsarbeiten in der Psychologie der Personalauswahl. (Basierend auf Anderson & Ones [3], weiter entwickelt und erweitert nach Anderson et al. [4])

- a. Die Studie liefert neues Wissen basierend auf angemessenem Design
- b. Die Studie basiert auf relevanter Theorie und früheren Erkenntnissen
- c. „Was kommt als nächstes?“-Fragen (für die Forschung) werden adäquat behandelt
- d. Forscher auf dem Gebiet der Personalauswahl finden die Studie nützlich.

vielleicht übereilt publiziert, um ein „heißes Thema“ innerhalb der Personalpsychologie aufzugreifen oder sie können unzulässigerweise von persönlichen Interessen, der Relevanz eines aktuellen Trends oder von der Intention eines bevorzugten psychometrischen Verfahrens zu „beweisen“, beeinflusst worden sein. Zweifelsohne zeigt dieser Quadrant der Forschung in der Personalpsychologie auf, wie wichtig es ist, dass unabhängige Experten die zur Publikation eingereichten Manuskripte begutachten. Jedoch sind natürlich nicht alle Publikationen unseres Fachgebiets aus referierten Fachzeitschriften, sodass populistischen Erkenntnissen der Weg in die Öffentlichkeit durchaus gelingt.

Man kann argumentieren, dass schlecht konzipierte bzw. durchgeführte Studien aus dieser Kategorie möglicherweise die derzeit größte Gefahr für die wissenschaftsbasierte Praxis der Personalauswahl darstellen. Wenn organisationale Praktiken in der Mitarbeiterauswahl auf derart unzuverlässigen Ergebnissen beruhen, ist die Wahrhaftigkeit unseres Anspruches, dass es sich um eine wissenschaftsbasierte Praxis handelt, infrage gestellt. Leser dieses Kapitels waren gewiss schon mit derartigen Ansichten von zahlreichen Personal- und Linienmanagern konfrontiert. Oft basiert dies auf einem blinden Glauben an eine besonders dubiose Methode der Personalauswahl, die keine ordentliche Validierung erfahren hat oder auf methodisch umstrittenen Studien. Kritisch betrachtet ist es also die Aufgabe unseres Faches einzufordern, dass jeder einzelnen Methode der Personalauswahl und in der Eignungsdiagnostik profunde Forschungsarbeit vorangeht und dass unseriöse Techniken und vorübergehend Erfolg versprechende Methoden gewissenhafter Überprüfung unterzogen werden, auch wenn diese Methoden lukrative Beratungshonorare versprechen.

Im zweiten Quadranten der Grafik finden sich Studien, die sowohl eine hohe methodologische Gründlichkeit als auch eine hohe praktische Relevanz aufweisen, wobei der resultierende Forschungstypus als pragmatische Wissenschaft bezeichnet werden kann. Dieser Quadrant sollte in Auswahl-situationen, ebenso wie in anderen Bereichen der Arbeits-, Personal- und Organisationspsychologie unsere Arbeit dominieren und, wenn möglich, die Grundlage unseres professionellen Handelns darstellen. Einige bemerkenswerte Beispiele für diese pragmatische Wissenschaft sind in Abb. 1.1 dargestellt. Sie veranschaulichen vor allem eine erheblich weitläufigere wissenschaftliche Basis dieses Bereichs. So wird z. B., wie weiter unten im Text noch ausgeführt werden wird, die weit verbreitete Verwendung von kognitiven Fähigkeitstests (General Mental Ability, GMA) in der Personalauswahl nunmehr von namhaften Studien befürwortet. Diese Evidenz ist generalisierbar auf verschiedene Berufsgruppen, Organisationen und Länder [60, 61, 66]. Das ist nur ein Beispiel dafür, wie die Personalpsychologie legitimerweise den Status einer wissenschaftsbasierten Disziplin für sich beanspruchen kann; weitere Beispiele folgen an anderer Stelle. Der zentrale Punkt ist, dass von den vier in unserem Modell dargestellten Quadranten nur dieser den langfristigen Interessen der Personalpsychologie nützlich ist – und zwar für ihre Wissenschaft, ihre Praxis und die Interdependenz zwischen den beiden Bereichen.

Als logische Schlussfolgerung daraus müssen sich Eignungsdiagnostiker folgende Fragen stellen:

- a) Wie kann dieser Quadrant hinsichtlich des Verhältnisses zwischen Forschungsbemühungen und Ergebnissen maximiert werden?
- b) Wie können die Verbindungen und Beziehungen zwischen Forschung und Praxis optimiert werden, um sicherzustellen, dass Schlüsselerkenntnisse auch in der organisationalen Praxis umgesetzt werden?
- c) Wie kann das Feedback aus der Praxis an die Forschung sichergestellt werden, um zu gewährleisten, dass Forscher an Untersuchungsthemen arbeiten, die für Kunden und Organisationen im internationalen Kontext relevant und aktuell sind?

Wo methodologische Gründlichkeit hoch, die praktische Relevanz jedoch niedrig ist, sprechen wir von pedantischer Wissenschaft. In diesem Fall basieren die Studien auf einem anspruchsvollen Design oder wurden unter Beachtung von Einzelaspekten durchgeführt, behandeln jedoch kein Thema, das für die organisationale Auswahlpraxis von aktueller Bedeutung ist. Beispiele für diese Forschungsart umfassen längerfristige Untersuchungen zu Forschungsfragen von marginaler oder peripherer Bedeutung für die heutige organisationale Praxis oder solche, in welchen Erweiterungs- und Wiederholungsstudien nichts Neues zu bereits bestehenden Erkenntnissen in der Personalauswahl beitragen. Dies ist der Quadrant des „sicheren Hafens“, in welchem eine Minderheit von Wissenschaftlern ihren persönlichen Lieblingsthemen nachgeht, unter Verwendung konventioneller Methodologien in einem wissenschaftlichen Teilbereich, der aufgrund organisationaler Veränderungen bereits als obsolet anzusehen ist [29]. Hier ist wieder der Prüfprozess entscheidend, um die oben erwähnte pedantische Wissenschaft herauszufiltern. Dies trifft ganz besonders zu, wenn man bedenkt, dass auf Kosten anderer Möglichkeiten der Personalauswahl solche Elfenbeinturmforschung uneingeschränkt weitergeführt wird, trotz aller Veränderungen im Bereich der Personalauswahlpraxis und aktuellen Trends in der Personalauswahl [39]. Nach einer gewissen Zeit treten die Gefahren eines zu großen Anteils an pedantischer Wissenschaft ganz deutlich zutage: die Forschung verliert ihre Relevanz für die Praxis und kann eigennützige und höchst dysfunktionale Züge annehmen, indem sie sich immer mehr von den Anforderungen der praktischen Relevanz entfernt und keine Gültigkeit mehr als Grundlage für organisationale Auswahlpraktiken hat.

Schließlich stellt der vierte Quadrant das „Worst-Case-Szenario“ der Forschung dar, da sie sowohl inakzeptable Methoden und Forschungsdesigns anwendet und zudem von zweifelhafter praktischer Relevanz ist. Wir nennen sie die puerile bzw. infantile Wissenschaft. Es muss ganz klar das Ziel sein, das Ausmaß dieses Quadranten zu minimieren oder sogar zu eliminieren, da er weder von wissenschaftlicher noch von praktischer Bedeutsamkeit ist. Dennoch existieren Beispiele dieser Art von Forschung. Zum Beispiel, wenn schon im Ansatz nicht fundierte psychologische Beurteilungskriterien mittels nicht fundierter Methoden oder Forschungsdesigns „validiert“ wurden (eine untaugliche Untersuchung der kriteriumsbezogenen Validität wie etwa der Graphologie).

In unserer ursprünglichen Darstellung des Modells stellen wir die vier Quadranten in Größe und Umfang gleich vor. Tatsächlich dominiert die pragmatische Forschung die Bereiche der Personalauswahl und der Eignungsdiagnostik, wobei vielleicht weniger

vorteilhafte Aspekte der populistischen und pedantischen Wissenschaft weniger offensichtlich, aber dennoch vorhanden sind und einen gewissen Einfluss auf unseren Arbeitsbereich ausüben. Wir (Anderson et al.) [5] argumentierten durch Rekurs auf verschiedene Evidenzquellen, dass die zwei letzteren Wissenschaftsarten aufgrund dysfunktionalen Erfolgsdrucks auf Wissenschaftler und Praktiker der Personalauswahl zugenommen haben. Solchen Trends sollte auf jeden Fall entgegengewirkt werden, da die erstere (populistische Wissenschaft) wahrscheinlich zu theoretisch nicht fundierten und nicht validierten Praktiken führt, während die letztere (pedantische Wissenschaft), wie bereits dargestellt, Gefahr läuft, zu Isolierung und Diskreditierung der Forschung bei Auswahl- und Assessmentmethoden in Organisationen zu führen.

---

#### **1.4 Beziehungen zwischen Wissenschaft und Praxis: Ein neuer Forschungsbereich in der Personal- und Organisationspsychologie**

Als eines der Ziele des vorliegenden Kapitels sollten die Beziehungen zwischen Wissenschaft und Praxis als eigenständiger Forschungsbereich in der Personalauswahl etabliert werden. Der aktuelle internationale Stand der Dinge ist, dass Wissenschaftler weiterhin Themen erforschen, die am ehesten als „Inhaltsbereiche“ in der Personalauswahl bezeichnet werden können (z. B. kriteriumsbezogene Validität, „adverse impact“, Reaktionen der Kandidaten usw.), während Praktiker weiterhin mit kommerziell beliebten Ansätzen, Methoden und Entscheidungshilfen arbeiten. Der Umstand, dass es den Ergebnissen unserer Forschungsbestrebungen nicht gelingt, Einfluss auf die Praxis auszuüben, verdient ein Augenmerk der Forscher sein [22, 56, 72]. Vielleicht sollten wir dieses Thema als Herausforderung für die Forschung im Bereich der Organisationsentwicklung auf Mikroebene betrachten – ein neuer „Prozessbereich“ für Forschungsstudien. Anders ausgedrückt: Die Ergebnisse von Studien zur Personalauswahl können die Grundlage für Interventionsprogramme in Organisationen darstellen. Kunden für diese Interventionen sind HRM-Experten und Praktiker, die nicht unbedingt qualifizierte Psychologen sind, und es ist das allgemeine Ziel solcher Forschung, die Effizienz derartiger Interventionen zu steigern. Wie später noch dargestellt wird, ist eine strikt wissenschaftliche, rational-wirtschaftliche Denkweise nicht unbedingt der beste Ansatz für den Transfer unserer Ergebnisse in die allgemeine Praxis der Personalauswahl. Bevor sich jedoch auch nur irgendetwas diesbezüglich tun kann, müssen Forscher im Bereich der Personalauswahl selbst noch davon überzeugt werden, dass dieser Themenbereich es wert ist, als wissenschaftliche Sparte ernst genommen zu werden. Welche Argumente können also angeführt werden, damit dies geschieht?

Zum Ersten werden Psychologen, die sich mit Forschung im Bereich der Personalauswahl befassen, sicherlich zustimmen, dass die psychologischen Aspekte, inwieweit Leute aus der Praxis unsere Ergebnisse annehmen, durch sie beeinflusst werden und sich nach ihnen richten, relevante Fragen für die Forschung sind. Im Gegensatz zu dem von Sackett [57] auf so eloquente Weise verfochtenen und bahnbrechendem Wissenschaft-

ler-Praktiker-Modell kann dies vielleicht mit dem Begriff „Wissenschaftler-Wissenschaftler-Modell“ bezeichnet werden, wobei Wissenschaftler ermutigt werden, über die Auswirkungen ihrer Forschungsergebnisse zu reflektieren.

Zum zweiten sind die psychologischen und wirtschaftlichen Kosten für Wissenschaftler, die dies nicht als neues Forschungsfeld eröffnen wollen, einfach zu hoch [55, 56]. Die derzeitige internationale Situation sieht bestenfalls so aus, dass es eine fehlende Synergie zwischen Forschung und Praxis in der Personalauswahl gibt. Der Forschung gelingt es nicht, die Praxis zu beeinflussen und umgekehrt, und so dreht sich das Rad der Isolation immer weiter. Die entscheidende Frage lautet: „Warum kann fundierte Forschung die Praxis nicht beeinflussen (und natürlich umgekehrt)?“ Es könnte argumentiert werden, dass von allen Bereichen der Psychologie der Personalauswahl unser Wissen gerade hier besonders bruchstückhaft und unvollständig ist – wir wissen sehr viel über die wichtigen Inhaltsbereiche, aber viel weniger über den Prozessbereich bezüglich dessen, warum es nicht gelingt, die Praxis innerhalb von Organisationen in gewissen Situationen zu beeinflussen. Es stimmt zwar, dass es nicht die Aufgabe der Wissenschaftler ist, ihre Forschungsergebnisse tagesaktuell in Organisationen umzusetzen, aber es kann argumentiert werden, dass es in der Verantwortung der Wissenschaftler liegt, die Auswirkungen ihrer Forschungsergebnisse zu prüfen. Sollte das denn nicht das eigentliche Herzstück der Forschung in der Personal- und Organisationspsychologie sein?

Zum dritten sind diese Verhältnisse schlichtweg an sich interessant. Warum begnügen sich Organisationen damit, nicht die gültigsten und zuverlässigsten Prädiktoren berufsbezogener Kriterien heranzuziehen? Warum sind HR-Manager vom nachweislich riesigen Nutzen einer Verbesserung der kriteriumsbezogenen Validität ihrer Auswahlverfahren offenbar unbeeindruckt? Wie kann die Entwicklung psychometrischer Tests für die Mitarbeiterauswahl besser auf psychologischer Theorie und empirischer Forschung begründet werden? Warum weisen Forscher Trends in den Personalauswahlmethoden in Organisationen als unwissenschaftlich, und somit aus dem Bereich seriöser Untersuchungen herausfallend, zurück?

All das sind wichtige Fragen, welche die angewandte Forschung in Zukunft stimulieren sollten und dabei behilflich sein könnten, Brücken zwischen Wissenschaft und Praxis zu bauen, indem die Gründe für den mangelnden Wissenstransfer innerhalb unserer Disziplin besser verstanden werden. Durch das Erschließen dieses sogenannten Prozessbereiches für die zukünftige Forschung wäre es nützlich, zuerst die Lektionen aus der Vergangenheit zu untersuchen. Genau damit beschäftigt sich der folgende Abschnitt, in welchem vier historische Szenarien des Forschungs-Praxis-Verhältnisses in der Psychologie der Personalauswahl erörtert werden.

---

## 1.5 Synergie zwischen Forschung und Praxis: Vier historische Szenarien der psychologischen Personalauswahl

Wie der Titel dieses Kapitels schon andeutet, steht es um das Verhältnis zwischen Wissenschaft und Praxis in der psychologischen Personalauswahl dann am schlechtesten, wenn die beiden Bereiche der Disziplin sich dessen nicht bewusst sind, woran der jeweils

andere gerade arbeitet, oder dies absichtlich ignoriert (sprichwörtlich gesagt, wenn „die linke Hand nicht weiß, was die rechte tut“). Auch wenn es schwer ist, diese beiden Szenarien im Nachhinein genau zu unterscheiden, sei angemerkt, dass ein Unterschied darin besteht, ob man sich aktueller Entwicklungen einfach nicht bewusst ist oder diese absichtlich nicht berücksichtigt. Es ist problematischer, wenn sich einer der beiden Bereiche unserer Disziplin der Entwicklungen im anderen Bereich einfach nicht bewusst ist, da solch fehlendes Wissen bedeuten würde, dass strukturelle Defizite im Informationsaustausch und Wissenstransfer zwischen Wissenschaftlern und Praktikern oder umgekehrt in derartigem Ausmaß bestehen – dass die eine „Hand“ tatsächlich nicht weiß, was die andere tut [22, 37]. Auch wenn der Informationsfluss zwischen den beiden Feldern der psychologischen Personalauswahl nicht unbedingt perfekt funktioniert, muss eingeräumt werden, dass dank der breiten Palette von aktuell bestehenden Zeitschriften und Newslettern, der bestehenden Kontakte zwischen Praktikern und Wissenschaftlern sowie regelmäßiger Konferenzen und Fachtagungen die Psychologie der Personalauswahl im Allgemeinen Nutzen aus schon fest verankerten Kontakten und Kanälen für Informationsaustausch zieht. Hinzu kommt, dass Forschungsförderungseinrichtungen in mehreren Ländern (USA, Großbritannien, Niederlande, Australien) verstärkt die Praxisbezogenheit als Kriterium für Förderungsprogramme hervorheben, womit in den letzten Jahren der Finanzierungsdruck auf dem Gebiet der psychologischen Personalauswahl weltweit in Richtung angewandte Forschung geht. Diese Mechanismen und dieser Finanzierungsdruck werden wahrscheinlich dazu führen, dass sich die Vernetzung zwischen Forschung und Praxis verbessert, zumindest was den Wissenstransfer und den seitens der Interessensgruppen auf den Wissenschaftlern lastenden Druck betrifft, für die Praxis relevante, angewandte Forschung (d. h. pragmatische Forschung) zu betreiben.

Ein Blick in die Geschichte der psychologischen Personalauswahl ermöglicht die Identifikation von vier grundlegenden Szenarien des Verhältnisses zwischen Wissenschaft und Praxis, wobei die ersten drei höchst funktional und nützlich, die vierte als kontraproduktiv für den Berufsstand angesehen werden muss. Diese sind:

1. Fundierte Forschung informiert die Personalauswahlpraxis.
2. Unzuverlässige Forschung verfehlt ihren Einfluss auf die Personalauswahlpraxis.
3. Praxistrends beeinflussen empirische Forschungsbemühungen.
4. Fundierte Forschung verfehlt ihren Einfluss auf die Personalauswahlpraxis.

Im vorliegenden Abschnitt dieses Kapitels werden alle Szenarien sequenziell beleuchtet und Schlüsselbeispiele gegeben, wie sich im historischen Rückblick Forschung und Praxis in unterschiedlichen Ländern im internationalen Vergleich wechselseitig beeinflusst haben oder eben nicht (für solch eine historische Perspektive siehe auch [30]). Es war nicht zu vermeiden, dass diese Beispiele und die Interpretation, ob die Forschung die Praxis entsprechend beeinflusst hat oder nicht, im Ermessen des Autors lagen. Nichtsdestotrotz ist diese Übersicht hilfreich und zur Illustration der oben genannten unterschiedlichen Szenarien berechtigt.

### **Szenario 1: Fundierte Forschung informiert die Personalauswahlpraxis**

Die Standardsituation, an deren Existenz alle Personal-, Betriebs- und Organisationspsychologen im Allgemeinen und Psychologen im Bereich der Personalauswahl im Besonderen in Bezug auf das Forschung-Praxis-Verhältnis gern überzeugt sein würden, ist, dass eine fundierte Forschung die Personalauswahlpraxis in angemessener Weise informiert. In diesem Szenario bildet die pragmatische Wissenschaft für gewöhnlich die Basis für Beratungsinterventionen. Wir alle halten uns an Einschränkungen, die den Wissenschaftler-Praktiker Strukturen zueigen sind [57], und alle Forschungsergebnisse werden automatisch auf die Berufspraxis umgelegt, mit nur geringer Verzögerung oder Spannung zwischen den Anforderungen der wissenschaftlichen und der „realen“ Welt. Es erübrigt sich anzumerken, dass es sich dabei um ein idealisiertes Szenario handelt. In der Realität werden Auswahlpraktiken eher von nur wenigen der wichtigsten Forschungsergebnisse beeinflusst werden, es wird Verzögerungen zwischen wissenschaftlichen Publikationen und deren Umsetzung in der Auswahlpraxis geben, mit Kompromissen zwischen den Erkenntnissen der theoretischen und angewandten Disziplin und den tagesaktuellen Anforderungen der Personalauswahlpraxis zu rechnen sein [55, 56].

Ein Beispiel für dieses Szenario aus der Geschichte der Personalauswahlpsychologie ist die Verwendung von Tests zur Messung der kognitiven Fähigkeiten (GMA) in verschiedenen Ländern. Dieses Beispiel illustriert auch sehr anschaulich die offensichtlichen Spannungen zwischen wissenschaftlichen Forschungsmethoden – in diesem Fall der zunehmenden Verwendung von Metaanalyseverfahren zur Herstellung generalisierbarer kriteriumsbezogener Validität – und den tagesaktuellen Anforderungen von Personalauswahlpraktikern, welche im spezifischen Kontext ihrer jeweiligen Organisation und ihrer jeweiligen Auswahl-situation Validität aufzeigen müssen. Es gibt heute eine überwältigende Fülle von Hinweisen dafür, dass GMA-Tests die besten Prädiktoren für die berufliche Leistung und den Ausbildungserfolg sowohl in den USA (z. B. [36, 66, 67]) als auch in Europa [60, 61, 64, 65] sind. Schmidt und Hunter stellen in ihrem 85 Jahre umfassenden Überblick zu bisherigen Studien in den USA fest, dass die durchschnittliche operationale Validität von GMA-Messungen ,44 für die berufliche Leistung und ,58 für die Vorhersage des Ausbildungserfolgs beträgt. Salgado und Anderson [60, 61] berichten, dass im Kontext der Personalauswahl in Europa die operationalen Validitäten von GMA-Tests etwas höher ausfallen als in den USA. Ihre Metaanalyse in elf EU-Ländern führte zu korrigierten beobachteten Validitätskoeffizienten ( $\rho$ ) von ,62 zur Vorhersage der Leistung und ,54 zur Prognose des Ausbildungserfolgs. Die operationale Validität wurde nicht durch die Kultur der jeweiligen Länder moderiert, aber durch die Aufgabenkomplexität: die operationale Validität war höher bei Tätigkeiten von hoher Komplexität. Dieser Befund deckt sich mit früheren Metaanalysen in den USA. Die Autoren präsentieren außerdem zusammenfassende Evidenz für die Popularität von GMA-Tests in verschiedenen Ländern Europas. Sie stellen fest, dass GMA-Tests in europäischen Ländern weiter verbreitet sind als in den USA und weisen darauf hin, dass diese Popularität in Europa nicht durch organisationale Bedenken über mögliche



von den Bewerbern eingeklagte Diskriminierungsfälle eingeschränkt wird, da in den meisten europäischen Ländern die Antidiskriminierungsgesetze weniger streng sind als in den USA. Rynes et al. [56] kamen in ihrer Studie über HR-Praktiker in den USA zu dem Schluss, dass es die allgemein vertretene Überzeugung ist, dass GMA-Tests weniger aussagekräftige Prädiktoren sind als Forschungsergebnisse dies nahe belegen. Dies führt wohl dazu, dass GMA-Tests in Amerika weniger populär sind als in Europa. Interessanterweise geht aus zahlreichen Untersuchungen zur Verwendung von GMA-Tests in europäischen Organisationen auch hervor, dass diese Testverfahren im Zeitverlauf immer populärer geworden sind [53]. Man kann durchaus argumentieren, dass die steigende Popularität im Zeitverlauf durch die Publikation von fundierten Forschungsergebnissen grundlegend beeinflusst worden ist. Natürlich beeinflussen auch andere Faktoren, wie zum Beispiel die Kosten, Trainingsanforderungen, kommerzielle Verfügbarkeit, „adverse impact“ etc. die Entscheidungen der Personalabteilungen bezüglich der anzuwendenden Methoden. Dennoch wäre es unangebracht zu behaupten, dass derart deutliche Belege für die kriteriumsbezogene Validität keinen Einfluss auf die wachsende Verwendung von Testverfahren zu kognitiven Fähigkeiten in der Auswahlpraxis ausüben. Dieses Beispiel kann als archetypische Illustration der Vorteile der pragmatischen Forschung bezüglich ihres internationalen Einflusses auf die Auswahlpraxis herangezogen werden. Es gibt jedoch auch Spannungen zwischen Wissenschaftlern und ihrer Tendenz zur Verwendung von Metaanalyseverfahren für die Zusammenfassung von kriteriumsbezogenen Validitätskoeffizienten über verschiedene Organisationen und Berufsfelder hinweg, und Personalauswahlpraktikern, welche direkte und situationsspezifische Belege dafür, dass GMA-Tests die Validität und Reliabilität für ihre spezifische Situation darstellen, bevorzugen [16, 25, 48]. Dennoch ist die steigende Verwendung von GMA-Tests in Organisationen im internationalen Vergleich zumindest teilweise darauf zurückzuführen, dass es nunmehr eine derartige Fülle an Belegen gibt, die die Verwendung dieser Tests im angewandten Kontext der Personalauswahl befürworten. Salgado und Anderson schließen daraus: „Die Höhe der festgestellten operationalen Validitäten lässt darauf schließen, dass das Ergebnis von GMA-Tests der beste Einzelprädiktor für die Personalauswahl für alle Berufe ist“ (S. 16).

Andere, weniger positive Forschungsergebnisse haben die Auswahlpraxis ebenfalls beeinflusst, in diesem Fall jedoch dahin gehend, dass so genannte „alternative“ Auswahlmethoden eher nicht verwendet werden. Evidenz über die nicht vorhandene kriteriumsbezogene Validität von Methoden wie z. B. Graphologie oder in einem gewissen Ausmaß auch Arbeitszeugnisse und Empfehlungsschreiben haben die Verwendung dieser Methoden verringert oder zumindest das Vertrauen, das von Praktikern in sie gesetzt wird [53]. Ein weiteres Beispiel für fundierte Forschung, die die Personalauswahlpraxis informiert, kann der Forschung zu Unterschieden zwischen Geschlechtern und ethnischen Untergruppen und die potenziell gegenteilige Wirkungen von Auswahlmethoden entnommen werden [6, 13]. Insbesondere in den USA haben sich Organisationen aufgrund der Forschungsergebnisse in diesem Bereich eingehend damit beschäftigt, nicht-diskriminierende Praktiken in der Personalauswahl zu gewährleisten, und offensichtlich ziehen auch



britische Organisationen nach, um zeigen zu können, dass auch ihre Auswahlverfahren frei von Diskriminierungen sind [49, 53]. Ein weiterer Bereich, dem in der aktuellen Forschung viel Aufmerksamkeit zu Teil wurde, ist jener der Bewerberreaktionen und der Entscheidungsfindung [54]. Auch wenn dieser Bereich erst in den Anfängen steht, wird auch er wahrscheinlich ein Beispiel für die die organisationale Praxis beeinflussenden pragmatischen Forschungsbemühungen werden.

Zusammenfassend kann für mehrere Bereiche der Forschung in der Personalauswahl festgehalten werden, dass die organisationale Praxis von den jeweiligen Forschungsergebnissen grundlegend beeinflusst wird. Auch wenn es Verzögerungen zwischen der Publikation von wissenschaftlichen Ergebnissen und deren Umsetzung in der organisationalen Praxis gibt, wobei die Interessen der Wissenschaftler und Praktiker nur mangelhaft auf einen Nenner gebracht werden können, kann man doch klar sagen, dass die professionelle Personalauswahl einige der besten Elemente der forschungsbasierten Praxis repräsentiert [22, 69].

Um es noch einmal zusammenzufassen – dieses erste Szenario stellt in vielerlei Hinsicht den Idealfall des Forschung-Praxis-Verhältnisses dar. Glücklicherweise gibt es im Bereich der psychologischen der Personalauswahl im historischen Rückblick eine Fülle von Beispielen, in denen dieses Szenario tatsächlich gegeben war. Um diese Feststellung zu veranschaulichen, werden hier nur einige der möglichen Beispiele dafür angeführt, wie stringente Forschung die Personalauswahlpraxis positiv beeinflusst hat. Aufgrund dieses Sachverhalts entstand in unserem Bereich eine eher naive und nicht weiter hinterfragte Reihe von Annahmen, dass dieses Szenario automatisch der Standard im Verhältnis zwischen Wissenschaft und Praxis ist. Es gibt jedoch keine Garantie, dass dieses Szenario tatsächlich Standard ist – ganz im Gegenteil. Es obliegt den Psychologen im Bereich der Personalauswahl, die der Forschung bzw. der Praxis angehören, sicherzustellen, dass zwischen Wissenschaft und Praxis eine wechselseitige, symbiotische Beziehung besteht. Wir können das nicht einfach als Gegebenheit voraussetzen. Um auf diesen Punkt weiter einzugehen, können die drei anderen festgestellten und im Folgenden beschriebenen Szenarien als typische Beispiele für die Beziehungen zwischen Forschung und Praxis in unserer Disziplin herangezogen werden, wie zahlreiche Beispiele im Laufe der Geschichte der Personalauswahlpsychologie anschaulich zeigen.

### **Szenario 2: Unzuverlässige Forschung verfehlt ihren Einfluss auf die Personalauswahlpraxis**

Das zweite Forschung-Praxis-Szenario ist jenes, wo unzuverlässige Forschungsergebnisse aufgrund eines Zufalls daran scheitern, die Praxis der Personalauswahl in Organisationen zu beeinflussen. In diesem Fall erweist sich eine historische Perspektive als besonders informativ [59]. Es liegt in der Natur der Wissenschaft, dass sie sich mit der Zeit entwickelt, wobei die auf weniger stringenter Forschung und weniger genauen Analysetechniken basierenden Ergebnisse früherer Studien von gegenwärtiger Forschung infrage gestellt und falsifiziert werden [40, 50]. Was die Wahl des Begriffes „unzuverlässige“ Forschung betrifft, sollte an dieser Stelle klar gestellt werden, dass sich dies

wahrscheinlich erst im Nachhinein und in Kenntnis der späteren Fortschritte in den Forschungsmethoden und Analysetechniken als solches herausstellt (Ryan, persönliche Korrespondenz). Zu dem Zeitpunkt, als die Originalstudien durchgeführt worden sind, haben die Forscher wahrscheinlich die damals am weitesten entwickelten Zugänge, die ihnen zur Verfügung standen, verwendet. Dies ist eine inhärente Eigenschaft jeglichen Forschungsfortschrittes in allen Bereichen [40]. Die Falsifikation von früheren empirischen Erkenntnissen und theoretischen Modellen stellt einen der wichtigsten Grundsätze wissenschaftlicher Untersuchungen dar, und die „normale“ Wissenschaft in der Personal-, Betriebs- und Organisationspsychologie erzielt ihre Fortschritte aufgrund eben dieser Gegebenheiten [1]. Wir sollten daher nicht überrascht sein, wenn frühe Erkenntnisse aus unserem Bereich von späteren Forschungsarbeiten, welche ein fundierteres Forschungsdesign und analytischere Methoden verwenden, infrage gestellt oder sogar modifiziert werden. Die logische Folge dieses Sachverhalts ist interessanterweise, dass sich die Auswahlpraktiken auf unserem Gebiet historisch betrachtet nicht immer nach wissenschaftlichen Ergebnissen richteten. Das unter mehreren Beispielen wohl auffälligste ist die fortwährende Verwendung aller möglichen Arten von Interviews, auch völlig unstrukturierter Natur, trotz früherer kritischer Befunde, welche an ihrem Wert zweifeln ließen (in chronologischer Reihenfolge: [74, 45, 71, 75, 7, 8]). Erst ab Mitte der 1980er Jahre begann sich diese Auffassung von der offenbar inhärenten Unzuverlässigkeit und Ungültigkeit von Interviews durch eine Reihe von veröffentlichten Metaanalysen zur prädiktiven Validität von Interviews ändern [35, 47, 62]. In der Metaanalyse von Salgado und Moscoso [62] zum Beispiel wurde festgestellt, dass unstrukturierte Interviews eine durchschnittliche korrigierte operationale Validität von ,20 haben, während für hoch strukturierte Interviews eine beachtliche Validität in Höhe von ,56 empirisch belegt werden konnte (d. h. nur knapp unter der Marke der Validität, die, wie oben angeführt, auf internationaler Ebene für GMA-Tests festgestellt wurde). McDaniel et al. [47] berichteten von durchschnittlichen operationalen Validitäten von ,37 für alle Arten von Interviews, ,44 für strukturierte und ,33 für unstrukturierte Interviews. Diese Reihe von metaanalytischen Ergebnissen hat die Glaubwürdigkeit von Interviews als Messinstrument im Wesentlichen wieder hergestellt. Sogar die Einbeziehung unstrukturierter Interviews in den Auswahlprozess hat einen gewissen Wert [28]. Hinsichtlich des Szenarios des Verhältnisses zwischen Forschung und Praxis ist es unumstritten, dass die Popularität von Interviews auch in der früheren Phase, bevor die oben genannten Metaanalysen bekannt wurden, nicht zurückging. Vielmehr blieb das Interview für alle Berufsarten nahezu universell beliebt, ungeachtet der herkömmlichen Annahme seitens der Personalauswahlpraktiker, dass es dieser Methode an Validität und Zuverlässigkeit mangle [21]. Dies war ein gutes Beispiel dafür, dass Forschung die Praxis nicht beeinflusst, auch wenn Personalexperten diese Forschungsergebnisse offensichtlich kannten. Noch immer scheinen sich diese Überzeugungen in den letzten Jahren bei HR-Praktikern hartnäckig zu halten, trotz der Publikation der oben genannten bedeutsamen metaanalytischen Ergebnisse [23]. Das heißt, dass ein Teil der Praktiker allem Anschein nach noch immer glaubt, dass alle Arten von Interviews inhärent fehlerhaft sind. In diesem Zusammenhang

ist anzunehmen, dass die aktuelleren Forschungsergebnisse Praktikern im HR-Bereich vielleicht noch nicht ausreichend bekannt sind, um die Praxis positiv zu beeinflussen.

Ein weiteres Beispiel für unzuverlässige Forschung, welche die Praxis nicht beeinflusst, gibt es im Bereich der Persönlichkeitstests, wobei dieser Bereich der Forschung enorme Auswirkungen in den USA, nicht aber in Europa hatte. Barrick et al. [11] weisen in ihrer retrospektiven Übersicht über die Persönlichkeitsforschung und Persönlichkeitstests im letzten Jahrhundert auf diese unterschiedlichen Auswirkungen hin. Indem sie die maßgebliche Überblicksarbeit von Guion und Gottier [27] zitieren, schließen die Autoren: „Es gibt keine generalisierbaren Hinweise darauf, dass Persönlichkeitsmessungen als gute oder praktische Hilfsmittel für die Personalauswahl empfohlen werden können“ (S. 159). Dieses vernichtende Urteil hatte großen Einfluss auf die Popularität von Persönlichkeitstests in den USA, und wie Barrick et al. [11] meinen, wurde dieser Überzeugung, die sich in weiterer Folge als falsch herauskristallisierte, über einen Zeitraum von rund 25 Jahren kaum widersprochen. In der Europäischen Union hatte Guions und Gottiers negative Einschätzung jedoch nicht einmal annähernd gleich starke Auswirkungen auf die Auswahlpraxis und die dortige Verwendung von Persönlichkeitsinventaren zum Zwecke der Personalauswahl. Tatsächlich ist vielmehr das Gegenteil geschehen [29]. In späteren Untersuchungen zur Popularität von Persönlichkeitstests in diesem Zeitraum von 25 Jahren gibt es vielmehr Hinweise auf eine Zunahme von derartigen Messverfahren für die Personalauswahl in Organisationen [12, 31, 52, 68]. Wie kommt es zu diesen beachtlichen Unterschieden zwischen den USA und Europa? Eine Erklärung dafür ist, dass die Arbeit von Guion und Gottier [27] und die darin enthaltenen Schlussfolgerungen zum damaligen Zeitpunkt einfach nicht zu den europäischen Wissenschaftlern durchdrangen. Vor ca. 40 Jahren wurden wissenschaftliche Zeitschriften weit weniger international berücksichtigt als es heute aufgrund des elektronischen Zugangs und der viel stärkeren Vernetzung von Wissenschaftlern in den unterschiedlichen Ländern der Fall ist. Eine einfache, plausible Erklärung wäre also, dass die europäischen Wissenschaftler geringere Kenntnis von Guions und Gottiers Schlussfolgerung hatten oder zumindest nicht so sehr davon beeinflusst waren. Ganz sicher erfuhr diese apokalyptische Schlussfolgerung in den europäischen akademischen Reviews bezüglich des Nutzens der Persönlichkeitstests viel geringere Akzeptanz, und ihre Botschaft, die im Grunde genommen ein Moratorium für die Verwendung von Persönlichkeitstests darstellte, wurde in den wichtigsten Texten zum HR-Management in Großbritannien nicht einmal erwähnt [9, 70]. Eine weitere plausible Erklärung ist, dass dieser Beitrag kurz vor einer Phase des enormen Wachstums, vor allem während des Booms der 80er Jahre, im Bereich der Personalberatungsfirmen in Europa erschien, als viele dieser Beratungsfirmen Persönlichkeitstests als Teil ihres Produktmix und Beratungsservice verwendeten. Was auch immer die richtige Erklärung sein mag, fest steht, dass die damalige ständig zunehmende Popularität der Persönlichkeitstests für die Personalauswahl in ganz Europa in starkem Kontrast zu den Entwicklungen in Amerika stand [11]. Ironischerweise könnte dies ein Fall dafür sein, dass die rechte Hand [in Europa] nicht wusste, was die linke Hand [in den USA] tat, auch wenn dieses Unwissen im Nachhinein betrachtet unvorhergesehene Vorteile mit sich brachte. Seither

wurde in den zahlreichen Metaanalysen zur prädiktiven Validität von Persönlichkeitstests natürlich zweifelsfrei belegt, dass sorgfältig ausgearbeitete Persönlichkeitsinventare verlässliche Prädiktoren sind, im kulturellen Kontext der USA [10] und Europas [58].

Was kann also aus diesen Beispielen des zweiten Szenarios, in welchem unzuverlässige Forschung die professionelle Praxis aufgrund eines Zufalls nicht beeinflusst, geschlossen werden? Erstens das Eingeständnis, dass Forschung kein unfehlbares Allheilmittel für jeglichen Aspekt der Personalauswahlpraxis in Organisationen darstellen muss. Die Wissenschaft entwickelt sich ständig weiter, ihre älteren Erkenntnisse und Schlussfolgerungen werden manchmal (wenn auch seltener als man vielleicht annehmen würde) revidiert, und ihre Methoden und analytischen Richtlinien verbessern sich ständig. Man sollte nicht vergessen, dass die Personalpsychologie eine relativ junge Wissenschaft ist, und dass unser Verständnis in den letzten zwei Jahrzehnten vor allem durch die Einführung und Popularisierung von Metaanalysetechniken verbessert wurde [66]. Zweitens sollten wir als Forscher dafür dankbar sein, dass Praktiker in der Personalauswahl unsere Ergebnisse oft mit einer gesunden Portion Skepsis aufnehmen. Es sei nochmals darauf hingewiesen, dass es in der Natur der Forschung liegt, sich immer auf einen Teilaspekt des Gesamten zu spezialisieren, während Praktiker im Bereich des HR-Managements, die nur über begrenzte Kenntnis wissenschaftlicher Belege verfügen und eine Fülle unterschiedlicher Anforderungen erfüllen müssen, die Beliebtheit der einzelnen Auswahlmethoden bestimmen [41]. Drittens unterstützt dieses Szenario die schon dargelegte Äußerung, dass es eine natürliche Distanz zwischen Forschung und Praxis in der psychologischen Personalauswahl geben sollte. Es wäre für das Prozedere der Personalauswahl in jeder Organisation überaus störend, wenn man sich dabei unkritisch an die Ergebnisse jeder einzelnen publizierten Studie halten oder auch nur versuchen würde, mit der ganzen Fülle von veröffentlichten Forschungsergebnissen sofort Schritt zu halten. Vielmehr ist es günstiger, eine gewisse Zeit verstreichen zu lassen, um wichtige Ergebnisse zu reflektieren, sie kritisch zu hinterfragen und dann die wissenschaftliche Meinung in der richtigen Balance in die berufliche Praxis zu übertragen. Schließlich hebt dieses zweite Szenario auch hervor, wie wichtig es für Forschung ist, auf internationaler Ebene für die Entwicklungen in der internationalen beruflichen Praxis offen zu sein und sich gegebenenfalls auch von diesen beeinflussen zu lassen. Die Frage, warum diese Ergebnisse die Praxis nicht verändern konnten, ist überaus interessant, und es kann durchaus behauptet werden, dass diese in der Vergangenheit von der Forschung zu wenig Aufmerksamkeit erhielten.

### **Szenario 3: Praxistrends beeinflussen empirische Forschungsbemühungen**

Es wäre natürlich unangemessen zu behaupten, dass das Verhältnis zwischen Forschung und Praxis in der Personalauswahl nur in einer Richtung besteht. Es können mehrere Beispiele angeführt werden, wo neuere Trends in der Praxis der Personalauswahl in Organisationen neue Ausrichtungen für die angewandte Forschung, Metaanalysen und die Theoriebildung angeregt haben [44]. Bereiche, wo dies beispielsweise geschieht, sind Forschungsarbeiten zu Kompetenzansätzen, Leistungsbewertung durch mehrere unabhängige Beurteiler, emotionale Intelligenz, computerbasierte Testverfahren, Ehrlichkeits- und Integritätstests, Drogen- und Alkoholtests, internetbasierte Personalauswahl, telefonbasierte

Interviews und adaptives Computertesten. Von all diesen Bereichen gilt die größte Aufmerksamkeit derzeit der internetbasierten Personalauswahl und der internetbasierten Eignungsdiagnostik. Diese Regsamkeit bei Forschungsaktivitäten ist ein direktes Ergebnis davon, dass Organisationen die Methoden der webbasierten Personalauswahlverfahren auffallend schnell aufgreifen (vgl. z. B. [2] für ein aktuelles Interview). Vor allem in den USA verlassen sich inzwischen viele große Organisationen sowohl für Auswahl- als auch für Screening-Zwecke auf das Internet. Lievens und Harris [43] führen an, dass 88 % von 500 global agierenden Unternehmen in den USA nunmehr webbasierte Auswahlverfahren nutzen. Dieses Wachstum ist der wichtigste Faktor, der zurzeit Forschung in diesem Bereich auslöst; man kann also sagen, dass es sich dabei um ein deutliches Beispiel dafür handelt, wie Entwicklungen in der Praxis die Forschungsbemühungen beeinflussen.

Es lässt sich nicht vermeiden, dass es zu einer Verzögerung zwischen schon umgesetzten Veränderungen in der Personalauswahlpraxis in Organisationen und der darauf folgenden diesbezüglichen Forschung kommen kann (Lievens, persönliche Korrespondenz). Im Fall internetbasierter Verfahren geschah dies sicherlich aufgrund der enormen Geschwindigkeit, mit welcher Organisationen neue Technologie aufgriffen. Sollte nun die Forschung eine solch reaktive Position einnehmen oder sollte im Vorfeld dieser Entwicklungen spekulativeren Ansätzen in der Theoriebildung und in empirischen Studien der Vorzug gegeben werden? Ich plädiere für Letzteres. Besonders im Fall des oben genannten Beispiels war es vorhersehbar, dass einige Organisationen webbasierte Lösungen aufgrund der ihnen innewohnenden Vorteile im Bereich der Kosteneffizienz, der unmittelbaren Reaktion, der Bequemlichkeit für Bewerber usw. übernehmen würden. Was vielleicht nicht vorhersehbar war, war das Ausmaß, in welchem diese neue Technologie übernommen wurde. Dies hat dazu geführt, dass es einen großen Mangel an Forschungsarbeiten zu den Auswirkungen von webbasierter Personalauswahl und Berufseignungsdiagnostik gibt, der einer enorm gestiegenen Verwendung dieser Methoden in der Praxis gegenübersteht [43]. Weitere Forschung hierzu hat nun begonnen, wobei noch unklar ist, wie spekulativere Forschungsarbeiten, die solchen Entwicklungen in der Praxis quasi vorausgreifen, in Zukunft gefördert werden können. Vielleicht wurden Wissenschaftler im Bereich der Personalauswahl tatsächlich zu konservativ in ihren Ansätzen, und es hätte schon vor Jahren einer visionärerem Denkweise bedurft, um den Vorsprung der Erkenntnisgrundlagen gegenüber diesen Veränderungen in der Praxis zu gewährleisten.

In Zukunft wäre es für die Forschung im Bereich der Personalauswahl ganz klar von Vorteil, Wegweiser in solchen Entwicklungen in der Praxis zu sein, das heißt, eine verstärkt aktive anstatt der traditionellerweise reaktiven Position einzunehmen. Eine Voraussetzung dafür ist, wie gesagt, das Bestehen ausreichender Feedback-Kontakte und Informationskanäle von der Praxis zurück zur Forschung, damit Wissenschaftler fortwährend mit dem Pragmatismus der Praktiker konfrontiert und durch diesen gefordert werden. Möglicherweise ist es nun an der Zeit, dass eine der führenden Fachzeitschriften im Bereich der Personalauswahl eine Sonderausgabe zur Schnittstelle zwischen Forschung und Praxis und umgekehrt in Auftrag gibt, ähnlich der von Rynes et al. [55] herausgegebenen Sonderausgabe des *Academy of Management Journals* (AMJ). Es könnten auch einige rein spekulative Workshops für Praktiker und Forscher ins Leben

gerufen werden, zur Ausarbeitung möglicher Zukunftsszenarien und -trends, als Anregung für eine verstärkt visionäre Perspektive und innovative Denkweise (und wenn ein derartiger Vorschlag absurd klingt, dann ist das auch wieder ein Hinweis darauf, wie sehr es unserem Arbeitsbereich an Visionen mangelt). Abschließend kann also gesagt werden, dass es zwar Beispiele dafür gibt, dass Trends in der Praxis die Forschung stimulieren, dass jedoch der Bereich der Personalauswahlpsychologie insgesamt auch von praxisstimulierter Forschung erheblich profitieren würde.

#### **Szenario 4: Fundierte Forschung verfehlt ihren Einfluss auf die Personalauswahlpraxis**

Das vierte und letzte Szenario, in welchem es fundierter Forschung aus welchem Grund auch immer nicht gelingt, in die Personalauswahlpraxis zu diffundieren, ist vielleicht das schwierigste für die psychologische Personalauswahl als Berufsfeld. In diesem Szenario mussten alle Kosten für die Generierung einer pragmatischen Forschung getragen werden, und dennoch wurde keinerlei Nutzen aus deren praktischer Anwendung gezogen. Wie im Fall des oben beschriebenen zweiten Szenarios ist auch dieses vierte Szenario dort seltener anzutreffen, wo fundierte Forschung die professionelle Praxis beeinflusst hat (Szenario 1). Dennoch gibt es eindeutige Beispiele, in welchen die Forschung offensichtlich zu wenig auf die Praxis der Personalauswahl in Organisationen einwirkte [55, 56]. Dieser Beitrag konzentriert sich insbesondere auf zwei Beispiele – das Fünf-Faktoren-Modell (FFM) der Persönlichkeit und die Entwicklung von berufsbezogenen Persönlichkeitstests sowie den Einfluss von Nutzenanalysen auf die Entscheidung für eine Auswahlmethode seitens der HR-Praktiker.

In den letzten drei Jahrzehnten wurde eine Fülle an Evidenz zugunsten der taxonomischen Struktur des FFM (auch „Big Five“ genannt) als latentes Modell für Persönlichkeit und individuelle Unterschiede hervorgebracht [20]. Beeindruckenderweise indiziert dieses Forschungsmaterial die Anwendbarkeit des FFM für ein wesentliches Spektrum von Messungen inhärenter Persönlichkeitsmerkmale [17, 24] und über viele unterschiedliche Länder [46] und Sprachen und Kulturen [75] hinweg. In Bezug auf die Personalauswahl stellte sich jedoch als wichtigste Erkenntnis in der Forschung heraus, dass FFM-basierte Persönlichkeitsinventare eine größere kriteriumsbezogene Validität aufweisen als Persönlichkeitsmessungen, welche auf alternativen Persönlichkeitsmodellen basieren [58]. Dieser metaanalytische Befund, der aus Studien in Europa, nicht in den USA, gewonnen wurde, ließe eigentlich erwarten, dass seitens kommerzieller Testherausgeber in ganz Europa ein dramatischer Wechsel zu FFM-basierten Persönlichkeitsinventaren erfolgte. Dies war jedoch keineswegs der Fall, sondern, ganz im Gegenteil, wie Hough [34] prägnant anmerkt: „Personal- und Organisationspsychologen haben es verabsäumt, der taxonomischen Struktur ihrer Variablen Beachtung zu schenken, teilweise vielleicht aufgrund ihrer exzessiven Empirie, vielleicht auch als Ergebnis ihrer pragmatischen Ausrichtung auf ein sofort anwendbares Ziel!“ (S. 21).

Während diese Belege für die Überlegenheit von FFM-basierten Messungsmethoden erst vor relativ kurzer Zeit in Europa veröffentlicht wurden, gibt es in den USA bereits seit



einigen Jahren wegweisende Ergebnisse [10, 18]. Angesichts der nunmehr umfangreichen empirischen Evidenz könnte man erwarten, dass kommerziell vermarktete Persönlichkeits-testverfahren zur Personalauswahl diese Ergebnisse in ihr Design und das ihnen zugrunde liegende Konstruktionsmodell miteinbezogen haben. Dies ist jedoch bis jetzt nicht erfolgt. Einige geschützte Testverfahren geben zwar vor, dass Verbindungen zum FFM bestehen, aber auf psychometrischer Ebene wurden seit ihrer Einführung relativ wenige Messverfahren entwickelt, die auf dieser taxonomischen Struktur basieren (eine Ausnahme bildet natürlich das überarbeitete NEO Persönlichkeitsinventar nach Costa und McCrae; siehe auch Anderson & Ones [4, 18]). Es wird weiterhin darüber debattiert, ob das FFM die umfassendste und sparsamste Typologie der normalen erwachsenen Persönlichkeit darstellt, womit angedeutet wird, dass kommerzielle Testverlage nur Vorsicht walten ließen, „solange das letzte Wort noch nicht gesprochen ist“, bevor man zu diesem Zugang wechselt [34]. Es gibt zwangsläufig auch eine zeitliche Verzögerung, bevor Forschungsergebnisse Eingang in die Praxis finden, was wiederum die Möglichkeit offen hält, dass sich kommerziell vermarktete Persönlichkeitsinventare in den nächsten Jahren in Richtung des FFM als typologischem Rahmen bewegen werden. Aus welchen Gründen auch immer scheint das FFM trotz des umfangreichen, international zusammengetragenen empirischen Belegmaterials für seine Konstruktion und kriteriumsbezogene Validität weniger Einfluss auf kommerzielle Testverlage ausgeübt zu haben als ihm zugestanden hätte [33]. Es stimmt, dass für einige geschützte Persönlichkeitstests durch das Zusammenzählen von Rohwerten auf Primärdimensionen Faktoren zweiter Ordnung berechnet werden können, und dass diese Faktoren zweiter Ordnung dem FFM ähneln (z. B. der Hogan-Persönlichkeitstest HPI, der 16 PF5, einige Versionen der OPQ-Testgruppe), aber für viele andere Tests wurden Verhältnisse zum FFM noch nicht ausgearbeitet. Dies hat dazu geführt, dass Testverlage dazu aufgerufen werden, das ihren kommerziell veröffentlichten Messverfahren zugrunde liegende Persönlichkeitsmodell vollständig zu dokumentieren und Verbindungen zum FFM zu beschreiben [4]. Außerdem scheinen HR-Praktiker jene Persönlichkeitsmessungen attraktiver zu finden, die feinere Analysemöglichkeiten bieten, das heißt, beträchtlich mehr als nur die fünf übergeordneten Dimensionen des FFM [19]. Ungeachtet dieser möglichen Spannungen zwischen Forschung und Praxis, oder vielleicht gerade ihretwegen, steht der Einfluss des FFM auf die Persönlichkeitsmodelle, auf welchen die geschützten Persönlichkeitstests zur Personalauswahl derzeit basieren, offensichtlich noch aus.

Das zweite Beispiel für dieses letzte Szenario, wo fundierte Forschungsergebnisse offensichtlich zu wenig Einfluss auf die Praxis ausgeübt haben, betrifft das Thema der Nutzenanalyse in der Personalauswahl. In der Tat ist dieses Beispiel besonders geeignet, da Nutzenanalysemodelle auf der Annahme beruhen, dass Praktiker im HR-Bereich unter verschiedenen Auswahlmethoden auswählen, größtenteils aufgrund rational-wirtschaftlicher Kriterien [14]. Diese Annahmen sind ein Mikrokosmos von Annahmen, dass Forschung im Allgemeinen die Tätigkeit der Praktiker beeinflussen sollte, und zwar basierend auf objektiven und rationalen Belegen, die den Mehrwert solcher Interventionen bestätigen [3, 41]. In der Studie von Latham und Whyte [41] evaluierten 143 erfahrene HR- und Linienmanager die Überzeugungskraft des Outputs von Nutzenanalysen (Dollar-Einsparungen) im Vergleich zu eher allgemeinen Informationsberichten über die Vorteile

unterschiedlicher Auswahlmethoden. Entgegen einiger Hauptannahmen der Nutzenanalyse beeinflussten die finanzorientierten Informationen über die Kosteneinsparungen die Manager negativ, was die Autoren zu folgender Schlussfolgerung veranlasste: „Diejenigen, die sich auf die Nutzenanalyse verlassen und denen es gelingt, dass ihre Empfehlungen akzeptiert werden, verbuchen diese als Erfolge trotz und nicht wegen ihrer Zuversicht in diese Technik“ (S. 43). Es scheint, dass die Nutzenanalyse dahin gehend wertlos wurde, was ihre Funktion betrifft, HR-Praktiker bezüglich finanzieller Vorteile gültigerer Prädiktoren zu überzeugen, da ihre Berechnung angeblich enormer finanzieller Einsparungen in den Augen der Praktiker unglaubwürdig ist [15]. Problematischer in Bezug auf den vorliegenden Beitrag ist die Frage, ob die Übermittlung von wissenschaftlichen Belegen an Praktiker der überzeugendste Weg ist, um Vorteile zu kommunizieren. Im Gegenteil, Praktiker werden vielleicht eher von den Machtverhältnissen und der Politik in den Organisationen, für die sie gerade arbeiten, beeinflusst, wobei Mitarbeiter im HR-Bereich nicht gerade die höchsten Positionen in der organisationalen Hierarchie einnehmen. Sicherlich vergleicht die Nutzenanalyse in ihrer derzeitigen Struktur nicht Like-for-like-Interventionen auf einer organisationalen Analyseebene, und so sind Praktiker eher skeptisch hinsichtlich der angeblichen riesigen finanziellen Erträge. Sind also unsere Annahmen bezüglich der Überzeugungskraft rational-wirtschaftlicher Belege in der Personalauswahl fehl am Platz? Und wenn ja, können wir jemals erwarten, dass wissenschaftliche Erkenntnisse jene Auswirkungen auf die Praxis haben werden, die ihnen (auf rational-wirtschaftlichen Kriterien basierend) zustehen? Das sind die Schlüsselfragen und die größten Herausforderungen, die den Kern der Forschungs-Praxis-Beziehungen in der psychologischen Personalauswahl treffen. Anderson et al. [3] schließen folgendermaßen: „(...) wir legten typischerweise zu viel Augenmerk auf die Durchführung der Personalauswahl als rationale technische Interventionen und haben deshalb oft keinen Einfluss auf Organisationen ... Forscher im Bereich der Personalauswahl sollten ihre Interventionen als organisationale Interventionen sehen, welche demselben Druck unterliegen wie andere organisationale Innovationen“ (S. 11; s. auch Johns [38]).

Zusammenfassend stellt das vierte Szenario das schwierigste und anspruchsvollste Szenario für Psychologen im Bereich der Personalauswahl dar. Sicherlich, es unterminiert unser Selbstbild als wissenschaftliche Praktiker [57], aber aufschlussreicherweise ist es dort, wo bedeutende wissenschaftliche Erkenntnisse die Praxis nicht beeinflussen können, ein eindeutiges Zeugnis dafür, dass es in unserem Bereich eine Kluft zwischen Praxis und Forschung gibt. Auch wenn viele idealistischerweise davon überzeugt sein wollen, dass Szenario 1 das Verhältnis zwischen Forschung und Praxis in der Personal-, Betriebs- und Organisationspsychologie immer typifiziert, können doch die oben genannten Beispiele in einer Anwendung der Selbsttäuschung, in der wir uns einreden, dass in dieser Hinsicht alles immer rosig ist, nicht einfach fallen gelassen werden. Am schwierigsten sind Annahmen, dass die rational-wissenschaftliche Logik selbst das Interesse der Praktiker an wissenschaftlichen Erkenntnissen nicht wecken kann. Vielmehr müssen wir Wege erforschen, wie Schlüsselerkenntnisse der Forschung im Bereich der Personalauswahl so kommuniziert werden, dass die politischen Gegebenheiten und die



Denkweise der Praktiker in den Organisationen berücksichtigt werden. Vielleicht hält ein kleiner Teil der Forscher bereits zu lange an der Einstellung fest, dass Praktiker diesen Botschaften ungeachtet des Mediums, das diese vermittelt, Gehör schenken müssen [55].

---

## 1.6 Anregungen für die zukünftige Forschung im Bereich des Forschung-Praxis-Verhältnisses in der Personalauswahl

Nachdem nun diese vier Szenarien des Verhältnisses zwischen Forschung und Praxis in der Personalauswahl auf internationaler Ebene identifiziert und beschrieben worden sind, sollen im letzten Teil des Beitrags einige Anregungen für die zukünftige Forschung zu Fragen rund um das Forschung-Praxis-Verhältnis in der Personal-, Betriebs- und Organisationspsychologie gegeben werden. Von den vier Szenarien sind die ersten drei eindeutig positiv, was ihren Effekt und ihre Resultate betrifft, während das vierte negativ ist, möglicherweise äußerst negativ, wenn zum Beispiel umfangreiche Evidenz die Personalauswahlpraxis in Organisationen nicht beeinflussen kann. Aus diesen Szenarien lässt sich eine Reihe von wichtigen Fragen ableiten:

1. Wie können wir als Berufsgruppe sicherstellen, dass Szenario 1 und 2 im Forschungs-Praxis-Verhältnis vorherrschen und dass gleichzeitig Szenario 3 minimiert oder sogar eliminiert wird?
2. Warum konnte die fundierte Wissenschaft Personalauswahlpraxis in einigen Fällen nicht beeinflussen, während in anderen Fällen weniger zuverlässige wissenschaftliche Erkenntnisse von Praktikern in der Personalauswahl richtigerweise ignoriert wurden?
3. Wie können wir eine Situation herbeiführen, in der es einerseits eine „natürliche Distanz“ zwischen Forschung und Praxis gibt, aber andererseits fundierte Forschungsergebnisse die Tätigkeit der Praktiker dennoch beeinflussen (d. h. Szenario 1), nachdem sie in pragmatische Terminologie „übersetzt“ wurden?
4. Wie können wir den Fortbestand von starken Vernetzungen zwischen Forschung und Praxis durch strukturelle Maßnahmen und Kommunikationskanäle (Zeitschriften, Newsletter, Konferenzen, Interessensgruppen etc.) sicherstellen?
5. Wie können wir auf internationaler Ebene die Weitergabe von Informationen und Best-Practice-Erfahrungen in der Personalauswahl über nationale Grenzen und Kulturen hinweg sicherstellen?

Alle fünf Fragen sind für die Förderung der Prozessbereich-Forschung hinsichtlich des Forschungs-Praxis-Verhältnisses in der Personalauswahl von zentraler Bedeutung. Im letzten Abschnitt des vorliegenden Kapitels schlage ich zwei Hauptausrichtungen für zukünftige Forschung vor. Paradoxerweise lässt jedoch schon ein oberflächlicher Blick auf die Beiträge zu diesem Thema in der Fachliteratur erkennen, dass Forschungsarbeiten zum Forschungs-Praxis-Verhältnis sehr rar sind. Abgesehen von einigen bemerkenswerten Ausnahmen [5, 22, 55, 57], haben Wissenschaftler diese überaus wichtigen Themen größtenteils vernachlässigt, vielleicht deshalb, weil sie es vorzogen, ihre eigenen

Forschungsagenden voranzutreiben, statt inne zu halten und zu reflektieren, ob ihre Ergebnisse auch den erwünschten Einfluss auf die Praxis haben. Dies war keine gesunde Situation für die Personal-, Betriebs- und Organisationspsychologie im Allgemeinen und die psychologische Personalauswahl im Besonderen [73]. Die Auswirkungen unserer Forschungsergebnisse auf die Praxis sind ebenso wichtig wie der Inhalt der Ergebnisse per se, und dennoch hat die Psychologie der Personalauswahl diese Transferfragen in den letzten Jahren erstaunlicherweise vernachlässigt. Forscher meinen vielleicht, dass ihre Ergebnisse auf taube Ohren stoßen, wobei man andererseits auf internationalen Konferenzen immer wieder verbitterte Beschwerden von Praktikern hört, nämlich darüber, dass Studien in einem Stil verfasst sind, der Reaktionen des Unverständnisses, der Langeweile oder des Spotts hervorruft, da umständlich konzipierte Studien zur empirischen Bestätigung einleuchtender und auf der Hand liegender Tatsachen durchgeführt wurden (d. h. pedantische Wissenschaft, siehe Abb. 1.1). Es ist daher an der Zeit dazu aufzurufen, dass dem Verhältnis zwischen Forschung und Praxis in der psychologischen Personalauswahl mehr Aufmerksamkeit gewidmet wird, wobei der vorliegende Beitrag ein Versuch ist, die Bedeutung dieser Themen in den Vordergrund zu stellen. Zwei Ausrichtungen für die zukünftige Forschung und wichtige Untersuchungsansätze erscheinen zum derzeitigen Zeitpunkt besonders wertvoll – und zwar die Forschung bezüglich der Anschauungen und Überzeugungsstrategien der Praktiker sowie die international durchgeführte Validitätsforschung zur Effektivität von Trainingsinterventionen der Personal-, Betriebs- und Organisationspsychologie innerhalb des „Continuing Professional Development“ (CPD).

Im Vergleich zu den geballten Bemühungen der Wissenschaftler, die den Inhaltsbereich der Personalauswahl (i. e. prädiktive Validität, Konstruktvalidität, gegenteilige Wirkungen etc.) betreffen, gibt es bisher sehr wenig Forschung zu den Anschauungen der Praktiker bezüglich Schlüsselergebnissen aus der Forschung einerseits, und andererseits zu den Strategien, welche die Anschauungen der Praktiker und ihre tagesaktuellen Tätigkeiten ändern könnten. Unserem Verständnis zufolge in Bezug auf die Frage, wie Wissenschaft die Praxis beeinflussen kann und dies auch tut, ist dies ein großes Versäumnis. So zum Beispiel ist, im Gegensatz zur bestehenden Fülle von Untersuchungen zur Verwendung von Auswahlmethoden in Organisationen [12, 52, 68], unser Wissen über die Anschauungen der Praktiker zu Validität, Reliabilität und negativen Effekten der Methoden bestenfalls rudimentär (siehe auch [53]). De facto stellen die Anschauungen der Praktiker hinsichtlich diverser Auswahlmethoden und -verfahren einen unermesslich großen Bereich dar, in den die Forschung bisher kaum bzw. gar nicht vorgedrungen ist [38]. Nur in Bezug auf die Auswirkungen der Nutzenanalyse wurde dieser Bereich anfänglich betreten, insbesondere wenn es darum ging, Praktiker von der Verwendung validerer Prädiktormethoden zu überzeugen [14, 41]. Grundlegende Fragen werden jedoch nach wie vor nicht angesprochen. Dazu gehören folgende:

- Welche Auswahlmethoden halten HR-Managern für valider und warum?
- Wie wählen Praktiker in Auswahl-situationen in der Praxis zwischen unterschiedlichen Prädiktoren und Quellen der Evidenz für ihre Bewerber aus?

- Wie können wir die maßgeblichsten Forschungsergebnisse am besten formulieren, wenn sie den Praktikern präsentiert werden, um sie davon zu überzeugen, sich für validere, zuverlässigere und fairere Prädiktortechniken zu entscheiden?

Dieses mangelnde Interesse vonseiten der Forschung steht auch im Widerspruch zur bereits seit mehreren Jahrzehnten bestehenden Konzentration der Forschung auf den Entscheidungsfindungsprozess von Recruitern innerhalb des Auswahlprozesses, und zwar in jedem Stadium dieses Prozesses. Inwieweit die Entscheidung der Recruiter auf unterschiedlichen Prädiktortechniken basiert, wurde hingegen von der Forschung kaum beachtet.

Andererseits sollte die zukünftige Forschung auf die Evaluation der Effizienz unterschiedlicher Methoden der Praktikerfortbildung als Teil professioneller CPD-Events abzielen. Heute ist in mehreren Ländern (z. B. in USA, Großbritannien, Australien) ein verpflichtendes CPD ein integraler Bestandteil für die Beibehaltung der Lizenz oder Berufsberechtigung, wie von den relevanten nationalen Berufsverbänden für Psychologen (APA, BPS und APS) festgelegt. Zusätzlich werden in vielen anderen Ländern, in denen es noch kein verpflichtendes CPD-Training gibt, Veranstaltungen durchgeführt, um Praktiker auf den neuesten Stand zu bringen und weitere praktizierende Psychologen auszubilden. In Anbetracht dessen, dass eines unserer Fachgebiete die Evaluation von Trainingseffizienz ist [26], wäre es angebracht, die Effektivität und den Transfer solcher CPD-Veranstaltungen auf die Praxis zu untersuchen, insbesondere im Bereich der Personalauswahl und der Eignungsdiagnostik. CPD ist einer der zentralen Kanäle, die Anschauungen und Arbeitsweisen der Praktiker direkt beeinflussen, und deshalb ist es von zentraler Bedeutung für die diesbezügliche Evaluation der Effektivität alternativer Trainingsinterventionen. Verpflichtendes CPD ist in den meisten Ländern ein relativ neuer Zugang. Bestimmt könnten Psychologen in Zukunft verstärkt in die Validierung dieser Interventionen eingebunden werden. Dies gilt ganz besonders für CPD-Trainingsveranstaltungen im Bereich der Personalauswahlpraxis. Viele dieser Veranstaltungen werden ein erklärtes pragmatisches Element beinhalten, meist in Bezug auf das Anliegen, wie man Praktiker über aktuelle Entwicklungen in der Forschung informieren kann, und wie diese Entwicklungen in der Praxis angewandt werden können. Der Transfer dieser Punkte in die Praxis durch die anwesenden Psychologen ist ein Themenbereich, der unmittelbare, vordergründige Fragen aufwirft. Resultieren CPD-Veranstaltungen in wirklichen Veränderungen in der Praxis, oder sind die Teilnehmer dieser Veranstaltungen nur körperlich anwesend und hören sich die Informationen höflich an, um CPD-Credits zu bekommen? Es ist ganz klar, dass solche Veranstaltungen eine interessante Fallstudie für die Evaluation des Verhältnisses zwischen Forschung und Praxis bieten, wobei Personal- und Organisationspsychologen selbst Gegenstand der Untersuchung sind. Wenn solche Veranstaltungen die Kollegenschaft von Personal-, Betriebs- und Organisationspsychologen nicht davon überzeugen können, ihre Berufspraxis weiter zu entwickeln und auf den neuesten Stand zu bringen, darf es wohl nicht überraschen, dass sich HR- und Linienmanager weiterhin hartnäckig weigern, sich von unseren gemeinsamen Forschungsbestrebungen und von der Verbreitung unserer Schlüsselergebnisse beeinflussen zu lassen.

### Fragen

1. Kennzeichnen Sie das Verhältnis zwischen Forschung und Praxis in der Personalauswahl.
2. Welche Szenarien lassen sich im historischen Verlauf der psychologischen Personalauswahl identifizieren und anhand welcher Beispiele lassen sie sich verdeutlichen?

### Anmerkung

Ich danke David Chan, Gerard Hodgkinson, Filip Lievens, Rob Ployhart, Ann Marie Ryan, Sonja Schinkel, Neal Schmitt für ihre wertvollen Anmerkungen zu einer früheren Version dieses Kapitels.

### Literatur

1. Anderson, N. (1998). The people make the paradigm. *Journal of Organizational Behaviour*, *19*, 323–328.
2. Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment*, *11*, 121–136.
3. Anderson, N., & Ones, D. S. (2003). The construct validity of three entry level personality inventories used in the UK: Cautionary findings from a multiple-inventory investigation. *European Journal of Personality*, *17*, 39–66.
4. Anderson, N., Herriot, P., & Hodgkinson, G. P. (2001). The practitioner-researcher divide in Industrial, Work and Organizational (IWO) psychology: Where are we now, and where do we go from here? *Journal of Occupational and Organizational Psychology*, *74*, 391–411.
5. Anderson, N., Lievens, F., Dam, K. v., & Ryan, A. M. (2004). Future perspectives on employee selection: Key directions for future research and practice. *Applied Psychology: An International Review*, *53*, 487–501.
6. Arvey, R. D. (1979a). *Fairness in selecting employees*. Reading: Addison-Wesley.
7. Arvey, R. D. (1979b). Unfair discrimination in the employment interview: Legal and psychological aspects. *Psychological Bulletin*, *86*, 739–765.
8. Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology*, *35*, 281–322.
9. Barber, D. (1973). *Basic personnel procedures*. London: Institute of Personnel Management.
10. Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26.
11. 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*, 9–29.
12. Bartram, D., Lindley, P. A., Marshall, L., & Foster, J. (1995). The recruitment and selection of young people by small businesses. *Journal of Occupational and Organizational Psychology*, *68*, 339–358.
13. Borman, W. C., Hansen, M., & Hedge, J. W. (1997). Personnel selection. *Annual Review of Psychology*, *48*, 299–337.
14. Boudreau, J. W., Sturman, M. C., & Judge, T. A. (1997). Utility analysis: What are the black boxes, and do they affect decisions? In N. Anderson & P. Herriot (Hrsg.), *International handbook of selection and assessment* (S. 303–321). Chichester: Wiley.

15. Cascio, W. F. (1993). Assessing the utility of selection decisions: Theoretical and practical considerations. In N. Schmitt, W. C. Borman, & Associates (Hrsg.), *Personnel selection in organizations* (S. 310–340). San Francisco: Jossey-Bass.
16. Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal means and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods, 1*, 421–483.
17. Costa, P. T., Jr., & McCrae, R. R. (1988). From catalogue to classification: Murray's needs and the five factor model. *Journal of Personality and Social Psychology, 55*, 258–265.
18. Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI) professional manual*. Odessa: Psychological Assessment Resources.
19. De Fruyt, F., & Salgado, J. F. (2003). Applied personality psychology: Lessons learned from the IWO field. *European Journal of Personality, 17*, 123–131.
20. Digman, J. M. (1990). Personality structure: Emergence of the Five Factor Model. *Annual Review of Psychology, 41*, 417–440.
21. Dipboye, R. L. (1997). Structured selection interviews: Why do they work? Why are they underutilized? In N. Anderson & P. Herriot (Hrsg.), *International handbook of selection and assessment* (S. 455–473). Chichester: Wiley.
22. Dunnette, M. D. (1990). Blending the science and practice of industrial and organizational psychology: Where are we now and where are we going? In M. D. Dunnette & L. M. Hough (Hrsg.), *Handbook of industrial and organizational psychology* (2. Aufl., Bd. 1, S. 1–37). Palo Alto: Consulting Psychologists Press.
23. Eder, R. W., & Harris, M. M. (1999). Employment interview research: Historical update and introduction. In R. W. Eder & M. M. Harris (Hrsg.), *The employment interview handbook* (S. 1–27). Thousand Oaks: Sage.
24. Ferguson, E., Payne, T., & Anderson, N. (1994). Occupational personality assessment: An evaluation of the psychometric properties of the Occupational Personality Questionnaire (OPQ). *Personality and Individual Differences, 17*, 217–225.
25. Goldstein, H. W., Zedeck, S., & Goldstein, I. L. (2002). Is this your final answer? *Human Performance, 15*, 123–142.
26. Goldstein, I. L. (1997). Interrelationships between the foundations for selection and training systems. In N. Anderson & P. Herriot (Hrsg.), *International handbook of selection and assessment* (S. 529–542). Chichester: Wiley.
27. Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*, 135–164.
28. Herriot, P. (1989). Selection as a social process. In M. Smith & I. T. Robertson (Hrsg.), *Advances in staff selection* (S. 171–187). Chichester: Wiley.
29. Herriot, P., & Anderson, N. (1997). Selecting for change: How will personnel and selection psychology survive? In N. Anderson & P. Herriot (Hrsg.), *International handbook of selection and assessment* (S. 1–34). Chichester: Wiley.
30. Highhouse, S. (2002). Assessing the candidate as a whole: A historical and critical analysis of individual psychological assessment for personnel decision making. *Personnel Psychology, 55*, 363–396.
31. Hodgkinson, G. P., & Payne, R. L. (1998). Graduate selection in three European Countries. *Journal of Occupational and Organizational Psychology, 71*, 359–365.
32. Hodgkinson, G. P., Herriot, P., & Anderson, N. (2001). Re-aligning the stakeholders in management research: Lessons from industrial, work and organizational psychology. *British Journal of Management, 12*, 41–48.
33. Hogan, R. T., & Roberts, B. W. (2001). Introduction: Personality and industrial-organizational psychology. In B. W. Roberts & R. T. Hogan (Hrsg.), *Personality psychology in the workplace* (S. 3–16). Washington: American Psychological Association.

34. Hough, L. M. (2001). I/O owes its advances to personality. In B. W. Roberts & R. Hogan (Hrsg.), *Personality psychology in the workplace* (S. 19–44). Washington: American Psychological Association.
35. Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*, 459–473.
36. Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72–98.
37. Hyatt, D., Cropanzano, R., Funder, L. A., Levy, P., Ruddy, T. M., Vandaveer, V., et al. (1997). Bridging the gap between academics and practice: Suggestions from the field. *The Industrial-Organizational Psychologist, 35*(1), 29–32.
38. Johns, G. (1993). Constraints on the adoption of psychology-based personnel practices: Lessons from organizational innovation. *Personnel Psychology, 46*, 569–592.
39. Krause, D. E. (2011). *Trends in der internationalen Personalauswahl*. Göttingen: Hogrefe.
40. Kuhn, T. S. (1970). *The structure of scientific revolutions* (2. Aufl.). Chicago: University of Chicago Press.
41. Latham, G. P., & Whyte, G. (1994). The futility of utility analysis. *Personnel Psychology, 47*, 31–46.
42. Levy-Leboyer, C. (1988). Success and failure in applying psychology. *American Psychologist, 43*, 779–785.
43. Lievens, F., & Harris, M. M. (2003). Research on Internet recruitment and testing: Current status and future directions. In C. L. Cooper & I. T. Robertson (Hrsg.), *International review of industrial and organizational psychology* (S. 131–165). Chichester: Wiley.
44. Lievens, F., Dam, K. van, & Anderson, N. (2002). Recent trends and challenges in personnel selection. *Personnel Review, 31*, 580–601.
45. Mayfield, E. C. (1964). The selection interview: A re-evaluation of published research. *Personnel Psychology, 17*, 239–260.
46. McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52*, 509–516.
47. 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.
48. Murphy, K. R. (1996). Individual differences and behaviour in organizations. Much more than g. In K. R. Murphy (Hrsg.), *Individual differences and behaviour in organizations* (S. 3–30). San Francisco: Jossey-Bass.
49. Ones, D. S., & Anderson, N. (2002). Gender and ethnic group differences on personality scales in selection: Some British data. *Journal of Occupational and Organizational Psychology, 75*, 255–276.
50. Pfeffer, J. (1993). Barriers to the advancement of organizational science: Paradigm development as a dependent variable. *Academy of Management Review, 18*, 599–620.
51. Rice, E. E. (1997). Scenarios: The scientist-practitioner split and the future of psychology. *American Psychologist, 52*, 1173–1181.
52. Robertson, I. T., & Makin, P. J. (1986). Management selection in Britain: A survey and critique. *Journal of Occupational Psychology, 59*, 45–57.
53. Robertson, I. T., & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology, 74*, 441–472.
54. Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management, 26*, 565–606.
55. Rynes, S. L., Bertunek, J. M., & Draft, R. L. (2001). Across the great divide: Knowledge creation and transfer between practitioners and academics. *Academy of Management Journal, 44*, 340–355.



56. Rynes, S. L., Brown, K. G., & Colbert, A. E. (2002). Seven common misconceptions about human resource practices: Research findings versus practitioner beliefs. *Academy of Management Executive*, *16*, 92–102.
57. Sackett, P. R. (1994). The content and process of the research enterprise within industrial and organizational psychology. Nashville, presidential address to the society for industrial and organizational psychology conference.
58. Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, *82*, 30–43.
59. Salgado, J. F. (2001). Some landmarks of 100 years of scientific personnel selection at the beginning of the new century. *International Journal of Selection and Assessment*, *9*, 3–8.
60. Salgado, J. F., & Anderson, N. (2002). Cognitive and GMA testing in the European Community: Issues and evidence. *Human Performance*, *15*, 75–96.
61. Salgado, J. F., & Anderson, N. (2003). Validity generalization of GMA tests across countries in the European Community. *European Journal of Work and Organizational Psychology*, *12*, 1–17.
62. 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.
63. Salgado, J. F., Viswesvaran, C., & Ones, D. S. (2001). Predictors used for personnel selection: An overview of constructs, methods and techniques. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Hrsg.), *Handbook of industrial, work and organizational psychology* (S. 165–199). London: Sage.
64. Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities as predictors of work behaviours: A European contribution and comparison with American findings. *Personnel Psychology*, *56*, 573–605.
65. Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology*, *88*, 1068–1081.
66. Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection research methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.
67. Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, *37*, 402–422.
68. Shackleton, V. J., & Newell, S. (1994). European management selection methods: A comparison of five countries. *International Journal of Selection and Assessment*, *2*, 91–102.
69. Tenopyr, M. L. (2002). Theory versus reality: Evaluation of g in the workplace. *Human Performance*, *15*, 107–122.
70. Torrington, D., & Chapman, J. (1979). *Personnel management*. London: Prentice-Hall International.
71. Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin*, *63*, 100–116.
72. Viswesvaran, C., Sinangil, H. K., Ones, D. S., & Anderson, N. (2001). Introduction to the handbook and volume 1 – Personnel psychology: Where we have been, where we are, (and where we could be). In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Hrsg.), *Handbook of industrial, work and organizational psychology* (S. 1–9). London: Sage.
73. Wagner, R. (1949). The employment interview: A critical review. *Personnel Psychology*, *2*, 17–46.
74. Wright, O. R. (1969). Summary or research on the selection interview since 1964. *Personnel Psychology*, *22*, 391–413.
75. Yang, K., & Bond, M. H. (1990). Exploring implicit personality theories with indigenous or imported constructs: The Chinese case. *Journal of Personality and Social Psychology*, *58*, 1087–1095.

## Zusammenfassung

In diesem Kapitel wird der Frage nachgegangen, was unter beruflicher Eignung zu verstehen ist und wie berufliche Eignung erfasst werden kann. Neben einer Definition beruflicher Eignung werden wichtige Eignungsmerkmale vorgestellt. Die Bedeutung der Passung von Person und Organisation wird in einem eigenen Abschnitt diskutiert. Neben der Betrachtung *was* berufliche Eignung ausmacht, liegt im zweiten Teil des Kapitels der Fokus auf der Frage *wie* diese erfasst werden kann. So werden Methoden der beruflichen Eignungsdiagnostik vorgestellt. Außerdem stellen wir mit der Balanced Scorecard eignungsdiagnostischer Verfahren eine neue Idee für eine integrative Bewertungssystematik vor.

## Inhaltsverzeichnis

2.1	Berufliche Eignung	32
2.1.1	Definition beruflicher Eignung	32
2.1.2	Merkmale beruflicher Eignung	32
2.2	Diagnostik der beruflichen Eignung	35
2.2.1	Definition berufliche Eignungsdiagnostik	36
2.2.2	Aufgaben und Anwendungsgebiete beruflicher Eignungsdiagnostik	36
2.2.3	Nutzen beruflicher Eignungsdiagnostik	38

C. Palmer (✉) · M. Kersting  
Institut für Psychologie und Sportwissenschaft  
Justus-Liebig-Universität Gießen, Gießen, Deutschland  
E-Mail: Carolin.Palmer@psychol.uni-giessen.de

M. Kersting  
E-Mail: Martin.Kersting@psychol.uni-giessen.de



2.3	Passung von Person und Organisation .....	41
2.4	Methoden der Eignungsdiagnostik .....	44
2.5	Balanced Scorecard eignungsdiagnostischer Verfahren. ....	48
2.6	Fazit .....	53
	Literatur .....	54

---

## 2.1 Berufliche Eignung

„Getestet und für gut befunden.“ Dieses Urteil wünschen sich beide sowohl die Bewerber als auch die auswählenden Organisationen im Bewerbungsprozess. Doch welche Merkmale sind seitens der Bewerber/-innen überhaupt zu prüfen? Und wie hoch (oder auch niedrig) sollten diese Merkmale ausgeprägt sein? Nach einer Definition des Begriffs der beruflichen Eignung wird diesen Fragen im ersten Abschnitt dieses Kapitels nachgegangen.

### 2.1.1 Definition beruflicher Eignung

Auf Basis der DIN-Norm 33430 [10], die als Prozessnorm die Anforderungen an berufsbezogene Eignungsdiagnostik beschreibt, wird berufliche Eignung wie folgt definiert: „Grad der Ausprägung in dem eine Person über die Eignungsmerkmale verfügt, die Voraussetzung für die jeweils geforderte berufliche Leistungshöhe sind und zur Zufriedenheit mit dem zu besetzenden Arbeitsplatz, dem Aufgabenfeld, der Ausbildung bzw. dem Studium oder dem Beruf beitragen“ (S. 5.). Berufliche Eignung ist folglich aus dem Zusammenhang von Personmerkmalen mit Anforderungen und Kriterien beruflichen Erfolgs abzuleiten. Im Bestreben diesen Zusammenhang zwischen Personmerkmalen und Anforderungen in der beruflichen Tätigkeit bestmöglich beschreiben, erklären und vorhersagen zu können, integriert die berufliche Eignungsdiagnostik u. a. die Ergebnisse der differenziellen bzw. Persönlichkeits-Psychologie, der Sozialpsychologie, der psychologischen Diagnostik sowie der Arbeits- und Organisationspsychologie. Die Identifikation und Betrachtung von Eignungsmerkmalen beschränkt sich jedoch nicht auf aktuelle Ausprägungen von Personmerkmalen, sondern umfasst auch das entsprechende Entwicklungspotenzial sowie den Einfluss situativer Bedingungen auf das gezeigte Verhalten.

### 2.1.2 Merkmale beruflicher Eignung

Unter Eignungsmerkmalen sind keinesfalls nur „harte Faktoren“ wie Eigenschaften (z. B. allgemeine Intelligenz), Fähigkeiten (z. B. Sprachverständnis), Fertigkeiten (z. B. Gutachten schreiben können) oder Kenntnisse (z. B. rechtliche Bestimmungen) subsumiert. Mit der Berücksichtigung von Interessen, Motiven, Bedürfnissen und Wünschen der Personen werden ebenso „weiche Faktoren“ (z. B. emotionale Intelligenz) in die

Bewertung der beruflichen Eignung einbezogen. Zudem wird auf Seiten der beruflichen Erfolgskriterien nicht allein die Kategorie der Leistungsmaße betrachtet, sondern werden zusätzlich Arbeitszufriedenheit, psychische und physische Gesundheit und andere Aspekte des Wohlbefindens sowie Werte des persönlichen Wachstums und Erfolgserlebens berücksichtigt [53, 54]. Damit werden beispielsweise auch das Gefühl, gefordert zu sein, oder individuelles Sinnerleben sowie andere subjektive Kriterien, zum Beispiel gesellschaftlichen Status zu erlangen, als Kriterien betrachtet, die beruflichen Erfolg markieren.

Um den Zusammenhang zwischen Personmerkmalen und beruflichen Anforderungen zu verdeutlichen, bietet sich die in Abb. 2.1 dargestellte Zuordnung an.

Berufliche Tätigkeiten lassen sich hinsichtlich dreier Merkmalsebenen beschreiben. Zunächst sind die grundlegenden Anforderungen zu nennen. Jenen Merkmalen, die tätigkeitsspezifische Voraussetzungen zur erfolgreichen Ausübung einer Tätigkeit darstellen, stehen aufseiten der Person Fähigkeiten, Fertigkeiten und Kenntnisse gegenüber. Die Anforderungen einer Tätigkeit können in einer Arbeits- und Anforderungsanalyse spezifiziert und dann in Personmerkmale übersetzt werden (s. Kapitel zur Anforderungsanalyse von Nerdinger in diesem Band). Die Diagnose von Fähigkeiten, Fertigkeiten und Kenntnissen des Bewerbers wiederum stellt den Fokus der psychologischen Eignungsdiagnostik dar. Die Frage, inwiefern Bewerber oder auch Stelleninhaber die Anforderungen einer Tätigkeit erfüllen, kann anhand eignungsdiagnostischer Methoden beantwortet werden. Ein Überblick über diese Verfahren wird in diesem Kapitel vorgestellt. Ausgewählten Ansätzen, wie etwa Testverfahren zu kognitiven Fähigkeiten bzw. personenbezogenen Attributen oder simulativen Verfahren wie etwa Assessment Centern, sind in diesem Band eigene Kapitel gewidmet (s. Kapitel von Kersting und Palmer; Hossiep und Weiss; Krause). Mit der ersten Ebene der Anforderungen werden fachliche und überfachliche Anforderungen formuliert und damit die Voraussetzungen für beruflichen Erfolg definiert.

Auf der zweiten Ebene zur Beschreibung von Tätigkeiten rückt nun stärker die Person in den Fokus: Welche individuellen Interessen, Bedürfnisse und Werthaltungen hat sie und inwiefern lassen sich diese bei Ausübung der spezifischen Tätigkeit realisieren? Das Befriedigungspotenzial einer Arbeitsstelle, welches über Arbeits- und Organisationsanalysen ermittelt werden kann, wirkt sich jedoch nicht nur auf die Arbeitszufriedenheit aus, sondern beeinflusst indirekt auch Leistungsmaße. Die Passung individueller Interessen

Tätigkeit	Person
tätigkeitsspezifische Anforderungen	Fähigkeiten, Fertigkeiten und Kenntnisse
Befriedigungs- und Motivierungspotenzial	Interessen, Bedürfnisse, Werthaltungen
tätigkeitsübergreifende Anforderungen und Veränderung	generell erfolgsrelevante Eigenschaften, Entwicklungspotenzial

**Abb. 2.1** Zusammenhang zwischen Personmerkmalen und Tätigkeit. (Basierend auf Schuler [52])

und Motive, Bedürfnisse, Wünsche und Werthaltungen der Person mit dem Tätigkeitsumfeld erhöht das arbeitsbezogene Engagement, kann darüber hinaus aber auch das sogenannte Umfeld bezogene Engagement (in der englischsprachigen Literatur: organizational citizenship behavior, OCB; [40]) steigern. Damit wird ein Verhalten bezeichnet, das zur Produktivität einer Organisation beiträgt, dessen Ausführung aber, da formal nicht vorgeschrieben, im Ermessen der Mitarbeiter liegt. Hiermit wird also Verhalten erfasst, das einen freiwilligen Beitrag zum Unternehmenserfolg und vor allem zu einem positiven Organisationsklima leistet. Informationen über die Interessen, Bedürfnisse, Wünsche und Werthaltungen einer Person gewinnt die Eignungsdiagnostik vornehmlich durch Fragebogen, Interessenstests oder im Rahmen von Interviews; Verfahren, die ebenfalls in das weiter unten vorgestellte Rahmenmodell eignungsdiagnostischer Methoden integriert sind.

Inwiefern Personen Potenzial zur weiteren beruflichen und persönlichen Entwicklung haben bzw. sich an veränderte Arbeitsanforderungen anpassen können, ist Gegenstand der dritten Ebene zur Beschreibung beruflicher Tätigkeiten. Hier werden Personmerkmale nicht auf ihre Passung zu den tätigkeitsspezifischen Anforderungen bewertet, sondern allgemein (berufs-)erfolgsrelevante Merkmale fokussiert. Die DIN 33430 [10] definiert Potenzial als die „Fähigkeit einer Person, ihr bislang nicht vertraute Aufgaben zu bewältigen und Kompetenzen zu entwickeln“ (S. 7). Eine bislang nicht vertraute Aufgabe wäre z. B. für eine Sachbearbeiterin die Übernahme von Führungsaufgaben – häufig werden derartige neue Aufgaben dann im Rahmen der Eignungsdiagnostik mit Hilfe von speziellen Übungen im Development-Center simuliert. Auch in dem Fall, dass eine Person die gleiche Leistung die sie bislang schon erbracht hat, auf einem höheren Niveau erbringt, spricht man von Potenzial. Ein Beispiel wäre die Beherrschung und Anwendung einer Fremdsprache auf einem höheren Niveau (z. B. die Entwicklung vom Schulenglisch zu verhandlungssicherem „business English“). Schließlich gibt es auch bestimmte Personmerkmale, die man allgemein als Potenzialindikatoren ansieht. Als valider Prädiktor über verschiedene Erfolgskriterien und Berufsgruppen hinweg hat sich beispielsweise das Persönlichkeitsmerkmal Gewissenhaftigkeit [3, 38] erwiesen. Unter Gewissenhaftigkeit werden unter anderem Eigenschaften wie Pflichtbewusstsein, Selbstdisziplin und Leistungsstreben zusammengefasst [41]. Damit umfasst dieser Merkmalsbereich Eigenschaften, die verlässliche und engagierte Personen auszeichnen. Neben Gewissenhaftigkeit als Persönlichkeitsmerkmal kommt vor allem auch der allgemeinen Intelligenz eine tätigkeitsübergreifende Bedeutung zu. Wie Lang et al. [33] sowie Salgado et al. [43] zeigen, steigt die Vorhersagekraft von allgemeiner Intelligenz für beruflichen Erfolg mit zunehmendem Komplexitätsgrad der Tätigkeit an. Durch den Einsatz von Potenzialanalysen bzw. auf Basis allgemeiner Fähigkeitstests sowie Leistungs- und Persönlichkeitsinventare können Einschätzungen getroffen werden, inwiefern sich eine Person auf veränderte Herausforderungen einlassen könnte und möchte. Ob die fragliche Tätigkeit überhaupt Möglichkeiten zur Entwicklung bietet (oder gar erfordert) oder ob diese Entwicklungsperspektiven nur durch eine berufliche Veränderung realisiert werden können, lässt sich allenfalls schätzen. Zu erwartende technologische Fortschritte,

Branchen- oder Marktentwicklungen bzw. strategische Entscheidungen erlauben jedoch eine Prognose des Veränderungsgrades einer Tätigkeit.

Die Merkmale beruflicher Eignung lassen sich also danach unterscheiden, ob sie berufs- oder gar tätigkeitsspezifische Anforderungen zu bewältigen helfen oder aber tätigkeitsübergreifende, generell berufserfolgsrelevante Eigenschaften darstellen. In Abhängigkeit der diagnostischen Fragestellung werden folglich auch unterschiedliche personbezogene Merkmale zu betrachten sein. So wird bei der Besetzung einer konkreten Stelle wohl jene Person aus der Gruppe an Bewerbern ausgewählt werden, deren Fähigkeiten, Fertigkeiten, Kenntnisse und Eigenschaften am besten zu den Anforderungen der Tätigkeit passen. Für eine Ausbildungs-, Berufs- oder Karriereberatung hingegen liegt der Fokus wohl besser auf allgemeineren Fähigkeiten und der Herausarbeitung individueller Motive, Wünsche und Erwartungen um Empfehlungen für künftige Entwicklungen und Tätigkeitsfelder aussprechen zu können.

---

## 2.2 Diagnostik der beruflichen Eignung

Bereits vor dem Antritt der ersten Arbeitsstelle sind wichtige Weichen für das spätere (Berufs-)Leben gestellt worden. Zunächst muss die grundlegende Entscheidung getroffen werden, überhaupt am Arbeitsmarkt teilzunehmen. Das dies keinesfalls selbstverständlich ist, zeigen aktuelle Statistiken: Im Jahr 2012 etwa waren 14,9 % der deutschen Bevölkerung im Alter von 20 bis 64 Jahren ohne generellen Arbeitswunsch, entsprechend weder erwerbstätig noch erwerbssuchend und damit der Gruppe der Nichterwerbspersonen zugeordnet [57]. Vom generellen Arbeitswunsch hin zu einem konkreten Berufswunsch ist es jedoch nochmals ein ganz eigener Weg, wie viele Jugendliche und junge Erwachsene zu berichten wissen und zahlreiche Ausbildungs-, Studienberatungs- und Berufsberatungsstellen dokumentieren. Den Traumberuf vor Augen, gilt es, die eigenen Fähigkeiten mit den zu erwartenden Anforderungen abzugleichen und eventuelle illusorische Vorstellungen über die Ausübung des Berufs und die damit verbundenen Konsequenzen (etwa das Gehaltsniveau, aber auch physische und psychische Belastungen) zu korrigieren. Wie bedeutsam eine frühe Kenntnis der eigenen beruflichen Eignung ist, zeigen Befunde aus der Forschung zum Studienerfolg. Bereits im Studium lässt sich die Anzahl an Studienabbrechern durch eine psychometrisch fundierte Analyse der Fähigkeiten und Interessen im Rahmen sogenannter Studienberatungstests (in ihrer modernen Form als Online-Self-Assessments bekannt, s. Beitrag von Ott, Ulfert und Kersting in diesem Band) reduzieren [4].

Vergleichbare Effekte zeigen sich auch im Unternehmenskontext: Passen die beruflichen Bedingungen nicht (oder nicht mehr) zur Person, sind vielfältige Probleme zu beobachten. Gründe für mangelnde Passung können unter anderem eine unzulängliche Information über die Anforderungen und Inhalte von Tätigkeiten oder ein aufgrund persönlicher Lebensumstände verändertes Leistungsniveau sein. Altersbedingt, durch Krankheit oder Übernahme familiärer Aufgaben verändern sich berufsrelevante

Fertigkeiten, die Bedeutung von Arbeit und einhergehend auch die Leistungsmotivation. Während die Ausübung einer als unpassend empfundenen Tätigkeit oder sogar schon allein der Gedanke daran auf Mitarbeiterseite zu Unbehagen oder gar Erkrankung bzw. nicht integrem Verhalten führen kann, sieht sich das Unternehmen einer geringeren Leistungsbereitschaft seitens der Mitarbeiter, folglich Produktivitätseinbußen und erheblichen Folgekosten gegenüber. Kündigen Mitarbeiter zuvor, lässt sich zwar beiderseits Schaden vermeiden; zufriedenstellend dürfte diese Entscheidung jedoch weder für die Mitarbeiter noch für die Arbeitgeber sein.

Die Passung von Person und Organisation wird an späterer Stelle in diesem Kapitel wieder aufgegriffen, in diesem Abschnitt soll es vornehmlich um die Diagnostik von Eignungsmerkmalen gehen. Idealerweise findet ein Abgleich der individuellen Fähigkeiten und Wünsche mit Tätigkeitsmerkmalen über die gesamte berufliche Lebensspanne hinweg statt. Damit können etwaige Veränderungen aufseiten der Person oder auch seitens der Tätigkeit nicht nur zeitnah identifiziert werden, sondern auch früh genug Maßnahmen eingeleitet werden, um die Passung zwischen Person und Tätigkeit wiederherzustellen.

### **2.2.1 Definition berufliche Eignungsdiagnostik**

Berufliche Eignungsdiagnostik befasst sich mit dem Zusammenhang zwischen Personenmerkmalen, Tätigkeitsmerkmalen und beruflichem Erfolg sowie beruflicher Zufriedenheit. Neben der Beschreibung dieser Beziehungen liegt das Hauptaugenmerk der beruflichen Eignungsdiagnostik auf der Messung relevanter Personenmerkmale und der Optimierung der hierfür eingesetzten Instrumente und Vorgehensweisen zur bestmöglichen Vorhersage beruflicher Erfolgskriterien. Qualitativ hochwertige Eignungsdiagnostik ist förderorientiert und gibt sowohl wertvolle Hinweise für die kurzfristige Gestaltung des so genannten „on-boarding“ als auch für die langfristige Personalentwicklung.

### **2.2.2 Aufgaben und Anwendungsgebiete beruflicher Eignungsdiagnostik**

Die wesentlichen Aufgaben beruflicher Eignungsdiagnostik liegen in der Identifikation und Beschreibung beruflicher Anforderungen (Was ist zu tun?), deren Übersetzung in Personenmerkmale (Was muss man dafür können und wollen? Wie sollte man sein bzw. sich zu verhalten wissen?) und der Konstruktion und Auswahl geeigneter diagnostischer Verfahren (Wie können relevante Personenmerkmale erfasst werden? Vgl. Abschn. 2.4). Mit der Ableitung von Vorschlägen für individuelle Fördermaßnahmen stellt die Eignungsdiagnostik das Fundament der Personalentwicklung dar.

Sind die Anforderungen einerseits und die Kompetenzen und Wünsche andererseits bekannt, können Aussagen über die Passung von Tätigkeit und Person getroffen

werden. Gerade in dynamischen Arbeitsumwelten und bei vielseitig begabten und interessierten Personen kommt hierbei der Berücksichtigung sich verändernder Bedingungen eine große Bedeutung zu. So sind Unternehmen in der Regel schlecht beraten, wenn sie eignungsdiagnostische Analysen nur im Moment der Einstellung vornehmen und spätere Entwicklungen nicht berücksichtigen. Arbeitgeber, die Bewerbern aktive Personalentwicklung versprechen, sollten auch eine eignungsdiagnostische bzw. personalpsychologische Begleitung des neu gewonnenen Mitarbeiters über die Dauer seiner Unternehmenszugehörigkeit hinweg umsetzen, so etwa durch die Bereitstellung geeigneter Instrumente.

Wissenschaftlich fundiert und empirisch abgesichert kann Eignungsdiagnostik den Nutzen für beide Seiten – Person und Unternehmen – nachhaltig steigern. Unter Berücksichtigung strategischer, gesellschaftlicher, juristischer oder technologischer Entwicklungen als Rahmenbedingungen der Arbeit ist der Fokus der Eignungsdiagnostik in der Ableitung des Entwicklungspotenzials von Mitarbeitern nicht nur auf fachliche Kompetenzentwicklung und harte Karrieremaßstäbe (Gehalt, hierarchischer Aufstieg etc.) zu legen, sondern auch die Berücksichtigung sich ändernder privater Bedürfnisse und Wünsche (Alter, Vereinbarkeit von Familie und Beruf, Freizeitorientierung) angezeigt.

Aus diesem Verständnis heraus ergeben sich zwei Kernaufgaben für die psychologische Eignungsdiagnostik: die Unterstützung in der Auswahl geeigneter Mitarbeiter sowie die Begleitung von Veränderungsprozessen. Entsprechend wird in der eignungsdiagnostischen Literatur nach Selektions- und Modifikationsdiagnostik unterschieden. Selektionsdiagnostik umfasst die Prüfung von Personen im Hinblick darauf, ob und inwiefern sie den Anforderungen einer Tätigkeit oder Stelle gerecht werden (Personenselektion). Zugleich wird unter Selektionsdiagnostik auch die Suche nach einer zur Person passenden Aufgabe oder Situation bzw. Umgebung (Bedingungsselektion) gefasst. Die Modifikationsdiagnostik hingegen klärt weniger Fragen der aktuellen Passung, sondern unterstützt dabei, individuelles Verhalten oder spezifische Bedingungen zu verändern (Personen- oder Bedingungsmodifikation). Hier geht Eignungsdiagnostik also über die reine Personalauswahl hinaus und wird um Themen der Personal-, Tätigkeits- und Organisationsentwicklung erweitert.

Für die Erfüllung der eignungsdiagnostischen Aufgaben kann auf eine große Auswahl an Verfahren und Methoden zurückgegriffen werden. Aber nicht alle halten auch einer wissenschaftlich fundierten Prüfung stand. Obwohl beispielsweise die Grafologie, der (vermeintlich mögliche) Rückschluss aus Handschriften auf Persönlichkeitseigenschaften, in der Praxis noch immer eingesetzt wird, ist ihre Prognosekraft zur Beschreibung von Persönlichkeit oder gar Vorhersage von beruflichem Erfolg gleich Null [46]. Interessanterweise erkennen Bewerber die mangelnde Aussagekraft der Grafologie und attestieren ihr nur geringe wissenschaftliche Fundierung, niedrige Augenscheinvalidität und das Fehlen der Möglichkeit, sein wahres Potenzial zu zeigen [1, 2].

Die Eignungsdiagnostik an der Schnittstelle aus differenzieller Psychologie, psychologischer Diagnostik und Arbeits- und Organisationspsychologie leistet den wertvollen Beitrag, zur Diagnostik eingesetzte Verfahren hinsichtlich verschiedener Gütekriterien zu

überprüfen, zu bewerten und gegebenenfalls zu optimieren. Zur Personalauswahl und -entwicklung eingesetzte Instrumente sollten nicht nur objektive, d. h. von den durchführenden Personen unabhängige, Ergebnisse liefern, sondern auch zugleich möglichst messgenau (oder: messfehlerfrei, reliabel) die intendierten Eigenschaften, Fähigkeiten oder Einstellungen erfassen (Konstruktvalidität), mittels derer eine möglichst genaue Vorhersage des relevanten Verhaltens bzw. relevanter beruflicher Erfolgskriterien (Kriteriumsvalidität) erfolgen kann.

Damit ergeben sich als Anwendungsgebiete der Eignungsdiagnostik u. a. die folgenden Bereiche:

- Personalauswahl
- Personalentwicklung
- Studiums-, Ausbildungs-, Karriere- bzw. Laufbahnberatung
- Coaching
- Forschung zum Zusammenhang von Personmerkmalen und beruflichen Erfolgskriterien und damit verbunden
- Entwicklung und Bewertung eignungsdiagnostischer Verfahren unter psychometrischen Gesichtspunkten und praxisrelevanten Kriterien.

### 2.2.3 Nutzen beruflicher Eignungsdiagnostik

Zwei wesentliche Vorbedingungen psychologischer Eignungsdiagnostik wurden bereits angesprochen: 1) Leistungsunterschiede sind auf unterschiedliche Ausprägungen bestimmter Personmerkmale rückführbar und 2) jene Personmerkmale sind über die Zeit hinweg ausreichend stabil, um Aussagen zu beruflicher Eignung über den aktuellen Beobachtungsmoment hinaus treffen zu können. Beide Annahmen haben sich in umfassenden Validierungsstudien bestätigen lassen. In einer der ersten Meta-Analysen zur Vorhersagekraft verschiedener Auswahlverfahren ergab sich beispielsweise ein Zusammenhang von  $r = ,51$  zwischen Tests zur Erfassung allgemeiner Intelligenz und beruflichem Erfolg in Tätigkeiten mittlerer Komplexität [46]. Die Stärke dieses Zusammenhangs wurde später in einer weiteren Meta-Analyse unter Einbezug europäischer Studien repliziert ( $r = ,53$ ; [43]). Wird hingegen auf komplexere Aufgaben fokussiert, so zeigt sich ein noch stärkerer Einfluss der Intelligenz auf den Berufserfolg ( $r = ,64$ ; [43]). Die Stabilität dieses folglich bedeutsamen Merkmals offenbart sich in einer Korrelation von  $r = ,66$  (attenuationskorrigiert:  $,73$ ) zwischen Intelligenzwerten, die im Abstand von fast 70 Lebensjahren erfasst wurden (erste Messung im Alter von 11 Jahren, zweite Messung im Alter von 80 Jahren; [9]). Für den Einfluss von Persönlichkeit auf den Berufserfolg finden sich ebenfalls generalisierbare Zusammenhänge in zwar geringerer, aber immer noch beachtenswerter Höhe. Gewissenhaftigkeit etwa korreliert mit beruflichem Erfolg in Abhängigkeit der eingesetzten Instrumente zwischen  $\rho = ,22$  [3] und  $\rho = ,40$  [44]. Die Stabilität von Persönlichkeitseigenschaften variiert zwischen  $,50$  und  $,70$  über mehrere Jahrzehnte hinweg [42, 60].



Hierbei handelt es sich allerdings um Aussagen zur differenziellen Stabilität, das bedeutet, dass sich die Rangplätze von Personen bezüglich ihrer individuellen Ausprägung von Personmerkmalen im Vergleich zu einer bestimmten Gruppe über die Jahre hinweg nicht verändern. Das Ausprägungsniveau einzelner Eigenschaften kann sich im Laufe des Lebens damit aber verändern. Absolute Stabilität gemäß Caspi und Roberts [8] ist weder für kognitive noch für non-kognitive Eigenschaften gegeben.

Die Zusammenhänge zwischen persönlichen Kompetenzen und der erfolgreichen Bewältigung beruflicher Anforderungen sind demnach hinlänglich bekannt. Leistungsunterschiede lassen sich also in gewissem Maße auf Ausprägungen individueller Kompetenzen zurückführen. Und genau auf Basis dieser beobachtbaren Leistungsunterschiede offenbart sich der Nutzen der Eignungsdiagnostik, wie es sich beispielsweise in der Arbeit von Hunter et al. [19] zeigt. Die Autoren verglichen bei Tätigkeiten mittlerer Komplexität (z. B. Beschwerdemanagement) die Leistung der Top 1 % der Arbeiter mit der Leistung der 1 % schwächsten Arbeiter und errechneten ein Leistungsverhältnis von 1:12,33. In anderen Worten: die leistungsstärksten Personen zeigen 185 % der durchschnittlichen Leistung, die leistungsschwächsten Personen gerade einmal 15 %. Für komplexe Tätigkeiten (z. B. Rechtsanwälte, Ärzte) wurde der Beitrag der leistungsstärksten Personen gar auf 227 % der durchschnittlichen Leistung geschätzt. Neben der Komplexität dürfte auch die stärkere Autonomie in diesen Berufen für die großen Leistungsspannen verantwortlich sein [53]. Sofern nun das Gehalt als Gegenwert für die Erfüllung der Anforderungen einer Stelle (durchschnittliche Leistung) gewertet wird, lässt sich errechnen, wie viel finanziellen Mehrwert die Einstellung besonders leistungsstarker Bewerber für das Unternehmen erbringt. Für komplexe Tätigkeiten mit einem Durchschnittsgehalt von 100.000 EUR, schafft damit eine gelungene Auswahl einen jährlichen Zusatznutzen von 127.000 EUR. De facto wird der ökonomische Mehrwert noch weit über dieser theoretischen Summe liegen, denn eine Gleichsetzung von Gehalt und unternehmerischem Nutzen, oder treffender: Gewinn, widerspricht der originären Gewinnerzielungsabsicht privatwirtschaftlicher Unternehmen.

Durch Leistungsunterschiede allein ist jedoch noch kein Mehrwert geschaffen. Erst wenn es bei der Selektion von Bewerbern gelingt, die spätere Berufsleistung möglichst genau vorherzusagen, können Unternehmen leistungsstarke Personen identifizieren und einstellen oder befördern. Zur exakten Nutzenberechnung eignungsdiagnostischer Entscheidungen liegen unterschiedliche Modelle vor, die sich im Wesentlichen durch die Differenzierung im Einbezug betriebswirtschaftlicher Stellgrößen (z. B. Verweildauer, Diskontierung, Inflationsrate) unterscheiden. Allen gemein ist aber: Erst der Einsatz von Verfahren, die geeignet sind Leistungsunterschiede aufzudecken (Validität), ist nutzenstiftend, wobei der erzielte Nutzen proportional zur Validität des Verfahrens steigt [26].

Natürlich bringt jedoch das unter eignungsdiagnostischen Gesichtspunkten beste Verfahren nichts, sofern sich nicht überhaupt geeignete Personen unter den Bewerbern finden (Grundquote). Je mehr Bewerbungen vorliegen, desto kleiner ist (bei einer fixen Anzahl zu besetzender Positionen) der relative Anteil derer, die letztlich ausgewählt



werden (Selektionsquote). Eine Erhöhung der Grundquote sowie eine Reduktion der Selektionsquote führen in Kombination mit einem kriterienvaliden Auswahlverfahren zu einer hohen Erfolgsquote in dem Sinne, dass möglichst viele tatsächlich geeignete Bewerber auch ausgewählt werden („selektiver Eignungsquotient“). Wie hoch der Anteil der tatsächlich geeigneten (erfolgreichen) Personen unter allen eingestellten Personen ist, kann für verschiedene Konstellationen von Grund- und Selektionsquoten den Taylor-Russell-Tafeln [58] entnommen werden. Bei einer Selektionsquote von 20 % und einem sehr guten Auswahlverfahren mit einer Kriteriumsvalidität von  $r = ,50$  (was der prä-diktiven Validität allgemeiner Fähigkeitstests entspricht) sind bei einer Grundquote von ebenfalls 20 % beispielsweise später auch 44 % der eingestellten Personen erfolgreich. Beträgt der Anteil der Geeigneten unter allen Bewerbern (Grundquote) bei sonst gleichen Bedingungen 70 % (Selektionsquote 20 %, Kriteriumsvalidität  $r = ,50$ ), werden hingegen 91 % der eingestellten Personen beruflich erfolgreich sein. Eine gezielte Ansprache geeigneter Personen innerhalb des potenziellen Bewerberkreises im Rahmen des Personalmarketings ist folglich Vorbedingung ökonomischer Personalgewinnung [13].

Im Wettbewerb um Budgets ist die Abbildung des finanziellen Nutzens eignungsdiagnostisch basierter Personalentscheidungen sicherlich förderlich. Der tatsächliche Wert fundierter Eignungsdiagnostik umfasst jedoch zusätzlich weitere Aspekte, die sich ebenfalls finanziell auswirken können.

Ergänzend zur Identifikation geeigneter oder gar besonders leistungsstarker Bewerber, wirkt sich eine professionelle, wissenschaftlich fundierte Personalauswahl und -entwicklung auch auf das Image des auswählenden Unternehmens aus [15]. Als Gegenbeispiel wurde bereits die Grafologie erwähnt, wobei auch Körpersprachanalysen oder astrologische Gutachten als Negativbeispiele dienen können. Unternehmen, die um die Bedeutung eines professionellen Erstkontakts zu potenziellen Mitarbeitern wissen, sind folglich gut beraten, den gesamten Auswahlprozess und damit auch den Verfahrenseinsatz an Qualitätsstandards wie der DIN 33430 (DIN 2014) [10] auszurichten.

Neben dem Gewinn auf Seiten der auswählenden Unternehmen wirkt fundierte Eignungsdiagnostik auch individuell auf Bewerber- bzw. Mitarbeiterseite. Sehen sich Bewerber Auswahlverfahren gegenüber, die sie als akzeptabel, angenehm und wertschätzend empfinden, beeinflusst das nicht nur ihr Bild vom auswählenden Unternehmen positiv, sondern führt auch mit höherer Wahrscheinlichkeit zur Annahme eines Stellenangebots [13, 15, 31]. Auswahlprozesse werden dann akzeptiert, wenn nicht nur von den Bewerbern Informationen eingeholt werden, sondern auch Informationen über die zu besetzende Position und das suchende Unternehmen gegeben werden (1. Information), Bewerber den Eindruck haben, die Situation kontrollieren zu können (z. B. indem sie Antworten auf Interviewfragen auch auslassen können; 2. Partizipation/Kontrolle), Klarheit darüber herrscht, wie der Bewertungsprozess abläuft, welche Personen daran beteiligt sind und welche Anforderungen an Bewerber gestellt werden (3. Transparenz) und wenn über den gesamten eignungsdiagnostischen Prozess hinweg klar kommuniziert wird (z. B. auch Ergebnisrückmeldungen gegeben werden; 4. Kommunikation, Feedback). Alle vier Punkte zur Förderung der Akzeptanz von Auswahlprozessen (Information, Partizipation/

Kontrolle, Transparenz, Kommunikation/Feedback) sind im Konzept der sozialen Validität nach Schuler [50] enthalten (s. Kapitel von Kanning in diesem Band). Kersting [25] weist darauf hin, dass sich das Urteil über die Akzeptanz von Auswahlverfahren auch auf das (Test-)Verhalten und damit die Validität der Verfahren auswirken könnte. Inwiefern Bewerber ein konkretes Verfahren akzeptieren, lässt sich beispielsweise anhand des Akzept!-Fragebogens und den vier Dimensionen Augenscheinvalidität, Belastungsfreiheit, Kontrollierbarkeit und Messqualität erfassen [25]. In entsprechenden Studien (ebd.) zeigte sich, dass insbesondere die „Augenscheinvalidität“ entscheidend für das Akzeptanzurteil ist.

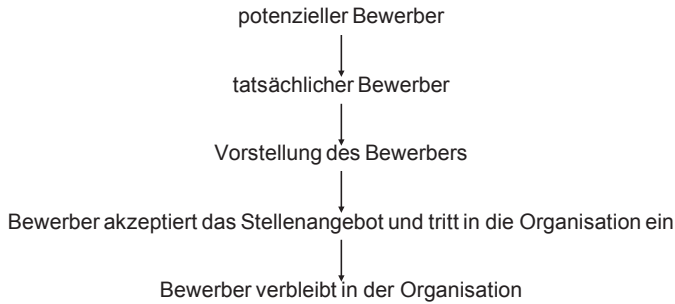
Der große Nutzen von Eignungsdiagnostik zeigt sich vor allem aber in der mittel- und langfristigen Betrachtung. Wenn es gelingt, nicht nur die bezüglich der Anforderungen geeigneten Mitarbeiter zu gewinnen, sondern eine generelle Passung zwischen Person und Organisation herbeizuführen, kann der Grundstein für eine andauernde, vertrauensvolle Arbeitsbeziehung gelegt werden. Aspekte der guten Übereinstimmung von Personen und Organisationen werden daher nachfolgend erläutert.

---

## 2.3 Passung von Person und Organisation

Wie bereits beschrieben, gehen die Aufgaben der beruflichen Eignungsdiagnostik über die Identifikation tätigkeitsspezifisch relevanter Fähigkeiten, Fertigkeiten und Kenntnisse hinaus und schließen auch die Messung von Interessen, Bedürfnisse, Werthaltungen sowie die Diagnostik generell erfolgsrelevanter Eigenschaften und des beruflichen Entwicklungspotenzials ein. Ziel ist es, das berufliche Potenzial von Bewerbern möglichst umfassend zu erkennen und so die Auswahlentscheidung seitens des Unternehmens absichern zu können. Idealerweise steht dem auswählenden Unternehmen hierfür ein ausreichend großer Bewerberpool zur Verfügung, in dem sich viele grundsätzlich geeignete Kandidaten befinden (Grundquote). Dabei können die Unternehmen im Rahmen eines abgestimmten Personalmarketings versuchen, die Grundquote positiv zu beeinflussen. Die Umsetzung eignungsdiagnostischer Informationen in Auswahl oder Ablehnung von Bewerbern stellt einen vergleichsweise späten Schritt im Personalgewinnungsprozess dar (s. Abb. 2.2). Ihm vorgelagert ist die Entscheidung der Bewerber, sich überhaupt bei einem bestimmten Unternehmen zu bewerben.

Diese Entscheidung kann auf Basis ganz unterschiedlicher Erwägungen getroffen werden. So spielt die Übereinstimmung der Anforderungen mit den eigenen Fähigkeiten und Vorstellungen vordergründig sicher die größte Rolle. Auch Faktoren wie der Unternehmensstandort, die Organisationsform und -größe, die Sicherheit des Arbeitsplatzes oder die Dotierung der Stelle werden bei der Organisationswahl berücksichtigt [34]. Neben diesen Kriterien tragen aber auch weniger offensichtlich zugängliche Faktoren zur individuellen Organisationswahl bei. Ein Beispiel hierfür ist das Image von Unternehmen und Branchen. „Image“ fasst nach Meffert et al. [36] die subjektiven Assoziationen und Bewertungen eines Bezugsobjekts zusammen. In die Beurteilung eines Unternehmens fließen



**Abb. 2.2** Personalgewinnungsprozess. (Moser und Zempel [37])

so unterschiedliche Aspekte ein wie die wirtschaftliche Reputation, soziales Engagement, Branche und Produkt, Möglichkeiten zur persönlichen Weiterentwicklung, Arbeitszeitflexibilität, Organisationskultur und die persönliche Bekanntheit mit dem Unternehmen (z. B. durch Verwandte oder Freunde oder frühere Kontakte; [13]). „Big is beautiful“ gilt also im Wettbewerb um die besten Arbeitsplätze nur bedingt. Wie Watzka [61] zeigt, sind Studierenden viele große, international agierende Unternehmen gar nicht bekannt. Als (regional) attraktive Arbeitgeber werden vornehmlich jene Unternehmen genannt, die vor der „eigenen Haustür“ beheimatet sind. Damit sind „Platzhirsche“ hinsichtlich ihrer Attraktivität für Bewerber/-innen den „big playern“ häufig überlegen. Auch wenn die subjektive Vertrautheit mit Unternehmen nur schwer messbar ist, erweist sie sich als wichtige Einflussgröße auf die Reputation eines Unternehmens und damit auch auf die Bewerbungsabsicht von Arbeitssuchenden [6].

Die Einbettung der Eignungsdiagnostik in den Personalgewinnungsprozess verdeutlicht, dass eine Bewerbungssituation immer auch ein gegenseitiges Vorstellen und Umwerben ist. Unternehmen sind bestrebt, passende Bewerber anzuziehen, zu erkennen, zu gewinnen und bestmöglich platzieren zu können; Bewerber wiederum wählen potenzielle Arbeitgeber aus, gleichen ihre Vorstellungen und Werte mit den aufseiten der Organisation vermuteten ab und entscheiden sich für das attraktivste Angebot. Beide Seiten sind daran interessiert, bestmögliche „Passung“ zu erzielen, denn hiervon versprechen sie sich nicht nur unmittelbar Leistung (und Gegenleistung), sondern auch eine motivierende, langfristige Perspektive.

Das Konzept der Passung zwischen Person und Umwelt greift Holland [17] in seiner Interessentypologie auf. In seinem weit verbreiteten RIASEC-Modell unterscheidet er sechs Person- bzw. Umweltypen: 1) praktisch-technische Orientierung R, 2) intellektuell-forschende Orientierung I, 3) künstlerisch-sprachliche Orientierung A, 4) soziale Orientierung S, 5) unternehmerische Orientierung E und 6) konventionelle Orientierung C. Gemäß Holland suchen Menschen nach Umwelten und damit auch Berufen, in denen sie ihre Fähigkeiten einbringen und ihre Werte leben können. Bei manchen Personen

mag eine Orientierung ganz besonders stark ausgeprägt sein, für die sich dann passende Berufe empfehlen lassen (z. B. soziale Orientierung → Lehrer/-in). Andere Personen hingegen weisen eine breitere Orientierung auf, sodass der gewählte Beruf die verschiedenen Interessen idealerweise simultan befriedigt (z. B. praktisch-technische, intellektuell-forschende und unternehmerische Orientierung → Werkstoffingenieur/-in). Neben den (beruflichen) Interessen gemäß Holland spielen natürlich aber noch weitere Motive, die Ausprägung von Fähigkeiten bzw. deren Entwicklungspotenzial und Persönlichkeitsaspekte eine Rolle bei der Passung von Person und Umwelt.

Studien an der Schnittstelle zwischen Arbeits- und Organisationspsychologie und Gesundheitspsychologie zeigen die Bedeutung des Person-Umwelt-Fits auf [7]. Passen Person und Arbeitsumfeld gut zusammen, fördert das die Arbeitszufriedenheit und das organisationale Commitment und kann sich zugleich auch direkt auf den Umsatz auswirken [39]. Bei einem „Misfit“ von Person und (beruflicher) Umwelt andererseits lassen sich Langeweile bzw. Stressreaktionen wie Ärger, Depressivität, Angst oder auch körperliche Beschwerden beobachten und die allgemeine Arbeitszufriedenheit und die Zufriedenheit mit der Arbeitsbelastung und dem Arbeitsinhalt sinken [11].

Wenn die mangelnde Passung von Person und Umwelt so weitreichende Konsequenzen hat, kommt es nicht nur darauf an, die richtigen Bewerber anzuziehen (Attraktion), sondern sie auch möglichst rasch und nachhaltig im Unternehmen zu integrieren (Integration), um Demotivierung und Fluktuation vorzubeugen und einen längeren, engagierten Verbleib der „Neuen“ im Unternehmen zu ermöglichen (Retention). Maassen [35] fasst die Stufen Attraktion – Integration – Retention in der AIR-Formel zusammen und betont die Wichtigkeit eines ganzheitlichen Personalmarketing-Konzepts. So können auf Basis von Tätigkeits- und Organisationsanalysen die Anforderungen an Personen definiert oder aus Kompetenzmodellen abgeleitet werden, was zur Identifikation potenzieller Bewerber genutzt werden kann. Wenn sich dann die Kernaufgabe der Eignungsdiagnostik – die Beurteilung der Eignung der Bewerber – nicht nur auf die rein fachlichen, sondern gerade auch auf die motivationalen und längerfristig entwicklungsrelevanten Eigenschaften und Einstellungen von Personen bezieht, ist ein wichtiger Grundstein gelegt, um die Passung zwischen Person und Umwelt sicherzustellen und auch zukünftig (etwa durch Personalentwicklungsmaßnahmen) zu wahren.

Passung kann als „Ähnlichkeit“ interpretiert werden, aber auch als „Ergänzung“ in dem Sinne, das neue Mitarbeiter/-innen bislang „Fehlendes“ ergänzen sollen („complementary fit“) [16]. Ein einseitig im Sinne des „supplementary fit“ interpretierter Passungs-Begriff führt in Kombination mit dem genannten AIR-Zyklus dazu, dass eine Personalauswahl nach dem „Schmidt-sucht-Schmidtchen-Prinzip“ sich als Hemmnis für Diversity auswirkt [29].

Was aber bleibt, ist die Grundfrage, wer nun denn eigentlich wen auswählt? Suchen Unternehmen passende Mitarbeiter, die sie gegebenenfalls noch entsprechend entwickeln oder suchen Bewerber von vorneherein bestimmte Umgebungen bzw. Unternehmen aus? Die Arbeits- und Organisationspsychologie unterscheidet entsprechend die Sozialisationstheorie, deren gemäß Unternehmen sich die passenden Mitarbeiter formen, von

der Selektionstheorie, nach der Personen ohnehin nur mit Umgebungen in Kontakt treten, die von ihnen als attraktiv, da passend, empfunden werden. Schuler [54] führt aus, dass die Selektionseffekte den Sozialisationseffekten überlegen scheinen. So findet sich bereits im Bewerbungsprozess bzw. beim Einstieg in Organisationen eine Übereinstimmung zwischen der individuellen und der organisationalen „Persönlichkeit“ [20, 21]. Personen mit bestimmten Ausprägungen in den fünf Persönlichkeitsdimensionen („big five“: Emotionale Stabilität, Extraversion, Offenheit, Verträglichkeit und Gewissenhaftigkeit) werden von Unternehmen mit entsprechenden Charakteristika angezogen (Attraktion) und ausgewählt (Selektion). Dabei kommt es neben der objektiven Passung vor allem auch auf die subjektiv empfundene Passung von Person und Unternehmen an [49]. Nach einiger Zeit verbleiben vornehmlich jene Personen in Unternehmen, die dort auch „hinpassen“; erweist sich hingegen die anfänglich empfundene Passung zum Unternehmen als Trugschluss, scheiden diese Personen bei Vorliegen „besser passender“ Angebote wieder aus der Organisation aus (Attrition). Weit vor der oben bereits vorgestellten AIR-Formel hat Schneider [48] diesen dreistufigen Prozess dargestellt (ASA-Theorie) und so die personalistische Sicht auf organisationspsychologische Fragestellungen nachhaltig gestärkt. Die ASA-Theorie beschreibt dabei nicht nur den Auswahl- und Integrationsprozess für neue Mitarbeiter, sondern erklärt auch die Formung von Organisationskultur und -klima, Prozessen und Führungsstilen [56].

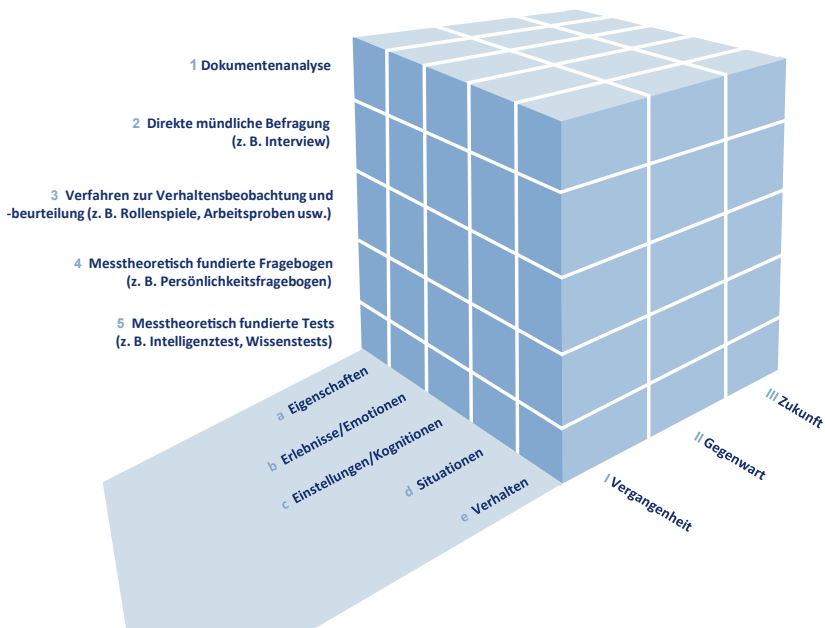
---

## 2.4 Methoden der Eignungsdiagnostik

Die Gegenüberstellung der verschiedenen Anforderungen einer Tätigkeit mit den entsprechenden personenbezogenen Eignungsmerkmalen liefert bereits einen Hinweis darauf, wie unterschiedlich die Informationsquellen über die Ausprägung der Kompetenzen einer Person sein können. Tatsächlich existiert eine große Anzahl an Verfahrenstypen: Tests, Fragebogen, Interviews, Referenzen, Bewerbungsunterlagen, Probezeiten und viele mehr. Deren Einsatzhäufigkeiten in der Praxis variieren allerdings stark im internationalen Vergleich und in Abhängigkeit der Bewerbergruppe [32]. Während vor 15 Jahren nur in gerade einmal vier Prozent der befragten deutschen Unternehmen psychologische Tests zur Auswahl von Führungskräften zum Einsatz kamen, fanden sie in den benachbarten Benelux-Ländern in über 60 Prozent der Unternehmen Anwendung [51]. Positionsübergreifend werden zur Auswahl externer Bewerber in deutschen Unternehmen vor allem die Bewerbungsunterlagen analysiert und Interviews bzw. Einstellungsgespräche mit den Bewerbern/-innen durch die Personal- und Fachabteilungen geführt [55].

Eine Möglichkeit zur Klassifikation der verschiedenen Zugänge zu den diagnostisch relevanten Informationen bietet das CUBE-System von Kersting [24, 27], das hier in einer aktualisierten Fassung präsentiert wird. Anhand der Art der Informationen, ihrem inhaltlich Fokus und ihrem zeitlichen Bezug wird der in Abb. 2.3 dargestellte Würfel aufgespannt.

Zur Verfahrenseinteilung stehen nach Kersting [28] die folgenden Kategorien zur Verfügung:



**Abb. 2.3** CUBE-System

- 1 Dokumenten- und Faktenanalysen  
(z. B. die Analyse und Interpretation von Schul-, Hochschul- und Arbeits-Zeugnissen, vom Lebenslauf, von Beurteilungen [auch 360-Grad-Beurteilungen], von Internet-Funden über die Kandidaten/-innen usw.)
- 2 Direkte mündliche Befragungen  
(z. B. Interviews, Gespräche mit Referenzgeber/-innen)
- 3 Verfahren zur Verhaltensbeobachtung und -beurteilung  
(z. B. Rollenspiele, Gruppendiskussionen, Präsentationsübungen usw.)
- 4 Messtheoretisch fundierte Fragebogen  
(z. B. Persönlichkeitsfragebogen)
- 5 Messtheoretisch fundierte Tests  
(z. B. Intelligenztests, Wissenstests, schriftliche Fachprüfungen)

Diese Einteilung liegt auch der aktuellen DIN 33430 (DIN, 2014) [10] zugrunde. Die jeweils gestellten Anforderungen orientieren sich an den Verfahrensklassen. So erfordern Verfahren der Kategorie 2 (z. B. Interviews) und 3 (z. B. Assessment-Center-Übungen wie Rollenspiele) beispielsweise Verhaltensbeobachtungen und -bewertungen, entsprechend werden hohe Anforderungen an die Gestaltung der Situation und an die Beobachter/-innen gestellt. Verfahren der Kategorien 4 (messtheoretisch fundierte Fragebogen) und 5

(messtheoretisch fundierte Tests) sind hingegen in der Regel bereits hochgradig standardisiert, diesbezüglich stellt die DIN 33430 vor allem die Forderung, dass eine Vielzahl an genau definierten Informationen zu den Verfahren vorliegen müssen. Viele dieser Informationen, wie z. B. Informationen über die Validität der auf Basis der Verfahren getroffenen Entscheidungen, setzen empirische Studien voraus, die wiederum – so die Forderung der DIN 33430 – detailliert und nachvollziehbar beschrieben werden müssen.

Die in der DIN 33430 spezifizierten Informationen sind die Voraussetzung, um entscheiden zu können, ob man ein bestimmtes Verfahren für die interne und externe Eignungsbeurteilung nutzen kann. Verfahrensanbieter(innen), die – z. B. unter Verweis auf vermeintliche Verschwiegenheitsverpflichtungen oder Wettbewerbsgründe – den Informationsanforderungen laut DIN 33430 nicht gerecht werden, diskreditieren ihr Verfahren. Nur Verfahren zu denen Informationen vorliegen, können überhaupt unter Qualitätsgesichtspunkten bewertet werden (z. B. mithilfe des Testbeurteilungssystems des Diagnostik- und Testkuratoriums [59]).

Über die Einordnung von Verfahren entscheidet demnach nicht deren Oberfläche (z. B. Arbeitsprobe), sondern u. a. der Standardisierungsgrad ihrer Auswertung (im Beispiel: Kategorie 3 bei einer Bewertung anhand von Personen vs. Kategorie 5 bei einer vollstandardisierten, computerbasierten Auswertung). Multiple Verfahrenstypen, wie Assessment Center oder Development Center, werden nicht als einzelne Verfahrensklasse angeführt, sondern setzen sich aus Elementen zusammen, die gemäß dem CUBE-System eingeordnet werden können. Damit lassen sich Unterschiede in Assessment Centern beschreiben; etwa, ob psychometrische Leistungstests (z. B. zur sprachlichen Intelligenz oder anderen kognitiven Fähigkeitsbereichen) Bestandteil des Auswahlverfahrens sind. Abb. 2.4 zeigt die Zuordnung gängiger Verfahren der Personalpsychologie zur ersten Dimension des CUBE auf.

Eine zweite Kategorisierungsvariable im CUBE-System trennt Verfahren nach deren „Inhalt“. Sollen mit dem Verfahren a) Eigenschaften, b) Erlebnisse/Emotionen, c) Einstellungen Kognitionen, d) Situationen oder e) Verhalten erfasst werden? So kann zur (aus Anschauungsgründen vereinfachten) Erfassung ordentlicher und pflichtbewusster Arbeitsweise entweder nach der Zustimmung zu konkreten Verhaltensweisen (z. B. „Ich komme immer pünktlich zur Arbeit“) oder aber direkt nach der Zustimmung zu Eigenschaftseinschätzungen gefragt werden (z. B. „Ich bin sehr gewissenhaft“). Selbstverständlich kommt es häufig vor, dass innerhalb eines Verfahrens mehrere Inhalte angesprochen werden, dass ein Teil der Fragen des Interviews oder der Items eines Fragebogens beispielsweise verhaltensorientiert ist, ein anderer aber eigenschaftsorientiert.

Die dritte Dimension im CUBE-System ermöglicht eine Beschreibung der Verfahren danach, ob sie sich vornehmlich an der Vergangenheit, der Gegenwart oder der Zukunft orientieren. So können beispielsweise in Persönlichkeitsfragebogen (Selbst- oder Fremd-)Einschätzungen von Verhalten in der Vergangenheit eingeholt werden oder aber hypothetische (situative) Fragen enthalten sein (etwa: „Stellen Sie sich vor, Sie sind Vorgesetzte/r und ihr Mitarbeiter teilt Ihnen mit ...“).



Auswahlverfahren	Datenquelle <sup>1</sup>				
	1	2	3	4	5
Sachinformationen wie Berufsqualifikationen, Schulabschluss, bisherige Arbeitgeber, Auslandsaufenthalte usw.	■				
Befragungen von Referenzgeber(inne)n		■			
Interviews		■			
Arbeitsproben mit Verhaltensbeobachtungen			■		
Rollenspiele			■		
Präsentationsübungen			■		
Gruppendiskussionen			■		
Fallstudien: mündlich (Präsentation der Ergebnisse)			■		
Postkorbübungen: mündlich (Präsentation der Ergebnisse)			■		
Persönlichkeitsfragebogen				■	
Interessentests in Fragebogenform				■	
Situational-Judgement-Tests					■
Tests zur kognitiven Kompetenz (Intelligenztests)					■
Wissenstests					■
Fallstudien: schriftlich mit hochstandardisierter Auswertung					■
Postkorbübungen: schriftlich mit hochstandardisierter Auswertung					■

<sup>1</sup> 1 Dokumenten-/Faktenanalyse; 2 Direkte Befragungen; 3 Verfahren zur Verhaltensbeobachtung und -beurteilung; 4 Messtheoretisch fundierte Fragebogen; 5 Messtheoretisch fundierte Tests

**Abb. 2.4** Zuordnung gängiger Verfahren der Personalpsychologie zur ersten Dimension des CUBE. (Basierend auf Kersting [27]; geringfügig modifiziert)

Das CUBE-System ist offen für Erweiterungen. Innerhalb der Dimensionen kann es weitere Kategorien geben, beispielsweise spielen außerhalb der Eignungsdiagnostik psychophysiologische Verfahren eine große Rolle. Eine weitere Dimension könnte z. B. die Unterscheidung darstellen, ob Verfahren eher „maximales“ oder „typisches“ Verhalten fokussieren [30]. Assessment Center und Leistungstests erfassen – geht man von leistungsmotivierten Teilnehmern aus – maximales Verhalten, die Teilnehmer strengen sich besonders an, sie verhalten sich anders als „gewöhnlich“. Demgegenüber gewinnt man über die zielgerichtete Betrachtung und Analyse von Lebenslauf-Dokumenten eines Bewerbers eher einen Eindruck von dem „gewöhnlichen“, typischen Verhalten dieser Person.

Das CUBE-System skizziert den Möglichkeitsraum diagnostischer Verfahren. Mit dem System sind keine Aussagen zur Häufigkeit oder Sinnhaftigkeit bestimmter Kategorien verbunden. So ist es durchaus denkbar, dass bestimmte Kombinationen nicht realisiert werden oder nicht realisiert werden können.

In den nachfolgenden Beiträgen in diesem Band werden die verschiedenen Verfahrensklassen besprochen. Übersichten über eignungsdiagnostische Verfahren in deutscher



Sprache liefern beispielsweise die Handbücher von Erpenbeck und von Rosenstiel [12], Hossiep und Mühlhaus [18], Kanning und Holling [22] und Sarges und Wottawa [45].

---

## 2.5 Balanced Scorecard eignungsdiagnostischer Verfahren

Die Auswahl von Verfahren zur Messung beruflichen Potenzials kann unter verschiedenen Gesichtspunkten erfolgen. Wissenschaftler achten vornehmlich auf psychometrische Gütekriterien wie Reliabilitätskennwerte oder Aussagen über die Gültigkeit der gemessenen Werte im Sinne der Konstruktvalidität. Für Anwender sind darüber hinaus Akzeptanz- und Praktikabilitätsaspekte entscheidend. Zudem rücken in der eignungsdiagnostischen Praxis auch verstärkt die Kosten des Verfahrenseinsatzes in den Vordergrund.

Die in Lehrbüchern zur psychologischen Diagnostik bzw. Personalauswahl wiedergegebene „klassische“ Differenzierung der vielen verfügbaren Kennwerte rund um den Verfahrenseinsatz in Hauptgütekriterien (Objektivität, Reliabilität, Validität) und Nebengütekriterien (z. B. Akzeptanz und Nutzen) erscheint äußerst künstlich und unpassend, wenn man diese Einteilung als eine Priorisierung interpretiert. Welche Kennzahlen als wichtig oder nebensächlich erachtet werden, lässt sich nicht pauschal entscheiden, sondern hängt von der jeweiligen Situation und dem diagnostischen Ziel ab und wird maßgeblich durch die Personengruppe beeinflusst, welche die Verfahren zu bewerten hat.

Aus dieser Überlegung heraus empfiehlt sich daher weniger eine Trennung der Bewertungskriterien eignungsdiagnostischer Verfahren, als vielmehr deren parallele Betrachtung in multiattributiven Nutzenmodellen, wie sie z. B. bei Görlich und Schuler [14] vorgeschlagen werden.

Wir möchten hier erstmals vorschlagen, eine bewährte Logik aus den Wirtschaftswissenschaften zu entlehnen und diese auf die Eignungsdiagnostik anzuwenden: die Balanced Scorecard. Das Konzept der Balanced Scorecard geht auf Robert S. Kaplan und David P. Norton zurück [23], die nach einer Möglichkeit suchten, wie zur langfristigen strategischen Ausrichtung von Unternehmen neben dem klassischen Rückgriff auf Finanzkennzahlen auch weitere wichtige Merkmale berücksichtigt werden können. Die Ergänzung der finanzwirtschaftlichen Perspektive um eine kundenorientierte, eine prozessorientierte und eine potenzialorientierte Perspektive ermöglicht die umfassende Berücksichtigung unterschiedlicher Aspekte der Unternehmensausrichtung und damit eine wertorientierte Unternehmensführung [5]. In Abb. 2.5 ist ein Beispiel einer Balanced Scorecard abgebildet.

Die Balanced Scorecard stellt eine strukturierte und ausgewogene Sammlung unterschiedlicher Bewertungsaspekte dar, wobei die Bewertungsaspekte zu vier gleichrangigen Perspektiven geordnet werden. Diese vier Blickwinkel bieten sich auch zur Bewertung eignungsdiagnostischer Verfahren an, wenngleich die Terminologie an die Personalauswahl und -entwicklung angepasst werden muss. Wichtig ist hierbei, dass zur umfassenden Bewertung von Verfahren auch die direkten und indirekten Konsequenzen ihres praktischen Einsatzes herangezogen werden. Die Balanced Scorecard eignungsdiagnostischer Verfahren wird in Abb. 2.6 dargestellt.



**Abb. 2.5** Balanced Scorecard



**Abb. 2.6** Balanced Scorecard eignungsdiagnostischer Verfahren

Eignungsdiagnostische Verfahren lassen sich aus vier Perspektiven bewerten. Als „harte“ Kennwerte dienen die psychometrischen Gütekriterien wie Objektivität, Reliabilität und Validität. Bei der Frage, inwiefern Verfahrensergebnisse unabhängig von den durchführenden Personen (Testgeber) sind, werden üblicherweise die Durchführungs-, die Auswertungs- und die Interpretationsobjektivität unterschieden. Zur Beschreibung der Messgenauigkeit von Verfahren können Kennwerte der Konsistenz (z. B. interne

Konsistenz oder „split-half“), der Stabilität oder der Parallelität herangezogen werden. Inwiefern die Testergebnisse Gültigkeit haben, also das messen, was sie messen sollen, und auch Zusammenhänge zu Außenkriterien aufweisen, wie beispielsweise (späteren) Leistungsbeurteilungen oder Karriereverläufe, wird mittels Maßen der Augenschein-, der Konstrukt- und der Kriteriumsvalidität beurteilt.

Während psychometrische Hauptgütekriterien die Verfahrensgüte relativ abstrakt beschreiben, wird die im konkreten Anwendungsfall erzielte Qualität eine Interaktion aus dem spezifischen Verfahren und den beteiligten Personen sein. Als Kunden gemäß der Nomenklatur der ursprünglichen Balanced Scorecard werden im eignungsdiagnostischen Kontext die Anwender gefasst. Anwender von Testverfahren sind einerseits jene Personen, die direkt mit dem Verfahrenseinsatz betraut sind, etwa weil sie die Durchführung anleiten oder bei der Auswertung der Ergebnisse unterstützen. In der Literatur werden diese Personen häufig unter dem Begriff der Testgeber zusammengefasst. Entsprechend zählen andererseits auch die Testnehmer zu den Anwendern von Testverfahren, also jene Personen, die entscheiden, ob sie an einem Auswahl- bzw. Beurteilungsverfahren teilnehmen wollen oder nicht. Aus Anwenderperspektive sind beim Verfahrenseinsatz juristische Vorgaben und ethische Prinzipien zu berücksichtigen. Im weiteren Sinne gehört hierzu auch die Abstimmung mit dem Betriebsrat und weiteren Stakeholdern. Weiterhin kommt, wie bereits beschrieben, der Akzeptanz des Verfahrens eine wesentliche Bedeutung für den Erfolg der Maßnahme zu. Aspekte wie die Augenscheinvalidität, Belastungsfreiheit, Kontrollierbarkeit und Messqualität eines diagnostischen Instruments, aber auch der Grad der Partizipation bzw. Kontrolle, die das Instrument erlaubt, fließen neben Kriterien wie dem Informationsgewinn, der Transparenz und der Form der Kommunikation (und damit auch der Gestaltung des Tests selbst) in die Bewertung und Annahme eines Verfahrens ein [14, 25]. Sicherlich ebenfalls relevant für die Akzeptanz eines Verfahrens, hier aber aufgrund der Bedeutung als eigener Punkt erwähnt, ist die Fairness beim Einsatz eines diagnostischen Verfahrens. Zunächst sollte zwischen tatsächlicher und empfundener Fairness unterschieden werden. Für die Bewertung der Fairness beim Verfahrenseinsatz liegen unterschiedliche Methoden vor [47]. Wann die Teilnahme an einer eignungsdiagnostischen Untersuchung allerdings vom einzelnen Teilnehmer als fair oder benachteiligend und damit ungerecht empfunden wird, mag noch von weiteren subjektiven Wahrnehmungen und Einstellungen abhängen. Eng mit dem Fairnesskonzept verbunden ist die Interpretation der Ergebnisse. Bei normorientierter Rückmeldung, also beim Vergleich individueller Werte mit den Werten einer Referenzgruppe, muss nicht nur die Entscheidung über die Referenzgruppe sinnvoll aus der diagnostischen Fragestellung abgeleitet werden, sondern sollten vor allem auch aktuelle Normwerte vorliegen. In diesem Zusammenhang gilt es auch zu bedenken, dass das Vorliegen besonders günstiger Hauptgütekriterien keinesfalls zwangsläufig eine differenzierte Diagnostik bei der nun vorliegenden Testnehmergruppe garantiert. Sind etwa Boden- oder Deckeneffekte zu erwarten, weil das Verfahren sehr schwer bzw. besonders leicht für die Zielgruppe ist, differenziert das Verfahren in den Randbereichen nicht und ist somit für die Zielgruppe ggf. untauglich, auch wenn es bei anderen Gruppen gut

funktioniert. Ein weiterer Aspekt, der aus Anwendersicht in die Bewertung eines eignungsdiagnostischen Verfahrens einfließt, ist die Dokumentation rund um den Verfahrenseinsatz. Hierunter fällt zum einen die Brauchbarkeit, Effizienz und Hilfestellung durch die Materialien selbst, zum anderen die Möglichkeit, zu einem späteren Zeitpunkt den Verfahrenseinsatz und seine Ergebnisse nachvollziehen zu können. Als hilfreich empfunden werden beispielsweise die Trennung von umfangreichem Manual und knappem Instruktionssheft für die Testgeber/-in oder auch kommentierte (ggf. automatisch elektronisch erstellte) Profilbogen zur Auswertung und Interpretation der Ergebnisse. Selbstverständlich sind bezüglich der Dokumentation von Teilnehmerdaten oder auch expliziten Ergebnissen die Vorgaben des Datenschutzes einzuhalten.

Gerade in der Praxis wird die Entscheidung über die Auswahl eines eignungsdiagnostischen Verfahrens und die Ausgestaltung des damit einhergehenden eignungsdiagnostischen Prozesses häufig auf Basis von Effizienzüberlegungen getroffen. Die Nützlichkeit bzw. Effizienz von Verfahren spiegelt sich zunächst im Kosten-Nutzen-Verhältnis wieder. Neben monetären Berechnungen dieser Relation, zu der verschiedene Berechnungsweisen vorliegen [14], sollte, wie bereits in diesem Kapitel diskutiert, auch der nicht-monetäre Nutzen durch den Einsatz personalpsychologischer Verfahren in die Bewertung aufgenommen werden. Die Möglichkeit zur Selbstselektion von Bewerber/-innen oder die empfundene Wertschätzung durch die Teilnahme an einem Bewertungsprozess im Rahmen der Personalentwicklung sind nur zwei (zunächst) nicht-monetäre Aspekte, die jedoch ganz wesentliche Erfolgsmerkmale eines gelungenen Einsatzes eignungsdiagnostischer Instrumente verkörpern. Eine fundierte Bewertung des Kosten-Nutzen-Verhältnisses berücksichtigt den gesamten Prozess des Verfahrenseinsatzes, also Kosten und Nutzen in der Auswahl- bzw. Vorbereitungsphase, bei der Durchführung des Verfahrens, bei der Auswertung und Interpretation der Ergebnisse und vor allem auch die Kosten- und Nutzenelemente, die im Anschluss an die Verfahrensdurchführung generiert werden. Damit kommt auch der Relevanz des Gemessenen eine hohe Bedeutung zu. Aus einer psychometrisch-diagnostischen Sichtweise ergibt sich die Relevanz der analysierten Fähigkeiten oder Merkmale aus einer Anforderungsanalyse, also der vorangehenden Prüfung, welche Anforderungen überhaupt an Bewerber/-innen oder Mitarbeiter/-innen gestellt werden. Aus praktischer Sicht sind unter Umständen auch jene Verfahren relevant, die Merkmale in einer moderneren oder aktuell angesagten Weise erfassen bzw. anderweitig im Fokus stehen. Um einen wirklichen Mehrwert zu schaffen, sollte mit einem eignungsdiagnostischen Verfahren dabei nicht nur ein ökonomischer Einsatz von Zeit-, Human- und Material-Ressourcen einhergehen, sondern auch die Praktikabilität des Verfahrenseinsatzes gewährleistet sein. Sofern bereits andere eignungsdiagnostische Verfahren im Einsatz sind (oder geplant werden), ist der Mehrwert abzuschätzen, den das betrachtete Verfahren im Vergleich bzw. in Ergänzung zu den bestehenden Instrumenten leisten kann. Ein Mehrwert ergibt sich dann, wenn das Verfahren das interessierende Merkmal entweder reliabler und/oder valider messen kann oder sein Einsatz mit geringeren Kosten bzw. einer höheren Praktikabilität einhergeht. Wesentlich entscheidend für die tatsächliche Durchführung eines eignungsdiagnostischen Verfahrens ist die Frage,

ob es kostenmäßig insgesamt realisierbar ist. So mag in der betrieblichen Praxis bisweilen die Entscheidung gegen ein eigentlich effizientes Verfahren zu treffen sein, da seine absoluten Kosten das vorhandene Budget übersteigen und so ein Verfahren mit ungünstigerer Kosten-Nutzen-Ratio zu bevorzugen ist. Für die Praxis sind unter anderem daher besonders jene Verfahren interessant, die hinsichtlich ihres Einsatzzweckes optimiert werden können. So lassen sich beispielsweise Zeit und damit auch Kosten sparen, wenn Teile (z. B. Subtests, einzelne Skalen oder auch Kurzversionen) aus Verfahren herausgenommen und einzeln eingesetzt werden können oder sich Instrumente in bereits bestehende Verfahren (z. B. Assessment Center) integrieren lassen. Wenn, wie in größeren Unternehmen üblich, die Entscheidung über ein eignungsdiagnostisches Verfahren nicht allein von den Personalexperten getroffen wird, sondern auch andere Personen beteiligt sind, kommt der Kommunizierbarkeit des Verfahrens eine wesentliche Bedeutung zu. So sind allgemein bekanntere Verfahrenstypen oder auch positiv besetzte Merkmale sicherlich einfacher gegenüber Stake- und Shareholdern zu vermitteln, während für ungewöhnlichere Verfahrenseinsätze erst Überzeugungsarbeit geleistet werden muss.

Die vierte Perspektive auf eignungsdiagnostische Verfahren beschäftigt sich mit der „Passung des Verfahrenseinsatzes zur Organisation“ und den Möglichkeiten zur Weiterentwicklung des Instruments. In manchen Fällen mag bei der Auswahl von Verfahren neben dem diagnostischen Nutzen im Sinne einer Vorhersagemöglichkeit von späterer Leistung bzw. Verhalten besonders auch die Außenwirkung des Verfahrenseinsatzes im Fokus stehen. Eine attraktive, weil zeitgemäße Oberfläche eignungsdiagnostischer Verfahren dürfte die Akzeptanz für diese Verfahren steigern. So wäre in der heutigen Zeit sicherlich ein Großteil der Bewerber irritiert, wenn sie im Vorfeld eines Vorstellungsgesprächs umfangreiche Tests in Papierform zugeschickt bekämen und diese wiederum ausgefüllt zurücksenden müssten. Ein persönlicher Log-in zu einem gestalterisch und technisch hochwertig umgesetzten Online-Verfahren wäre sowohl zeitgemäßer als auch komfortabler für alle Beteiligten. Nicht zuletzt dürfte eine attraktive Ausgestaltung personalpsychologischer Verfahren von Vorteil sein, wenn sich Bewerber beispielsweise bei mehreren Unternehmen im Auswahlprozess befinden. Demgemäß scheint es sinnvoll, die Aktualität der Verfahrensoberfläche bei der Bewertung eignungsdiagnostischer Verfahren zu berücksichtigen. Aber auch an anderer Stelle spielt Aktualität eine Rolle. So wurde bereits angesprochen, dass bei der Rückmeldung normierter Ergebnisse die Aktualität der Vergleichsdaten von Relevanz ist. So erscheint es beispielsweise wenig sinnvoll, den Wissensstand heutiger Gymnasiasten im Vergleich zu fünfzig Jahre alten Wissensdaten einer damaligen Schülergruppe zu bewerten. Damit ist es also auf jeden Fall empfehlenswert, den zeitlichen Kontext des Verfahrenseinsatzes zu berücksichtigen. Im Extrem mag dies jedoch dazu führen, dass nicht nur Verfahren ausgewählt werden, die bezüglich ihrer Oberfläche oder der vorhandenen Normwerte aktuell sind, sondern die in erster Linie „angesagte“ Merkmale zu testen beanspruchen. Solange diese Merkmale sich als anforderungsanalytisch relevant erweisen und valide gemessen werden, ist dem sicher nichts entgegenzusetzen. Neben der Aktualität ist bei der Betrachtung eines Verfahrens aus der kontextualen Perspektive vor allem auch die von Interesse, inwieweit es auf spezifische

Anforderungen der eignungsdiagnostischen Fragestellung abgestimmt ist. So könnte eine Anforderung sein, ein für die Zielgruppe (z. B. Topmanager) angemessenes Verfahren auszuwählen oder zu berücksichtigen, dass es in ein bestehendes Assessment Center integriert werden soll. Ferner fällt hierunter auch die Vereinbarkeit des Verfahrenseinsatzes mit unternehmenseigenen Leitlinien bzw. der Mission und Vision der Organisation. Darüber hinaus sollten eignungsdiagnostische Maßnahmen darauf hin bewertet werden, inwiefern sie zur Statusdiagnostik (Ist) durchgeführt werden oder ob sie darüber hinaus auch zur Modifikation (Veränderung) von Verhalten, Einstellungen oder Fähigkeiten dienen sollen. Konkret ist etwa zu berücksichtigen, ob Verfahren spezifische Anforderungen an Bewerber/-innen bzw. Mitarbeiter abprüfen oder viel eher globalere Potenzialaussagen liefern. Letztlich wird damit in die Entscheidung für oder gegen ein personalpsychologisches Verfahren einfließen, in welchem Grad sich bereits aus der Durchführung des Verfahrens Entwicklungsmaßnahmen (Trainings) ableiten lassen.

Die hier vorgestellte Balanced Scorecard der Eignungsdiagnostik erhebt, wie bei der Arbeit mit Balanced Scorecards üblich, keinesfalls den Anspruch auf Vollständigkeit. Ebenso können einzelne Punkte für manche Anwendungssituationen nicht passend sein. Viel eher ist sie als Rahmen zu verstehen, in der wissenschaftlichen wie praktischen personalpsychologischen Arbeit und beim Umgang mit eignungsdiagnostischen Verfahren zunächst bewertungsrelevante Kriterien herauszuarbeiten und so die Auswahl von Verfahren zu professionalisieren. Erst durch das Zusammenspiel der vier Perspektiven – Psychometrische Hauptgütekriterien, Passung zum Kunden, Nützlichkeit und Effizienz sowie Passung zur Organisation – lassen sich eignungsdiagnostische Verfahren umfassend bewerten. Die Balanced Scorecard der Eignungsdiagnostik ermöglicht hierbei den jeweiligen Anwendungshintergrund zu berücksichtigen.

---

## 2.6 Fazit

Berufliche Eignungsdiagnostik beschäftigt sich mit den Tätigkeitsanforderungen einerseits und der Identifikation und Messung berufserfolgsrelevanter Personmerkmale andererseits. Neben der Personalauswahl unterstützt die Eignungsdiagnostik aber auch Personalentwicklungsprozesse. Ziel ist es, eine Passung zwischen Person und Tätigkeit bzw. Organisation herzustellen, die dann für alle Beteiligten nutzensteigernd ist.

---

### Fragen

1. In welchen Bereichen finden eignungsdiagnostische Maßnahmen und Ergebnisse Anwendung?
2. Worin liegt der Nutzen der Eignungsdiagnostik?
3. Wie lassen sich unterschiedliche eignungsdiagnostische Verfahren klassifizieren?
4. Welche Kennwerte können zur Bewertung eines eignungsdiagnostischen Verfahrens herangezogen werden?

## Literatur

1. Anderson, N., & Witvliet, C. (2008). Fairness reactions to personnel selection methods: An international comparison between the Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment*, *16*, 1–13.
2. Anderson, N., Salgado, J. F., & Hülshager, 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.
3. Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26.
4. Bergmann, C. (2008). Beratungsorientierte Diagnostik zur Unterstützung der Studienentscheidung studierwilliger Maturanten. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 67–77). Göttingen: Hogrefe.
5. Bruhn, M. (1998). Balanced Scorecard: Ein ganzheitliches Konzept der Wertorientierten Unternehmensführung? In M. Bruhn, M. Lusti, W. Müller, H. Schierenbeck, & T. Studer (Hrsg.), *Wertorientierte Unternehmensführung* (S. 145–167). Wiesbaden: Gabler.
6. Cable, D. M., & Graham, M. E. (2000). The determinants of job seekers' reputation perceptions. *Journal of Organizational Behavior*, *21*, 929–947.
7. Caplan, R. D., & Van Harrison, R. (1993). Person-Environment Fit Theory. Some History, Recent Developments, and Future Directions. *J Soc Issues*, *49*, 253–275.
8. Caspi, A., & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry*, *12*, 49–66.
9. Deary, I. J., Whiteman, M. C., Starr, J. M., et al. (2004). The impact of childhood intelligence on later life: Following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*, 130–147.
10. Din. (2014). *Norm-Entwurf (Gelbdruck): DIN 33430: Anforderungen an berufsbezogene Eigenschaftsdiagnostik. Ref. Nr. NA 159-02-09 AA N 196*. Berlin: Beuth.
11. Edwards, J. R., & Van Harrison, R. (1993). Job demands and worker health. Three-Dimensional reexamination of the relationship between person-environment fit and strain. *Journal of Applied Psychology*, *78*, 626–648.
12. Erpenbeck, J., & Von Rosenstiel, L. (Hrsg.). (2007). *Handbuch Kompetenzmessung: Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*. Stuttgart: Schäffer-Poeschel.
13. Felser, G. (2009). *Personalmarketing*. Göttingen: Hogrefe.
14. Görlich, Y., & Schuler, H. (2006). Personalentscheidungen, Nutzen und Fairness. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 797–840). Göttingen: Hogrefe.
15. 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.
16. Hoefert, H.-W. (2013). Rollen des Diagnostikers – sozialpsychologische Aspekte. In W. Sargers (Hrsg.), *Management-Diagnostik* (S. 72–82). Göttingen: Hogrefe.
17. Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Odessa: Psychological Assessment Resources.
18. Hossiep, R., & Mülhhaus, O. (2015). *Personalauswahl und -entwicklung mit Persönlichkeits-tests*. Göttingen: Hogrefe.
19. Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology*, *75*, 28–42.
20. Iyer, P. V., Serlie, A. W., & Oostrom, J. K., et al. (2015). Personality-based Person-Organization (PO) fit: A new direction for personality assessments. In der 13th European conference on psychological assessment. Zürich.



21. Judge, T. A., & Cable, D. M. (1997). Applicant personality, organizational culture, and organization attraction. *Personnel Psychology*, *50*, 359–394.
22. Kanning, U. P., & Holling, H. (Hrsg.). (2002). *Handbuch personaldiagnostischer Instrumente*. Göttingen: Hogrefe.
23. Kaplan, R. S., & Norton, D. P. (1997). *Balanced Scorecard. Strategien erfolgreich umsetzen*. Stuttgart: Schäffer-Poeschel.
24. Kersting, M. (2006). Stand, Herausforderungen und Perspektiven der Managementdiagnostik. *Personalführung*, *10*, 16–27.
25. Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, *33*, 420–433.
26. Kersting, M. (2009). Profit durch Personalauswahl – warum sich eine qualitativ hoch-wertige Personalauswahl langfristig rechnet. In A. Gourmelon, C. Kirbach, & S. Etzel (Hrsg.), *Personalauswahl im öffentlichen Sektor* (2 neubearbeitete Aufl., S. 33–54). Baden-Baden: Nomos.
27. Kersting, M. (2010). Personalauswahl: Methoden für erfolgreiche Unternehmen. In Psychologen BDPu (Hrsg.), *Psychologische Expertise für erfolgreiches Unternehmertum in Deutschland* (S. 57–65). Berlin: Deutscher Psychologen Verlag.
28. Kersting, M. (2011). Managementdiagnostik: Verfahren und Qualitätsaspekte. In C. Niedereichholz, J. Niedereichholz, & J. Staude (Hrsg.), *Managementdiagnostik: Verfahren und Qualitätsaspekte* (S. 1–18). Berlin: Schmidt
29. Kersting, M., & Ott, M. (2015). Diversity-gerechte Personalauswahl – Wie man die Personalauswahl gestalten muss, um Potenziale in allen Gesellschaftsgruppen zu erkennen. In T. Ringeisen & P. Genkova (Hrsg.), *Diversity Kompetenz: Perspektiven und Anwendungsfelder* (S. 1–11). Berlin: Springer.
30. Klehe, U.-C., & Anderson, N. (2007). Working hard and working smart: Motivation and ability during typical and maximum performance. *Journal of Applied Psychology*, *92*, 978–992.
31. Köchling, A. C. (2000). Bewerberorientierte Personalauswahl : Ein effektives Instrument des Personalmarketing. In *Europäische Hochschulschriften: Reihe 6, Psychologie = Psychologie = Psychology*. Frankfurt a. M.: Lang.
32. Krause, D. E. (2010). *Trends in der internationalen Personalauswahl*. Göttingen: Hogrefe.
33. Lang, J. W. B., Kersting, M., Hülshager, U. R., et al. (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.
34. Lievens, F., Decaestecker, C., Coetsier, P., et al. (2001). Organizational attractiveness for prospective applicants: A person-organisation fit perspective. *Applied Psychology: An International Review*, *50*, 30.
35. Maassen, O. (2004). Die AIR-Formel als ganzheitliches Personalmarketing-Konzept. In W. Fröhlich (Hrsg.), *Nachhaltiges Personalmarketing. Strategische Ansätze und Erfolgskonzepte aus der Praxis* (S. 79–100). Frechen: Datakontext.
36. Meffert, H., Burmann, C., & Kirchgeorg, M. (2008). *Marketing. Grundlagen marktorientierter Unternehmensführung. Konzepte – Instrumente – Praxisbeispiele*. Wiesbaden: Gabler.
37. Moser, K., & Zempel, J. (2006). Personalmarketing. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 63–87). Göttingen: Hogrefe.
38. Mount, M. K., & Barrick, M. R. (1995). The big five personality dimensions: Implications for research and practice in human resources management. *Research In Personnel And Human Resources Management*, *13*, 153–200.
39. O'reilly, C. A., Chatman, J., & Caldwell, D. (1991). People and organizational culture: A profile comparison approach to assessing person-organization fit. *Academy of Management Journal*, *34*, 487–516.
40. Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology*, *48*, 775–802.



41. Ostendorf, F., & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung*. Göttingen: Hogrefe.
42. Roberts, B. W., & Delvecchio, W. F. (2000). The rank-order consistency of personality Traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3–25.
43. Salgado, J. F., & Tauriz, G. (2014). The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology*, *1*, 3–30.
44. Salgado, J. F., Anderson, N., Moscoso, S., et al. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology*, *88*, 1068–1081.
45. Sarges, W., & Wottawa, H. (Hrsg.) (2004). *Handbuch wirtschaftspsychologischer Testverfahren – Band I: Personalpsychologische Instrumente* (2. Aufl., überarb. u. erw.). Lengerich: Pabst.
46. 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.
47. Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
48. Schneider, B. (1987). The people make the place. *Personnel Psychology*, *40*, 437–453.
49. Schneider, B. (2008). The people still make the place. In D. B. Smith (Hrsg.), *The people make the place: Dynamic linkages between individuals and organizations* (S. 267–290). New York: Lawrence Erlbaum Associates.
50. Schuler, H. (1990). Personalauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, *34*, 184–191.
51. Schuler, H. (2000). Personalauswahl im europäischen Vergleich. In E. Regnet & L. M. Hofmann (Hrsg.), *Personalmanagement in Europa* (S. 129–139). Göttingen: Verlag für Angewandte Psychologie.
52. Schuler, H. (2006). Arbeits- und Anforderungsanalyse. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 45–68). Göttingen: Hogrefe.
53. Schuler, H. (2007). Berufseignungstheorie. In H. Schuler & K. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 429–440). Göttingen: Hogrefe.
54. Schuler, H. (2014). *Psychologische Personalauswahl: Eignungsdiagnostik für Personalentscheidungen und Berufsberatung*. Göttingen: Hogrefe.
55. Schuler, H., Hell, B., Trapmann, S., et al. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. *Zeitschrift für Personalpsychologie*, *6*, 60–70.
56. Smith, D. B. (Hrsg.) (2008). *The people make the place: Dynamic linkages between individuals and organizations*. New York: Lawrence Erlbaum Associates.
57. Statistisches Bundesamt. (2014). Fast 2 Millionen Nichterwerbspersonen mit Wunsch nach Arbeit. Pressemitteilung Nr. 073 vom 28.02.2014
58. Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, *23*, 565–585.
59. Testkuratorium. (2010). TBS-TK – Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 09. September 2009. *Psychologische Rundschau*, *61*, 52–56.
60. Viswesvaran, C., & Ones, D. S. (2000). Measurement error in “Big Five Factors” personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement*, *60*, 224–235.
61. Watzka, K. (2003). Hochschulmarketing: Arbeitgeberattraktivität und Rekrutierungskanäle. *Personal*, *55*, 8–11.

## Zusammenfassung

Im Rahmen dieses Kapitels werden wichtige Qualitätsstandards für die Personalauswahl vorgestellt, die in verschiedenen Ländern und von verschiedenen Organisationen bzw. Verbänden erarbeitet wurden, und es wird erläutert, welche zentralen Qualitätsaspekte in diesen Standards enthalten sind. Anschließend wird dargelegt, welche Befunde es gibt, dass sich die Berücksichtigung dieser Qualitätsaspekte auf verschiedene Indikatoren für die diagnostische Güte von Auswahlverfahren auswirkt, wie etwa die Reliabilität und die Konstrukt- und Kriteriumsvalidität der verwendeten Verfahren. Schließlich wird beschrieben, wie dies den finanziellen Nutzen beeinflusst, der für ein Unternehmen mit einem qualitativ guten Auswahlverfahren verbunden ist.

## Inhaltsverzeichnis

3.1	Einleitung	58
3.2	Indikatoren für die diagnostische Güte von Auswahlverfahren	58
3.3	Qualitätsstandards für die Personalauswahl	60
3.3.1	Wichtige deutschsprachige und internationale Standards	60
3.3.2	Weitere Qualitätsaspekte jenseits der Standards	62
3.4	Wie wirkt sich die Einhaltung zentraler Qualitätsaspekte auf die Qualität der Personalauswahl aus?	62
3.4.1	Durchführung einer Anforderungsanalyse	62
3.4.2	Eignung des Auswahlverfahrens für die Erfassung der interessierenden Fähigkeiten, Fertigkeiten, Kenntnisse und Merkmale der Bewerber	63

K.G. Melchers (✉)

Institut für Psychologie und Pädagogik, Universität Ulm, Ulm, Deutschland

E-Mail: klaus.melchers@uni-ulm.de

3.4.3	Standardisierung der Durchführung eines Auswahlverfahrens . . . . .	64
3.4.4	Einheitliche Bewertung von Bewerbern in Auswahlverfahren . . . . .	65
3.4.5	Qualifikation der am Auswahlverfahren beteiligten Personen. . . . .	65
3.4.6	Einhaltung rechtlicher Vorgaben . . . . .	66
3.4.7	Behandlung der Bewerber . . . . .	66
3.5	Welchen Nutzen haben auswählende Organisationen von der Berücksichtigung der Qualitätsaspekte? . . . . .	66
3.6	Fazit . . . . .	68
	Literatur . . . . .	69

---

### 3.1 Einleitung

Bei jeder Personalentscheidung, die man bei der Besetzung einer Stelle trifft, hofft man angesichts des Aufwands für die Auswahl und der erheblichen Kosten und Probleme, die mit Fehlbesetzungen verbunden sind, sich richtig entschieden zu haben. Leider sind aber selbst die besten Personalauswahlverfahren und die erfahrensten Entscheider kein 100%iger Schutz, hin und wieder eine falsche Personalentscheidung zu treffen. Allerdings können qualitativ gute Auswahlverfahren erheblich dazu beitragen, solche Fehler zu reduzieren. Vor diesem Hintergrund stellt sich deswegen die Frage, woran sich die Qualität eines Auswahlverfahrens erkennen lässt?

Wie Kersting [16] kürzlich betont hat, bezeichnet Qualität dabei keine dichotome und absolut messbare Eigenschaft eines Auswahlverfahrens, sondern das Ausmaß, in dem eine Reihe wichtiger Qualitätsaspekte bei der Planung, Durchführung und Auswertung eines Auswahlverfahrens sowie bei der Verwendung der Ergebnisse des Auswahlverfahrens beachtet werden. Je mehr dieser Qualitätsaspekte erfüllt sind, desto mehr spricht dies für die Qualität eines Auswahlverfahrens.

---

### 3.2 Indikatoren für die diagnostische Güte von Auswahlverfahren

Aus psychometrischer Sicht zeigt sich die Güte eines Auswahlverfahrens u. a. in seiner Reliabilität, seiner Validität sowie seiner statistischen Fairness (d. h. dem Ausmaß, in dem es vergleichbar valide Vorhersagen für unterschiedliche Gruppen von Bewerbern macht). In der Personalauswahl ist zudem relevant, ob ein Verfahren auch auf Akzeptanz bei den Bewerbern stößt.

Reliabilität meint das Ausmaß, in dem ein Test bei vergleichbarer Durchführung zu einem vergleichbaren Ergebnis führt. Zur Schätzung der Reliabilität gibt es dabei verschiedene Möglichkeiten: Die interne Konsistenz bei Tests, die aus mehreren Items bestehen, die Retest-Reliabilität bei wiederholten Durchführungen des gleichen Tests, die Paralleltest-Reliabilität bei der Durchführung eines parallelen Tests oder die Interrater-Reliabilität bei der Beurteilung ein und derselben Testleistung durch unterschiedliche

Beurteiler. Diese verschiedenen Möglichkeiten zur Schätzung der Reliabilität helfen, unterschiedliche Fehlerquellen zu berücksichtigen [26].

Bezüglich der Validität eines Testverfahrens lassen sich drei Teilaspekte unterscheiden: die Inhaltsvalidität, die Konstruktvalidität und die Kriteriumsvalidität. Inhaltsvalidität meint dabei, inwiefern ein Testverfahren die unterschiedlichen Teilaufgaben und -aspekte inhaltlich umfassend und vollständig abdeckt, die zum interessierenden Konstrukt gehören. Im Gegensatz zu den anderen Validitätsaspekten kann die Inhaltsvalidität nicht durch einen Zahlenwert ausgedrückt werden, sondern durch ein inhaltliches (Experten-)Urteil. Konstruktvalidität meint die Frage, ob ein Testverfahren tatsächlich das Konstrukt misst, das es zu messen vorgibt. Auf eine gute Konstruktvalidität wird im Sinne des Multi-Trait-Multi-Method-Ansatzes dann geschlossen, wenn ein Testverfahren hoch mit Indikatoren korreliert, die das gleiche Konstrukt messen sollen (z. B. im Falle eines Intelligenztests mit anderen Intelligenztests) und niedrig mit Indikatoren, die ein anderes Konstrukt messen sollen (im Falle des Intelligenztests z. B. mit Persönlichkeitsvariablen wie etwa Gewissenhaftigkeit oder Offenheit für Erfahrungen). Als Unterfacetten der Konstruktvalidität spricht man im ersteren Fall von konvergenter Validität und im zweiten Fall von diskriminanter Validität. Die Kriteriumsvalidität beinhaltet schließlich das Ausmaß, in dem ein Verfahren mit wichtigen Kriteriumsvariablen zusammenhängt. In der Personalauswahl ist dabei vor allem gemeint, ob z. B. die Leistung eines Bewerbers in einem Auswahlverfahren mit der tatsächlichen Arbeitsleistung korreliert oder auch mit anderen Indikatoren für beruflichen Erfolg (z. B. Beförderungen oder Lohnanstieg über die Zeit). Im Rahmen der Kriteriumsvalidität können die Kriterien dabei zeitgleich erhoben werden (konkurrente Validität) oder zeitlich später erhoben werden (prädiktive oder prognostische Validität).

Statistische Fairness als weiterer Indikator der diagnostischen Güte meint das Ausmaß, in dem ein Testverfahren vergleichbar kriteriumsvalide Vorhersagen für unterschiedliche Gruppen von Testteilnehmern macht (z. B. für Männer und für Frauen oder für Ausländer und für deutsche Bewerber). Diese Fairness zeigt sich darin, dass das interessierende Kriterium für beide Gruppen auf Basis einer gemeinsamen Regressionsgleichung vorhergesagt werden kann. Dies ist dann der Fall, wenn die Kriteriumsvalidität eines Verfahrens gleich hoch für beide Teilgruppen ist – und wenn mögliche Leistungsunterschiede im Testverfahren Leistungsunterschiede im Kriterium korrekt abbilden [5, 22]. D. h. ein Mangel an statistischer Fairness zeigt sich nicht darin, dass es zwischen unterschiedlichen Gruppen Leistungsunterschiede in einem Test gibt, sondern darin, dass keine einheitliche Vorhersage des interessierenden Kriteriums für beide Gruppen möglich ist. In diesem Fall spricht man von differenzieller Validität oder auch von prädiktivem Bias. Dabei ist zu betonen, dass die statistische Fairness von der wahrgenommenen Fairness des Auswahlverfahrens durch die Bewerber zu unterscheiden ist (s. dazu das Kapitel von Kanning in diesem Band).

Neben diesen psychometrischen Aspekten ist die wahrgenommene Fairness bzw. die Akzeptanz von Test- und Auswahlverfahren auf Seiten der Testteilnehmer jedoch zusätzlich wichtig, da selbst ein hoch reliables und valides Auswahlverfahren kaum etwas

bringt, wenn es von Bewerbern nicht akzeptiert wird oder sie abschreckt. Die Akzeptanz eines Auswahlverfahrens wird dabei unter anderem durch die wahrgenommene Fairness bei der Durchführung dieses Auswahlverfahrens beeinflusst.

---

### **3.3 Qualitätsstandards für die Personalauswahl**

#### **3.3.1 Wichtige deutschsprachige und internationale Standards**

Es gibt eine Reihe von Richtlinien und Normen, die Qualitätsstandards für den Einsatz von Tests im Allgemeinen bzw. von Tests und anderen eignungsdiagnostischen Verfahren speziell im Rahmen der beruflichen Eignungsdiagnostik und der Personalauswahl beschreiben. In diesen Normen werden Regeln und Prinzipien formuliert, deren Einhaltung die Qualität der getroffenen eignungsdiagnostischen Entscheidungen sicherstellen soll. Diese Normen und Richtlinien wurden entweder von einschlägigen Berufs- oder Wissenschaftsverbänden formuliert, oder sie stammen von etablierten Norminstituten wie dem Deutschen Institut für Normung oder der Internationalen Organisation für Normung (ISO).

Im Folgenden sollen kurz einschlägige nationale bzw. internationale Normen vorgestellt werden sowie für die Personalauswahl besonders einschlägige Richtlinien. Umfassendere Übersichten mit vertiefenden Informationen zu diesen und weiteren Richtlinien finden sich bei Kersting [16] bzw. Kersting et al. [17].

Bzgl. der Normen etablierter Norminstitute ist in Deutschland die DIN 33430 zu nennen, die im Jahr 2002 verabschiedet wurde und die kürzlich überarbeitet wurde. Diese Norm enthält Anforderungen an „Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen“. Sie beschreibt zentrale Vorgehensweisen bei der Planung und der Durchführung eignungsdiagnostischer Verfahren, formuliert aber auch Standards zum Umgang mit den gewonnenen eignungsdiagnostischen Informationen sowie Anforderungen an die verwendeten eignungsdiagnostischen Verfahren oder die beteiligten Personen. Wie Kersting [16] betont, sieht die DIN 33430 „ausdrücklich keine isolierte Betrachtung der Qualität von Verfahren (z. B. Tests) vor, sondern umfasst Festlegungen und Leitsätze für den Einsatz dieser Verfahren bei berufsbezogenen Eignungsbeurteilungen“. Dementsprechend können auch keine einzelnen Tests oder andere Verfahren wie Interviews oder Assessment Center gemäß dieser DIN zertifiziert werden, jedoch Personen, die über die in der Norm beschriebenen Kenntnisse und Fähigkeiten verfügen, bzw. Organisationen, deren Qualitätsmanagementsystem für eignungsdiagnostische Prozesse die in der Norm beschriebenen Standards erfüllt.

Ähnlich der DIN 33430 beschreibt die ISO-Norm 10667 Standards für den Einsatz eignungsdiagnostischer Verfahren und Methoden. Diese Norm enthält dabei zwei parallel aufgebaute Teile, die sich zum einen an Auftraggeber und zum anderen an Auftragnehmer richten. An vielen Stellen werden in der ISO-Norm vergleichbare Standards formuliert wie in der DIN-Norm. Im Vergleich zur DIN-Norm wird allerdings die Kom-

munikation zwischen den am eignungsdiagnostischen Prozess beteiligten Personen und Organisationen deutlich stärker betont und formalisiert.

Neben diesen Normen gibt es zahlreiche weitere Richtlinien von Berufs- oder Wissenschaftsverbänden. Für die Personalauswahl im deutschsprachigen Raum sind dabei vor allem die Assessment-Center-Standards und die Interview-Standards des Arbeitskreis' Assessment Center e. V. (vgl. [www.arbeitskreis-ac.de](http://www.arbeitskreis-ac.de)) relevant. In ihnen werden acht bzw. neun Standards formuliert, die einzelne Phasen bei der Entwicklung und Durchführung dieser Verfahren und die Verwendung der gewonnenen Informationen betreffen, deren Nutzen im Rahmen der beruflichen Eignungsdiagnostik erläutert und Hinweise zur konkreten Umsetzung gegeben. Vergleichbare – wenn auch deutlich detailliertere – Standards gibt es auch von internationalen Verbänden (z. B. für Assessment Center von der International Task Force on Assessment Center Guidelines [15]), sowie für den Einsatz von Tests (z. B. von der American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [1]) oder Personalauswahlverfahren im Allgemeinen (z. B. von der Society for Industrial and Organizational Psychology [29]) sowie für zahlreiche weitere Spezialbereiche wie die Übersetzung und Anpassung von Tests für den Einsatz in anderen Sprachen und Kulturen oder den Einsatz von computer- bzw. internetbasierten Tests.

Gemeinsam ist den verschiedenen Normen und Richtlinien, dass in ihnen Anforderungen beschrieben werden, die im Rahmen der Personalauswahl von zentraler Bedeutung sind, da sie „best practices“ für die einzelnen Schritte des Personalauswahlprozess enthalten. Hierbei sind insbesondere die folgenden Aspekte zu nennen:

#### Zentrale Anforderungen an Personalauswahlverfahren:

- Die Durchführung einer Anforderungsanalyse zur Ermittlung der für eine Tätigkeit relevanten Kenntnisse, Fähigkeiten, Fertigkeiten und weiteren Merkmale von Bewerbern.
- Die Auswahl des bzw. der Auswahlverfahren, sodass sie tatsächlich geeignet sind, die ermittelten Kenntnisse, Fähigkeiten, Fertigkeiten und anderen Merkmale valide, reliabel und ohne Benachteiligung bzw. Bevorzugung einzelner Bewerber zu ermöglichen.
- Eine standardisierte Durchführung der Personalauswahlverfahren sowie eine konsistente Bewertung der Bewerber.
- Die Sicherstellung der notwendigen Qualifikation der an der Entwicklung, Durchführung und Auswertung beteiligten Personen.
- Die Einhaltung einschlägiger rechtlicher Vorgaben in allen Phasen des Auswahlprozess'.
- Eine Evaluation der Reliabilität, Validität und Fairness der verwendeten Personalauswahlverfahren.

### 3.3.2 Weitere Qualitätsaspekte jenseits der Standards

Die in den verschiedenen Standards genannten Anforderungen zielen in erster Linie auf Aspekte ab, die die Entwicklung, Vorbereitung, Durchführung und Evaluation von Personalauswahlverfahren betreffen. Viele dieser Anforderungen sprechen damit auch Aspekte an, die sich auf die Akzeptanz der Auswahlverfahren bei Bewerbern auswirken können (vgl. das Kapitel von Kanning in diesem Band). Gemäß den etablierten Modellen im Bereich Bewerberreaktionen ist vor allem die wahrgenommene Fairness eines Auswahlverfahrens für die Akzeptanz auf Seiten der Bewerber relevant. Diese Fairnesswahrnehmung wird u. a. durch solche Aspekte beeinflusst wie den Tätigkeitsbezug eines Auswahlverfahrens oder eine einheitliche Durchführung für alle Bewerber, d. h. durch Aspekte, die durch Anforderungen abgedeckt werden, die in den verschiedenen Qualitätsstandards enthalten sind.

Zusätzlich ist für die Akzeptanz relevant, wie Bewerber sich behandelt fühlen (prozedurale Fairness). Das Modell von Gilliland [9], bei dem es sich um das einflussreichste Modell im Rahmen der Forschung zu Bewerberreaktionen handelt, formuliert diesbezüglich z. B. eine Reihe von Fairness-Regeln, die über Aspekte aus den oben aufgeführten Qualitätsstandards hinausgehen und die beeinflussen, wann ein Auswahlverfahren als fair wahrgenommen wird. Bei diesen zusätzlichen Aspekten handelt es um Dinge wie zusätzliche Erklärungen zum Auswahlverfahren, Ehrlichkeit gegenüber den Bewerbern, eine wertschätzende Behandlung, die Angemessenheit der Fragen, die Bewerber im Rahmen des Auswahlprozesses beantworten müssen, und die Möglichkeit, selbst Fragen zum Auswahlprozess zu stellen (s. dazu das Kapitel von Kanning in diesem Band).

---

### 3.4 Wie wirkt sich die Einhaltung zentraler Qualitätsaspekte auf die Qualität der Personalauswahl aus?

Viele der in den Qualitätsstandards oder in den Modellen zu Bewerberreaktionen genannten Anforderungen und Regeln machen die Durchführung eines Auswahlverfahrens auf den ersten Blick erheblich aufwendiger. Vor diesem Hintergrund erscheinen sie gelegentlich als wünschenswerte aber zu umfangreiche Listen von Aspekten, die zwar „nice to have“ sind, die aber sonst wenig Mehrwert bringen. Das Ziel dieses Abschnitts ist es zu zeigen, welche Belege es gibt, dass eine solche Sichtweise überdenkenswert ist und eine Berücksichtigung der genannten Anforderungen und Regeln sehr wohl an vielen Stellen positive Effekte auf die Qualität von Personalentscheidungen hat.

#### 3.4.1 Durchführung einer Anforderungsanalyse

Anforderungsanalysen helfen, ein umfassendes Bild über die für eine Tätigkeit notwendigen Fähigkeiten, Fertigkeiten, Kenntnisse und sonstigen Merkmale der Bewerber zu bekommen und ein genaueres Verständnis zu haben, welche Teilaufgaben zu dieser



Tätigkeit gehören. Auf Grundlage einer Anforderungsanalyse (s. dazu das Kapitel von Nerdinger in diesem Band) können Personalauswahlverfahren dementsprechend zielgerichteter ausgewählt bzw. ausgestaltet werden. D. h. für ein Unternehmen ist es leichter zu entscheiden, welche Anforderungen Bewerber tatsächlich erfüllen müssen, aber auch, welche Fähigkeiten o. ä. für eine bestimmte Tätigkeit weniger relevant sind. Zudem ist eine Anforderungsanalyse eine zentrale Grundlage für die Auswahl und Ausgestaltung von Arbeitsproben, strukturierten Einstellungsgesprächen, Assessment-Center-Übungen oder Situational Judgment Tests.

In Übereinstimmung mit diesen inhaltlichen Überlegungen haben zahlreiche Studien empirisch belegt, dass die Durchführung einer Anforderungsanalyse positive Effekte auf die Kriteriumsvalidität unterschiedlicher Auswahlverfahren haben kann: Für Persönlichkeitstests zum Beispiel haben Tett et al. [31] bereits vor mehr als 20 Jahren in ihrer Metaanalyse bestätigt, dass Persönlichkeitsdimensionen, die auf Basis einer Anforderungsanalyse ausgewählt wurden, eine deutlich höhere Kriteriumsvalidität aufweisen als Persönlichkeitsdimensionen, die auf gut Glück erhoben werden. Tett et al. haben dabei gezeigt, dass selbst Persönlichkeitsaspekte wie Offenheit für Erfahrung oder Verträglichkeit, die in anderen Studien oft keine guten Prädiktoren für die Arbeitsleistung von Kandidaten waren [2], eine beachtliche Kriteriumsvalidität aufweisen können, wenn sie dort gezielt erfasst werden, wo eine zuvor durchgeführte Anforderungsanalyse ergeben hat, dass sie für eine spezifische Tätigkeit tatsächlich relevant sind.

Weitere Metaanalysen haben auch für andere Auswahlverfahren gezeigt, dass die Durchführung einer Anforderungsanalyse helfen kann, diese Auswahlverfahren so auszugestalten, dass sie eine bessere Vorhersage beruflicher Leistung ermöglichen: Wiesner und Cronshaw [32] fanden solche positiven Effekte auf die Kriteriumsvalidität z. B. von strukturierten Einstellungsgesprächen und McDaniel et al. [21] für Situational Judgment Tests.

### **3.4.2 Eignung des Auswahlverfahrens für die Erfassung der interessierenden Fähigkeiten, Fertigkeiten, Kenntnisse und Merkmale der Bewerber**

Wenn auf Grundlage einer Anforderungsanalyse entschieden wurde, über welche Fähigkeiten etc. Bewerber verfügen sollen, müssen in einem nächsten Schritt Auswahlverfahren ausgewählt werden, die geeignet sind, diese Fähigkeiten möglichst gut zu erfassen. Hier geht es einerseits darum, ein – oder auch mehrere – geeignete Auswahlverfahren auszuwählen, um diese Fähigkeiten zu erfassen (z. B. ein Interview, unterschiedliche Tests, Arbeitsproben o. ä.) und zum anderen um die konkrete Ausgestaltung dieser Verfahren.

Die Frage, mit welchem Auswahlverfahren bestimmte Fähigkeiten etc. erhoben werden können, stellt sich überall dort, wo es mehrere Auswahlverfahren gibt, die hierfür in Frage kommen. Als grundsätzliche Entscheidungshilfe sollte hierbei die Frage die-



nen, welches Verfahren die beste Erfassung ermöglicht. Geht es z. B. darum, Intelligenz oder andere kognitive Fähigkeiten zu erfassen, so empfehlen sich sorgfältig konstruierte Tests, die für die Erfassung dieser Fähigkeiten entwickelt worden sind. Im Gegensatz dazu erscheint es weniger sinnvoll, solche Fähigkeiten mit anderen Verfahren wie z. B. mittels Interviews zu erfassen, da entsprechende psychometrische Tests in der Regel eine deutlich höhere Kriteriumsvalidität haben als diese anderen Verfahren [25]. Zudem lassen es die bisher vorliegenden Daten zur Konstruktvalidität anderer Verfahren wie von Assessment Centern (ACs) oder Interviews zweifelhaft erscheinen, dass eine vergleichbar konstruktvalide Erfassung dieser kognitiven Fähigkeiten mit ihnen möglich ist [23, 35].

Bzgl. der Ausgestaltung der gewählten Auswahlverfahren gibt es weitere Befunde für positive Effekte der verschiedenen Richtlinien und Normen auf die Kriteriumsvalidität der gewählten Verfahren. Im Kontext von Assessment Centern hat sich z. B. gezeigt, dass sich eine größere Anzahl unterschiedlicher AC-Übungen bzw. Instrumente positiv auf die Kriteriumsvalidität dieser ACs auswirkt [4, 8]. Dies spricht dafür, dass eine umfassendere Abdeckung der unterschiedlichen Teilaspekte einer Tätigkeit günstig für die Kriteriumsvalidität der ACs ist. Zudem hat sich gezeigt, dass eine inhaltliche Ausrichtung auf die Zielposition, für die die künftige Leistung vorhergesagt werden soll, ebenfalls zu einer höheren Kriteriumsvalidität führt als die Verwendung eines Standard-ACs, das nicht auf die Zielposition abgestimmt ist [3].

Darüber hinaus empfiehlt es sich, den Kontext zu berücksichtigen, für den das zukünftige Verhalten bzw. die zukünftige Leistung von Bewerbern vorhergesagt werden soll. Für die Erfassung relevanter Persönlichkeitseigenschaften im Rahmen der Personalauswahl bedeutet dies z. B., dass ein arbeitsbezogener Persönlichkeitstest verwendet werden sollte und kein generischer oder gar klinischer Persönlichkeitstest. Auch hier gibt es wiederum eindeutige Belege, die zeigen, dass arbeitsbezogene Persönlichkeitstests eine deutlich höhere Kriteriumsvalidität haben als nicht-arbeitsbezogene Tests [28].

### **3.4.3 Standardisierung der Durchführung eines Auswahlverfahrens**

Eine in den verschiedenen Qualitätsstandards stark betonte Forderung ist, dass Auswahlverfahren für alle Bewerber auf die gleiche standardisierte Art und Weise durchgeführt werden. Bei der Verwendung von Tests ist dies ein Aspekt, den sicherlich niemand bezweifeln würde – bei anderen Auswahlverfahren gibt es allerdings z. T. Vorbehalte gegenüber einer allzu standardisierten Durchführung [11]. Am auffallendsten ist dies vermutlich bei Vorstellungsgesprächen, die nach wie vor häufig nur sehr wenig standardisiert sind. Dies ist umso überraschender, da bei diesen Vorstellungsgesprächen schon sehr lange sehr deutliche Vorteile einer standardisierten Durchführung für die Reliabilität [6] und Kriteriumsvalidität [12] belegt sind: Hoch standardisierte Interviews, bei denen alle Bewerber die gleichen Fragen in der gleichen Reihenfolge gestellt bekommen und

nach dem gleichen Beurteilungsmaßstab bewertet werden – und die in der Regel auch auf der Basis einer Anforderungsanalyse entwickelt wurden – erreichen dabei Kriteriumsvaliditäten in vergleichbarer Höhe wie Intelligenztests, also von  $\rho > ,50$ .

### 3.4.4 Einheitliche Bewertung von Bewerbern in Auswahlverfahren

Neben einer standardisierten Durchführung eines Auswahlverfahrens ist eine einheitliche Bewertung der Bewerber ebenfalls wichtig. Hier geht es darum, dass für alle Bewerber der gleiche Beurteilungsmaßstab angelegt wird, anhand dessen der Grad ihrer Eignung bewertet wird. Auch im Hinblick auf diese Anforderung aus den Richtlinien und Normen gibt es vor allem im Bereich von Einstellungsgesprächen eindeutige Belege, dass eine Standardisierung, die einen einheitlichen Bewertungsmaßstab vorgibt, Vorteile hat. Bei derartigen Interviews werden in der Regel für jede einzelne Frage Beurteilungsanker für die Bewertung der jeweiligen Antworten vorgegeben [19]. Dies trägt dazu bei, dass die Reliabilität und Kriteriumsvalidität der durchgeführten Interviews höher ausfällt [30]. Außerdem lässt sich ein erheblicher Teil der höheren Kriteriumsvalidität dieser Interviews auf ihre höhere Reliabilität und damit auf die einheitlichere Bewertung der Antworten der Bewerber zurückführen [27].

Neben diesen positiven Aspekten einer einheitlichen Bewertung in Bezug auf Reliabilität und Kriteriumsvalidität von Interviews gibt es ebenfalls Belege, dass das Ausmaß von möglicher Diskriminierung und von Bewertungsunterschieden, die z. B. mit der Hautfarbe oder dem Geschlecht von Bewerbern oder auch mit physischen Merkmalen wie Übergewicht zusammen hängen, kleiner ausfällt, wenn ein einheitlicher Bewertungsmaßstab vorgegeben wird [13, 18].

### 3.4.5 Qualifikation der am Auswahlverfahren beteiligten Personen

Wie wichtig es ist, dass die am Auswahlverfahren beteiligten Personen ausreichend qualifiziert sind, zeigt sich anhand von Befunden, dass Trainingsmaßnahmen zur Verbesserung der Qualifikation positive Effekte auf verschiedene Indikatoren der diagnostischen Güte haben: Training von Interviewern wirkt sich z. B. günstig auf die Reliabilität und Kriteriumsvalidität von Einstellungsinterviews aus [6, 13] sowie auf die Konstrukt- und Kriteriumsvalidität von ACs [8, 35].

Vergleichbare Effekte der Qualifikation zeigen sich im AC-Bereich ebenfalls für den Ausbildungshintergrund der AC-Beobachter. ACs werden häufig zur Erfassung interpersoneller Fähigkeiten bzw. Kompetenzen verwendet. Deswegen erstaunt es nicht, dass AC-Beobachter, die Psychologie studiert haben, konstrukt- und kriteriumsvalidere Beurteilungen abgeben als Beobachter mit einem anderen Ausbildungshintergrund, der sie weniger auf die Beobachtung und differenzierte Bewertung entsprechender Verhaltensweisen vorbereitet [8, 35].

Neben diesen direkten Effekten auf die diagnostische Güte von Auswahlverfahren hat die Qualifikation der am Auswahlverfahren beteiligten Personen zudem Effekte bzgl. der Berücksichtigung weiterer Qualitätsaspekte. Lievens und De Paepe [20] fanden diesbezüglich z. B., dass die Wahrscheinlichkeit, bei der Durchführung von Interviews auf eine höhere Standardisierung und eine einheitlichere Bewertung der Antworten zu achten, bei solchen Interviewern höher war, die in der Vergangenheit an speziellen Trainingsworkshops teilgenommen hatten.

### **3.4.6 Einhaltung rechtlicher Vorgaben**

Die Einhaltung rechtlicher Vorgaben hat zwar keine direkten Vorteile für die psychometrische Güte von Auswahlverfahren, aber sie beeinflusst naheliegender Weise das Ausmaß, in dem Auswahlverfahren bei möglichen juristischen Auseinandersetzungen vor Gericht anfechtbar sind. Dies ist im deutschen Sprachraum zwar bisher kein allzu großes Thema, aber für multinationale Unternehmen von erheblicher Relevanz, die z. B. in den USA Niederlassungen haben, für die sie Personal auswählen müssen. In zahlreichen Gerichtsverfahren hat sich dort wiederholt gezeigt, dass die Einhaltung der Anforderungen aus den verschiedenen Richtlinien und Normen die Chancen vor Gericht erheblich erhöht, ein Auswahlverfahren erfolgreich zu verteidigen [33, 34]. Außerdem hat die Einhaltung rechtlicher Vorgaben auch auf Wahrnehmungen von Bewerbern deutliche Effekte. Werden z. B. im Rahmen von Einstellungsgesprächen verbotene Fragen gestellt (z. B. ob und wann Bewerber vorhaben, Kinder zu bekommen), so führt dies u. a. zu einer negativeren Wahrnehmung des Gesprächs und einer niedrigeren Bereitschaft, ein mögliches Stellenangebot anzunehmen [24].

### **3.4.7 Behandlung der Bewerber**

Die Akzeptanz eines Auswahlverfahrens hat zwar in der Regel keine Effekte auf die Kriteriumsvalidität, wirkt sich aber auf Reaktionen auf Seiten der Bewerber aus. Hausknecht et al. [10] konnten diesbezüglich in ihrer Metaanalyse u. a. positive Effekte auf die Wahrnehmung des auswählenden Unternehmens oder auf Intentionen zur Annahme eines möglichen Stellenangebots belegen.

---

## **3.5 Welchen Nutzen haben auswählende Organisationen von der Berücksichtigung der Qualitätsaspekte?**

Die positiven Effekte auf die Kriteriumsvalidität, die mit der Einhaltung der Anforderungen aus den verschiedenen Richtlinien und Normen einhergehen, wirken sich direkt auf den finanziellen Nutzen der auswählenden Unternehmen aus. Nimmt man hier zur

Abschätzung dieses finanziellen Nutzens das klassische Nutzenmodell von Cronbach und Gleser [7], so ergibt sich der finanzielle Nutzenzuwachs durch Personalauswahlverfahren im Vergleich zu einer Zufallsauswahl nach folgender Formel:

$$\Delta U = (N_E \times T \times SD_Y \times r_{XY} \times \bar{Z}_X) - (N_B \times K). \quad (3.1)$$

Die erste Klammer in dieser Formel steht dabei für den Nutzen und die zweite für die Kosten, die durch ein Auswahlverfahren entstehen. Zum Nutzen tragen die Anzahl der eingestellten Bewerber  $N_E$  bei, die multiplikativ mit ihrer Verweildauer  $T$  im Unternehmen verknüpft sind sowie mit der Standardabweichung des verwendeten Kriteriums in Geldeinheiten  $SD_Y$ , der Kriteriumsvalidität  $r_{XY}$  des verwendeten Auswahlverfahrens  $X$  für die Vorhersage des Kriteriums  $Y$  und der durchschnittlichen z-standardisierten Leistung  $\bar{Z}_X$  der eingestellten Bewerber im Auswahlverfahren. Von diesem Nutzen werden die Kosten abgezogen, die sich anhand der Kosten  $K$  pro Bewerber sowie der Anzahl der getesteten Bewerber  $N_B$  ergeben (siehe dazu auch das nachfolgende Beispiel).

### Beispiel

Beispiel zur Bestimmung des finanziellen Nutzens eines Personalauswahlverfahrens:

Ein großes Finanzdienstleistungsunternehmen stellt jedes Jahr 100 (=  $N_E$ ) neue Kundenbetreuer ein. Dazu führt es im Lauf des Jahres regelmäßig eintägige ACs für insgesamt  $N_B = 200$  Bewerbern durch. Dieses AC ist auf Basis einer Anforderungsanalyse entwickelt worden, simuliert zentrale Aufgaben der Kundenbetreuer und wird mit trainierten Beobachtern durchgeführt, sodass seine Kriteriumsvalidität  $r_{XY} = ,35$  beträgt. Die Kosten des ACs belaufen sich pro Bewerber auf  $K = 1000$  EUR. Die eingestellten Bewerber sind mit  $\bar{Z}_X = 0,75$  im Durchschnitt eine dreiviertel Standardabweichung besser als der Mittelwert aller Bewerber im AC. Bei einem durchschnittlichen Verbleib im Unternehmen von 5 Jahren ergibt sich folgender Nutzenzuwachs im Vergleich zu einer Zufallsauswahl:

$$\Delta U = (100 \times 5 \times 10.000 \text{ EUR} \times 0,35 \times 0,75) - (200 \times 1000 \text{ EUR})$$

$$\Delta U = 1.312.500 \text{ EUR} - 200.000 \text{ EUR} = 1.112.500 \text{ EUR}$$

Da jedoch auch ohne Verwendung eines ACs keine Zufallsauswahl unter den Bewerbern getroffen würde, ist es sinnvoll, den Nutzen durch das AC im Vergleich zu anderen Auswahlverfahren wie z. B. herkömmlichen, unstrukturierten Einstellungsgesprächen bestimmen. Für solche unstrukturierten Einstellungsgespräche kann eine Kriteriumsvalidität von  $r_{XY} = ,20$  angenommen werden [12]. Nimmt man in diesem Fall Kosten von 400 EUR pro Bewerber an und lässt alle übrigen Werte gleich wie bei der Beispiel-Rechnung für das AC, ergibt sich im Vergleich zu einer Zufallsauswahl folgender Nutzenzuwachs für die Auswahl auf Basis unstrukturierter Einstellungsgespräche:

$$\Delta U = (100 \times 5 \times 10.000 \text{ EUR} \times 0,20 \times 0,75) - (200 \times 400 \text{ EUR})$$

$$\Delta U = 750.000 \text{ EUR} - 80.000 \text{ EUR} = 670.000 \text{ EUR}$$

D. h. im Vergleich zum AC sind die Kosten zwar um 120.000 EUR niedriger, der Nutzen ist aufgrund der niedrigeren Kriteriumsvalidität des Auswahlverfahrens jedoch auch erheblich niedriger. Im Vergleich zu unstrukturierten Einstellungsgesprächen entgeht dem Unternehmen auf diese Art und Weise sogar ein finanzieller Nutzen von 442.500 EUR (= 1.112.500 EUR – 670.000 EUR). Diesen hätte es in 5 Jahren erwirtschaftet, da die mit dem valideren Verfahren ausgewählten Kundenberater besser für ihre Tätigkeit geeignet gewesen wären.

Wie oben aufgeführt, wirkt sich die Berücksichtigung zahlreicher Qualitätsaspekte aus den verschiedenen Standards und Richtlinien positiv auf die Kriteriumsvalidität aus und damit auf einen der Faktoren, die den Nutzen in der ersten Klammer in Formel 3.1 ausmachen. Überall dort, wo eine größere Anzahl von Bewerbern getestet und eingestellt wird, führt dies dazu, dass der Effekt der zusätzlichen Kosten, die sich anhand der zweiten Klammer der Formel ergeben, sehr schnell und sehr deutlich übertroffen wird – und dies umso mehr, je länger die mit einem Verfahren ausgewählten Bewerber im Unternehmen verbleiben. Oder andersherum gedacht: Das Geld, das man einsparen kann, indem man auf Qualitätsaspekte wie die Durchführung einer Anforderungsanalyse, die Standardisierung von Auswahlverfahren oder die Schulung von Beurteilern oder Interviewern verzichtet, geht in der Regel mit einer niedrigeren Kriteriumsvalidität einher. – Dadurch wird der gesparte Betrag sehr schnell durch eine Senkung des finanziellen Nutzens 'mehr als nur aufgefressen'. D. h. ein billigeres, aber qualitativ schlechteres Auswahlverfahren spart meist nur auf den ersten Blick Geld, kostet Unternehmen aber häufig mittelfristig deutlich mehr als sie im ersten Augenblick eingespart haben.

---

### 3.6 Fazit

Wie in diesem Kapitel aufgezeigt wurde, lässt sich die Frage nach der Qualität eines Auswahlverfahrens nicht mit einer dichotomen Ja- oder Nein-Antwort beantworten. Stattdessen muss hierfür das Ausmaß berücksichtigt werden, in dem unterschiedliche Qualitätsaspekte berücksichtigt werden, wie sie in einschlägigen Richtlinien und Normen zusammengestellt und erläutert werden. Dass die Berücksichtigung dieser Qualitätsaspekte im Rahmen von Personalauswahlprozessen positive Effekte auf die Reliabilität, die Validität und sowie auf weitere Indikatoren der diagnostischen Güte von Auswahlverfahren hat, konnte in zahlreichen Studien empirisch belegt werden. Diese positiven Effekte auf die diagnostische Güte von Personalauswahlverfahren wiederum führen zu einem positiven Effekt für den finanziellen Nutzen, den Unternehmen von einer qualitativ guten Personalauswahl haben.

### Fragen

1. Welche Indikatoren für die diagnostische Güte von Tests und Auswahlverfahren gibt es?
2. Erläutern Sie, inwiefern und aus welchen Gründen sich die Berücksichtigung der verschiedenen Qualitätsaspekte positiv auf die diagnostische Güte von Personalauswahlverfahren auswirkt.
3. Warum zahlt sich die Berücksichtigung der unterschiedlichen Qualitätsaspekte oftmals finanziell für Unternehmen aus?

### Literatur

1. American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2015). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
2. 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.
3. Bartels, L. K., & Doverspike, D. (1997). Effects of disaggregation on managerial assessment center validity. *Journal of Business and Psychology*, 12, 45–53.
4. 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.
5. Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115–124.
6. Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology*, 80, 565–579.
7. Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana, IL: University of Illinois Press.
8. Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III, & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493–511.
9. Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, 18, 694–734.
10. 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.
11. Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 333–342.
12. Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79, 184–190.
13. Huffcutt, A. I., & Woehr, D. J. (1999). Further analysis of employment interview validity: A quantitative evaluation of interviewer-related structuring methods. *Journal of Organizational Behavior*, 20, 549–560.
14. 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.
15. International Task Force on Assessment Center Guidelines. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management*, 41, 1244–1273.
16. Kersting, M. (2014). Qualitätsstandards der Personalauswahl. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (3. Aufl.). Göttingen: Hogrefe.

17. Kersting, M., Häcker, H., & Hornke, L. F. (2011). Qualitätsstandards in der Diagnostik. In L. F. Hornke, M. Amelang, & M. Kersting (Hrsg.), *Grundfragen und Anwendungsfelder psychologischer Diagnostik. Enzyklopädie der Psychologie, Serie Psychologische Diagnostik* (Bd. 1). Göttingen: Hogrefe.
18. Kutcher, E. J., & Bragger, J. D. (2004). Selection interviews of overweight job applicants: Can structure reduce the bias? *Journal of Applied Social Psychology, 34*, 1993–2022.
19. Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*, 241–293.
20. Lievens, F., & De Paepe, A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high structure interviews. *Journal of Organizational Behavior, 25*, 29–46.
21. McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730–740.
22. Meade, A. W., & Tonidandel, S. (2010). Not seeing clearly with Cleary: What test bias analyses do and do not tell us. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3*, 192–205.
23. Melchers, K. G., Klehe, U.-C., Richter, G. M., Kleinmann, M., König, C. J., & Lievens, F. (2009). “I know what you want to know”: The impact of interviewees’ ability to identify criteria on interview performance and construct-related validity. *Human Performance, 22*, 355–374.
24. Saks, A. M., & McCarthy, J. M. (2006). Effects of discriminatory interview questions and gender on applicant reactions. *Journal of Business and Psychology, 21*, 175–191.
25. 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.
26. Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence, 27*, 183–198.
27. 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.
28. Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65*, 445–493.
29. Society for industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.
30. Taylor, P. J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology, 75*, 277–294.
31. Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703–742.
32. Wiesner, W. H., & Cronshaw, S. F. (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.
33. Williams, K. Z., Schaffer, M. M., & Ellis, L. E. (2013). Legal risk in selection: An analysis of processes and tools. *Journal of Business and Psychology, 28*, 401–410.
34. Williamson, L. G., Campion, J. E., Malos, S. B., Roehling, M. V., & Campion, M. A. (1997). Employment interview on trial: Linking interview structure with litigation outcomes. *Journal of Applied Psychology, 82*, 900–912.
35. Woehr, D. J., & Arthur, W., Jr. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*, 231–258.



---

## Zusammenfassung

Anforderungsanalysen bilden eine Form der Arbeitsanalyse, bei der Personenmerkmale ermittelt werden, die für die erfolgreiche Bewältigung einer Tätigkeit notwendig sind. Die für Anforderungsanalysen relevanten Personenmerkmale werden gewöhnlich in fachliches Wissen, Fähigkeiten, Fertigkeiten und Interessen gegliedert, in letzter Zeit werden sie auch häufiger in Form geforderter Kompetenzen beschrieben. Drei methodische Ansätze zur Durchführung von Anforderungsanalysen lassen sich unterscheiden: Die erfahrungsgelernt-intuitive Methode, die arbeitsplatzanalytisch-empirische Methode und die personenbezogen-empirische Methode. Die erfahrungsgelernt-intuitive Methode basiert auf der Expertise des Anforderungsanalytikers, entsprechend große Erfahrungen werden dabei vorausgesetzt. Bei der arbeitsplatzanalytisch-empirischen Methode wird zunächst eine Arbeitsanalyse durchgeführt und anschließend werden die relevanten persönlichen Attribute systematisch durch psychologisch geschulte Experten ermittelt. Die personenbezogen-empirische Methode ermittelt die Anforderungen aufgrund von statistischen Zusammenhängen zwischen den Merkmalen von Arbeitsplatzinhabern sowie individuellen Erfolgskriterien der beruflichen Tätigkeit. Entscheidend ist, dass die gewählte Methode für das jeweilige Ziel der Anforderungsanalyse geeignet ist.

---

F.W. Nerdinger (✉)  
Wirtschafts- und Sozialwissenschaftliche Fakultät,  
Universität Rostock, Rostock, Deutschland  
E-Mail: [friedemann.nerdinger@uni-rostock.de](mailto:friedemann.nerdinger@uni-rostock.de)

## Inhaltsverzeichnis

4.1	Anforderungen und Anforderungsanalysen.....	72
4.2	Konzepte anforderungsrelevanter Personenmerkmale .....	74
4.3	Methoden und Verfahren zur Bestimmung von Anforderungen .....	76
4.3.1	Die erfahrungsgeleitet-intuitive Methode .....	76
4.3.2	Die arbeitsplatzanalytisch-empirische Methode .....	77
4.3.3	Die personenbezogen-empirische Methode .....	82
4.4	Fazit.....	85
	Literatur.....	87

---

### 4.1 Anforderungen und Anforderungsanalysen

Unter „Anforderungen“ sind diejenigen Merkmale zu verstehen, über die ein Mitarbeiter verfügen muss, damit er eine berufliche Aufgabe erfolgreich bewältigen kann [3] [31]. Die Analyse der so verstandenen Anforderungen kann daher auch als eine Form der Arbeitsanalyse betrachtet werden [7]. Unter Arbeitsanalysen versteht man „alle Methoden, Verfahren und Instrumente, die dazu dienen, Informationen über die Arbeitstätigkeiten, die organisatorisch-technischen Arbeitsbedingungen, die Arbeitsmittel und Werkzeuge sowie deren Auswirkungen auf den Menschen zu sammeln, zu verarbeiten und zu interpretieren“ [10, S. 117]. Werden Arbeitsanalysen eingesetzt, um die Anforderungen zu ermitteln, die eine Tätigkeit an die Qualifikationen des Positionsinhabers stellt, spricht man von einer Anforderungsanalyse. Der Begriff „Qualifikation“ bezeichnet dabei in einem weiten Sinn die Gesamtheit der leistungsbezogenen Merkmale einer Person, die dabei behilflich sind, die Anforderungen im Arbeitsprozess erfolgreich zu bewältigen. Unter Qualifikation im engeren Sinn versteht man dagegen den Befähigungsnachweis für ein bestimmtes Tätigkeits- bzw. Berufsbild.

Bei Arbeitsanalysen, die z. B. das Ziel der Arbeitsgestaltung verfolgen, werden die Merkmale der Arbeit in situativen Begriffen ausgedrückt. Dagegen wird das Ergebnis einer Anforderungsanalyse in Personenmerkmalen formuliert. In dieser Form bilden Anforderungsanalysen die Voraussetzung für viele personalpsychologische Interventionen, wobei Personalplanung, -auswahl und -entwicklung besonders wichtige Ziele solcher Interventionen sind [31]. Im vorliegenden Zusammenhang steht die Beziehung zur Personalauswahl im Zentrum des Interesses, daher sei diese etwas näher beleuchtet.

Um eine sinnvolle, psychologisch fundierte Personalauswahl betreiben zu können, ist es notwendig zu wissen, welche Vorbildung, welchen Werdegang, welche Verhaltensweisen, Fähigkeiten, Eigenschaften etc. die Personen kennzeichnen, die bisher in einer zu besetzenden Tätigkeit erfolgreich waren [3]. Die Summe dieser Merkmale in ihrer jeweils spezifischen Ausprägung bezeichnet man als Anforderungsprofil. Zum Beispiel können

die Anforderungen an Führungskräfte der mittleren Führungsebene und an Führungsnachwuchskräfte wie in Abb. 4.1 dargestellt zusammengefasst werden (in Anlehnung an [2]).

Anhand eines solchen Anforderungsprofils, das Merkmale von Stelleninhabern beschreibt, die in der Vergangenheit die Tätigkeit erfolgreich ausgeübt haben, lassen sich gezielt neue Stelleninhaber suchen und unter den Bewerbern geeignete Personen auswählen, indem zur Erfassung der erfolgsnotwendigen Personenmerkmale angemessene Erhebungsmethoden eingesetzt werden. Eine Anforderungsanalyse identifiziert also solche Personenmerkmale, die u. a. eine methodisch abgesicherte Grundlage für die Personalplanung, -suche und -auswahl darstellen [30].

1. Wissen und Erfahrung:
  - a. Wissen:
    - Art und Ergebnis von Studien- und Schulabschlüssen
    - Art und Ergebnis der berufsorientierten Ausbildung
    - Spezial- und Zusatzkenntnisse (Fremdsprachen, EDV usw.)
  - b. Erfahrungen und Können:
    - Dauer der allgemeinen Berufserfahrung
    - Dauer der branchenübergreifenden Organisations- und Personalführung
    - Dauer der branchenbezogenen Berufserfahrung
2. Fertigkeiten:
  - Reproduktive Fertigkeiten zur Lösung von Routineaufgaben
  - Planungs- und Organisationsfertigkeiten
  - Delegations- und Koordinationsfertigkeiten
  - Persönliches Auftreten und Repräsentationsverhalten
  - Kooperationsfertigkeiten
  - Informationsfertigkeiten
  - Ausdrucksfertigkeiten
3. Fähigkeiten:
  - Problemsensitivität und Fähigkeit zur Problemlösung
  - Kreativität und Innovationsfähigkeit
  - Personalführungsfähigkeiten
4. Persönlichkeitseigenschaften:
  - Zielstrebigkeit
  - Selbstbewusstsein
  - Zuverlässigkeit
  - Verantwortungsbereitschaft
  - Kritikfähigkeit
  - Risikoeinstellung
  - Entscheidungsfreudigkeit
  - Belastbarkeit
  - Stressresistenz
  - Vitalität
  - Flexibilität/Anpassungsfähigkeit

**Abb. 4.1** Anforderungsprofil für Führungskräfte der mittleren Führungsebene (nach [2])

## 4.2 Konzepte anforderungsrelevanter Personenmerkmale

Gegenstand einer Anforderungsanalyse sind die Leistungsvoraussetzungen von Personen, die für die erfolgreiche Bewältigung der jeweiligen betrieblichen Aufgaben benötigt werden. Solche Leistungsvoraussetzungen werden in Abhängigkeit von der wissenschaftlichen Bezugsdisziplin sowie den professionellen Tätigkeitsfeldern unterschiedlich bezeichnet [3, 26]. In der Personalwirtschaftslehre, im Personalmanagement sowie in der amerikanischen Personalpsychologie werden gewöhnlich folgende Bezeichnungen gewählt:

- Fachliches Wissen (z. B. Rechtskenntnisse, Maschinenkenntnisse, Wissen über fachliche Abläufe etc.);
- Fertigkeiten, i. S. der Beherrschung bestimmter fachlicher Abläufe und Verhaltensweisen (z. B. eine Kundenbestellung korrekt aufnehmen, eine Software kompetent bedienen können);
- Fähigkeiten, d. h. relativ grundlegende und stabile individuelle Handlungsgrundlagen, die für die Bewältigung sehr verschiedenartiger Situationen und Aufgaben eingesetzt werden können, wie z. B. Konzentrationsfähigkeit, räumliches Orientierungsvermögen oder numerische Intelligenz;
- Personenmerkmale, dazu zählen u. a. Interessen (z. B. Interesse an der Arbeit mit Computern oder aber mit Menschen), Persönlichkeitseigenschaften (z. B. Gewissenhaftigkeit, Offenheit für neue Erfahrungen, Extravertiertheit etc., s. u. 4.3.3), Werthaltungen (i. S. der Orientierung an Konzeptionen des Wünschenswerten wie z. B. Gerechtigkeit).

In der Arbeits- und Personalpsychologie wird dagegen oft zwischen Verhaltensweisen, die für den beruflichen Erfolg erforderlich sind – das entspricht gewöhnlich den Fertigkeiten –, erfolgsnotwendigen Eigenschaften (Fähigkeiten, Persönlichkeitsmerkmale, Interessen, Werthaltungen) sowie Aufgaben- und Ergebnisanforderungen (Wissen, berufliche Qualifikationsnachweise, Berufserfahrung, Leistungsergebnisse) unterschieden. Schließlich wird im deutschen Sprachraum insbesondere in der Betriebspädagogik und der Berufssoziologie, neuerdings auch in der Organisationspsychologie häufig der Begriff „Kompetenz“ zur Erfassung notwendiger Personenmerkmale verwendet [8]. Dabei werden gewöhnlich vier Kompetenzen unterschieden. Kenntnisse, Fertigkeiten und Fähigkeiten, die der Bewältigung bekannter beruflicher Problemsituationen dienen, werden unter dem Begriff „Fachkompetenz“ zusammengefasst [38]. Zur Bewältigung neuartiger und komplexer beruflicher Aufgaben bedarf es der sogenannten „Methodenkompetenz“ [35]. Sie resultiert aus flexibel einsetzbaren kognitiven Fähigkeiten, die auf Intelligenz und Expertise beruhen. Die „Sozialkompetenz“ umfasst kooperative und kommunikative Fertigkeiten zur Realisierung von Zielen in sozialen Interaktionssituationen [13]. Schließlich werden mit dem Begriff „Selbstkompetenz“ Aspekte der

motivationalen und emotionalen Steuerung im Sinne der Selbstregulation des beruflichen Handelns bezeichnet [16]. Diese Kompetenzform resultiert aus Einstellungen, Werthaltungen, Bedürfnissen und Motiven.

Darauf aufbauend spielen im Human Resource Management vieler Unternehmen sogenannte „Kompetenzmodelle“ eine wichtige Rolle [31, 37]. Es handelt sich dabei um den Versuch, unternehmensweit einheitliche Anforderungen und Erwartungen an das Handeln der Führungskräfte bzw. des Personals zu formulieren. Die mit einem solchen Modell beschriebenen Kompetenzen sollen möglichst alle relevanten Aktivitäten abdecken, die an einem Arbeitsplatz gefordert sind. Zum Kompetenzmodell einer Führungskraft zählen z. B. kontinuierliches Lernen, Trainieren der Mitarbeiter, Teamarbeit befördern, Diversität stärken, Visionen verwirklichen, Wissen anwenden, Komplexität steuern, effektiv kommunizieren und Haltung zeigen [5]. Aufgrund des Anspruchs, unternehmensweit und über die hierarchischen Ebenen hinweg Anforderungen abzudecken, sind solche Kompetenzen oft unscharf formuliert, sie können aber durch spezifischere fachliche Konzepte für spezielle Anwendungszwecke konkretisiert werden [18].

Methoden-, Sozial- und Selbstkompetenz werden häufig unter dem Begriff der „Schlüsselqualifikationen“ zusammengefasst [24]. Stellen Anforderungen Personenmerkmale zur Kennzeichnung von Tätigkeiten dar, so versteht man unter Qualifikationen Personenmerkmale, die Individuen beschreiben sollen. Schlüsselqualifikationen sind dann Kenntnisse und Fähigkeiten/Fertigkeiten, welche die Bewältigung vieler alternativer, momentan bestehender Positionen und Funktionen sowie von sich verändernden Anforderungen ermöglichen. Schlüsselqualifikationen sind demnach überfachliche Qualifikationen, die zum Handeln befähigen und zusammen mit der Fachkompetenz den beruflichen Erfolg bestimmen.

Schließlich ist noch auf eine weitere begriffliche Schwierigkeit zu verweisen. Der Verband für Arbeitsstudien und Betriebsorganisation (REFA [25]) untergliedert Anforderungen in die Bereiche

- Können (Kenntnisse und Geschicklichkeit),
- Verantwortung (z. B. für die Sicherheit anderer oder für Betriebsmittel),
- Belastung (Nervenbelastungen, Muskelbelastungen, zusätzliche Denkprozesse),
- Arbeitsbedingungen (z. B. Schmutz, Staub, Nässe etc.).

Mit dem von der REFA verwendeten Begriff der Arbeitsanforderungen werden also zum Teil ganz andere Sachverhalte angesprochen (Verantwortung, Arbeitsbedingungen) als in der psychologischen Anforderungsanalyse. Die Konzeptionen von Belastungen (Denkanforderungen) überschneiden sich teilweise, nur der Begriff des Könnens bezeichnet Ähnliches wie die psychologische Anforderungsanalyse. Arbeitsanforderungen im psychologischen Sinne und Arbeitsanforderungen im Sinne der REFA müssen also genau unterschieden werden, da dieselben Begriffe zum Teil Unterschiedliches bezeichnen.

### 4.3 Methoden und Verfahren zur Bestimmung von Anforderungen

Schuler und Höft [3, 32] haben drei grundsätzliche Zugänge zur Bestimmung von Anforderungen differenziert: die erfahrungsgeleitet-intuitive Methode, die arbeitsplatzanalytisch-empirische Methode sowie die personenbezogen-empirische Methode. Diese werden im Folgenden etwas genauer betrachtet.

#### 4.3.1 Die erfahrungsgeleitet-intuitive Methode

Die erfahrungsgeleitet-intuitive Methode setzt bei den Personen, welche die Anforderungsanalyse durchführen, eine langjährige Erfahrung mit bestimmten Tätigkeiten oder Berufen voraus. Dies betrifft v. a. die Merkmale der Tätigkeit im engeren Sinn, also die typischen Arbeitsaufgaben, die Arbeitsmittel, die Art der Arbeitsumgebung, die einzuhaltenden Qualitätsstandards, die auftretenden Belastungen usw. Aufgrund der großen Erfahrung mit dem Arbeitsplatz, die der Anforderungsanalytiker aufweisen sollte, wird bei dieser Vorgehensweise auf eine vorausgehende, systematische Arbeitsanalyse verzichtet. Ein typisches Beispiel für dieses Vorgehen bildet die Methode der kritischen Ereignisse.

Mit der Methode der kritischen Ereignisse („critical incident technique“ [9]) lassen sich Anforderungen speziell im Verhaltensbereich ermitteln. Um die Verhaltensanforderungen einer Zielposition genauer zu explorieren, können alle Personen herangezogen werden, die sich in engem Arbeitskontakt mit dem jeweiligen Vertreter der Stelle befinden (dazu zählen Vorgesetzte, Arbeitsplatzinhaber, Kollegen oder auch Kunden). Diese Arbeitsplatzexperten sollen dann das Arbeitsverhalten eines Mitarbeiters reflektieren und sich dabei auf solche Verhaltensweisen konzentrieren, die in der Vergangenheit zu einer besonders erfolgreichen oder einer besonders wenig erfolgreichen Aufgabenerfüllung geführt haben. Alle dabei gelieferten Beschreibungen werden als (erfolgs-)kritische Ereignisse bezeichnet. Mit Blick auf jedes kritische Ereignis sollen die Befragten die Umstände, die zu dem Verhalten geführt haben, möglichst umfassend beschreiben (die auslösenden Bedingungen). Dann ist das kritische Verhalten so konkret wie möglich zu beschreiben und dabei muss erläutert werden, was an dem Verhalten besonders erfolgreich bzw. wenig erfolgreich war. Schließlich sind die Konsequenzen des als kritisch bewerteten Verhaltens zu skizzieren. Ein Beispiel für die Formulierung solcher Fragen zeigt Abb. 4.2.

Nachdem alle Arbeitsplatzexperten ihre Beschreibungen der kritischen Ereignisse erstellt haben, werden im nächsten Schritt gleiche oder ähnliche Fälle gruppiert und dann gemeinsam durchgesprochen. Damit soll erreicht werden, dass die jeweiligen Ereignisse mit Blick auf die auslösenden Bedingungen, Manifestationen und Resultate der kritischen Vorfälle jeweils möglichst übereinstimmend beurteilt werden. Neben der Klärung der Abweichungen zwischen den einzelnen Arbeitsplatzexperten ist zu diesem Zweck darauf zu achten, dass möglichst alle Ereignisse, die für den Erfolg in einer bestimmten Tätigkeit kritisch sind, erfasst werden. Wichtig sind also immer die Aspekte, die eine Beziehung zum Tätigkeitserfolg aufweisen. In jeder Tätigkeit gibt es viele Routineaufgaben, die

Denken Sie an ein Beispiel für das Arbeitsverhalten eines Mitarbeiters, das eine besonders effektive oder besonders ineffektive Arbeitsweise veranschaulicht. Beschreiben Sie die Situation und das fragliche Verhalten möglichst konkret. Stellen Sie sich dazu die folgenden Fragen:

- Was waren die Umstände oder Hintergrundbedingungen, die zu diesem Verhalten führten?
- Beschreiben Sie das konkrete Verhalten des Mitarbeiters. Was war besonders effektiv oder ineffektiv an diesem Verhalten?
- Was waren die Konsequenzen seines Verhaltens?

**Abb. 4.2** Formulierung zur Erfassung kritischer Ereignisse. (Verfahrensbeispiel; nach [31, S. 69])

zwar oft vorkommen und häufig auch einen großen Teil der Arbeitszeit in Anspruch nehmen, aber den Tätigkeitserfolg nicht entscheidend beeinflussen. Die Methode der kritischen Ereignisse fordert jedoch die Konzentration auf erfolgskritische Vorkommnisse und Verhaltensweisen. Aus den im Konsens mit den Arbeitsplatzexperten gruppierten und möglichst umfassenden Situations- und Verhaltensbeschreibungen lässt sich dann ein Profil situativ bedingter Anforderungen an das jeweils erwünschte Verhalten erstellen. Ein Beispiel für ein mögliches Vorgehen zeigt die folgende Abb. 4.3 [3, S. 214].

Vor allem in neuen Arbeitsfeldern ist es häufig angemessen, zunächst mithilfe der Methode kritischer Ereignisse die Gesamtmenge erfolgsrelevanter Verhaltensweisen zu sammeln und darauf aufbauend ein standardisiertes Instrument (i. S. der arbeitsplatzanalytisch-empirischen Methode) zu entwickeln. Das Hauptproblem beim Einsatz der erfahrungsgelitet-intuitiven Methode bildet der Einsatz von Experten. So haben Kersting und Birk [14] herausgefunden, dass die Einstufungen, die von „internen Experten“ (Arbeitsplatzinhaber und ihre Vorgesetzten) vorgenommen werden, hochgradig mit der sozialen Erwünschtheit der Fragen zusammenhängen. Externe Experten (Arbeits- und Organisationspsychologen) dagegen konnten die Anforderungen besser beurteilen. Andere Untersuchungen zeigen, dass Arbeitsplatzinhaber zwar brauchbare Einschätzungen ihrer Tätigkeit auf einem ganzheitlichen Niveau abgeben können, dagegen in der Beurteilung von Details Schwierigkeiten haben [34]. Daher kommt Schuler [31] zum Ergebnis, dass eine erfahrungsgelitet-intuitive Methode zur Bestimmung von Arbeitsanforderungen nur selten ausreichend ist, da eine zufriedenstellende Ermittlung grundlegender Dimensionen und deren Gewichtung letztlich nur auf quantitativem Wege möglich ist.

### 4.3.2 Die arbeitsplatzanalytisch-empirische Methode

Die arbeitsplatzanalytisch-empirische Methode geht davon aus, dass sich spezifische Merkmale von Tätigkeiten auch in spezifischen Anforderungen widerspiegeln. Die Anforderungsanalyse beruht dabei auf einer in einem ersten Schritt durchgeführten Arbeitsanalyse, wobei sich eine Vielzahl von Methoden unterscheiden lässt [28]. Im zweiten Schritt



Beispiel für einen ineffizienten Kundenkontakt:

Auslösende Bedingung: Ein Kunde ruft am Montagnachmittag aggressiv und empört an und beschwert sich, dass die zugesagte Lieferung bei ihm immer noch nicht eingetroffen ist.

- Das Verhalten des Verkäufers auf diese Situation war: Der Verkäufer unterbricht den Kunden rasch und sagt, dass er gute Ohren habe und dass das Telefon auch in Ordnung sei. Der Kunde brauche also nicht so laut zu reden.
- Negativ daran ist: Der Verkäufer lässt den Kunden nicht ausreden. Er zeigt kein Verständnis für die Empörung des Kunden, sondern übt sogar Kritik an dessen Verhalten.
- Konsequenz: Der Kunde wird ausfallend. Der Verkäufer legt den Hörer auf.

Beispiel für einen effizienten Kundenkontakt:

Auslösende Bedingung: Von einem Kunden ist noch keine schriftliche Bestellung eingegangen, obwohl dem Kunden vorher erklärt wurde, dass nur bei Vorliegen eines schriftlichen Auftrags die Ware versandt werden kann. Der Kunde hat bei dem entsprechenden Telefonat eine schnelle Auslieferung dringend gefordert. Wenn die Lieferung raus soll, muss der Kunde aber binnen Kurzem die Bestellung faxen.

- Der Verkäufer reagiert darauf mit folgendem Verhalten: Er lässt die Sache nicht auf sich beruhen, nach dem Motto „Der Kunde ist ja selbst schuld!“, sondern ruft von sich aus bei dem Kunden an. Der Verkäufer erklärt ihm kurz das Problem und fragt nach, ob er helfen kann.
- Positiv daran ist: Der Verkäufer ist proaktiv. Für ihn zählt das Ergebnis und nicht die Frage, wer schuld hat. Er macht dem Kunden keine Vorwürfe, sondern versucht ihn freundlich dazu zu bringen, zu tun, was für beide Seiten gut ist.
- Konsequenz: Der Kunde entschuldigt sich für sein Versäumnis und schickt im letzten Moment die Bestellung per E-Mail.

**Abb. 4.3** Beispiele für die Analyse kritischer Ereignisse. (Nach [3, S. 2014])

werden dann die persönlichen Attribute (z. B. Konzentrationsfähigkeit, Kommunikativität oder Interesse an ökonomischen Prozessen) systematisch ermittelt, die für eine erfolgreiche Ausführung der betreffenden Tätigkeit relevant sind. Zur Ermittlung dieser Attribute werden gewöhnlich zunächst psychologische Experten, die zugleich mit den Arbeitsaufgaben und -bedingungen vertraut sein müssen, Hypothesen formulieren. Die Gültigkeit der Hypothesen kann anschließend in Labor- oder Felduntersuchungen überprüft werden. Beispielsweise kann die Hypothese aufgestellt werden, dass für die erfolgreiche Bearbeitung bestimmter Überwachungsaufgaben, wie sie z. B. bei der Steuerung komplexer industrieller Prozesse (u. a. in Kraftwerken) häufig vorkommen, die Konzentrationsfähigkeit eine wichtige Rolle spielt [3]. Dazu kann in einem ersten Schritt eine solche Überwachungs-

aufgabe im Labor simuliert und überprüft werden, ob im Beruf bewährte Ingenieure, die mit Überwachungsaufgaben betraut sind, die simulierten Aufgaben besser bewältigen als unerfahrene Personen. Ist dies der Fall, können Personen mit sonst gleichen Eignungsvoraussetzungen zunächst einem Aufmerksamkeitstest unterzogen werden, der die individuelle Konzentrationsfähigkeit misst. Anschließend bearbeiten diese Personen die Überwachungsaufgabe. Stellt Konzentrationsfähigkeit tatsächlich ein kritisches persönliches Attribut dar, sollten sich signifikante Zusammenhänge zwischen dem Abschneiden im Konzentrationstest und bei der Bewältigung der simulierten Überwachungsaufgabe ergeben. Auf statistischem Wege kann schließlich diejenige Ausprägung der Konzentrationsfähigkeit ermittelt werden, die für eine hinreichende Aufgabenerfüllung notwendig ist.

Damit ist die prinzipielle Logik des arbeitsplatzanalytisch-empirischen Vorgehens beschrieben. In der Praxis finden sich hier vielfältige Variationen. Das sei an einem konkreten Beispiel verdeutlicht (vgl. dazu [37, S. 637 ff.]), bei dem für ein Dienstleistungsunternehmen – die Schweizerische Post – nach einem Veränderungsprozess ein Kompetenzmodell entwickelt wurde mit dem Ziel, dieses als Grundlage der künftigen Mitarbeiterauswahl und -förderung zu etablieren. Die Schweizerische Post hat seit dem Jahre 2000 einen gravierenden Veränderungsprozess durchlaufen: vom Staatsbetrieb hin zu einem staatsnahen Unternehmen. Marktöffnung, Lösungsorientierung, Kundenorientierung und das Wahrnehmen von Verantwortung wurden in diesem Prozess zu entscheidenden Zielgrößen. Die grundlegende Umstrukturierung und Neuorganisation des Unternehmens erforderte es, die Aufgaben und Anforderungen der Stelleninhaber (Poststellenleiter, Schalterpersonal etc.) neu zu definieren und in entsprechende Kompetenzmodelle zu transformieren. Diese sollten letztlich in das Human Resource Management integriert und künftig der Mitarbeiterauswahl und -förderung zugrunde gelegt werden. Zu diesem Zweck wurde eine aufgabenbezogene Anforderungsanalyse durchgeführt, d. h. auf der Grundlage der Ergebnisse standardisierter und strukturierter Erhebungsverfahren wurden Aufgaben ermittelt, die bei den Arbeitstätigkeiten tatsächlich zu bewältigen sind. Über Schlussfolgerungen wurden dann die erforderlichen Leistungsvoraussetzungen wie Kenntnisse, Fähigkeiten und Fertigkeiten etc. zu relevanten Kompetenzen verdichtet.

In einem Kick-off-Workshop mit der Geschäftsleitung und dem strategischen Management wurden zunächst die notwendigen Informationen über Aufgaben und Tätigkeiten der Stelleninhaber erhoben. Anschließend wurden mit 28 Stelleninhabern, Vorgesetzten und Mitgliedern des strategischen Managements halbstandardisierte Interviews durchgeführt, u. a. um Informationen über relevante Anforderungsbereiche mit Blick auf die aktuelle Situation und die künftigen Entwicklungen zu erheben. Die Ergebnisse dieser Befragungen dienen dazu, ein bereits bestehendes Instrument der Arbeitsanalyse – den „Leitfaden für qualitative Personalplanung bei technisch-organisatorischen Innovationen“ (LPI; vgl. [36]) – an die Situation des Auftrag gebenden Unternehmens anzupassen. Der LPI ist ein strategisch ausgerichtetes Arbeitsanalyseverfahren, das es ermöglicht, sowohl aktuelle als auch künftige Aufgaben und Anforderungen für unterschiedliche Funktionen und Stellen zu erfassen. Nach einer empirischen Prüfung der Messqualitäten des modifizierten Verfahrens (LPI-PV) wurden mit der Endversion Aufgaben- und Anforderungsanalysen durchgeführt (eine Auswahl

der für die Aufgabenanalyse eingesetzten Items zeigt Abb. 4.4). Analysegegenstand waren die erforderlichen Leistungsvoraussetzungen für das Ausführen von Arbeitstätigkeiten.

Mit diesem Instrument wurden dann 1098 Funktionsträgern in den über die ganze Schweiz verteilten Poststellen befragt. Die Daten wurden anschließend ausgewertet und in Form von Aufgaben- und Anforderungsprofilen aufbereitet. Im Sinne einer Organisationsentwicklungsmaßnahme [23] wurden die so ermittelten Istwerte in einem Workshop den Vertretern des strategischen Managements des Unternehmens vorgestellt. Die Manager mussten die aktuelle Situation mit Blick auf zukünftige mittelfristige Veränderungen (Soll) einstufen. Damit können sowohl verschiedene Stellengruppen in den zentralen Anforderungen als auch aktuelle und zukünftige Anforderungen der einzelnen Stelleninhaber verglichen werden (vgl. Abb. 4.5). Mit Blick auf die Mitarbeiterauswahl ermöglicht ein solcher Vergleich die Auswahl derjenigen unter den Bewerbern, die über das (Lern-)Potenzial verfügen, um auch die zu erwartenden Änderungen von Stellenan-

Erhebungsinstrument LPI-PV: Aufgaben	
1. Zustelltour	
1.1 Zustelltour vorbereiten	6 Items
1.2 Zustelltour durchführen	11 Items
2. Schalterdienst	
2.1 Zahlungsverkehr durchführen	5 Items
2.2 Brief- und Paketpost abwickeln	17 Items
3. Reklamation	7 Items
4. Organisation und Administration	17 Items
5. Verkauf	
5.1 Verkauf planen und Kontrollieren	7 Items
5.2 Verkauf fördern	7 Items
5.3 Verkauf realisieren	6 Items
6. Kundenberatung und -betreuung	12 Items
7. Koordination und Kommunikation	13 Items
8. Budgetierung	
8.1 Statistiken erstellen	4 Items
8.2 Statistiken analysieren	4 Items
8.3 Controlling-Maßnahmen ableiten	3 Items
8.4 Budgets erstellen	3 Items
8.5 Buchhaltung durchführen	4 Items
9. Personalführung	
9.1 Zielvereinbarungsgespräche führen	4 Items
9.2 Personalfriedenheit sicherstellen	4 Items
9.3 Personal betreuen und führen	14 Items
...	...

**Abb. 4.4** Aufgabenliste (Auszug) und jeweilige Item-Zahl des LPI-PV. (Nach [37, S. 640])

forderungen zu bewältigen [30]. Zudem lassen sich die Aufgaben und Anforderungen zu Stellenbeschreibungen aggregieren.

Zur Entwicklung von Kompetenzmodellen wurden auf der Basis der Ergebnisse der Aufgaben- und Anforderungsanalysen jeweils notwendige Kompetenzen für die Bewältigung der mit einer Stelle verbundenen Aufgabe abgeleitet (in diesem Fall waren es insgesamt 36), die sich den vier Kompetenzbereichen zuordnen lassen: Fach-, Methoden-, Sozial- und Selbstkompetenz. Diese Kompetenzen wurden dann zu 36 Kompetenzlisten zusammengefasst. Die Abb. 4.5 zeigt eine Liste für die Kompetenz „Teamfähigkeit“.

**Kompetenz „Teamfähigkeit“**

Definition: Fähigkeit zur konstruktiven Zusammenarbeit in einer Arbeitsgruppe/im Team im Rahmen der gemeinsamen Leistungserbringung. Dies beinhaltet die Fähigkeit, sich und sein Können im Sinne einer Gruppenaufgabe optimal einzubringen, ohne den Beitrag anderer zu beeinträchtigen, sowie einen freundlichen und höflichen Umgang mit den anderen Teammitgliedern.

Hoher Ausprägungsgrad (7): Konzentriert seine Handlungen auf Aktivitäten, die zum Erreichen der Ziele der Gruppe führen; stellt die Interessen des Teams über die Bedürfnisse der eigenen Person und geht respektvoll mit den anderen Teammitgliedern um. Tritt aktiv für gemeinsame Ziele ein und denkt auch bei internen Meinungsverschiedenheiten oder Streitigkeiten zuallererst an die Teamleistung.

Niedriger Ausprägungsgrad (1): Ist in der Lage, bei Bedarf in einem gut funktionierenden Team zu arbeiten und seinen Teil zur Gruppenleistung beizutragen.

Kompetenz nicht erforderlich (0)

Frage: Welcher Ausprägungsgrad der Kompetenz Teamfähigkeit ist erforderlich, um die Aufgaben eines PP-Filial-Verantwortlichen erfolgreich ausüben zu können?

Bitte nehmen Sie die Einstufung durch Ankreuzen auf der siebenstufigen Skala (zwischen 0 = Kompetenz nicht erforderlich und 7 = Hoher Ausprägungsgrad) und beachten Sie hierbei die unten aufgeführten Beispiele.

Um Ihnen die Einstufung zu erleichtern, sind im Folgenden Beispiele für Aufgaben aus dem Tätigkeitsbereicheines PP-Filial-Verantwortlichen aufgeführt:

- Sich im Team gegenseitig Auskünfte erteilen
- An Meetings- und Teamsitzungen teilnehmen
- Teamgespräche führen
- Kontakte mit Mitarbeitenden pflegen
- Betriebsklima fördern

**Abb. 4.5** Kompetenzliste am Beispiel „Teamfähigkeit“. (Nach [37, S. 643])

Die 36 Kompetenzlisten wurden dann im Rahmen von Workshops zur Kompetenzmodellierung jeweils Vorgesetzten der betreffenden Funktionsträger sowie Mitarbeitern des strategischen Managements vorgelegt. Sie mussten letztlich einschätzen, welcher Ausprägungsgrad der jeweils diskutierten Kompetenz erforderlich ist, um die jeweiligen Aufgabenstellungen erfolgreich ausüben zu können (eingestuft auf einer siebenstufigen Skala). Für jede der vier großen Kompetenzklassen wurden die Ausprägungen auf den einzelnen, darunter jeweils subsumierten Kompetenzen in Form eines Profilizuges eingetragen. Die Zusammenstellung für eine Funktionsgruppe bildet dann ein Kompetenzmodell, das u. a. der Auswahl von Mitarbeitern zugrunde gelegt wird.

Dieses Beispiel zeigt anschaulich, wie aufwendig eine gründlich durchgeführte Anforderungsanalyse ausfallen kann (zu begrifflichen Problemen von Kompetenzmodellen vgl. [31]). Der Aufwand war aber in diesem Fall auch dringend geboten, da die so entwickelten Kompetenzmodelle den strategischen Wandel der Unternehmenspolitik fundieren sollten.

### 4.3.3 Die personenbezogen-empirische Methode

Die personenbezogen-empirische Methode ermittelt Anforderungen aufgrund von statistischen Zusammenhängen zwischen den Merkmalen von Personen, die in einem bestimmten Bereich oder Beruf tätig sind, und Kriterien der Leistungsbeurteilung oder anderen individuellen Erfolgskriterien der beruflichen Tätigkeit. Liegen keine Daten zum beruflichen Erfolg vor, werden die Merkmale möglichst vieler in einem bestimmten Beruf oder einer bestimmten Tätigkeit aktiver Personen berücksichtigt.

Für die erste der beiden beschriebenen Formen, bei der Daten zum Berufs- bzw. Tätigkeitserfolg vorliegen, kann auf Metaanalysen zurückgegriffen werden, welche die Ergebnisse einzelner einschlägiger Studien zusammenfassen. Das sei am Beispiel verdeutlicht. Mount et al. [21] haben eine Metaanalyse der bis zu diesem Zeitpunkt vorliegenden Untersuchungen über den Zusammenhang zwischen Leistung und Persönlichkeit – erfasst über das Fünf-Faktoren-Modell – in Tätigkeiten, die interpersonale Interaktionen erfordern, durchgeführt. Im Fünf-Faktoren-Modell werden folgende Persönlichkeitsmerkmale unterschieden [4]:

1. Gewissenhaftigkeit: unterscheidet ordentliche, zuverlässige, hart arbeitende, disziplinierte, pünktliche, penible, ehrgeizige und systematische von nachlässigen und gleichgültigen Personen.
2. Extraversion: extravertierte Menschen sind gesellig, aktiv, energisch, gesprächig, personenorientiert, herzlich, optimistisch und heiter, sie mögen Anregungen und Aufregungen.
3. Neurotizismus: hohe Ausprägungen verweisen darauf, nervös, ängstlich, traurig, unsicher und verlegen zu sein und sich Sorgen um seine Gesundheit zu machen.
4. Verträglichkeit: kennzeichnet altruistische, mitfühlende, verständnisvolle und wohlwollende Menschen.
5. Offenheit für Erfahrung: erfahrungsoffene Menschen zeichnen sich durch hohe Wertschätzung für neue Erfahrungen aus, bevorzugen Abwechslung, sind wissbegierig, kreativ, fantasievoll und unabhängig in ihrem Urteil.

In der Metaanalyse von Mount et al. [21] wurden Untersuchungen in industriellen Teams und an Mitarbeitern in Dienstleistungstätigkeiten mit direktem Kundenkontakt berücksichtigt. Die Leistung wurde dabei über Beurteilungen des Leistungsverhaltens („job performance“) durch Vorgesetzte erhoben (vgl. Tab. 4.1).

Während in Tätigkeiten, die Teamarbeit erfordern, vor allem die Persönlichkeitsfaktoren Verträglichkeit (,33) und emotionale Stabilität (,27) mit der Leistung korrelieren, sind es bei Dienstleistungstätigkeiten mit Kundenkontakt in erster Linie Gewissenhaftigkeit (,29) und Offenheit für Erfahrung (,17). Demnach finden sich gravierende Unterschiede in den Anforderungen zwischen den beiden Tätigkeitsbereichen, obwohl letztlich beide durch hohe Kontaktintensitäten und persönliche Interaktion gekennzeichnet sind [22]. Dabei lässt sich die Leistung in Tätigkeiten mit Teamarbeit durch die fünf Faktoren sehr viel besser vorhersagen als in Dienstleistungstätigkeiten mit direktem Kundenkontakt. Zudem finden sich in den beiden Bereichen deutlich unterschiedliche Zusammenhänge zwischen den einzelnen Persönlichkeitsmerkmalen und der Leistung. Während es bei Teamarbeiten wichtig ist, dass die Mitarbeiter gut miteinander auskommen (Verträglichkeit), erfordern Dienstleistungen mit Kundenkontakt von den Mitarbeitern v. a. Zuverlässigkeit und Engagement (Gewissenhaftigkeit) sowie die Offenheit, sich auf Neues einzulassen – vermutlich in Form des Kontakts mit ständig wechselnden Menschen. Extraversion dagegen ist nahezu unkorreliert mit der Leistung in Dienstleistungen (zu beachten ist allerdings, dass diese Werte aufgrund der eher geringen Zahl von Studien noch nicht sehr stabil sind).

**Tab. 4.1** Metaanalyse des Zusammenhangs von Persönlichkeit und Leistung in Team- bzw. Dienstleistungstätigkeiten (beobachtet: gemessene Werte; korrigiert: um mangelnde Reliabilität korrigierte Werte; nach Mount et al. [21]; vgl. Nerdinger [22])

Variable	Koeffizienten- Anzahl	Stichproben- umfang	Geschätzte mittlere Korrelation	
			Beobachtet	Korrigiert
<b>Team-Tätigkeiten</b>				
Gewissenhaftigkeit	4	678	,13	,21
Emotionale Stabilität	4	678	,16	,27
Verträglichkeit	4	678	,20	,33
Extraversion	4	678	,14	,22
Offenheit für Erfahrung	4	678	,10	,16
<b>Dienstleistungstätigkeiten mit Kundenkontakt</b>				
Gewissenhaftigkeit	7	908	,20	,29
Emotionale Stabilität	7	908	,08	,12
Verträglichkeit	7	908	,09	,13
Extraversion	6	829	,05	,07
Offenheit für Erfahrung	6	829	,11	,17

Dieses Beispiel verdeutlicht, wie wichtig die richtige Wahl der Konstrukte zur Erfassung von Persönlichkeitsmerkmalen im Rahmen von Anforderungsanalysen ist. Ausgehend von einem Kompetenzansatz würde es für beide Tätigkeitsfelder naheliegen, hohe Anforderungen an die soziale Kompetenz der Beschäftigten zu erwarten. Im Rahmen von Teamtätigkeiten nimmt die soziale Kompetenz aber eine deutlich andere Form an im Vergleich zu Dienstleistungstätigkeiten mit Kundenkontakt. Während es im Team darauf ankommt, sich auf die verschiedenen Persönlichkeiten der Teammitglieder einzustellen, muss im direkten Kundenkontakt die Interaktion im Sinne des Unternehmens gesteuert werden, und das erfordert aus Sicht der die Leistung beurteilenden Führungskräfte v. a. Gewissenhaftigkeit. Diese Merkmale lassen sich aber nur mit differenzierten Persönlichkeitsmodellen abbilden.

Ein Beispiel für die zweite der beiden beschriebenen Formen der personenbezogen-empirischen Methode der Anforderungsanalyse, bei der keine Daten zum Berufs- bzw. Tätigkeitserfolg vorliegen, bietet das Vorgehen von Bergmann und Eder [1, 3]. Vor dem Hintergrund der Theorie der Berufsinteressen von Holland [11] unterscheiden die Autoren sechs verschiedene berufliche Interessensrichtungen:

- Handwerklich-technische Interessen („realistic“ – R)
- Forschende Interessen („investigative“ – I)
- Künstlerische Interessen („artistic“ – A)
- Soziale Interessen („social“ – S)
- Unternehmerische Interessen („entrepreneurial“ – E)
- Interesse an administrativen Tätigkeiten („conventional“ – C)

Die Konstellation dieser Interessen haben die Autoren in 33 praktischen Berufen bei über 800 österreichischen Auszubildenden und Facharbeitern untersucht. Die Tab. 4.2 zeigt die interessenbezogenen Anforderungsprofile unterschiedlicher beruflicher Tätigkeiten.

**Tab. 4.2** Anforderungsprofil verschiedener Berufe auf der Basis von Interessen (AIST-R; nach Bergmann und Eder [1]; die Interessenskalen haben jeweils einen Mittelwert von 100 und eine Standardabweichung von 10 Punkten; die dominanten Interessen sind fett markiert)

Berufe	Mittelwerte der Interessenrichtungen					
	R	I	A	S	E	C
Schlosser	<b>112</b>	96	83	87	95	102
Chemielaborant	101	<b>117</b>	85	86	88	106
Fotograf	102	102	102	96	107	<b>108</b>
Masseur	90	98	90	<b>108</b>	99	101
Kellner	93	85	91	102	112	<b>115</b>
Buchhalter	84	90	80	89	97	<b>120</b>

R = realistic; I = investigative; A = artistic; S = social; E = entrepreneurial; C = conventional interests.



Demnach unterscheiden sich die Berufe hinsichtlich ihrer Interessenschwerpunkte doch recht deutlich: Bei Schlossertätigkeiten stehen handwerklich-technische Interessen im Vordergrund, bei Chemielaboranten forschende Interessen, bei Masseur-tätigkeiten soziale und bei der Tätigkeit von Kellnern unternehmerische und konventionelle Interessen. Wahrscheinlich würden sich die Profile noch deutlicher unterscheiden, hätten die Autoren nur solche Personen in die Stichprobe einbezogen, die bereits mehrere Jahre erfolgreich in dem jeweiligen Beruf tätig waren.

---

## 4.4 Fazit

Angesichts der vielfältigen Methoden und Verfahren zur Durchführung von Anforderungsanalysen kann die Frage nach der „richtigen“ Vorgehensweise gestellt werden (vgl. zum Folgenden [3]). Die Antwort auf diese Frage hängt zunächst immer von der Zielsetzung der jeweiligen Anforderungsanalyse ab. Je nach Zielsetzung – hier sind v. a. Personalplanung, -suche und -auswahl zu nennen, aber auch Stellenbewertung, Personalbeurteilung, Potenzialanalyse, Training, Personalentwicklung, Berufsberatung oder Bildungsplanung bilden in der Praxis wichtige Ziele – stehen andere Aspekte der Anforderungen im Vordergrund und entsprechend ist der Einsatz unterschiedlicher Verfahren zu empfehlen. Während die Methode der kritischen Ereignisse aufgrund ihres Schwerpunkts, der auf der Erfassung erfolgskritischen Verhaltens liegt, für die Training-sentwicklung häufig sehr hilfreich ist, wird beispielsweise in der Berufsberatung gerne die personenbezogen-empirische Methode eingesetzt. Vor der Entscheidung für eine bestimmte Methode muss also immer geprüft werden, ob das entsprechende Verfahren im Sinne der verfolgten Zielsetzung zweckdienlich ist.

Über die Angemessenheit mit Blick auf die Ziele hinaus sollten nach Schuler [31] bei Anforderungsanalysen immer auch Aufgaben- und Ergebnisanforderungen, Verhaltensanforderungen und Eigenschaftsanforderungen berücksichtigt werden. Zudem sollten zur Ermittlung dieser Anforderungen unterschiedliche Verfahren eingesetzt werden. Dies wird als „multimodaler-multimethodaler Ansatz“ bezeichnet. Dahinter steht die Annahme einer Mehrfachbestimmtheit menschlicher Leistungen, d. h. nach diesem Ansatz hat jede Leistung in der Regel nicht nur eine, sondern mehrere Ursachen und macht sich meist nicht nur bei einem, sondern bei mehreren unterschiedlichen Kennwerten bemerkbar. Empirisch konnte mehrfach gezeigt werden, dass ein solches Vorgehen bezüglich der verwendeten Verfahren und der Beurteilerquellen zu besseren Ergebnissen führt [15, 33].

Da Anforderungsanalysen auf Personenmerkmale zielen, sollten an der Erstellung von Anforderungsprofilen immer auch arbeitspsychologisch ausgebildete Fachkräfte beteiligt sein. Ansonsten besteht die Gefahr, auf alltagspsychologische Begriffe „hereinzufallen“. Für jede Anforderung muss der Anforderungsanalytiker angeben können, wie sie sich objektiv, zuverlässig und gültig messen lässt und in welcher begrifflichen und empirischen Beziehung sie zu anderen postulierten Anforderungen steht. Dies können in der Regel

psychologisch geschulte Analytiker besser [14]. Durch die (zusätzliche) Beteiligung von Arbeitsplatzexperten – dazu zählen Stelleninhaber, Vorgesetzte, Kollegen, Fachvertreter aus der Personalabteilung, Mitglieder einer möglicherweise vorhandenen Abteilung für Arbeitssicherheit oder der Personalvertretung – sollen zwei Ziele erfüllt werden: Zum einen helfen sie durch ihr Wissen über die tatsächlichen Gegebenheiten vor Ort sicherzustellen, dass die Anforderungsanalyse inhaltlich valide ist. Durch ihre Einbeziehung wird angestrebt, alle wesentlichen und relevanten Aspekte von Tätigkeiten abzudecken. Ihre Einbeziehung fördert aber auch – und das ist das zweite Ziel – die Akzeptanz der Vorgehensweise im Management, bei der Belegschaft sowie den Arbeitnehmervetretern.

Dabei ist aber nicht zu übersehen, dass dadurch auch neue Fehlermöglichkeiten erzeugt werden, denn Anforderungsanalysen stehen im Zentrum vielfältiger Interessen [3]. Arbeitsplatzinhaber neigen z. B. gerne dazu, bestimmte Aspekte ihrer Tätigkeit so darzustellen, dass sie für ihre Person vorteilhaft erscheinen [20]. Bei langwierigen Analysen können aber auch Ermüdungserscheinungen und damit verbundene Nachlässigkeiten auftreten. Nicht zuletzt kann auch ein Mangel an sprachlichen Fertigkeiten zu verzerrten Einschätzungen führen, weshalb es wichtig ist, die Aussagen von Arbeitsplatzexperten in Bezug auf Plausibilität und Widersprüche zu prüfen. Dafür können beispielsweise Beobachtungsinterviews am Arbeitsplatz durch psychologisch ausgebildete Fachexperten eingesetzt werden [6]. Aber auch psychologisch ausgebildeten Fachexperten können Fehler unterlaufen. Deshalb ist es ratsam, auch deren Übereinstimmung systematisch zu überprüfen [19]. Mit Blick auf psychologisch nicht ausgebildete Arbeitsplatzexperten im Prozess der Erstellung eines Anforderungsprofils bleibt aber festzuhalten, dass sie von der eigentlichen psychologischen Aufgabenstellung, die mit Anforderungsanalysen verbunden ist – die Übersetzung von Arbeitsbedingungen in ein begrifflich konsistentes und empirisch erfassbares Netzwerk von Personenmerkmalen sowie die empirische Prüfung der Gültigkeit – gewöhnlich überfordert sind.

Zwar können heute viele Anforderungsmerkmale durch gut geprüfte Erhebungsinstrumente erfasst werden [8, 27] – besonders zuverlässig lassen sich Anforderungen in den Bereichen Intelligenz, Persönlichkeit und Berufsinteressen erfassen (ein Überblick über Instrumente zur Erfassung von Anforderungen aus dem Bereich sozialer Fertigkeiten und Kompetenzen findet sich bei Kanning [12]). Dem stehen aber gravierende Änderungen im Arbeitsumfeld entgegen: Der durch die wirtschaftliche Globalisierung sowie die neuen Informations- und Kommunikationstechnologien bedingte Wandel des Arbeits- und Wirtschaftslebens hat auch für künftige Anforderungsanalysen große Auswirkungen [29]. Im Zuge dieser Entwicklungen werden die Arbeitsbedingungen zunehmend standardisiert, wobei die individuellen Handlungsspielräume zunehmen werden. Damit einher geht eine relativ schnelle Änderung der konkreten Arbeitsanforderungen, neue Berufe entstehen und die Ausführungsbedingungen traditioneller Berufstätigkeiten verändern sich. Aufgrund dieser Entwicklungen werden tätigkeitsübergreifende Arbeitsanforderungen wie Intelligenz und Lernfähigkeit, Leistungsmotivation, soziale Kompetenz,

Gewissenhaftigkeit und Selbstvertrauen immer wichtiger werden (zur Erfassung der Intensität der Änderungen in den Anforderungen vgl. [17]).

Eine bisher in der Praxis wenig genutzte Möglichkeit, für neue Berufe und Tätigkeiten Anforderungsprofile zu entwickeln, bietet die Synthetisierung eigenschaftsbezogener Anforderungsprofile. Das Vorgehen basiert auf folgenden Grundgedanken [31]: Für bekannte Arbeitstätigkeiten wird eine Zerlegung in deren einzelne Arbeitselemente vorgenommen (Tätigkeitskomponenten-Matrix). Die einzelnen Arbeitselemente werden anschließend von psychologischen Experten daraufhin beurteilt, wie wichtig bestimmte stabile Personenmerkmale (Fähigkeiten, Persönlichkeitsmerkmale, Interessen usw.) für die erfolgreiche Bewältigung dieser Arbeitselemente sind (Matrix aus Tätigkeitselementen und Personenattributen). Werden diese beiden Schritte miteinander kombiniert (Multiplikation der Matrizen), lässt sich daraus ableiten, wie wichtig bestimmte Personenmerkmale für den Erfolg in bestimmten Tätigkeiten sind. Soll eine Anforderungsanalyse für neue Tätigkeiten und Berufe durchgeführt werden, werden für die neue Tätigkeit zunächst die relevanten Arbeitselemente ermittelt. Wichtig ist, dass sich die neue Tätigkeit über eine andersartige Kombination der bereits bekannten Arbeitselemente darstellen lässt. Da für jedes Arbeitselement bereits bekannt ist, welche Relevanz ein bestimmtes Personenmerkmal für die erfolgreiche Bewältigung hat, lassen sich durch die Kombination der beiden Informationsmengen (Synthetisierung durch Matrizenmultiplikation) die Personenanforderungen in den neuen Tätigkeiten ableiten. Aber auch für diesen, als synthetische Validierung bezeichneten Prozess gilt, dass im strengen Sinn nur Anforderungshypothesen entwickelt werden, die noch empirisch zu überprüfen sind.

---

### Fragen

1. Worin besteht der Unterschied zwischen Arbeits- und Anforderungsanalysen?
2. Was versteht man unter einem Kompetenzmodell und welche Funktion erfüllen solche Modelle bei der Auswahl von Mitarbeitern?
3. Welche Art von Anforderungen kann mit der Methode der kritischen Ereignisse besonders gut erfasst werden?
4. Von welcher grundlegenden Annahme geht die arbeitsplatzanalytisch-empirische Methode aus?
5. Wie werden mit der personenbezogen-empirischen Methode Anforderungen ermittelt, sofern Daten zum beruflichen Erfolg von Arbeitsplatzinhabern vorliegen?

---

### Literatur

1. Bergmann, C., & Eder, F. (1992). *Allgemeiner Interessen-Struktur-Test – Umweltstrukturtest AIST/UST, Manual*. Weinheim: Beltz Test.
2. Bisani, F. (1989). Anforderungs- und Qualifikationsprofil. In H. Strutz (Hrsg.), *Handbuch Personalmarketing* (S. 230–243). Wiesbaden: Gabler.

3. Blickle, G. (2014). Anforderungsanalyse. In F. W. Nerdinger, G. Blickle, & N. Schaper (Hrsg.), *Arbeits- und Organisationspsychologie* (3. Aufl., S. 207–222). Heidelberg: Springer.
4. Borkenau, P., & Ostendorf, F. (2008). *NEO FFI – NEO-Fünf-Faktoren-Inventar nach Costa und McCrae* (2. Aufl.). Göttingen: Hogrefe.
5. Champion, M. A., Fink, A. A., Ruggenberg, B. J., Carr, L., Phillips, G. M., & Odman, R. B. (2011). Doing competencies well: Best practices in competency modelling. *Personnel Psychology*, *64*, 225–262.
6. Dunckel, H. (1999). *Handbuch psychologischer Arbeitsanalyseverfahren*. Zürich: Vdf.
7. Dunckel, H., & Resch, M. (2010). Arbeitsanalyse. In U. Kleinbeck & K.-H. Schmidt (Hrsg.), *Arbeitspsychologie. Enzyklopädie der Psychologie D/III/1* (S. 1111–1159). Göttingen: Hogrefe.
8. Erpenbeck, J., & Rosenstiel, L. von. (2007). *Handbuch Kompetenzmessung* (2. Aufl.). Stuttgart: Schäffer-Poeschel.
9. Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*, 327–358.
10. Frieling, E., & Buch, M. (2007). Arbeitsanalyse als Grundlage der Arbeitsgestaltung. In H. Schuler & K.-H. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 117–125). Göttingen: Hogrefe.
11. Holland, J. L. (1997). *Making vocational choices. A theory of vocational personalities and work environments*. Odessa: Psychological Assessment Resources.
12. Kanning, U. P. (2003). *Diagnostik sozialer Kompetenzen*. Göttingen: Hogrefe.
13. Kanning, U. P. (2005). *Soziale Kompetenzen. Entstehung, Diagnose und Förderung*. Göttingen: Hogrefe.
14. Kersting, M., & Birk, M. (2011). Zur zweifelhaften Validität und Nützlichkeit von Anforderungsanalysen für die Interpretation eignungsdiagnostischer Daten. In P. Gellèri & C. Winter (Hrsg.), *Potenziale der Personalpsychologie. Einfluss personaldiagnostischer Maßnahmen auf den Berufs- und Unternehmenserfolg* (S. 83–95). Göttingen: Hogrefe.
15. Koch, A., Strobel, A., Miller, R., Garten, A., Cimander, C., & Westhoff, K. (2012). Never use one when two will do: The effects of a multi-perspective approach on the outcome of job analysis. *Journal of Personnel Psychology*, *11*, 25–102.
16. König, C. J., & Kleinmann, M. (2013). Selbstmanagement. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (3. Aufl., S. 647–674). Göttingen: Hogrefe.
17. Kubicek, B., Paškvan, M., & Korunka, C. (2014). Development and validation of an instrument for assessing job demands arising from accelerated change: The intensification of job demands scale (IDS). *European Journal of Work and Organizational Psychology*, *24*, 898–913.
18. Lievens, F., Sanchez, J. I., Bartram, D., & Brown, A. (2010). Lack of consensus among competency ratings of the same occupation: Noise or substance? *Journal of Applied Psychology*, *95*, 562–571.
19. Morgeson, F. P., & Champion, M. A. (2000). Accuracy in job analysis: Toward an inference-based model. *Journal of Organizational Behavior*, *21*, 819–827.
20. Morgeson, F. P., Delaney-Klinger, K., Mayfield, M. S., Ferrara, P., & Champion, M. A. (2004). Self-presentation process in job analysis: A field experiment investigating inflation of abilities, tasks, and competencies. *Journal of Applied Psychology*, *89*, 674–686.
21. Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, *11*, 145–165.
22. Nerdinger, F. W. (2011). *Psychologie der Dienstleistung*. Göttingen: Hogrefe.
23. Nerdinger, F. W. (2014). Organisationsentwicklung. In F. W. Nerdinger, G. Blickle, & N. Schaper (Hrsg.), *Arbeits- und Organisationspsychologie* (3. Aufl., S. 159–170). Heidelberg: Springer.
24. Nerdinger, F. W., Blickle, G., & Schaper, N. (2014). *Arbeits- und Organisationspsychologie* (3. Aufl.). Heidelberg: Springer.
25. REFA – Verband für Arbeitsstudien und Betriebsorganisation. (1991). *Anforderungsermittlung (Arbeitsbewertung)* (2. Aufl.). München: Hanser.

26. Sackett, P. R., & Laczó, R. M. (2003). Job and work analysis. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Hrsg.), *Handbook of psychology. Industrial and organizational psychology* (12. Aufl., S. 21–37). Hoboken: Wiley.
27. Sarges, W. (2013). *Management-Diagnostik* (4. Aufl.). Göttingen: Hogrefe.
28. Schaper, N. (2014). Arbeitsanalyse und -bewertung. In F. W. Nerdinger, G. Blickle, & N. Schaper (Hrsg.), *Arbeits- und Organisationspsychologie* (3. Aufl., S. 347–370). Heidelberg: Springer.
29. Scherrer, C. (2011). *Globalisierung*. Göttingen: Vandenhoeck & Ruprecht.
30. Schuler, H. (2002). *Das Einstellungsinterview*. Göttingen: Hogrefe.
31. Schuler, H. (2014). *Psychologische Personalauswahl. Eignungsdiagnostik für Personalausscheidungen und Berufsberatung* (4. Aufl.). Göttingen: Hogrefe.
32. Schuler, H., & Höft, S. (2004). Berufseignungsdiagnostik und Personalauswahl. In H. Schuler (Hrsg.), *Organisationspsychologie – Grundlagen und Personalpsychologie. Enzyklopädie der Psychologie Bd D/III/3* (S. 439–532). Göttingen: Hogrefe.
33. Schuler, H., Becker, K., & Diemand, A. (2007). Potentialanalysen als Grundlage von Personalausscheidungen in einer Dienstleistungsorganisation. In H. Schuler (Hrsg.), *Assessment Center zur Potentialanalyse* (S. 297–312). Göttingen: Hogrefe.
34. Schumacher, S., Kleinmann, M., & König, J. C. (2012). Job analysis by incumbents and laypersons. Does item decomposition and the use of less complex items make the ratings of both groups more accurate? *Journal of Personnel Psychology, 11*, 69–76.
35. Sonntag, K. H., & Schaper, N. (1999). Personale Verhaltens- und Leistungsbedingungen. In C. Graf Hoyos & D. Frey (Hrsg.), *Arbeits- und Organisationspsychologie* (S. 298–312). Weinheim: Beltz.
36. Sonntag, K. H., & Schaper, N. (1999). Leitfaden zur qualitativen Personalplanung bei technisch-organisatorischen Innovationen. In H. Dunckel (Hrsg.), *Handbuch psychologischer Arbeitsanalyseverfahren* (S. 285–319). Zürich: Vdf.
37. Sonntag, K. H., Frieling, E., & Stegmaier, R. (2012). *Lehrbuch Arbeitspsychologie* (3. Aufl.). Bern: Huber.
38. Rosenstiel, L. von, & Nerdinger, F. W. von. (2011). *Grundlagen der Organisationspsychologie. Basiswissen und Anwendungshinweise* (7. Aufl.). Stuttgart: Schäffer-Poeschel.

Anja Strobel und Luise Franke-Bartholdt

---

## Zusammenfassung

Im vorliegenden Kapitel stellen wir Interviewverfahren und biografische Fragebogen als zwei valide Instrumente der beruflichen Eignungsdiagnostik vor. Zu Einstellungsinterviews werden Informationen über deren Verbreitung, Akzeptanz und Güte dargestellt. Ferner werden wesentliche Aspekte herausgearbeitet, die die Qualität von Einstellungsinterviews sichern können, die Rolle des Interviewers näher beleuchtet und strukturierte Interviewformen vorgestellt, um damit Empfehlungen zur Planung, Durchführung und Auswertung der Personalauswahlgespräche abzuleiten. Zu biografischen Fragebogen wird herausgearbeitet, dass diese geeignete Prädiktoren für eine Vielzahl unterschiedlicher Kriterien beruflicher Leistung darstellen und einen inkrementellen Erklärungsbeitrag über andere eignungsdiagnostische Maße hinaus liefern. Es wird gezeigt, dass diese wissenschaftlichen Erkenntnisse jedoch nicht mit den Wahrnehmungen in der Anwendungspraxis übereinstimmen, in der biografische Fragebogen vergleichsweise selten eingesetzt und sowohl von Anwendern als auch Bewerbern weniger gut akzeptiert werden.

---

A. Strobel (✉) · L. Franke-Bartholdt  
Persönlichkeitspsychologie und Diagnostik, Institut für Psychologie, TU Chemnitz,  
Chemnitz, Deutschland  
E-Mail: anja.strobel@psychologie.tu-chemnitz.de

L. Franke-Bartholdt  
E-Mail: luise.franke-bartholdt@psychologie.tu-chemnitz.de

## Inhaltsverzeichnis

5.1	Interviewverfahren.....	92
5.1.1	Einsatzhäufigkeit und Akzeptanz von Einstellungsinterviews .....	93
5.1.2	Güte von Interviews als eignungsdiagnostische Instrumente .....	95
5.1.3	Strukturierte Interviewformen.....	100
5.1.4	Die Rolle des Interviewers im Interviewprozess.....	105
5.2	Biografischer Fragebogen .....	106
5.2.1	Einsatzhäufigkeit und Akzeptanz.....	108
5.2.2	Konstruktion von biografischen Fragebogen.....	108
5.2.3	Güte von biografischen Fragebogen .....	115
5.3	Fazit.....	116
	Literatur.....	117

---

### 5.1 Interviewverfahren

Das Interview ist ein effizientes, breit akzeptiertes und daher häufig genutztes Verfahren in der Personalauswahl (z. B. [72, 120, 133]). Huffcut und Culbertson [44] unterstreichen dies mit der Aussage, dass es für Personen praktisch undenkbar sei, einen Auswahlprozess ohne irgendeine Art Interview zu durchlaufen. In der konkreten Anwendung lassen sich große Unterschiede in der Qualität der geführten Interviews finden, wobei Variablen wie der Grad der Strukturierung, die Art der Frageformulierung und das Verhalten des Interviewers eine wesentliche Rolle spielen [131].

Schuler [116, S. 281] konstatiert: „In allergrößter Verdichtung sind es nur zwei Prinzipien, von denen die Tauglichkeit eines Einstellungsinterviews abhängt: 1. Man muss die richtigen Fragen stellen. 2. Man muss die Antworten richtig bewerten.“ Nicht nur in seinen Arbeiten, sondern in der großen Anzahl verfügbarer Literatur über Planung, Durchführung und Auswertung von Einstellungsinterviews wird deutlich, welche Vielfalt von Aspekten im Interviewprozess beachtet werden sollten, um diese beiden Punkte zu erfüllen und damit auch den in der folgenden Definition umschriebenen Anliegen gerecht werden zu können (vgl. u. a. [70, 95, 116]).

► **Definition Einstellungsinterview** „Unter einem Interview als Methode der Personalauswahl ist eine Gesprächssituation zwischen zwei oder mehreren Personen – Repräsentanten der auswählenden Organisation einerseits und Stellenbewerber andererseits – zu verstehen, die Gelegenheit zum Austausch bewerbungsrelevanter person-, arbeits- und organisationsbezogener Information bietet und damit als Grundlage für Auswahlentscheidungen seitens der Organisation und der Organisationswahl seitens der Bewerber dient“ [117, S. 277].

Für die Planung, Durchführung und Auswertung von Einstellungsgesprächen in Deutschland sind wesentliche rechtliche Grundlagen zu beachten [133]. Ableitungen zu zulässigen bzw. unzulässigen Fragen lassen sich bereits aus dem Grundgesetz, Artikel 1



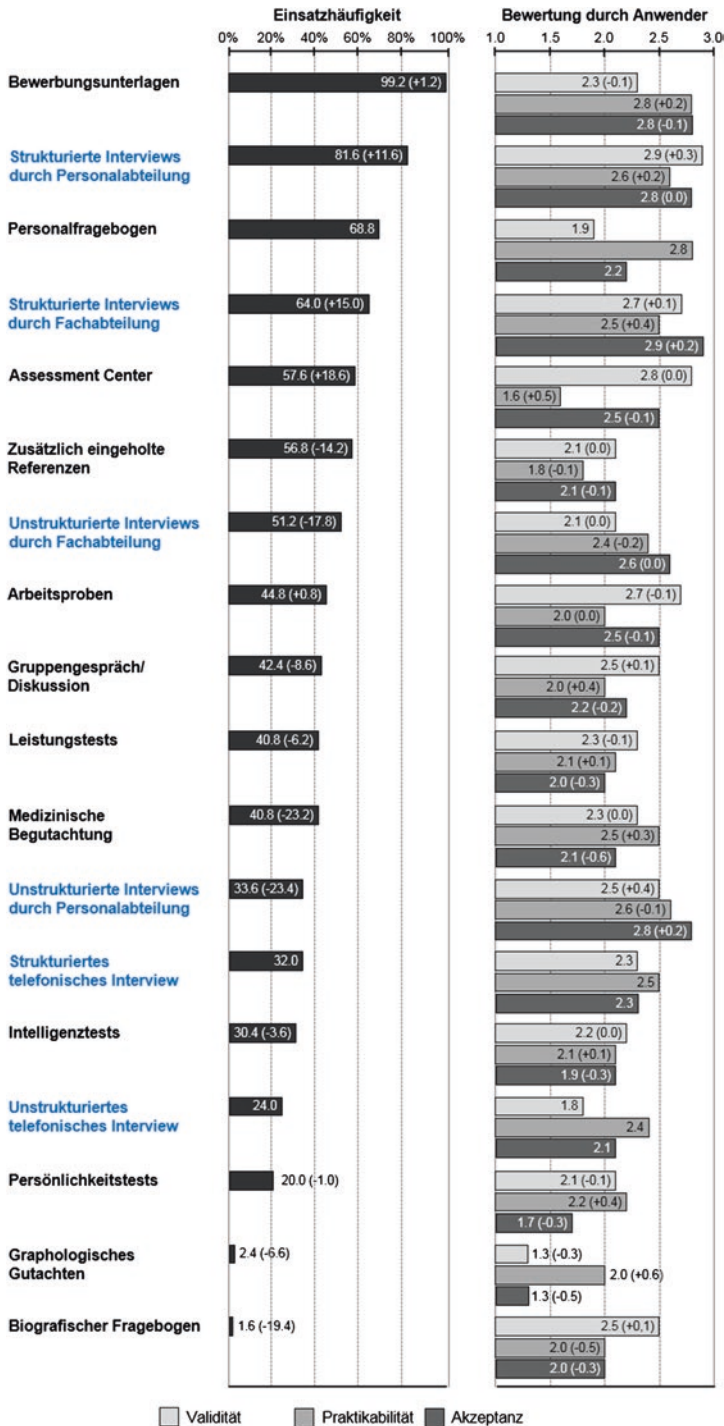
(„Schutz der Menschenwürde“) und Artikel 2 („Allgemeines Persönlichkeitsrecht“), ziehen. Weiterhin relevant sind das Strafgesetzbuch § 201 („Verletzung der Vertraulichkeit des Wortes“), § 202a („Ausspähen von Daten“) und § 203 („Verletzung von Privatgeheimnissen“) sowie das Bundesdatenschutzgesetz. Erfassung, Speicherung und Verwendung von Daten dürfen nur nach vorheriger Information der untersuchten Person und ihrem abgegebenen Einverständnis erfolgen. Die Verwendung darf lediglich zweckgerichtet erfolgen, die erhobenen Informationen grundsätzlich nur für den Zweck verwendet werden, der der Person bekannt gegeben wurde und dem sie zugestimmt hat. Einen umfassenden Überblick über Rechtsfragen im eignungsdiagnostischen Bereich sowie zur Zulässigkeit von Fragen gibt Schuler [116], rechtliche Hintergründe der Eignungsbeurteilung im Allgemeinen stellt Püttner [97] dar.

### 5.1.1 Einsatzhäufigkeit und Akzeptanz von Einstellungsinterviews

Wie bereits eingangs erwähnt, ist das Einstellungsinterview im Personalauswahlkontext weit verbreitet und erfreut sich bei Interviewern wie Interviewten großer Beliebtheit. Eine Übersicht über Einsatzhäufigkeit und Akzeptanz verschiedener Auswahlverfahren in deutschen Unternehmen findet sich in Abb. 5.1 basierend auf den Ergebnissen einer Befragung durch Schuler et al. [120]. Dargestellt ist die prozentuale Einsatzhäufigkeit von Einstellungsinterviews zum Zeitpunkt der Untersuchung und im Vergleich zu einer Befragung von 1993. Darüber hinaus wird die Einschätzung durch Anwender in Bezug auf Validität, Praktikabilität und vermutete Akzeptanz durch Bewerber, die auf einer dreistufigen Skala (gering, mittel, hoch) erfolgten, abgebildet.

Anderson et al. [3] stellen in einer Metaanalyse Ergebnisse von 38 Stichproben aus 17 Ländern bezogen auf die Bewerbereinschätzungen verschiedener Auswahlverfahren zusammen. Im Ergebnis zeigt sich auch hier die positive Beurteilung des Interviewverfahrens: Es findet sich in den übergreifenden Präferenzeinschätzungen der Kandidaten an zweiter Stelle hinter den Arbeitsproben. Auch bei der Beurteilung einzelner Akzeptanzdimensionen (z. B. interpersonelle Wärme, Augenscheinvalidität, Möglichkeit, eigene Fähigkeiten zu zeigen) wird das Interview jeweils äußerst positiv eingeschätzt. Was bedingt diese große Verbreitung und hohe Akzeptanz des Interviews? Als Vorteile des Interviews werden u. a. benannt, dass das Verfahren sehr breit einsetzbar ist und damit die Erhebung verschiedenster Informationen erlaubt und dass ein flexibles Eingehen auf die Kandidaten ebenso möglich ist wie ein gegenseitiger Informationsaustausch. Die Kandidaten können über die Organisation informiert werden, Bedingungen sind vereinbar, eine zweiseitige Entscheidungssituation wird geschaffen [117, 132].

Auch wenn damit im Interview, verglichen zum Beispiel mit standardisierten Testverfahren, bereits mehr Flexibilität und ein stärkeres Aufeinandereingehen möglich sind, kann die Interviewsituation selbst trotzdem als unangenehm und Angst auslösend empfunden werden (vgl. u. a. [76]). Dies kann die Aussagekraft des Interviews mindern [109] und selbst die wahrgenommene Attraktivität und Zufriedenheit mit dem potenziellen Arbeitgeber senken [73]. Trotz dieser problematischen Auswirkungen liegen bisher



**Abb. 5.1** Einsatzhäufigkeit und Akzeptanz verschiedener eignungsdiagnostischer Verfahren (basierend auf Schuler et al. [120]). Werte in Klammern geben jeweils die Veränderungen im Vergleich zu einer Befragung aus dem Jahr 1993 an. Online-Verfahren sind nicht dargestellt

noch vergleichsweise wenige Arbeiten vor, die das Thema umfassend und theoretisch fundiert erkunden. Unter anderem aus diesem Grund entwickelten McCarthy und Goffin [76] ein Instrument, um Interviewangst zu erfassen.

## 5.1.2 Güte von Interviews als eignungsdiagnostische Instrumente

### 5.1.2.1 Grad der Strukturierung

In der praktischen Anwendung von Einstellungsgesprächen lässt sich eine große Breite geführter Interviews von vollkommen unstrukturierten Formen bis hin zu standardisierten Gesprächen finden [138]. Für ein unstrukturiertes Vorgehen werden eine Reihe von Gründen in Betracht gezogen: So könnten Interviewer nicht ausreichend informiert sein, welche Vorteile ein strukturiertes Vorgehen haben kann, sie könnten sich in ihrer Flexibilität und Autonomie eingeschränkt fühlen oder auch die Umsetzung eines stärker strukturierten und standardisierten Vorgehens als zu monoton und anforderungsarm empfinden (vgl. zusammenfassend [138]). In der Außenwirkung nehmen Interviewte unstrukturierte Interviews teilweise positiver wahr als strukturierte, möglicherweise, da sie ein intensiveres Aufeinander Eingehen suggerieren. Nicht zuletzt ist ein strukturiertes Vorgehen in der Regel aufwändiger und kostenintensiver als ein unstrukturiertes Vorgehen, was sich in einer Präferenz für das weniger strukturierte Interview niederschlagen könnte (vgl. zusammenfassend [138]).

Die Strukturierung des Interviewprozesses hat sich jedoch als wesentlicher Aspekt erwiesen, um die Aussagekraft eines Interviews sicherzustellen. Bereits in frühen Analysen zur Qualität von Auswahlgesprächen (z. B. [75, 141]) wiesen die Autoren auf die zentrale Rolle des Interviewformates für die Validität hin. Auch in nachfolgenden metaanalytischen Untersuchungen konnte die Überlegenheit strukturierter Gespräche im Vergleich zu unstrukturierten Gesprächen systematisch belegt werden (vgl. Abschn. 5.1.2.3).

#### Was bedeutet „Strukturierung“ im Interviewprozess?

Campion et al. [15, S. 656] definieren Struktur als „any enhancement of the interview that is intended to increase psychometric properties by increasing standardization or otherwise assisting the interviewer in determining what questions to ask or how to evaluate responses“. Huffcutt und Arthur [43] wiederum begreifen Strukturierung als den Grad der Ermessensfreiheit, der Interviewern beim Führen von Interviews erlaubt ist. Sie unterteilen Interviews in vier verschiedene Strukturierungsstufen und nutzen dabei zwei Dimensionen: (1) den Grad der Standardisierung der Fragen und (2) den Grad der Standardisierung der Bewertung der Bewerberantworten. Wesentlich in beiden Definitionen ist, dass vorab festgelegte Regeln für den gesamten Interviewprozess als hilfreich angesehen werden, um die Prognosegüte des Interviews zu erhöhen [70].

#### Was trägt zur Strukturierung des Interviews bei?

Als wesentlicher Aspekt der Strukturierung wird in Metaanalysen das Stellen gleicher Fragen bei jedem Interviewten, d. h. die Erstellung und Nutzung eines Interviewleitfadens angesehen (vgl. [18, 43, 77]). Während ein unstrukturiertes Gespräch ohne Vorgaben zu Durchführung und Auswertung geführt wird, liegen bei einem teilstrukturierten Interview zumindest die zu stellenden Fragen explizit vor. Bei völliger Strukturierung wiederum

wird häufig auch die Reihenfolge der Fragen festgelegt. Ist zusätzlich auch jeder Schritt der Auswertung und Interpretation der so erhobenen Informationen festgelegt, so liegt ein standardisiertes Interview vor [133].

Campion et al. [15] haben die umfassendste Übersicht über insgesamt 15 Strukturkomponenten und deren Auswirkungen auf die Güte des Interviews vorgelegt. Betrachtet werden sowohl positive als auch potenzielle negative Auswirkungen auf die Reliabilität und Validität des Interviews sowie auf die Einschätzungen durch Anwender. Unterteilt werden die Komponenten in die Kategorien „Inhalt“ (z. B. das Durchführen einer Anforderungsanalyse, das Stellen gleicher Fragen, das Führen längerer Interviews) und „Auswertung“ (z. B. das Erstellen detaillierter Aufzeichnungen, der Einsatz mehrerer Interviewer, das Nutzen verankerter Einstufungsskalen). Während zum Beispiel der Einsatz mehrerer Interviewer eine ganze Reihe förderlicher Aspekte auf Reliabilität und Validität des Interviews haben kann, sind unter Umständen Einbußen in der Akzeptanz durch die Interviewten zu erwarten (vgl. [15]). In ihren u. a. an Campion et al. [15] anknüpfenden Arbeiten beschäftigen sich Chapman und Zweig [17] mit Bedingungen und Konsequenzen für die Strukturierung von Interviews und stellen die Notwendigkeit von Interviewtrainings als bedeutsame Voraussetzung für die Umsetzung von Strukturierungsmaßnahmen heraus.

### **Welche Einflüsse und Verzerrungen sind ohne Strukturierung wahrscheinlicher?**

Eine Reihe von Studien beschäftigt sich mit Einfluss- und Fehlerquellen in unstrukturierten Interviews (vgl. zusammenfassend [132]). Bereits frühe Untersuchungen von Anderson [2] und Mayfield [75] zeigten, dass besonders in unstrukturierten Gesprächen der Interviewer mehr spricht als der Befragte. Dadurch werden weniger Informationen erfasst und die Urteilsbildung folglich verzerrt. Des Weiteren zeigte sich in einer Reihe von Studien, dass Interviewer in unstrukturierten Gesprächen ein Stereotyp des „guten“ Bewerbers bilden und Informationen und Eindrücke infolgedessen nur eingeschränkt verarbeiten, um eigene Hypothesen zu bestätigen (vgl. [95]). Auch die Interpretation und Bewertung einzelner Informationen unterliegt in unstrukturierten Gesprächen stärker Fehlereinflüssen als in strukturiert geführten Interviews (u. a. [75, 137]): So werden Interviewer zum Teil durch negative Informationen stärker beeinflusst als durch positive, wobei eine Rolle spielt, welche Fähigkeiten oder Merkmale eingeschätzt werden (z. B. [80]). Ferner können die Einstellungen des Interviewten in unstrukturierten Gesprächen die Bewertungen der Antworten beeinflussen [26, 75]. Auch Fehler bei der Nutzung von Ratingskalen entstehen vor allem in unstrukturierten Gesprächen, z. B. wenn mehrere Interviewer die gleiche Ratingskala verschieden anwenden [36] oder wenn Ankerheuristiken genutzt werden [57]. Darüber hinaus lässt sich beobachten, dass in unstrukturierten Gesprächen die Interviewer ihre Entscheidung bereits zu einem sehr frühen Zeitpunkt während des Gesprächs treffen [75, 142, 143].

Für das Einstellungsinterview wurden auch Effekte von Impression Management (das Bestreben, einen möglichst guten Eindruck zu machen) bzw. Faking (das Verbergen negativer bzw. Überbetonen oder Erfinden positiver Persönlichkeitsaspekte oder Fähigkeiten, um persönliche Ziele zu erreichen) aus verschiedenen Perspektiven diskutiert. Van Iddekinge et al. [139] verglichen das Antwortverhalten und die resultierenden Ratings zu

Persönlichkeitsdimensionen in Interviews, in denen ehrlich geantwortet werden sollte, mit Interviews, in denen die Kandidaten gebeten wurden, so zu antworten, als seien sie hoch motiviert, die Stelle zu erhalten. Sie konnten zeigen, dass die Teilnehmer in dem Bewerbungsszenario höhere Ratings durch die Beurteiler erhielten als die Teilnehmer in der ehrlichen Bedingung. Des Weiteren zeigte sich in dieser Studie, dass die Angaben im Interview in einem geringeren Ausmaß verfälscht wurden als in Persönlichkeitsfragebogen, was möglicherweise darauf zurückzuführen ist, dass es Bewerbern im Interview schwerer fällt, eigene Angaben zu verfälschen als im Fragebogen (vgl. auch [70]).

Auch in diesem Zusammenhang wurde die Strukturierung von Interviews bereits diskutiert (vgl. [69, 70]) und wird als hilfreich angesehen: So können die Effekte von Impression Management auf die Beurteilung im Interview eingedämmt werden (z. B. [7]). In Bezug auf Faking stellen Levashina und Campion [69] in ihrem Modell die Strukturierung als eine Option dar, die gesamt betrachtet die Möglichkeit zur Verfälschung verringern sollte. Einzelne Strukturierungskomponenten (z. B. transparente Interviewfragen oder situative Fragen) könnten aber durchaus anfälliger für Faking sein (für das vollständige Modell und das Zusammenwirken der Komponenten siehe [69]).

Zusammengefasst stellt Strukturierung ein wesentliches Vorgehen dar, um die Güte im Interviewprozess zu erhöhen. Das Merkmal der Struktur zieht sich dabei durch Planung, Durchführung und Auswertung/Interpretation des Interviews und umfasst eine große Breite von Einzelmaßnahmen, die in der Summe zu einer größeren Kontrollierbarkeit im Gespräch beitragen und Urteilsfehler minimieren sollen. Was bedeutet dies für die psychometrischen Kennwerte des Interviews?

### 5.1.2.2 Inter-Rater-Reliabilität und interne Konsistenz

Bezogen auf die Inter-Rater-Reliabilität (also die Übereinstimmung von Beurteilern in einem Interview) findet sich in frühen Übersichtsarbeiten eine große Breite von Koeffizienten von sehr niedriger bis nahezu perfekter Übereinstimmung sowohl für die Beurteilung einzelner Kriterien als auch für die Gesamtbeurteilung der Eignung im Interview (vgl. u. a. [141]). Conway et al. [18] arbeiteten in einer Metaanalyse von 111 Übereinstimmungskoeffizienten (Mittelwert ,70, gewichtet nach Stichprobengröße) Variablen heraus, die die Objektivität von Auswahlinterviews mitbestimmen. Sie fanden einen signifikant positiven Einfluss des Strukturierungsgrades der Interviews auf die Objektivität. Für unstrukturierte Gesprächsformen wurden Koeffizienten um ,34 ermittelt, für teilstrukturierte Vorgehensweisen lagen die Werte bei ,56, bei strukturierten Interviews bei ,67. Auch in der aktuellen Metaanalyse von Huffcutt et al. [49] zeigte sich Strukturierung als entscheidender Einflussfaktor: Bei einer Einteilung in fünf verschiedene Level von Struktur stiegen über alle einbezogenen Interviewarten hinweg betrachtet die Koeffizienten für die Inter-Rater-Reliabilität von im Mittel ,36 auf Werte nahe ,80 für die beiden Gruppen mit der stärksten Strukturierung. Die Autoren weisen gleichzeitig darauf hin, dass es erhebliche Unterschiede in der Abschätzung der Kennwerte in Abhängigkeit davon gibt, ob Panel- oder Einzelinterviews betrachtet werden, wobei bei Panel-Interviews eine Überschätzung der Inter-Rater-Reliabilität zu beobachten ist [49].

Auch die interne Konsistenz der Interview-Ergebnisse wurde zur Beurteilung der Interviewgüte bereits vielfach herangezogen. Conway et al. [18] konnten in einer Metaanalyse von 49 Konsistenzkoeffizienten (Mittelwert ,39) wiederum den Strukturierungsgrad der Interviews als moderierende Variable feststellen (zu Reliabilitätskennwerten vgl. zusammenfassend [150]; die Ergebnisse zu spezifischen Interviewformen finden sich unter Abschn. 5.1.3).

### 5.1.2.3 Validität

#### Wie ist die Vorhersagekraft des Einstellungsinterviews einzuschätzen?

Die prädiktive Validität des Einstellungsinterviews wurde vielfach umfassend untersucht. Die Analysen bis zu Beginn der 1980er Jahre ergaben ein eher ernüchterndes Bild mit Kennwerten zur Vorhersage von Berufserfolg, die im Mittel um ,20 und damit deutlich niedriger lagen, als sie zum Beispiel mit ca. ,50 für kognitive Fähigkeitstests ermittelt werden konnten (vgl. u. a. [51, 75, 141]).

Später publizierte Metaanalysen nehmen den Grad der Strukturierung mit auf und ermöglichen so eine differenziertere Betrachtung: Während frühere Arbeiten einen geringen Zusammenhang zwischen Interviewer-Urteil und den späteren Leistungen des Kandidaten im Beruf nahe legen (vgl. [51]), zeigen später durchgeführte und auf wesentlich mehr Einzeluntersuchungen basierende Studien [43, 77, 108, 151], dass sich für das Einstellungsinterview prädiktive Validitäten finden lassen, die mit denen kognitiver Fähigkeitstests vergleichbar sind [144], wenn das Interview strukturiert durchgeführt wird. Die Koeffizienten in den Analysen liegen dabei im Bereich von ,40 bis ,60. Die aktuellste Metaanalyse von Schmidt und Hunter [108] berichtet für das Interview eine prädiktive Validität von ,51 und damit einen Koeffizienten in gleicher Höhe wie für Intelligenztests, die als valideste Prädiktoren für die spätere berufliche Leistung des Bewerbers gelten. Ähnlich hohe Werte (,54) wurden lediglich für die Arbeitsprobe berichtet. Huffcutt und Arthur [43] untersuchten verschiedene Stufen von Strukturierung und zeigten, dass die prädiktive Validität generell mit erhöhter Strukturierung wächst, es jedoch ein Optimum an Strukturierung zu geben scheint, über das hinaus keine Validitätssteigerung mehr erfolgt.

Jetter [55, S. 207] schlussfolgert entsprechend zurecht, „(...) dass nicht die Interviewmethode per se invalide ist, sondern die unter diesem Sammelbegriff praktizierte Vielfalt subjektiver ‚Subtechniken‘“. Notwendig für eine hohe prädiktive Validität ist die Umsetzung grundlegender Gestaltungsempfehlungen zur Strukturierung des Vorgehens (vgl. Abschn. 5.1.3).

#### Welchen Zugewinn hat das Einstellungsinterview über Intelligenztestergebnisse hinaus?

In der bereits angesprochenen Metaanalyse von Schmidt und Hunter [108] wurde ebenfalls die inkrementelle Validität des strukturierten Interviews über Intelligenztests hinaus bestimmt. Tatsächlich lässt sich durch die Kombination von Intelligenztest und strukturiertem Interview ein Validitätszuwachs auf insgesamt ,63 (also ein Zugewinn von ,12 über



die grundständige prädiktive Validität des Intelligenztests hinaus) erreichen. Schuler [116] berichtet zur inkrementellen Validität von Einstellungsinterviews, dass strukturierte Interviews vor allem bei Verwendung biografischer Fragen eine hohe inkrementelle Validität über die Vorhersagekraft von Intelligenztests hinaus erreichen können (v. a. auch in Abhängigkeit der jeweiligen Interviewer), während unstrukturierte Interviews lediglich einen mäßigen Zugewinn zeigen. Darüber hinaus lässt sich inkrementelle Validität vorrangig strukturierter Interviews auch gegenüber Schul- und Examensnoten, Gewissenhaftigkeitstests sowie allgemein Persönlichkeitsverfahren und biografischem Fragebogen (Abschn. 5.2) zeigen, letzteres insbesondere im Falle geringer Fragebogenwerte (vgl. zusammenfassend [116]).

### **Was wird im Einstellungsinterview erfasst?**

Die Konstruktvalidität des Einstellungsgespräches wurde vielfältig diskutiert. Ähnlich zur Befundlage der prädiktiven Validität wurden in den 1980er Jahren überwiegend niedrige Zusammenhänge von unter ,20 zwischen Beurteilungen aus dem Interview und den Ergebnissen kognitiver Fähigkeitstests berichtet. Ein ähnliches Bild zeigt sich auch für Zusammenhänge mit Persönlichkeitsverfahren oder Instrumenten zur Erfassung der Motivations- und Interessenlage (für einen Überblick vgl. [136]). Spätere Arbeiten wie die Metaanalyse von Huffcutt et al. [45] berichten höhere Zusammenhänge zwischen Interview-Ergebnissen und Intelligenz mit einer mittleren korrigierten Korrelation von ,40, Während Berry et al. [10] wiederum lediglich einen Zusammenhang von ,27 berichten, kommen Roth und Huffcutt [101] in einer Reanalyse dieser Arbeit zu dem Ergebnis, dass der korrigierte Zusammenhang zwischen Interviewbeurteilung und Ergebnissen zu kognitiven Fähigkeiten bei ,42 liegt, der bereits von Huffcutt et al. [45] benannte Koeffizient von ,40 also als guter Schätzer für diesen Zusammenhang angesehen werden kann [101].

Zur Frage, welche Merkmale in strukturierten und unstrukturierten Interviews erfasst werden, finden sich in der Literatur unterschiedliche Auffassungen [150]. Hunter und Hirsh [50] gehen davon aus, dass unstrukturiert geführte Interviews eher soziale Kompetenzen, strukturierte Gespräche hingegen auch kognitive Fähigkeiten erfassen würden. Salgado und Moscoso [107] zeigen in einem metaanalytischen Vergleich der Konstruktvalidität von konventionell geführten mit strukturierten Interviews, dass strukturierte Gespräche eher Fachkenntnisse und Berufserfahrung erfassen, konventionell geführte Gespräche wiederum eher Intelligenz und Persönlichkeitsmerkmale wie Extraversion und Neurotizismus oder aber soziale Kompetenzen. Huffcutt et al. [46] fanden in einer Metaanalyse, dass Persönlichkeitsmerkmale (vor allem Gewissenhaftigkeit) sowie soziale Fähigkeiten die am häufigsten erfassten Aspekte im Interview waren. Weiterhin standen intellektuelle Fähigkeiten sowie fachliches Wissen und Können im Fokus, während unstrukturierte Interviews eher die allgemeine kognitive Leistungsfähigkeit sowie Informationen zu Interessen, Training und Bildung erfassen. Die Informationen aus den strukturierten Interviews wiesen einen höheren Anforderungsbezug auf und hatten dadurch auch eine größere Aussagekraft für den Berufserfolg [150].

Insgesamt betrachtet gibt es zur Konstruktvalidität des Einstellungsinterviews noch eine Reihe offener Fragen (vgl. [44, 81]). Vor diesem Hintergrund stellt Huffcutt [42] ein Rahmenmodell zur Konstruktvalidität von Interviews vor, in dem als wesentliche Einflussgrößen für die abgegebenen Beurteilungen arbeitsbezogene Interviewinhalte (z. B. arbeitsbezogenes Wissen, Erfahrung), die Performanz der Interviewten (z. B. soziale Fertigkeiten, sprachliche Ausdrucksfähigkeit) sowie persönliche/demografische Charakteristika (z. B. Attraktivität, Geschlecht) benannt werden. Die bis dato vorhandenen Befunde zu Einzelaspekten des Modells werden umfassend von Huffcutt [42] dargestellt.

### 5.1.3 Strukturierte Interviewformen

Wie bereits herausgearbeitet wurde (vgl. Abschn. 5.1.2.1), hat der Grad der Strukturierung einen entscheidenden Einfluss auf die Güte der geführten Interviews. In der Literatur zum Einstellungsinterview finden sich mannigfaltige Gestaltungsempfehlungen, die jeweils einzelne Aspekte des Interviewprozesses näher beleuchten und Strukturierungsempfehlungen für die Planung, Durchführung oder Auswertung der Gespräche geben (vgl. zusammenfassend [44, 70, 132]). Eine Zusammenstellung von Handlungsempfehlungen zum strukturierten Vorgehen im gesamten Interviewprozess findet sich in den Interview-Standards des Arbeitskreises Assessment Center e. V. [4]. Die Standards basieren auf empirischen Befunden zu den Phasen des Interviews und leiten daraus praktikable Umsetzungshinweise ab, die jeweils in Form eines Kernsatzes mit der Hauptaussage, dem damit einhergehenden Nutzen, konkreten Umsetzungshinweisen und Beispielen für Verstöße aufgebaut sind [131]. Die Standards mit ihren jeweiligen Kernsätzen finden sich in Abb. 5.2.

Weiterhin liegen eine Reihe strukturierter Interviewformen vor, speziell für den Eignungsbereich. Die Interviews sind verschiedenen Grundannahmen verpflichtet, die sich in erster Linie in der Art zu stellender Fragen niederschlagen.

#### 5.1.3.1 Behavior Description Interview

Das (Patterned) Behavior Description Interview (PBDI; [53, 54]) ist ein biografieorientiertes Verfahren der Berufseignungsdiagnostik und folgt entsprechend der Grundannahme, dass vergangenes Verhalten den besten Prädiktor für Verhalten in zukünftigen Situationen darstellt. Basierend auf der Critical Incident Technique [29] werden Situationen und die darin vorkommenden Verhaltensweisen erfasst, die für die interessierende Tätigkeit erfolgsentscheidend sind (zu Anforderungsanalysen vgl. Kapitel von Nerdinger in diesem Band). Diese Ereignisse werden dann zu 5 bis 10 Leistungsdimensionen gruppiert, wobei zu jeder Dimension jeweils mindestens 10 bis 20 Ereignisse gesammelt werden [117]. Im Interview werden dann wiederum zu jeder Dimension mehrere Fragen gestellt, konkretisierende Nachfragen sollen ein möglichst umfassendes Bild der Situation und des erfolgsrelevanten Verhaltens zeichnen.

Das PBDI umfasst fünf Phasen: In Phase 1 werden überprüfbare Fakten (z. B. biografische Daten) abgefragt. Darauf folgen in Phase 2 Fachkenntnisse und Fertigkeiten. In Phase 3 sollen eigene Erfahrungen und bisherige Aktivitäten beschrieben werden,



<b>Standard 1: Auftragsklärung</b>
Vor dem Einsatz von Interviews werden Ziele, Rahmenbedingungen und Konsequenzen geklärt und an die beteiligten Personen kommuniziert.
<b>Standard 2: Arbeits- und Anforderungsanalyse</b>
Ein Interviewprozess lässt sich nur nach Analyse der konkreten Anforderungen sinnvoll gestalten.
<b>Standard 3: Interviewdesign</b>
Im Interviewdesign werden die Ergebnisse aus Auftragsklärung und Anforderungsanalyse in eine strukturierte und sozial angemessene Dialogsituation umgesetzt.
<b>Standard 4: Vorbereitung und Training der Interviewer</b>
Nur qualifizierte und gut vorbereitete Interviewer können fundierte Entscheidungen treffen und das Unternehmen angemessen repräsentieren.
<b>Standard 5: Durchführung</b>
Die konsequente Umsetzung des Interviewdesigns und die Anwendung geeigneter Techniken der Gesprächsführung gewährleisten einen transparenten und zielführenden Ablauf.
<b>Standard 6: Protokollierung und Auswertung</b>
Auf der Grundlage einer sorgfältigen Protokollierung der Gesprächsinhalte werden die erhobenen Informationen systematisch und anforderungsorientiert ausgewertet und zu einer Entscheidung verdichtet.
<b>Standard 7: Feedback</b>
Ein individuelles und anforderungsorientiertes Feedback ist unerlässlich.
<b>Standard 8: Dokumentation und Evaluation</b>
Ergebnisse werden so dokumentiert, dass diese für Dritte nachvollziehbar und für durchzuführende Evaluationen verfügbar sind.

**Abb. 5.2** Die Interview-Standards des Arbeitskreises Assessment Center e. V. [4]

während Phase 4 Bewertungen und Selbsteinschätzungen umfasst. In der fünften und zentralen Phase werden Fragen nach dem tatsächlichen Verhalten in vergangenen erfolgsentscheidenden Situationen gestellt (vgl. [118]). Die Antworten werden protokolliert und anschließend bewertet. Dabei sollte die Bewertung zunächst für jede der relevanten Dimensionen einzeln vorgenommen und mittels einer fünfstufigen Skala quantifiziert werden, wobei Taylor und O'Driscoll [134] Abstufungen von (1) „weit unter den Anforderungen“ über (3) „erfüllt gerade die Anforderungen“ bis (5) „weit über den Anforderungen“ empfehlen. In einem nächsten Schritt kann eine weitere fünfstufige Skala zum Einsatz kommen, die dem Vergleich mit anderen Bewerbern dient. Auf dieser Prozentrangskala wird die Leistung des Interviewten von (1) „gehört zu den untersten 20 % aller Bewerber“ bis (5) „gehört zu den obersten 80 bis 100 % aller Bewerber“ eingestuft [134].

Die Tatsache, dass Kandidaten auch unterschiedliche Fragen gestellt werden und konkretisierende Nachfragen angedacht sind, schlägt sich in einer verhältnismäßig geringen Inter-Rater-Reliabilität von im Mittel ,50 nieder (z. B. [52]). Ein deutlich positiver Befund lässt sich für die prognostische Validität finden: Verschiedene Studien berichten

Werte zwischen ,48 und ,54 für die prognostische Güte des PBDI (z. B. [36, 52, 53, 90]), was den aus Metaanalysen bekannten Kennwerten für strukturierte Interviews (z. B. [108]) entspricht.

### 5.1.3.2 Situational Interview

Im Gegensatz zum PBDI ist das Situational Interview (SI; [65]) nicht biografieorientiert, sondern erfasst Informationen darüber, wie sich der Bewerber in bestimmten zukünftigen Situationen verhalten würde [133]. Basierend auf der Goal-Setting-Theorie [71] wird im SI davon ausgegangen, dass im Gespräch geäußerte Absichten prädiktiv valide für die Vorhersage von Verhalten sind. Auch das SI basiert auf der Critical Incident Technique [29], d. h. auch hier basieren die Fragen auf vorher erhobenen erfolgsentscheidenden Situationen und dem darin relevanten Verhalten. Die Fragen sollen sehr sorgfältig konstruiert und vor dem tatsächlichen Einsatz getestet werden, um nicht differenzierende Fragen oder geringe Beurteilerübereinstimmungen zu vermeiden (zur vollständigen Entwicklung des SI siehe [64]).

Im Gespräch werden dem Bewerber ca. 20 bis 30 verschiedene situative Fragen gestellt, für die er jeweils angeben soll, wie er sich in der jeweiligen Situation verhalten würde. Diese werden nicht zu Dimensionen zusammengefasst, sondern die Auswertung erfolgt jeweils situationsbezogen, direkt nach der Frage. Über alle Fragen hinweg wird dann ein Gesamtwert gebildet. Für die Auswertung der Antworten werden von Experten vorab fünfstufige verhaltensverankerte Skalen (Behaviorally Anchored Rating Scales, BARS) entwickelt, die typische erwartete Antworten für gute, akzeptable und unzureichende Antwortmöglichkeiten enthalten [64].

Durch das hochstrukturierte Beurteilungssystem können für das SI sehr hohe Inter-Rater-Reliabilitäten erzielt werden. Latham [64] berichtet über zwischen ,76 und ,96 liegende Koeffizienten. Auch für die interne Konsistenz liegen Angaben von Latham [64] vor: Die berichteten Koeffizienten liegen zwischen ,61 und ,78 mit einem Median von ,71, was die Bildung des Summenwertes für die Gesamteignung rechtfertigt. Zur Frage der prädiktiven Validität werden bei Latham [64] Koeffizienten im Bereich um ,40 berichtet, in der Metaanalyse von Latham und Sue-Chan [66] findet sich ein Koeffizient von ,47 nach Korrektur für Reliabilitäts- und Varianzeinschränkungen.

#### Biografieorientiert oder situativ?

Eine große Anzahl von Studien hat sich mit der Frage beschäftigt, welche Frageform aussagekräftiger ist (für einen Überblick siehe Levashina et al. [70], ein Beispiel für die beiden Fragearten findet sich in Abb. 5.3): Im metaanalytischen Vergleich [135] lassen sich für beide Fragearten ähnliche Inter-Rater-Reliabilitäten finden (situative Fragen: ,79; biografische Fragen: ,76). Bezogen auf die erfassten Inhalte deuten eine Reihe von Studien darauf hin, dass situative Fragen eher arbeitsbezogenes Wissen und kognitive Fähigkeiten erfassen, während biografische Fragen eher Erfahrungen und Persönlichkeitsfacetten abbilden [70]. Weiterhin zeigte sich, dass biografische Fragen typisches Verhalten vorhersagten, während situative Fragen stärker maximales, in geringerem Maße aber auch typisches Verhalten präzisieren konnten [62]. Campion et al. [14] ebenso wie Huffcutt et al. [47] fanden für biografische Fragen eine etwas höhere prädiktive Validität als für situative Fragen. Campion et al. [14] und Krajewski et al. [63] ermittelten inkrementelle

**Situative Frage**

Ihre Arbeitsgruppe arbeitet an einem sehr wichtigen Projekt. Alle Mitglieder möchten eine gute Note für ihre Arbeit erhalten. Sie haben eine knappe Deadline für die Abgabe. Ein Gruppenmitglied war im vergangenen Semester in diesem Gebiet sehr erfolgreich. Mit Unterstützung von zwei anderen Gruppenmitgliedern übernimmt diese Person die Leitung der Gruppe. Sie führt ein Protokoll und lenkt die Diskussionen. Dennoch haben Sie den Eindruck, dass diese Person ausschließlich Ideen würdigt, die ihre eigene Position unterstützen, und Entscheidungen ohne Rücksprache mit den anderen Gruppenmitgliedern trifft. Was würden Sie tun?

---



---



---

**Biografische Frage**

Erzählen Sie mir vor einer Situation, in der eine Person die Leitung eines Gruppenprojektes übernahm und dabei Beiträge von Gruppenmitgliedern ignorierte, die nicht in Einklang mit ihrer eigenen Meinung standen. Wie kam es zu dieser Situation? Was genau haben Sie getan? Was war das Ergebnis?

---



---



---

**Bewertungshinweise**

- 5** andere Gruppenmitglieder einbeziehen; diese nach ihrer Meinung über das Diskussionsthema fragen; alle darum bitten, sich Notizen zu machen; Person mit Gruppenleitung darum bitten, ihre Mitschriften an alle zu senden, um Korrekturen und Ergänzungen vornehmen zu können
- 3** entweder: eine andere Person fragen, ob sie Protokoll führen und an die anderen Gruppenmitglieder versenden kann  
oder: Person mit Gruppenleitung nur dann konfrontieren, wenn ich unzufrieden damit bin, welche Richtung das Projekt nimmt, nichts unternehmen, wenn ich der Ansicht bin, dass die bisherige Lösung gut für das Projekt ist
- 1** nichts tun

**Bewertung**

1 -- 2 -- 3 -- 4 -- 5

**Abb. 5.3** Beispiele für Fragen aus einem Situational Interview und einem Behavior Description Interview. (entnommen aus: Klehe und Latham 2005, S. 111, Tab. 1, eigene Übersetzung) [61]

Validität der biografischen über die situativen Fragen hinaus, während Klehe und Latham [61] den gegenteiligen Befund berichten. Taylor und Small [135] fanden bei hohen Kennwerten für beide Fragetypen eine höhere Validität für biografische Fragen dann, wenn verhaltensverankerte Beurteilungsskalen genutzt wurden. In der Metaanalyse von Huffcutt et al. [48] fand sich bei zunehmender Komplexität der Arbeitsinhalte eine abnehmende Validität situativer, nicht aber biografischer Interviews. Letztlich sprechen die vorliegenden Informationen dafür, die Vorgehensweisen zu kombinieren, um die Vorteile beider Fragearten nutzen zu können.

### 5.1.3.3 Multimodales Interview

Das Multimodale Interview (MMI; [114]) kombiniert die Zugänge biografischer und situativer Fragen und strebt an, durch einen direkten Anforderungsbezug und ein großes Maß an Strukturierung eine hohe Aussagekraft zu erreichen [115]. Es besteht aus acht aufeinanderfolgenden Gesprächsphasen, von denen fünf Phasen zur diagnostischen Urteilsbildung beitragen [117]. Nach einem Gesprächsbeginn, der Selbstvorstellung des Kandidaten sowie einem freien Gesprächsteil werden in der vierten Phase

Berufsinteressen und Hintergründe der Berufs- und Organisationswahl erfasst. In Phase 5 werden biografieorientierte Fragen (vgl. Abb. 5.3) gestellt, dann folgt, basierend auf dem Ansatz der realistischen Tätigkeitsinformation, ein Informationsteil für den Kandidaten. In der siebten Phase werden dann situative Fragen (vgl. Abb. 5.3) gestellt. Danach endet das Gespräch mit einem Gesprächsabschluss, in dem seitens des Kandidaten die Möglichkeit für Rückfragen besteht und das weitere Vorgehen besprochen wird.

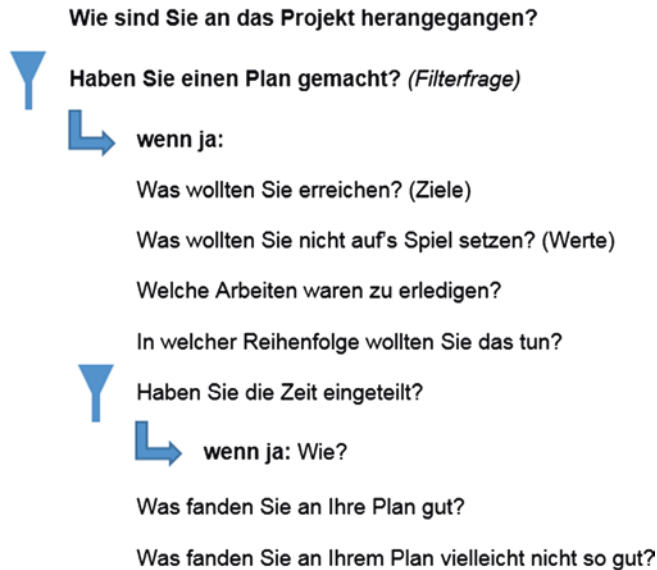
Für die Inter-Rater-Reliabilität des gesamten MMI wurden in verschiedenen Datensätzen Koeffizienten zwischen ,71 und ,92 ermittelt [118], während für die einzelnen Phasen teils erhebliche Unterschiede in der Beurteilerübereinstimmung gefunden wurden (z. B. zwischen ,32 für die Selbstvorstellung und ,90 für die Berufswahlfragen, [119]). Allgemein liegen die Koeffizienten bei trainierten Interviewern um etwa ,10 höher als bei untrainierten oder nur kurz eingewiesenen Personen [118]. Bezogen auf die interne Konsistenz berichtet Schuler [116] aus verschiedenen Arbeiten zum MMI über dimensionsbezogene Reliabilitätsschätzungen für verschiedene Anforderungsdimensionen zwischen ,70 und ,77 mit Gesamtwerten zwischen ,75 und ,87. Zur prädiktiven Validität des MMI lassen sich über eine Reihe Studien hinweg unkorrigierte Koeffizienten für den Gesamtwert zwischen ,30 und ,50 finden [116], im Bereich um ,50 liegen auch Zusammenhänge des MMI mit Außenkriterien in verschiedenen Anwendungskontexten (zusammenfassend s. [118]).

#### 5.1.3.4 Entscheidungsorientiertes Gespräch

Das Entscheidungsorientierte Gespräch (EOG; [145]) als Teil der Entscheidungsorientierten Diagnostik ([147], zuerst 1991) soll Informationen erheben, die es gestatten, Verhalten möglichst gut zu beschreiben, zu erklären oder vorherzusagen [149]. Es wird dabei auf gesicherte empirische Erkenntnisse zur Planung, Durchführung und Auswertung von Interviews Rückgriff genommen, sodass eine regelgestützte Technologie vorliegt, die ebenfalls für den Eignungsbereich anwendbar ist. Der Schwerpunkt des EOG liegt im biografiebezogenen Ansatz. Je nach Interviewten und Fragestellung können auch andere Fragearten wie zum Beispiel situative Fragen zum Einsatz kommen. Auch das EOG hat einen klaren Anforderungsbezug, die Critical Incident Technique [29] wird wiederum zur Erhebung erfolgskritischer Situationen und Verhaltensweisen genutzt.

Der Interviewte wird aufgefordert, sein konkretes Verhalten und Erleben so zu beschreiben, dass der Interviewer es sich möglichst genau vorstellen kann, wobei konkretisierende Nachfragen entlang des Verhaltensdreiecks (Situation – Verhalten – Ergebnis) vorgesehen sind. Ein Ausschnitt eines Leitfadens eines EOG findet sich in Abb. 5.4. Für die strukturierte Planung, Durchführung und Auswertung des EOG existiert eine Sammlung von sieben Checklisten, die zu einem systematischen und verhaltensorientierten Vorgehen im Interviewprozess anleiten. Die Besonderheit des EOG liegt in der Möglichkeit zur Erhebung und Auswertung qualitativer Informationen. Das Interview kann aber mit einer ergänzenden quantifizierenden Aussage über den Grad der Eignung eines Kandidaten enden. Dabei ist es möglich, Items als K.-o.-Kriterien zu definieren [150].

Empirische Studien zur Güte des EOG gibt es nur wenige: Die Inter-Rater-Reliabilität des EOG wurde im Kontext mündlicher akademischer Prüfungen geprüft und erbrachte



**Abb. 5.4** Leitfadenausschnitt aus einem Entscheidungsorientierten Gespräch. (modifiziert entnommen aus: Westhoff und Strobel 2009a, S. 81) [148]

eine Übereinstimmung von ,94 zwischen Prüfer und Beisitzer und ,68 zwischen der Selbsteinschätzung des Prüflings und der Note [146]. Die Studien zur Validität des EOG sprechen für die Vollständigkeit der Regelsysteme [58] sowie deren Lehr- und Lernbarkeit [59].

### 5.1.4 Die Rolle des Interviewers im Interviewprozess

In den vorangegangenen Abschnitten wurde eine Reihe von Einzelmaßnahmen und Interviewformen vorgestellt, die Fehlern im Interviewprozess durch ein hohes Maß von Strukturierung vorbeugen und eine hohe Qualität des Einstellungsgesprächs sicherstellen sollen. Damit – so könnte man zumindest annehmen – dürften Unterschiede zwischen verschiedenen Interviewern keine große Rolle mehr spielen. In der Tat werden bei der korrekten Umsetzung strukturierter Vorgehensweisen Interviewerdifferenzen im Entscheidungsprozess verringert [96], eliminiert werden sie jedoch keinesfalls [34]. Viele Studien zeigen, dass es auch bei gleicher verwendeter Interviewmethode individuelle Unterschiede zwischen Interviewern in der Informationsverarbeitung und der Validität ihrer Entscheidungen gibt (zusammenfassend [132]). Dipboye [24] wies darauf hin, dass Interviewer allmählich ihre eigenen Interviews modifizieren, um sie an ihre Bedürfnisse anzupassen und sich nicht mehr im gleichen Umfang an vorgegebene Durchführungsbestimmungen halten. Dipboye und Jackson [25] hielten fest, dass die verbreitete Annahme, Erfahrung führe zu Expertise, nicht konsistent bestätigt werden bzw. sogar gegenteilige Effekte haben kann (vgl. u. a. [31, 106]).

Enthalten die Gesprächssituationen bzw. die gesamte Informationserhebung jeweils Fragen oder Informationen zu...									
44	den Situationen, in denen das interessierende Verhalten auftritt?	<input type="radio"/>	immer	<input type="radio"/>	häufig	<input type="radio"/>	selten	<input type="radio"/>	nie
45	der genauen Beschreibung der Situationen?	<input type="radio"/>	immer	<input type="radio"/>	häufig	<input type="radio"/>	selten	<input type="radio"/>	nie
46	der Beschreibung, wie das Verhalten in den Situationen ablief?	<input type="radio"/>	immer	<input type="radio"/>	häufig	<input type="radio"/>	selten	<input type="radio"/>	nie
47	Wird dabei nach den beobachtbaren Handlungen gefragt?	<input type="radio"/>	immer	<input type="radio"/>	häufig	<input type="radio"/>	selten	<input type="radio"/>	nie

**Abb. 5.5** Ausschnitt aus dem Diagnoseinstrument zur Erfassung der Interviewerkompetenz in der Personalauswahl (Strobel und Westhoff 2009) [132]. (Copyright: Pearson Assessment)

Um Interviewern eine kontinuierliche Überprüfung des eigenen Vorgehens zu ermöglichen, wurde das Diagnoseinstrument zur Erfassung der Interviewerkompetenz in der Personalauswahl (DIPA; vgl. Abb. 5.5) [132] vorgelegt. Es soll dazu beitragen, die Kriterien an einen qualitativ hochwertigen Interviewprozess präsent zu halten und damit dem Abweichen von strukturierten Vorgehensweisen entgegenzuwirken. Das 146 Items umfassende DIPA besteht aus zwei inhaltlich äquivalenten Varianten: einer für die Beurteilung eines einzelnen Gespräches und einer für die zusammenfassende Beurteilung mehrerer Gespräche des Interviewers. Jede dieser Varianten besteht aus den drei Teilen Planung, Durchführung und Auswertung des bzw. der Interviews. Die Ergebnisse zur Prüfung des DIPA belegen, dass das Instrument eine weitgehend objektive, zuverlässige und valide Beurteilung der Interviewerkompetenz erlaubt und nutzbar ist, um Interviewern in ihrer praktischen Tätigkeit ein regelmäßiges Feedback über die Qualität ihres Vorgehens zu ermöglichen (zusammenfassend [132]).

## 5.2 Biografischer Fragebogen

Die Ursprünge des biografischen Fragebogens liegen mehr als 120 Jahre zurück: Colonel Thomas L. Peters unterbreitete bei einem Treffen amerikanischer Versicherungsgesellschaften im Jahre 1894 den Vorschlag, die Auswahl geeigneter Versicherungsvertreter durch die Verwendung standardisierter Fragen zu demografischen Variablen und bisherigen Erfahrungen der Bewerber zu verbessern. Zu seinen Vorschlägen zählten Fragen nach Beschäftigungsverhältnissen während der letzten zehn Jahre, dem Geburtsort und Familienstand (Ferguson 1961; zitiert nach [92]). Diese Sammlung standardisierter Fragen (im Englischen: application blank) wurde zum Ausgangspunkt des biografischen Fragebogens. Die Messmethode fand schnell Verbreitung, v. a. bei Versicherungsgesellschaften und im militärischen Kontext [40, 126], und wurde stetig weiterentwickelt. In den 1920er Jahren etablierten sich mit der Veröffentlichung von Goldsmith [32], die unterschiedlich gewichtete Items einsetzte, um erfolgreiche von weniger erfolgreichen Verkäufern zu unterscheiden, quantitative Auswertungsstrategien für die Fragebogen in Form von empirisch ermittelten Item-Gewichten (im Englischen: weighted application blank, WAB). In



zahllosen Studien erwiesen sich WAB als geeignete Instrumente, um verschiedene berufsrelevante Kriterien vorherzusagen, und prägten so das Image des biografischen Fragebogens als Messinstrument, dessen Fragen allein auf Basis empirischer Zusammenhänge zu berufsrelevanten Kriterien gewichtet und ausgewählt werden [88]. Dieses Image hat mit der heutigen Form der Messmethode jedoch nur noch wenig zu tun [126].

Biografische Fragebogen werden im Englischen als „biographical data“ (oder kurz: biodata), „background data“ oder „life histories“ bezeichnet. Sie können als Messmethode verstanden werden, da „(...) no two biodata instruments necessarily contain the same items, tap the same constructs, or are scored the same“ [21, S. 1]. Es handelt sich um standardisierte Selbstbeschreibungsinstrumente, bei denen alle Befragten dieselben Fragen beantworten, die zumeist in Form geschlossener Fragen im Multiple-Choice-Format vorgelegt werden. Typischerweise kommen die Instrumente als Papier- und Bleistift-Versionen ohne Zeitbegrenzung zum Einsatz [88].

► **Definition „biografischer Fragebogen“** „Mit biografischen Items werden Personen gebeten, sich an frühere Verhaltensweisen oder Erfahrungen aus Situationen zu erinnern, die wahrscheinlich in ihrem Leben aufgetreten sind, wobei die Personen jeweils aus einer vorgegebenen Liste von Antwortalternativen jene auswählen sollen, die am besten ihr typisches Verhalten und ihre Erfahrungen in den betreffenden Situationen beschreibt (...)“ [83, S. 119 f., eigene Übersetzung].

Die Grundannahme lautet, dass vergangenes Verhalten der beste Prädiktor für zukünftiges Verhalten ist [84, 92]. Dementsprechend erfassen die Instrumente vergangene Verhaltensweisen oder Erfahrungen, von denen angenommen wird oder bekannt ist, dass sie in (kausalem) Zusammenhang zu erfolgsrelevanten Kriterien stehen. Die vergangenen Verhaltensweisen oder Erfahrungen können sich dabei auf den Arbeits- oder Ausbildungskontext, aber auch auf den privaten Bereich beziehen [13]. Die Historizität der Items wird als definitorischer Kernaspekt betrachtet [84]: Erfragt werden vergangenheitsbezogene Inhalte, keine gegenwarts- oder zukunftsbezogenen Absichten und Erwartungen oder Verhaltensweisen in hypothetischen Situationen [74]. Da allen Befragten dieselben Fragen gestellt werden, steht nicht die einzigartige Lebensgeschichte der Bewerber im Vordergrund, sondern vielmehr Unterschiede in den Lebensgeschichten, die für Leistungskriterien aussagekräftig sind [88]. Darüber hinaus wird mit dem biografischen Fragebogen typisches, nicht maximales Verhalten in der Vergangenheit erfasst [84, 128].

Abgesehen von der Ausrichtung auf vergangenheitsbezogene Inhalte bestehen für biografische Fragebogen keine Einschränkungen bezüglich des Messgegenstandes: Die Instrumente sind häufig multidimensional [128] und eignen sich zur Erfassung von Fähigkeiten, Fertigkeiten, Kenntnissen, Eigenschaften, Einstellungen, Werten, Motivation und Interessen [117]. Obwohl sich die verwendeten Fragen von Items aus Persönlichkeitsfragebogen unterscheiden, erfassen beide Methoden häufig ähnliche Merkmale.

Trotz einer nahezu 100-jährigen Forschungstradition zum biografischen Fragebogen existiert keine einheitlich akzeptierte Definition der Messmethode: Viele Autoren haben unterschiedliche Ansichten zu deren Struktur, Inhalt und Auswertung [89, 128]. So wird

der definitorische Aspekt der Historizität keineswegs einhellig geteilt (z. B. [68, 92]). Nicht selten finden sich Fragen zu gegenwärtigen Einstellungen, Interessen oder Werten in biografischen Fragebogen, was es wesentlich erschwert, sie von anderen Selbstbeschreibungsinstrumenten zu unterscheiden (vgl. [5]). Der fehlende historische Bezug der Items einiger biografischer Fragebogen erschwert es zudem, verschiedene Instrumente zu vergleichen oder Studien metaanalytisch zusammenzufassen – auch vor dem Hintergrund, dass viele Autoren keine ausreichenden Angaben zu den verwendeten Items machen [13].

### 5.2.1 Einsatzhäufigkeit und Akzeptanz

In der bereits dargestellten Befragung von Schuler et al. [120] zur Einsatzhäufigkeit eignungsdiagnostischer Instrumente zur externen Personalauswahl in deutschen Unternehmen (Abb. 5.1) gaben lediglich 1,6 % bzw. 3,2 % der Befragten an, biografische Fragebogen zu nutzen. Der beobachtete Rückgang von mehr als 19 % im Vergleich zur Erhebung von 1993 beruht vermutlich auf die Aufnahme von Personalfragebogen als eigenständigem Auswahlinstrument. Auch im europäischen bzw. internationalen Vergleich kommen biografische Fragebogen selten zum Einsatz [105, 122].

Die geringe Einsatzhäufigkeit von biografischen Fragebogen geht vermutlich auf die vergleichsweise geringe Akzeptanz der Messinstrumente seitens der Anwender zurück (z. B. [122]). In der Untersuchung von Schuler et al. [120] schätzten die Befragten die Messmethode bezüglich der Validität, Praktikabilität und vermuteten Akzeptanz durch Bewerber auf einer dreistufigen Skala (gering, mittel, hoch) im niedrigen bis mittleren Bereich ein. In einer Untersuchung von Furnham [30] ordneten 255 Praktiker aus europäischen, überwiegend britischen Firmen mit mehr als 250 Mitarbeitern das Instrument biografischer Fragebogen im Vergleich zu elf anderen Auswahlinstrumenten (z. B. Interviews, kognitive Fähigkeitstests) in Bezug auf Validität, entstehende Kosten, Praktikabilität und Legalität im hinteren Drittel ein.

Aus Bewerbersicht reihen sich biografische Fragebogen in Abhängigkeit der betrachteten Beurteilungsdimension im mittleren [3] bzw. hinteren Feld [37] von Auswahlinstrumenten ein. So konnten Anderson et al. [3] in ihrer metaanalytischen Zusammenfassung zeigen, dass der biografische Fragebogen unter zehn Auswahlinstrumenten bei Kriterien wie der eingeschätzten prädiktiven Validität oder Augenscheinvalidität auf den Plätzen 4 bis 6 landete und in Bezug auf den Schutz der Privatsphäre auf Platz 7 eingestuft wurde.

### 5.2.2 Konstruktion von biografischen Fragebogen

Bei der Konstruktion sind sowohl Aspekte der Item-Generierung als auch Fragen der Gewichtung von Items bzw. der Skalenbildung zu berücksichtigen. Lange Zeit konzentrierten sich die Forschungsbemühungen auf Möglichkeiten der Item-Gewichtung, bevor mit dem wachsenden Bestreben, der Konstruktvalidität des biografischen Fragebogens stärkere Aufmerksamkeit zu schenken, auch Ansätze zur Skalenbildung Einzug in die



Literatur fanden (Abschn. 5.2.2.2). Richtlinien zur Item-Generierung wurden erst vergleichsweise spät thematisiert (erstmal bei [84]).

### 5.2.2.1 Item-Generierung

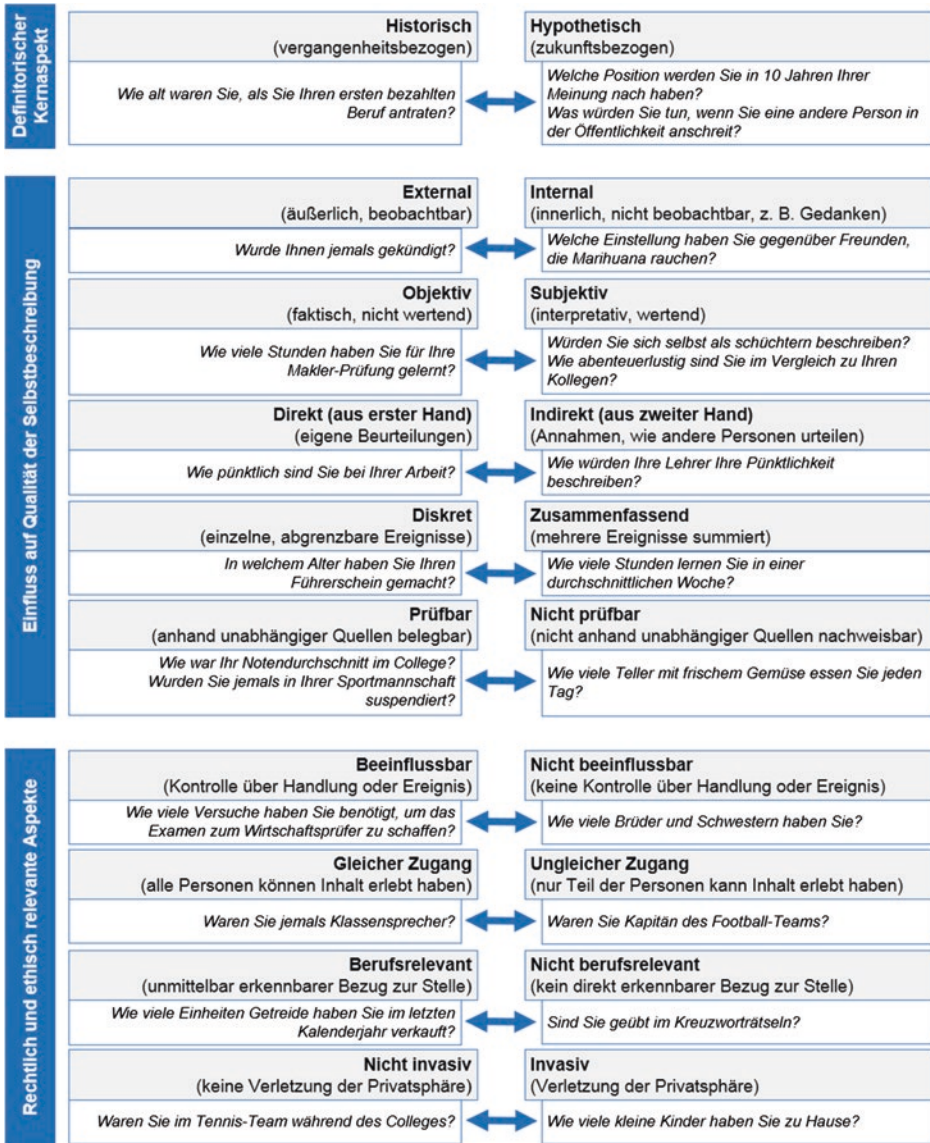
Bei der Generierung biografischer Items sind verschiedene Überlegungen anzustellen, beispielsweise bezüglich der Item-Inhalte, zu verwendeten Informationsquellen, Antwortformate und Kriterien für eine Item-Vorauswahl [84, 88].

#### 1. Item-Inhalt

Auf sehr allgemeiner Betrachtungsebene ist zunächst die grundlegende biografische Natur des Messinstruments zu berücksichtigen [88]. Es sind Verhaltensweisen und Erfahrungen abzubilden, die Bewerber mit großer Wahrscheinlichkeit in ihrem früheren Leben gemacht haben können, d. h. Entwicklungsgeschichte und Erfahrungshintergrund der Zielgruppe müssen berücksichtigt werden [84]. Darüber hinaus empfehlen neuere Ansätze, sich bei der Item-Konstruktion an der Struktur autobiografischer Gedächtnisinhalte zu orientieren, um deren Abruf zu erleichtern [87].

Auf einer spezifischeren Betrachtungsebene ist zu entscheiden, welcher Fokus mit den Item-Inhalten eingenommen werden soll, wobei eine berufsorientierte (job-oriented) und eine personenorientierte (worker-oriented) Perspektive zu unterscheiden sind [86, 127]. Beim berufsorientierten Ansatz werden in den generierten Items Situationen abgebildet, die dem zu erfassenden Kriterium ähnlich sind. So erfassten Barrick und Zimmerman [8] für die Vorhersage des Kriteriums Fluktuation, wie viele Monate ein Bewerber in seinem vorherigen Beruf gearbeitet hatte. Beim personenbezogenen Ansatz werden individuelle Merkmale ermittelt, die das Kriterium vermeintlich beeinflussen. Der Item-Inhalt spiegelt Manifestationen dieser Merkmale in früheren Verhaltensweisen bzw. Erfahrungen wider, wobei konkrete Hypothesen notwendig sind, wie frühere Verhaltensweisen bzw. Erfahrungen den späteren Ausdruck des Merkmals beeinflussen [84].

Auf einer sehr spezifischen Betrachtungsebene ist schließlich zu entscheiden, welche Charakteristika die Items aufweisen sollen. Hierzu wurden verschiedene Taxonomien vorgeschlagen (z. B. [5]), wobei die Taxonomie von Mael [74] die wohl am weitesten akzeptierte darstellt. Darin werden zehn Attribute biografischer Items in drei Kategorien definiert. Die erste Kategorie umfasst das Attribut vergangenheitsbezogen vs. zukunftsbezogen und stellt damit das definitorische Element biografischer Fragebogen dar [74]. Die zweite Kategorie beinhaltet methodisch relevante Attribute, welche die Genauigkeit der Selbstbeschreibung beeinflussen können. Die Attribute der dritten Kategorie betreffen rechtliche und ethische Fragen, die für die Akzeptanz biografischer Items seitens der Bewerber relevant sein können und ihre juristische Angreifbarkeit beeinflussen. Die einzelnen Dimensionen sind zusammen mit Beispiel-Items von Mael [74] in Abb. 5.6 dargestellt. Auch hier wird deutlich, dass die Unterscheidung zwischen Items aus biografischem und Persönlichkeitsfragebogen nicht immer trennscharf gelingt. Einige Studien untersuchten, welchen Einfluss die Manipulation der Attribute von Mael [74] auf die Kriteriumsvalidität und Verfälschung von biografischen Fragebogen hat [9, 33, 68, 78]: Zusammengefasst zeigt sich, dass sich externe, objektive, diskrete und prüfbare Items als besonders aussagekräftig erweisen.



**Abb. 5.6** Attribute biografischer Items nach Mael (1991) [74]. (modifiziert entnommen aus: Mael 1991, S. 773, Tab. 1, eigene Übersetzung) [74]

## 2. Informationsquellen für die Item-Generierung

Es können drei Informationsquellen für die Item-Generierung unterschieden werden [84, 88], die sich nicht gegenseitig ausschließen, sondern gewinnbringend kombiniert werden können (z. B. [127]):

- Anforderungsanalysen (z. B. [111, 112, 129]): Die anforderungsanalytische Fundierung von biografischen Fragebogen wird von verschiedenen Autoren gefordert (z. B. [28, 94]) und entspricht den Richtlinien der DIN 33430 [23] (zu Anforderungsanalysen vgl. Kapitel von Nerdinger in diesem Band).
- Bestehende Theorien und Empirie zur Vorhersagegüte einzelner Merkmale (z. B. [60, 86]): Die Konstruktion der Items beruht auf einer umfangreichen Sichtung theoretischer Grundlagen und empirischer Befunde zum Einfluss eines Konstruktes auf Verhaltensweisen und Erfahrungen über die Lebensspanne.
- Qualitative Methoden (z. B. [104]): Es werden erfolgreiche und weniger erfolgreiche Stelleninhaber gebeten, erfolgskritische Erfahrungen in Essays oder Interviews zu schildern. Einen Überblick über Studien, die diesen Ansatz verfolgten, gibt Russell [103].

### 3. Antwortformat

In biografischen Fragebogen werden Items mit geschlossenem Antwortformat eingesetzt. Dabei können die Anzahl der vorgegebenen (dichotom vs. mehrstufig) und auszuwählenden Antwortalternativen (Einfach- vs. Mehrfachauswahl) ebenso variieren wie deren Kontinuität [92], wobei nicht selten verschiedene Antwortformate in einem einzelnen Fragebogen zum Einsatz kommen. Ebenso wie bei der Generierung des Item-Stammes ist bei der Auswahl geeigneter Antwortalternativen der Erfahrungshintergrund der Zielgruppe zu berücksichtigen. So kann die Antwortoption „nicht anwendbar“ eingesetzt werden, wenn abzusehen ist, dass ein Teil der Bewerber die betreffende Erfahrungen nicht gemacht haben kann [88].

### 4. Item-Vorauswahl

Die Items eines biografischen Fragebogens können von einer einzelnen oder mehreren Personen generiert werden (z. B. [60, 129]). In aller Regel wird eine große Anzahl von Items erzeugt, bei Stokes und Cooper [127] beispielsweise 1000 Items, die jedoch nicht alle vorgegeben, sondern vorab anhand verschiedener Kriterien geprüft werden (sog. „pre-screening“; z. B. [111, 127]). Mumford et al. [86] ließen die generierten Items ihres biografischen Fragebogens u. a. hinsichtlich folgender Kriterien durch eine Expertengruppe bewerten: Relevanz für das Konstrukt, Überschneidungen zu anderen Konstrukten, soziale Erwünschtheit der Antwortoptionen, Verletzung der Privatsphäre und Berufsbezug. Auf Basis der Kriterien wird eine Auswahl geeigneter Items getroffen. Die typische Zahl verwendeter Items liegt nach Mumford et al. [88] zwischen 150 und 300 Fragen, bei Fragebogen mit mehreren Subskalen bei 10 bis 25 Items pro Skala.

#### 5.2.2.2 Item-Gewichtung und Skalenbildung

Der entwickelte biografische Fragebogen wird an einer Konstruktionsstichprobe erprobt. Auf Basis der Ergebnisse erfolgt eine Item-Auswahl, für die sich vier generelle Ansätze etabliert haben [13, 84, 88, 112]: der 1) empirische, 2) rationale, 3) faktorielle und 4) Subgruppen-Ansatz. Bei der Darstellung dieser Ansätze wird häufig nicht ausreichend zwischen der Gewichtung von Items einerseits und Skalenbildung andererseits

differenziert. Während der empirische Ansatz in seiner ursprünglichen Form ausschließlich die Gewichtung von Items betrifft, da eine Bildung von Subskalen nicht vorgesehen ist, muss beim rationalen Ansatz zwischen einer rationalen Form der Item-Gewichtung und einer rationalen Form der Skalenbildung unterschieden werden. Gleiches gilt für den faktoriellen Ansatz. Beim Subgruppen-Ansatz spielen beide Aspekte eine untergeordnete Rolle, da die Gruppierung von Personen (nicht Items) im Vordergrund steht.

### 1. Empirischer Ansatz

Bei diesem Ansatz werden Items oder Antwortoptionen auf Basis ihres statistischen Zusammenhangs zu Kriterien gewichtet bzw. ausgewählt (vgl. [40]). Ziel ist, das Ausmaß aufgeklärter Varianz im interessierenden Kriterium zu erhöhen, wobei dies voraussetzt, dass ein reliables und valides Kriterium vorliegt [83, 84]. Es wird ein Gesamtwert für den biografischen Fragebogen ermittelt, der in aller Regel die lineare Kombination gewichteter Items bzw. Antwortalternativen darstellt. Die vergebenen Gewichte beruhen auf dem Zusammenhang zwischen Item bzw. Antwortoption und Kriterium, der in einer Konstruktionsstichprobe ermittelt wurde. Da die beobachteten Zusammenhänge spezifisch für eine Stichprobe sind, wird eine Kreuzvalidierung an einer (im Idealfall) unabhängigen Stichprobe durchgeführt (vgl. [11]). Dabei ist zumeist eine Abnahme der ermittelten Validitätskoeffizienten zu beobachten [121], typischerweise in Höhe von ,10 [84].

Es gibt verschiedene Methoden, um Items bzw. Antwortalternativen nach dem empirischen Ansatz zu gewichten, z. B. auf Basis von Antworthäufigkeiten, Korrelationen oder Regressionsgewichten (für einen Überblick siehe [19, 40]), wobei die verschiedenen Methoden zumeist vergleichbare Validitätskoeffizienten erbringen [19, 22]. Die wohl bekannteste Methode ist die „Vertical Percent Method“ von England [27], die im Folgenden beispielhaft dargestellt wird.

#### Beispiel für die „Vertical Percent Method“ [27] als Methode zur Item-Gewichtung

Der biografische Fragebogen wird in zwei Extremgruppen eingesetzt, die sich anhand des interessierenden Kriteriums unterscheiden, um die Antworthäufigkeit beider Gruppen bei einzelnen Antwortoptionen zu ermitteln. Die Differenz der prozentualen Antworthäufigkeiten wird mittels Gewichtungstabellen (Strong 1926; dargestellt bei [27]) in sog. Nettogewichte umgewandelt, wobei das Gewicht mit zunehmender Differenz der Antworthäufigkeiten steigt. Mitchell und Klimoski [79] wählten als Kriterium zur Bildung von Extremgruppen den Erhalt einer Maklerlizenz (Lizenz erhalten vs. Lizenz nicht erhalten). In Tab. 5.1 ist ersichtlich, wie sich die Auswahlhäufigkeit einzelner Antwortalternativen bei einem Item zur aktuellen Wohnsituation zwischen den Extremgruppen unterschied und welche Gewichte den Differenzen zugewiesen wurden. Die Antwortoptionen, dass man ein eigenes Haus besitze (positives Gewicht) bzw. eine Wohnung miete (negatives Gewicht), erwiesen sich als besonders aussagekräftig, um beide Gruppen zu unterscheiden.

Nach einer Vereinfachung der Nettogewichte in sog. zugewiesene Gewichte wird der Gewichtungsschlüssel in einer weiteren Stichprobe Kreuzvalidiert. Items oder

**Tab. 5.1** Item-Gewichtung nach der „Vertical Percent Method“ (Quelle: Mitchell und Klimoski, S. 413, Tab. 1, eigene Übersetzung [79])

Ich ...	I	II	III	Nettogewicht
	Mit Lizenz	Ohne Lizenz	(I–II)	
Besitze ein Haus	81	60	21	5
Miete ein Haus	3	5	–2	–1
Miete eine Wohnung	9	25	–16	–4
Wohne bei Verwandten	5	10	–5	–2
	100 %	100 %		

Antwortalternativen, die nicht zwischen den Extremgruppen differenzieren, werden ausgeschlossen bzw. nicht für die Auswertung verwendet. Abschließend werden Cutoff-Werte für Minimalausprägungen bestimmt, wobei maximale Sensitivität bei der Ermittlung erfolgreicher Personen angestrebt wird [27].

Der empirische Ansatz war lange Zeit die am häufigsten verwendete Methode zur Auswahl von Items in biografischen Fragebogen [40]. Die strikte Orientierung an der Kriteriumsvalidität, ohne Berücksichtigung theoretischer, rechtlicher oder ethischer Fragen, wurde dabei als „dustbowl empiricism“ [112] verschiedentlich kritisiert [83, 94]. So führt die fehlende theoretische Fundierung der ausgewählten Items dazu, dass nicht gesagt werden kann, *warum* die Items zur optimalen Auswahl von Bewerbern beitragen (für ein praktisches Beispiel vgl. [99]). Es besteht ferner die Gefahr, Items einzuschließen, die rechtliche Grundlagen verletzen [94]. So ist das in Tab. 5.1 dargestellte Item zwar durchaus geeignet, erfolgreiche von weniger erfolgreichen Personen zu unterscheiden, rechtlich zulässig ist es allerdings nicht [124]. Schließlich kann die strikte empirische Fundierung auch zu Akzeptanzproblemen führen, da nicht immer Fragen mit hoher Augenscheinvalidität ausgewählt werden [98]. Die Kritik an der fehlenden theoretischen Fundierung des empirischen Ansatzes führte zur Entwicklung weiterer Ansätze zur Auswahl von Items, welche die Konstruktvalidität der Instrumente stärker betonen. Die rein empirisch basierte Item-Auswahl wird heute kaum noch eingesetzt [100].

## 2. Rationaler Ansatz

Beim rationalen Ansatz werden Items in der Regel bereits bei ihrer Konstruktion a priori Skalen zugeordnet, die verschiedene Konstrukte abbilden sollen [112]. Wie bei Persönlichkeitsfragebogen erfolgt die Item-Auswahl und Skalenbildung auf Basis einer klassischen Item-Analyse, bei der z. B. Varianz und Trennschärfe der Items sowie die interne Konsistenz der Skalen berücksichtigt werden (z. B. [86, 127]). Auch die Gewichtung der Items bzw. Antwortalternativen kann auf rationalem Wege erfolgen. Da bei rational konstruierten Skalen meist Items mit kontinuierlich angeordneten Antwortalternativen zum Einsatz kommen [88], wird wie bei Persönlichkeitsfragebogen der Antwortalternative,

die der stärksten Ausprägung des interessierenden Konstrukts entspricht, das höchste Gewicht zugewiesen (z. B. [56, 79]). Ein anderer Ansatz besteht darin, Gewichte auf Basis vermuteter Zusammenhänge zu einem Kriterium durch Experten vergeben zu lassen (z. B. [19]). Wie eine Untersuchung von Stokes et al. [130] zeigte, gelingt es Experten jedoch nur bedingt, die Kriteriumsvalidität einzelner Items vorab korrekt zu beurteilen.

### 3. Faktorieller Ansatz

Beim faktoriellen Ansatz (z. B. [6, 20]) erfolgt die Bildung von Skalen post hoc auf Basis beobachteter Item-Zusammenhänge mit Hilfe von Faktorenanalysen [113]. Es lässt sich zeigen, dass der faktorielle Ansatz geeignet ist, zeitlich stabile Faktorenstrukturen abzubilden (z. B. [67]). Allerdings ist der Anteil aufgeklärter Varianz durch die ermittelten Faktoren in aller Regel vergleichsweise gering. Eine faktorielle Gewichtung von Items kann beispielsweise auf Basis ihrer ermittelten Faktorladungen beruhen, nicht selten werden Items jedoch mit Einheitsgewichten zu einem Gesamt-Score pro Faktor summiert [88]. Ferner ist möglich, die Faktorwerte von Personen zu verwenden, um einen Gesamtwert für den biografischen Fragebogen zu ermitteln (z. B. [79]).

### 4. Subgruppen-Ansatz

Während sich empirische und rationale Ansätze auf die Gewichtung und Gruppierung von Items beziehen, zielt der Subgruppen-Ansatz [38, 93] auf die Gruppierung von Personen mit vergleichbaren Entwicklungsverläufen ab. Dazu werden biografische Fragebogen zunächst faktoriell ausgewertet und die resultierenden Faktorprofile einzelner Personen anschließend anhand von Cluster- bzw. Diskriminanzanalysen zusammengefasst. Es folgen Untersuchungen, ob sich die entstandenen Gruppen bezüglich sog. konkurrierender Marker (z. B. Persönlichkeitstests) unterscheiden und unterschiedliche prädiktive Zusammenhänge zu Kriterien aufweisen. Diese Informationen werden genutzt, um die Subgruppen weiter zu charakterisieren [93]. Der Subgruppen-Ansatz wird heute eher selten in der Personalauswahl eingesetzt [88]. Sein historischer Wert besteht v. a. in der Betrachtung verschiedener Entwicklungsverläufe über die Lebensspanne, die Ausgangspunkt für theoretische Überlegungen war, warum biografische Fragebogen valide Prädiktoren für zukünftiges Verhalten sein sollten (z. B. Ecological Model; [85]).

In ihrem Review zum Vergleich des empirischen, rationalen und faktoriellen Ansatzes konnten Hough und Paullin [41] keine klare Überlegenheit eines einzelnen Ansatzes ausmachen. Verschiedene Studien zeigen jedoch, dass v. a. der empirische Ansatz von einem Schrumpfen der Validitätskoeffizienten von Konstruktions- zu Kreuzvalidierungsstichprobe betroffen ist, während die Koeffizienten der beiden anderen Ansätzen vergleichsweise stabil bleiben [79, 112]. Darüber hinaus wurden empirisch ermittelte Item-Gewichtungen in der Vergangenheit bezüglich ihrer hohen Spezifität für einzelne Stichproben und mangelnden Generalisierbarkeit kritisiert (z. B. [51]). Mehrheitlich weisen die Befunde jedoch darauf hin, dass die Gewichtungsschlüssel sorgfältig konstruierter biografischer Fragebogen durchaus über Organisationen und Berufszweige [16, 102],



Länder [20, 39] und Kriterien [21, 91] generalisierbar sind. Die unterschiedlichen Ansätze zur Item-Gewichtung und Skalenbildung schließen sich ferner nicht gegenseitig aus, sondern kommen in neueren Studien häufig gemeinsam zum Einsatz [13, 41]. So kann eine rationale Skalenbildung sinnvoll mit einer empirischen Gewichtung von Items kombiniert werden [56, 99].

### 5.2.3 Güte von biografischen Fragebogen

Der langen Vorherrschaft des empirischen Ansatzes zur Konstruktion von biografischen Fragebogen mag es geschuldet sein, dass sich Studien zur Überprüfung der Güte der Messinstrumente vorrangig auf ihre Kriteriumsvalidität bezogen. Sie sind so zahlreich, dass sie bereits mehrfach metaanalytisch zusammengefasst wurden. Untersuchungen zur Konstruktvalidität (vorrangig rational konstruierter) biografischer Fragebogen liegen hingegen in weit geringerem Umfang vor (z. B. [12, 60, 86, 127]). Da biografische Fragebogen zudem als Messmethode geeignet sind, ganz unterschiedliche Konstrukte zu erfassen, beschränkt sich der folgende Abschnitt auf die Darstellung von Befunden zur Kriteriumsvalidität. Zunächst wird jedoch die Reliabilität der biografischen Fragebogen thematisiert.

#### 5.2.3.1 Reliabilität

Die Bestimmung der internen Konsistenz ist nur bei biografischen Fragebogen sinnvoll, die auf Basis des rationalen oder faktoriellen Ansatzes konstruiert wurden und über homogene Subskalen verfügen. Nach Angaben von Mumford und Owens [84] liegt die interne Konsistenz dieser Fragebogen im Bereich von ,40 bis ,80, nach Breagh [13] im Bereich von ,50 bis ,80. Eine abschließende Bewertung der internen Konsistenz von biografischen Fragebogen erscheint angesichts der Vielzahl unterschiedlich homogener Messgegenstände und stark differierenden Item-Zahlen pro Skala bzw. Faktor in verschiedenen Studien kaum möglich. Eine metaanalytische Zusammenfassung vorliegender Befunde wäre hier wünschenswert.

Bei empirisch konstruierten biografischen Fragebogen ist die Bestimmung der internen Konsistenz aufgrund der Heterogenität der verwendeten Items nicht sinnvoll [103]. Hier ist eine Ermittlung der Retest-Reliabilität angemessen (z. B. [123]), wird aber nur selten umgesetzt. Die ermittelten Stabilitätskoeffizienten liegen zumeist im Bereich von ,60 bis ,90 [13, 40].

#### 5.2.3.2 Kriteriumsvalidität

Biografische Fragebogen haben sich in verschiedenen Metaanalysen [11, 16, 35, 51, 98, 102, 108, 110, 140] als valide Messmethode zur Vorhersage zahlreicher Kriterien beruflichen Erfolges erwiesen: Die Validitätskoeffizienten variieren dabei zwischen ,20 und ,53.

Es ist kritisch hervorzuheben, dass die meisten Studien konkurrenente Designs, bei denen ein aktuell verfügbares Kriterium eingesetzt wird, zur Ermittlung der Kriteriumsvalidität

der biografischen Fragebogen nutzen [13]. Wie Bliesener [11] in seiner Metaanalyse bestätigte, zeigen Studien mit konkurrentem Design in Stichproben aktueller Stelleninhaber höhere Validitätskoeffizienten (im Mittel ,35) als Studien mit prädiktivem Design, die mit Bewerbern ohne (,29) bzw. mit anschließender Auswahlentscheidung (,21) durchgeführt wurden. Es liegen jedoch auch beeindruckende Beispiele für die prädiktive Validität von biografischen Fragebogen vor. So war ein faktoriell gebildeter Fragebogen von Owens und Schoenfeldt [93], der von Studierenden in ihrem ersten College-Semester bearbeitet worden war, in der Lage, bis zu 42 % der Varianz in der Berufswahl 12 bis 17 Jahre nach dem College-Abschluss zu erklären [125].

In verschiedenen Studien konnte gezeigt werden, dass biografische Fragebogen inkrementelle Validität über kognitive Fähigkeitstests bei der Vorhersage unterschiedlicher Kriterien besitzen [1, 21, 56, 82, 130]. Ausnahme bildet die metaanalytische Zusammenfassung von Schmidt und Hunter [108], in welcher der Anteil zusätzlich aufgeklärter Varianz für die Kriterien Vorgesetztenurteile und Trainingserfolg bei 2 % bzw. 0 % über kognitive Fähigkeitstests hinaus lag. Dies geht vermutlich auf die hohe angenommene Korrelation zwischen biografischen Fragebogen und kognitiven Fähigkeitstests von ,50 zurück [128], die sich in anderen Studien so nicht zeigte (z. B. ,19 bei Bobko et al. [12]). Seltener wurde der inkrementelle Beitrag biografischer Fragebogen über Persönlichkeits-tests hinaus untersucht (z. B. [8]), konnte aber v. a. über die Big Five hinaus in einigen Studien nachgewiesen werden [1, 78, 82].

---

### 5.3 Fazit

Interviews sind weit verbreitete und hoch akzeptierte Verfahren in der Eignungsdiagnostik. Über viele Jahrzehnte hinweg wurden in der Forschung zur Güte von Eignungsinterviews die wesentlichen Gestaltungsmerkmale herausgearbeitet, die eine objektive, reliable und valide Planung, Durchführung und Auswertung von Interviews ermöglichen. Ein wesentlicher übergreifender Aspekt, der die Aussagekraft für die Vorhersage von Berufserfolg maßgeblich bestimmt, ist die Strukturierung des Interviewprozesses, die sich u. a. durch eine anforderungsorientierte Planung, die Formulierung von Fragenkatalogen sowie eine regelgeleitete und kriterienorientierte Auswertung umsetzen lässt. In der Gestaltung strukturierter Interviews können sowohl biografische als auch situative Fragen genutzt werden, eine Kombination ermöglicht die Erhebung sowohl vergangenen, typischen als auch zukünftigen und eher maximalen Verhaltens. Auch wenn bereits umfassende Arbeiten zu verschiedenen Aspekten von Eignungsinterviews existieren, so sind noch eine Reihe von Fragen nur ansatzweise betrachtet oder noch unbeantwortet. Gerade die Wirkung einzelner Fragetypen wurde noch nicht umfassend und systematisch untersucht. Auch in der Betrachtung der Interviewten, von Bewerberreaktionen oder Impression-Management-Verhalten gibt es noch eine Reihe Fragestellungen und Herausforderungen für die kommende Arbeit zum Interview.

Auch biografische Fragebogen haben sich als geeignete Instrumente zur Vorhersage einer Vielzahl beruflicher Leistungskriterien erwiesen. Die meisten vorliegenden Studien



nutzten für die Validierung allerdings konkurrenente Designs, sodass weitere Forschung zur Vorhersagegüte der Messmethode notwendig erscheint. Darüber hinaus sollten weitere Anstrengungen unternommen werden, um die Konstruktvalidität der biografischen Fragebogen zu untersuchen. Wie der Beitrag zeigte, haben sich die Konstruktionsansätze für biografische Fragebogen in der nahezu 100-jährigen Forschungstradition weg von einem blinden „dustbowl empiricism“ hin zu rationalen, anforderungsbasierten oder theoriegeleiteten Strategien entwickelt. Diese Ansätze können gewinnbringend eingesetzt werden, um einerseits rechtlich zulässige und akzeptierte Personalauswahlinstrumente und andererseits inhaltlich gehaltvolle Verfahren zu entwickeln, die sich in einem nomologischen Netz verankern lassen und somit eine Prüfung von konvergenter und divergenter Validität erlauben. Dies könnte einen bedeutsamen Beitrag zur Theorieentwicklung und -prüfung in der Personalpsychologie leisten. Schließlich ist zwingend zu fordern, zukünftig eine deutliche Abgrenzung zwischen biografischen Fragebogen und Persönlichkeitsfragebogen vorzunehmen, um den genuinen Charakter der Messmethode zu erhalten. Dies kann erreicht werden, wenn der Historizität von biografischen Fragen als definitorischem Kernaspekt in jedem Falle Rechnung getragen wird.

---

### Fragen

1. Kennzeichnen Sie verschiedene Formen von strukturierten Interviews und legen Sie dar, welche Formen sich für die Praxis aus welchen Gründen als besonders empfehlenswert erweisen.
2. Erläutern Sie, welche Maßnahmen sich als förderlich zur Verbesserung der Aussagekraft von Interviews erwiesen haben.
3. Schildern Sie verschiedene Ansätze zur Item-Gewichtung und Skalenbildung bei biografischen Fragebogen und erläutern Sie deren Vor- und Nachteile.
4. Vergleichen Sie, inwieweit sich die wissenschaftlichen Erkenntnisse zur Güte biografischer Fragebogen in den Wahrnehmungen in der Anwendungspraxis wiederfinden und diskutieren Sie mögliche Ursachen.

---

### Literatur

1. Allworth, E., & Hesketh, B. (2000). Job requirements biodata as a predictor of performance in customer service roles. *International Journal of Selection and Assessment*, 8, 137–147. doi:10.1111/1468-2389.00142.
2. Anderson, C. W. (1960). The relation between speaking times and decision in the employment interview. *Journal of Applied Psychology*, 44, 267–268. doi:10.1037/h0042783.
3. Anderson, N., 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. doi:10.1111/j.1468-2389.2010.00512.x.
4. Arbeitskreis Assessment Center e. V. (2008). *Interview-Standards*. [http://www.arbeitskreis-ac.de/attachments/149\\_akac-interview-standards.pdf](http://www.arbeitskreis-ac.de/attachments/149_akac-interview-standards.pdf). Zugegriffen: 03. Dez. 2015.

5. Asher, J. J. (1972). The biographical item: Can it be improved? *Personnel Psychology*, *25*, 251–269. doi:[10.1111/j.1744-6570.1972.tb01102.x](https://doi.org/10.1111/j.1744-6570.1972.tb01102.x).
6. Baehr, M. E., & Williams, G. B. (1968). Prediction of sales success from factorially determined dimensions of personal background data. *Journal of Applied Psychology*, *52*, 98–103. doi:[10.1037/h0020587](https://doi.org/10.1037/h0020587).
7. 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. doi:[10.1037/a0016532](https://doi.org/10.1037/a0016532).
8. Barrick, M. R., & Zimmerman, R. D. (2005). Reducing voluntary, avoidable turnover through selection. *Journal of Applied Psychology*, *90*, 159–166. doi:[10.1037/0021-9010.90.1.159](https://doi.org/10.1037/0021-9010.90.1.159).
9. Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. *Personnel Psychology*, *45*, 389–406. doi:[10.1111/j.1744-6570.1992.tb00855.x](https://doi.org/10.1111/j.1744-6570.1992.tb00855.x).
10. Berry, C. M., Sackett, P. R., & Landers, R. N. (2007). Revisiting interview-cognitive ability relationships: Attending to specific range restriction mechanisms in meta-analysis. *Personnel Psychology*, *60*, 837–874. doi:[10.1111/j.1744-6570.2007.00093.x](https://doi.org/10.1111/j.1744-6570.2007.00093.x).
11. Bliesener, T. (1996). Methodological moderators in validating biographical data in personnel selection. *Journal of Occupational and Organizational Psychology*, *69*, 107–120. doi:[10.1111/j.2044-8325.1996.tb00603.x](https://doi.org/10.1111/j.2044-8325.1996.tb00603.x).
12. Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology*, *52*, 561–589. doi:[10.1111/j.1744-6570.1999.tb00172.x](https://doi.org/10.1111/j.1744-6570.1999.tb00172.x).
13. Breaugh, J. A. (2009). The use of biodata for employee selection: Past research and future directions. *Human Resource Management Review*, *19*, 219–231. doi:[10.1016/j.hrmr.2009.02.003](https://doi.org/10.1016/j.hrmr.2009.02.003).
14. Campion, M. A., Campion, J. E., & Hudson, J. P. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology*, *79*, 998–1002. doi:[10.1037/0021-9010.79.6.998](https://doi.org/10.1037/0021-9010.79.6.998).
15. Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, *50*, 655–702. doi:[10.1111/j.1744-6570.1997.tb00709.x](https://doi.org/10.1111/j.1744-6570.1997.tb00709.x).
16. Carlson, K. D., Scullen, S. E., Schmidt, F. L., Rothstein, H., & Erwin, F. (1999). Generalizable biographical data validity can be achieved without multi-organizational development and keying. *Personnel Psychology*, *52*, 731–755. doi:[10.1111/j.1744-6570.1999.tb00179.x](https://doi.org/10.1111/j.1744-6570.1999.tb00179.x).
17. 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. doi:[10.1111/j.1744-6570.2005.00516.x](https://doi.org/10.1111/j.1744-6570.2005.00516.x).
18. Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology*, *80*, 565–579. doi:[10.1037/0021-9010.80.5.565](https://doi.org/10.1037/0021-9010.80.5.565).
19. Cucina, J. M., Caputo, P. M., Thibodeaux, H. F., & Maclane, C. N. (2012). Unlocking the key to biodata scoring: A comparison of empirical, rational, and hybrid approaches at different sample sizes. *Personnel Psychology*, *65*, 385–428. doi:[10.1111/j.1744-6570.2012.01244.x](https://doi.org/10.1111/j.1744-6570.2012.01244.x).
20. Dalessio, A. T., Crosby, M. M., & McManus, M. A. (1996). Stability of biodata keys and dimensions across English-speaking countries: A test of the cross-situational hypothesis. *Journal of Business and Psychology*, *10*, 289–296. doi:[10.1007/BF02249604](https://doi.org/10.1007/BF02249604).

21. Dean, M. A. (2004). An assessment of biodata predictive ability across multiple performance criteria. *Applied H.R.M. Research*, 9, 1–12.
22. Devlin, S. E., Abrahams, N. M., & Edwards, J. E. (1992). Empirical keying of biographical data: Cross-validity as a function of scaling procedure and sample size. *Military Psychology*, 4, 119–136. doi:[10.1207/s15327876mp0403\\_1](https://doi.org/10.1207/s15327876mp0403_1).
23. DIN Deutsches Institut für Normung e. V. (2016). *DIN 33430: Anforderungen an berufsbezogene Eignungsdiagnostik*. Berlin: Beuth.
24. Dipboye, R. L. (1994). Structured and unstructured selection interviews: Beyond the job-fit model. In G. R. Ferris (Hrsg.), *Research in personnel and human resources management* (12. Aufl., S. 79–123). Greenwich: JAI Press.
25. Dipboye, R. L., & Jackson, S. (1999). Interviewer experience and expertise effects. In R. W. Eder & M. M. Harris (Hrsg.), *The employment interview handbook* (S. 259–278). Thousand Oaks: Sage.
26. Dipboye, R. L., Stramler, C., & Fontanelle, G. A. (1984). The effects of the application on recall of information from the interview. *Academy of Management Journal*, 27, 561–575. doi:[10.2307/256045](https://doi.org/10.2307/256045).
27. England, G. W. (1971). *Development and use of weighted application blanks* (überarb. Aufl.). Oxford: U. Minnesota Industrial Relations.
28. Fine, S. A., & Cronshaw, S. (1994). The role of job analysis in establishing the validity of biodata. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Hrsg.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (S. 39–64). Palo Alto: CPP Books.
29. Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327–358. doi:[10.1037/h0061470](https://doi.org/10.1037/h0061470).
30. Furnham, A. (2008). HR professionals' beliefs about, and knowledge of, assessment techniques and psychometric tests. *International Journal of Selection and Assessment*, 16, 300–305. doi:[10.1111/j.1468-2389.2008.00436.x](https://doi.org/10.1111/j.1468-2389.2008.00436.x).
31. Gehrlein, T. M., Dipboye, R. L., & Shahani, C. (1993). Nontraditional validity calculations and differential interviewer experience: Implications for selection interviews. *Educational & Psychological Measurement*, 53, 457–469. doi:[10.1177/0013164493053002015](https://doi.org/10.1177/0013164493053002015).
32. Goldsmith, D. (1922). The use of the personal history blank as a salesmanship test. *Journal of Applied Psychology*, 6, 149–155. doi:[10.1037/h0073592](https://doi.org/10.1037/h0073592).
33. Graham, K. E., McDaniel, M. A., Douglas, E. F., & Snell, A. F. (2002). Biodata validity decay and score inflation with faking: Do item attributes explain variance across items? *Journal of Business and Psychology*, 16, 573–592. doi:[10.1023/A:1015454319119](https://doi.org/10.1023/A:1015454319119).
34. Graves, L. M., & Karren, R. J. (1999). Are some interviewers better than others? In R. W. Eder & M. M. Harris (Hrsg.), *The employment interview handbook* (S. 243–258). Thousand Oaks: Sage.
35. Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 26, 463–488. doi:[10.1177/014920630002600305](https://doi.org/10.1177/014920630002600305).
36. Harris, M. M. (1989). Reconsidering the employment interview: A review of recent literature and suggestions for future research. *Personnel Psychology*, 42, 691–726. doi:[10.1111/j.1744-6570.1989.tb00673.x](https://doi.org/10.1111/j.1744-6570.1989.tb00673.x).
37. 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. doi:[10.1111/j.1744-6570.2004.00003.x](https://doi.org/10.1111/j.1744-6570.2004.00003.x).
38. Hein, M., & Wesley, S. (1994). Scaling biodata through subgrouping. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Hrsg.), *Biodata handbook: Theory, research, and use of*

- biographical information in selection and performance prediction* (S. 171–196). Palo Alto: CPP Books.
39. Hinrichs, J. R., Haanperä, S., & Sonkin, L. (1976). Validity of a biographical information blank across national boundaries. *Personnel Psychology*, *29*, 417–421. doi:[10.1111/j.1744-6570.1976.tb00425.x](https://doi.org/10.1111/j.1744-6570.1976.tb00425.x).
  40. Hogan, J. B. (1994). Empirical keying of background data measures. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Hrsg.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (S. 69–107). Palo Alto: CPP Books.
  41. Hough, L., & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Hrsg.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (S. 109–145). Palo Alto: CPP Books.
  42. Huffcutt, A. I. (2011). An empirical review of the employment interview construct literature. *International Journal of Selection and Assessment*, *19*, 62–81. doi:[10.1111/j.1468-2389.2010.00535.x](https://doi.org/10.1111/j.1468-2389.2010.00535.x).
  43. Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, *29*, 184–190. doi:[10.1037/0021-9010.79.2.184](https://doi.org/10.1037/0021-9010.79.2.184).
  44. Huffcutt, A. I., & Culbertson, S. S. (2010). Interviews. In S. Zedeck (Hrsg.), *APA handbook of industrial and organizational psychology* (S. 185–203). Washington: American Psychological Association.
  45. Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology*, *81*, 459–474. doi:[10.1037/0021-9010.81.5.459](https://doi.org/10.1037/0021-9010.81.5.459).
  46. 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. doi:[10.1037/0021-9010.86.5.897](https://doi.org/10.1037/0021-9010.86.5.897).
  47. Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Degroot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology*, *54*, 619–644. doi:[10.1111/j.1744-6570.2001.tb00225.x](https://doi.org/10.1111/j.1744-6570.2001.tb00225.x).
  48. Huffcutt, A. I., Conway, J. M., Roth, P. L., & Klehe, U. C. (2004). The impact of job complexity and study design of situational and behavior description interview validity. *International Journal of Selection and Assessment*, *12*, 262–273. doi:[10.1111/j.0965-075X.2004.280\\_1.x](https://doi.org/10.1111/j.0965-075X.2004.280_1.x).
  49. Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2013). Employment interview reliability: New meta-analytic estimates by structure and format. *International Journal of Selection and Assessment*, *21*, 264–276. doi:[10.1111/ijsa.12036](https://doi.org/10.1111/ijsa.12036).
  50. Hunter, J. E., & Hirsh, H. R. (1987). Applications of meta-analysis. In C. L. Cooper & I. T. Robertson (Hrsg.), *International review of industrial and organizational psychology* (S. 321–357). Chichester: Wiley.
  51. Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72–98. doi:[10.1037/0033-2909.96.1.72](https://doi.org/10.1037/0033-2909.96.1.72).
  52. Janz, T. (1982). Initial comparison of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology*, *67*, 577–580. doi:[10.1037/0021-9010.67.5.577](https://doi.org/10.1037/0021-9010.67.5.577).

53. Janz, T. (1989). Patterned behavior description interviews: The best prophet of the future is the past. In R. W. Eder & G. R. Ferris (Hrsg.), *The employment interview: Theory, research, and practice* (S. 158–168). Newbury Park: Sage.
54. Janz, T., Hellervik, L., & Gillmore, D. C. (1986). *Behavior description interviewing*. Boston: Allyn & Bacon.
55. Jetter, W. (1998). Strukturierte Interviews und PC-gestützte Auswahlverfahren. *Personalwirtschaft*, 6, 43–46.
56. Karas, M., & West, J. (1999). Construct-oriented biodata development for selection to a differentiated performance domain. *International Journal of Selection and Assessment*, 7, 86–96. doi:[10.1111/1468-2389.00109](https://doi.org/10.1111/1468-2389.00109).
57. Kataoka, H. C., Latham, G. P., & Whyte, G. (1997). The relative resistance of the situational, patterned behavior, and conventional structured interviews to anchoring effects. *Human Performance*, 10, 47–63. doi:[10.1207/s15327043hup1001\\_3](https://doi.org/10.1207/s15327043hup1001_3).
58. Kici, G., & Westhoff, K. (2000). Anforderungen an psychologisch-diagnostische Interviews in der Praxis. *Report Psychologie*, 25, 428–436.
59. Kici, G., & Westhoff, K. (2004). Evaluation of requirements for the assessment and construction of interview guides in psychological assessment. *European Journal of Psychological Assessment*, 20, 83–98. doi:[10.1027/1015-5759.20.2.83](https://doi.org/10.1027/1015-5759.20.2.83).
60. Kilcullen, R. N., White, L. A., Mumford, M. D., & Mack, H. (1995). Assessing the construct validity of rational biodata scales. *Military Psychology*, 7, 17–28. doi:[10.1207/s15327876mp0701\\_2](https://doi.org/10.1207/s15327876mp0701_2).
61. Klehe, U. C., & Latham, G. P. (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. doi:[10.1111/j.0965-075X.2005.00305.x](https://doi.org/10.1111/j.0965-075X.2005.00305.x).
62. Klehe, U. C., & Latham, G. P. (2006). What would you do – really or ideally? Constructs underlying the behavior description interview and the situational interview in predicting typical versus maximum performance. *Human Performance*, 19, 357–382. doi:[10.1207/s15327043hup1904\\_3](https://doi.org/10.1207/s15327043hup1904_3).
63. Krajewski, H. T., Goffin, R. D., McCarthy, J. M., Rothstein, M. G., & Johnston, N. (2006). Comparing the validity of structured interviews for managerial-level employees: Should we look to the past or focus on the future? *Journal of Occupational and Organizational Psychology*, 79, 411–432. doi:[10.1348/096317905X68790](https://doi.org/10.1348/096317905X68790).
64. Latham, G. P. (1989). The reliability, validity, and practicability of the Situational Interview. In R. W. Eder & G. R. Ferris (Hrsg.), *The employment interview: Theory, research, and practice* (S. 169–182). Thousand Oaks: Sage.
65. Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 17, 422–427. doi:[10.1037/0021-9010.65.4.422](https://doi.org/10.1037/0021-9010.65.4.422).
66. Latham, G. P., & Sue-Chan, C. (1999). A meta-analysis of the situational interview: An enumerative review of reasons for its validity. *Canadian Psychology*, 40, 56–67. doi:[10.1037/h0086826](https://doi.org/10.1037/h0086826).
67. Lautenschlager, G. J., & Shaffer, G. S. (1987). Reexamining the component stability of Owens's biographical questionnaire. *Journal of Applied Psychology*, 72, 149–152. doi:[10.1037/0021-9010.72.1.149](https://doi.org/10.1037/0021-9010.72.1.149).
68. Lefkowitz, J., Gebbia, M. I., Balsam, T., & Dunn, L. (1999). Dimensions of biodata items and their relationships to item validity. *Journal of Occupational and Organizational Psychology*, 72, 331–350. doi:[10.1348/096317999166716](https://doi.org/10.1348/096317999166716).
69. Levashina, J., & Campion, M. A. (2006). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment*, 14, 299–316. doi:[10.1111/j.1468-2389.2006.00353.x](https://doi.org/10.1111/j.1468-2389.2006.00353.x).

70. Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*, 241–293. doi:10.1111/peps.12052.
71. Locke, E. A., & Latham, G. P. (1990). *A theory of goal-setting and task performance*. Englewood Cliffs: Prentice Hall.
72. Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review, 19*, 203–218. doi:10.1016/j.hrmr.2009.03.006.
73. Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology, 47*, 715–738. doi:10.1111/j.1744-6570.1994.tb01573.x.
74. Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology, 44*, 763–792. doi:10.1111/j.1744-6570.1991.tb00698.x.
75. Mayfield, E. C. (1964). The selection interview: A reevaluation of published research. *Personnel Psychology, 17*, 599–616. doi:10.1111/j.1744-6570.1964.tb00065.x.
76. McCarthy, J., & Goffin, R. (2004). Measuring job interview anxiety: Beyond weak knees and sweaty palms. *Personnel Psychology, 57*, 607–637. doi:10.1111/j.1744-6570.2004.00002.x.
77. 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. doi:10.1037/0021-9010.79.4.599.
78. McManus, M. A., & Kelly, M. L. (1999). Personality measures and biodata: Evidence regarding their incremental predictive value in the life insurance industry. *Personnel Psychology, 52*, 137–148. doi:10.1111/j.1744-6570.1999.tb01817.x.
79. Mitchell, T. W., & Klimoski, R. J. (1982). Is it rational to be empirical? A test of methods for scoring biographical data. *Journal of Applied Psychology, 67*, 411–418. doi:10.1037/0021-9010.67.4.411.
80. Morton, K. S. (1994). *A schema model of dispositional attribution in the employment selection process*. Doctoral dissertation, Florida State University.
81. Moscoso, S. (2000). Selection interview: A review of validity evidence, adverse impact and applicant reactions. *International Journal of Selection and Assessment, 8*, 237–247. doi:10.1111/1468-2389.00153.
82. Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. *Personnel Psychology, 53*, 299–323. doi:10.1111/j.1744-6570.2000.tb00203.x.
83. Mumford, M. D. (1999). Construct validity and background data: Issues, abuses, and future directions. *Human Resource Management Review, 9*, 117–145. doi:10.1016/S1053-4822(99)00015-7.
84. Mumford, M. D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement, 11*, 1–31. doi:10.1177/014662168701100101.
85. Mumford, M. D., Stokes, G. S., & Owens, W. A. (1990). *Patterns of life history: The ecology of human individuality*. Hillsdale: Lawrence Erlbaum Associates.
86. Mumford, M. D., Costanza, D. P., Connelly, M. S., & Johnson, J. F. (1996). Item generation procedures and background data scales: Implications for construct and criterion-related validity. *Personnel Psychology, 49*, 361–398. doi:10.1111/j.1744-6570.1996.tb01804.x.
87. Mumford, M. D., Whetzel, D. L., Murphy, S. T., & Eubanks, D. L. (2007). Background data. In D. L. Whetzel & G. R. Wheaton (Hrsg.), *Applied measurement: Industrial psychology in human resources management* (S. 201–233). New York: Taylor & Francis Group & Lawrence Erlbaum Associates.



88. Mumford, M. D., Barrett, J. D., & Hester, K. S. (2012). Background data: Use of experiential knowledge in personnel selection. In N. Schmitt (Hrsg.), *The Oxford handbook of personnel assessment and selection* (S. 353–382). New York: Oxford University Press.
89. Nickels, B. J. (1994). The nature of biodata. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Hrsg.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (S. 1–16). Palo Alto: CPP Books.
90. Orpen, C. (1985). Patterned behavior description interviews versus unstructured interviews: A comparative validity study. *Journal of Applied Psychology*, *70*, 774–776. doi:[10.1037/0021-9010.70.4.774](https://doi.org/10.1037/0021-9010.70.4.774).
91. Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, *89*, 187–207. doi:[10.1037/0021-9010.89.2.187](https://doi.org/10.1037/0021-9010.89.2.187).
92. Owens, W. A. (1976). Background data. In M. D. Dunette (Hrsg.), *Handbook of industrial and organizational psychology* (S. 609–644). Chicago: Rand-McNally.
93. Owens, W. A., & Schoenfeldt, L. F. (1979). Toward a classification of persons. *Journal of Applied Psychology*, *64*, 569–607. doi:[10.1037/0021-9010.64.5.569](https://doi.org/10.1037/0021-9010.64.5.569).
94. Pace, L. A., & Schoenfeldt, L. F. (1977). Legal concerns in the use of weighted applications. *Personnel Psychology*, *30*, 159–166. doi:[10.1111/j.1744-6570.1977.tb02085.x](https://doi.org/10.1111/j.1744-6570.1977.tb02085.x).
95. Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, *55*, 1–81. doi:[10.1111/j.1744-6570.2002.tb00103.x](https://doi.org/10.1111/j.1744-6570.2002.tb00103.x).
96. Pulakos, E. D., Schmitt, N., Whitney, D., & Smith, M. (1996). Individual differences in interviewer ratings: The impact of standardization, consensus discussion, and sampling error on the validity of a structured interview. *Personnel Psychology*, *49*, 85–102. doi:[10.1111/j.1744-6570.1996.tb01792.x](https://doi.org/10.1111/j.1744-6570.1996.tb01792.x).
97. Püttner, I. (2014). Rechtliche Aspekte der Personalarbeit. In H. Schuler, & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (3. überarb. u. erw. Aufl., S. 1201–1226). Göttingen: Hogrefe.
98. Reilly, R. R., & Chao, G. R. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, *35*, 1–62. doi:[10.1111/j.1744-6570.1982.tb02184.x](https://doi.org/10.1111/j.1744-6570.1982.tb02184.x).
99. Reiter-Palmon, R., & Connelly, M. S. (2000). Item selection counts: A comparison of empirical key and rational scale validities in theory-based and non-theory-based item pools. *Journal of Applied Psychology*, *85*, 143–151. doi:[10.1037/0021-9010.85.1.143](https://doi.org/10.1037/0021-9010.85.1.143).
100. Robertson, I. T., & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology*, *74*, 441–472. doi:[10.1348/096317901167479](https://doi.org/10.1348/096317901167479).
101. Roth, P. L., & Huffcutt, A. I. (2013). A meta-analysis of interviews and cognitive ability: Back to the future? *Journal of Personnel Psychology*, *12*, 157–169. doi:[10.1027/1866-5888/a000091](https://doi.org/10.1027/1866-5888/a000091).
102. Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology*, *75*, 175–184. doi:[10.1037/0021-9010.75.2.175](https://doi.org/10.1037/0021-9010.75.2.175).
103. Russell, C. J. (1994). Generation procedures for biodata items: A point of departure. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Hrsg.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (S. 17–38). Palo Alto: CPP Books.



104. Russell, C. J., Mattson, J., Devlin, S. E., & Atwater, D. (1990). Predictive validity of biodata items generated from retrospective life experience essays. *Journal of Applied Psychology*, 75, 569–580. doi:[10.1037/0021-9010.75.5.569](https://doi.org/10.1037/0021-9010.75.5.569).
105. Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, 52, 359–391. doi:[10.1111/j.1744-6570.1999.tb00165.x](https://doi.org/10.1111/j.1744-6570.1999.tb00165.x).
106. Ryan, A. M., & Sackett, P. R. (1989). Exploratory study of individual assessment practices: Interrater-reliability and judgments of assessor effectiveness. *Journal of Applied Psychology*, 74, 568–579. doi:[10.1037/0021-9010.74.4.568](https://doi.org/10.1037/0021-9010.74.4.568).
107. 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. doi:[10.1080/13594320244000184](https://doi.org/10.1080/13594320244000184).
108. 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. doi:[10.1037/0033-2909.124.2.262](https://doi.org/10.1037/0033-2909.124.2.262).
109. Schmit, M. J., & Ryan, A. M. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology*, 77, 629–637. doi:[10.1037/0021-9010.77.5.629](https://doi.org/10.1037/0021-9010.77.5.629).
110. Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407–422. doi:[10.1111/j.1744-6570.1984.tb00519.x](https://doi.org/10.1111/j.1744-6570.1984.tb00519.x).
111. Schmitt, N., Jennings, D., & Toney, R. (1999). Can we develop measures of hypothetical constructs? *Human Resource Management Review*, 9, 169–183. doi:[10.1016/S1053-4822\(99\)00017-0](https://doi.org/10.1016/S1053-4822(99)00017-0).
112. Schoenfeldt, L. F. (1999). From dust bowl empiricism to rational constructs in biographical data. *Human Resource Management Review*, 9, 147–167. doi:[10.1016/S1053-4822\(99\)00016-9](https://doi.org/10.1016/S1053-4822(99)00016-9).
113. Schoenfeldt, L. F., & Mendoza, J. L. (1994). Developing and using factorially derived biographical scales. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Hrsg.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (S. 147–169). Palo Alto: CPP Books.
114. Schuler, H. (1992). Das Multimodale Interview. *Diagnostica*, 38, 281–300.
115. Schuler, H. (1994). Selektion und Selbstselektion durch das Multimodale Interview. In L. von Rosenstiel, T. Lang, & E. Sigl (Hrsg.), *Fach- und Führungsnachwuchs finden und fördern* (S. 97–112). Stuttgart: Schäffer-Poeschel.
116. Schuler, H. (2002). *Das Einstellungsinterview*. Göttingen: Hogrefe.
117. Schuler, H. (2014a). Biografieorientierte Verfahren der Personalauswahl. In H. Schuler, & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (3. überarb. u. erw. Aufl., S. 257–299). Göttingen: Hogrefe.
118. Schuler, H. (2014b). *Psychologische Personalauswahl – Eignungsdiagnostik für Personalentscheidungen und Berufsberatung* (4. vollst. überarb. u. erw. Aufl.). Göttingen: Hogrefe.
119. Schuler, H., & Funke, U. (1989). Berufseignungsdiagnostik. In E. Roth (Hrsg.), *Organisationspsychologie. Enzyklopädie der Psychologie D, Serie 3* (Bd. 3, S. 281–320). Göttingen: Hogrefe.
120. Schuler, H., Hell, B., Trapmann, S., Schaar, H., & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen: Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie*, 2, 60–70.
121. Schwab, D. P., & Oliver, R. L. (1974). Predicting tenure with biographical data: Exhuming buried evidence. *Personnel Psychology*, 27, 125–128. doi:[10.1111/j.1744-6570.1974.tb02068.x](https://doi.org/10.1111/j.1744-6570.1974.tb02068.x).

122. Shackleton, V., & Newell, S. (1994). European management selection methods: A comparison of five countries. *International Journal of Selection and Assessment*, 2, 91–102. doi:[10.1111/j.1468-2389.1994.tb00155.x](https://doi.org/10.1111/j.1468-2389.1994.tb00155.x).
123. Shaffer, G. S., Saunders, V., & Owens, W. A. (1986). Additional evidence for the accuracy of biographical data: Long-term retest and observer ratings. *Personnel Psychology*, 39, 791–809. doi:[10.1111/j.1744-6570.1986.tb00595.x](https://doi.org/10.1111/j.1744-6570.1986.tb00595.x).
124. Sharf, J. C. (1994). The impact of legal and equal employment opportunity issues on personal history inquiries. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Hrsg.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (S. 351–390). Palo Alto: CPP Books.
125. Snell, A. F., Stokes, G. S., Sands, M. M., & McBride, J. R. (1994). Adolescent life experiences as predictors of occupational attainment. *Journal of Applied Psychology*, 79, 131–141. doi:[10.1037/0021-9010.79.1.131](https://doi.org/10.1037/0021-9010.79.1.131).
126. Stokes, G. S. (1994). Introduction and history. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Hrsg.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (S. VX–VXII). Palo Alto: CPP Books.
127. Stokes, G. S., & Cooper, L. A. (2001). Content/construct approaches in life history form development for selection. *International Journal of Selection and Assessment*, 9, 138–151. doi:[10.1111/1468-2389.00170](https://doi.org/10.1111/1468-2389.00170).
128. Stokes, G. S., & Cooper, L. A. (2004). Biodata. In J. C. Thomas (Hrsg.), *Comprehensive handbook of psychological assessment, Vol. 4: Industrial and organizational assessment* (S. 243–268). Hoboken: Wiley.
129. Stokes, G. S., & Searcy, C. A. (1999). Specification of scales in biodata form development: Rational vs. empirical and global vs. specific. *International Journal of Selection and Assessment*, 7, 72–85. doi:[10.1111/1468-2389.00108](https://doi.org/10.1111/1468-2389.00108).
130. Stokes, G. S., Toth, C. S., Searcy, C. A., Stroupe, J. P., & Carter, G. W. (1999). Construct/rational biodata dimensions to predict salesperson performance: Report on the U.S. department of labor sales study. *Human Resource Management Review*, 9, 185–218. doi:[10.1016/S1053-4822\(99\)00018-2](https://doi.org/10.1016/S1053-4822(99)00018-2).
131. Strobel, A., Lammerskitten, E., & Glodek, R. (2010). Die Interview-Standards des Arbeitskreis Assessment Center e. V. – Entwicklung, Inhalte. Nutzen. *Wirtschaftspsychologie*, 2, 26–31.
132. Strobel, A., & Westhoff, K. (2009). *Diagnoseinstrument zur Erfassung der Interviewerkompetenz in der Personalauswahl (DIPA)*. Frankfurt a. M.: Pearson Assessment.
133. Strobel, A., & Westhoff, K. (2010). Eignungsinterview. In K. Westhoff, C. Hagemeyer, M. Kersting, F. Lang, H. Moosbrugger, G. Reimann, & G. Stemmler (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (3. überarb. Aufl., S. 84–101). Lengerich: Pabst Science Publishers.
134. Taylor, P. J., & O’Driscoll, M. P. (1995). *Structured employment interviewing*. Aldershot: Gower Publishing Company.
135. Taylor, P. J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology*, 75, 277–294. doi:[10.1348/096317902320369712](https://doi.org/10.1348/096317902320369712).
136. Trost, G. (1996). Interview. Grundlagen und Methoden der Differentiellen Psychologie. In K. J. Groffmann & L. Michel (Hrsg.), *Enzyklopädie der Psychologie* (Bd. 1, S. 463–505). Göttingen: Hogrefe.

137. Valenzi, E., & Andrews, I. R. (1973). Individual differences in the decision process of employment interviewers. *Journal of Applied Psychology*, 58, 49–53. doi:[10.1037/h0021696](https://doi.org/10.1037/h0021696).
138. Van der Zee, K. I., Bakker, A. B., & Bakker, P. (2002). Why are structured interviews so rarely used in personnel selection? *Journal of Applied Psychology*, 87, 176–184. doi:[10.1037//0021-9010.87.1.176](https://doi.org/10.1037//0021-9010.87.1.176).
139. Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology*, 90, 536–552. doi:[10.1037/0021-9010.90.3.536](https://doi.org/10.1037/0021-9010.90.3.536).
140. Vinchur, A. J., Schippmann, J. S., Switzer, F. S. I., & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, 83, 586–597. doi:[10.1037/0021-9010.83.4.586](https://doi.org/10.1037/0021-9010.83.4.586).
141. Wagner, R. (1949). The employment interview: A critical summary. *Personnel Psychology*, 2, 17–46. doi:[10.1111/j.1744-6570.1949.tb01669.x](https://doi.org/10.1111/j.1744-6570.1949.tb01669.x).
142. Webster, E. C. (1964). *Decision making in the employment interview*. Montreal: Industrial Relations Centre, McGill University.
143. Webster, E. C. (1982). *The employment interview*. Schomberg: SIP Publications.
144. Westhoff, K. (2000). Das psychologisch-diagnostische Interview – Ein Überblick über die Forschung für die Praxis. *Report Psychologie*, 25, 18–24.
145. Westhoff, K. (Hrsg.). (2009). *Das Entscheidungsorientierte Gespräch (EOG) als Eignungsinterview*. Lengerich: Pabst Science Publishers.
146. Westhoff, K., Hagemeister, C., & Eckert, H. (2002). On the objectivity of oral examinations in psychology. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 23, 149–157. doi:[10.1024/0170-1789.23.2.149](https://doi.org/10.1024/0170-1789.23.2.149).
147. Westhoff, K., & Kluck, M. L. (2014). *Psychologische Gutachten schreiben und beurteilen* (6. vollst. überarb. u. erw. Aufl.). Berlin: Springer.
148. Westhoff, K., & Strobel, A. (2009a). Leitfadenkonstruktion. In K. Westhoff (Hrsg.). *Das Entscheidungsorientierte Gespräch (EOG) als Eignungsinterview* (S. 79–86). Lengerich: Pabst Science Publishers.
149. Westhoff, K., & Strobel, A. (2009b). Strukturierung des Eignungsinterviews. In K. Westhoff (Hrsg.). *Das Entscheidungsorientierte Gespräch (EOG) als Eignungsinterview* (S. 27–36). Lengerich: Pabst Science Publishers.
150. Westhoff, K., & Strobel, A. (2011). Interview. In M. Amelang & L. F. Hornke (Hrsg.), *Enzyklopädie der Psychologie. Themenbereich B, Methodologie und Methoden. Serie II, Psychologische Diagnostik. Band II, Methoden* (S. 371–413). Göttingen: Hogrefe.
151. Wiesner, W. H., & Cronshaw, S. F. (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. doi:[10.1111/j.2044-8325.1988.tb00467](https://doi.org/10.1111/j.2044-8325.1988.tb00467).

## Zusammenfassung

Intelligenz ist ein kontrovers diskutiertes Eignungsmerkmal, insbesondere Intelligenztests rufen häufig emotionale (Abwehr-)Reaktionen hervor. Seit Beginn der Intelligenzforschung gab es kritische Stimmen, die Intelligenztests für bestenfalls nutzlos ggf. aber sogar für schädlich erklärten. Mit der „Erfindung“ der sozialen oder emotionalen Intelligenz war die Hoffnung verbunden, die herkömmlichen Intelligenztests begraben zu können. Dennoch ist kein Ende der Intelligenztestungen abzusehen. Nichts scheint so sehr die Unsterblichkeit zu verbürgen, wie die andauernde Debatte über den Tod. Zumal wenn, wie im Falle der Intelligenztests, die Befundlage so eindeutig ist: Kein anderes Verfahren erlaubt so treffsichere Vorhersagen über den Erfolg in Ausbildung und im Beruf. Wir skizzieren im Folgenden zunächst, was Intelligenz – einigen aktuellen Definitionen und Theorien zufolge – bedeutet und nennen beispielhaft Verfahren, mit denen sich Intelligenz messen lässt. Wir konzentrieren uns auf Theorien, die im deutschsprachigen Raum verbreitet sind und auf deren Basis deutschsprachige Intelligenztests entwickelt wurden, einige davon werden wir kurz vorstellen. Dabei gehen wir insbesondere auf Befunde zur Treffsicherheit von Aussagen (z. B. Personalentscheidungen) ein, die aus Intelligenztests abgeleitet wurden, wobei wir auch die Grenzen der Intelligenzdiagnostik ansprechen. Wir reflektieren über die Gründe für die geringe Einsatzhäufigkeit von Intelligenztests und gehen dabei besonders auf die Frage der

---

M. Kersting (✉) · C. Palmer  
Institut für Psychologie und Sportwissenschaft, Justus-Liebig-Universität Gießen,  
Gießen, Deutschland  
E-Mail: Martin.Kersting@psychol.uni-giessen.de

C. Palmer  
E-Mail: Carolin.Palmer@psychol.uni-giessen.de

Akzeptanz solcher Tests ein. Darüber hinaus geben wir Tipps, wie man die Qualität von Tests beurteilen kann. In weiteren Abschnitten behandeln wir die Kreativität und grenzen die traditionelle Intelligenzauffassung gegenüber neueren Konzepten, wie z. B. der emotionalen und praktischen Intelligenz ab. Auch für Kreativität und emotionale Intelligenz nennen wir ausgewählte Testverfahren und berichten über die Treffsicherheit entsprechend fundierter Personalentscheidungen.

## Inhaltsverzeichnis

6.1	Intelligenz: Bedeutung, Messung und Aussagekraft	128
6.1.1	Definition Intelligenz	128
6.2	Zur Treffsicherheit intelligenztestbasierter Aussagen	129
6.3	Intelligenzmodelle	131
6.3.1	Das Berliner Intelligenzstrukturmodell	132
6.3.2	Das hierarchische Rahmen- bzw. Protomodell der Intelligenzstrukturforschung	134
6.3.3	Modifiziertes Modell der „Primary Mental Abilities“	135
6.3.4	Gemeinsamkeiten aktueller Intelligenzmodelle	136
6.4	Ausgewählte Intelligenztests	136
6.5	Zur Einsatzhäufigkeit und zur Akzeptanz von Intelligenztests	139
6.6	Kriterien der Testauswahl	141
6.7	Kreativität	142
6.8	Kreativität und Intelligenz	143
6.8.1	Divergentes Denken	144
6.8.2	Kreativität im Berliner Intelligenzstrukturmodell	146
6.8.3	Kreativität und das Modell der fluiden und kristallinen Intelligenz nach Cattell (1943)	146
6.8.4	Cattell-Horn-Carroll-Modell	147
6.8.5	Eine prozessbasierte Sicht auf Kreativität	147
6.9	Kreativitätstests	148
6.10	Erweiterungen des Intelligenzbegriffs	149
6.10.1	Praktisch-mechanisch-technisches Verständnis	149
6.10.2	Praktische Intelligenz und Erfolgsintelligenz	150
6.10.3	Emotionale Intelligenz	150
6.11	Fazit	150
	Literatur	151

---

## 6.1 Intelligenz: Bedeutung, Messung und Aussagekraft

### 6.1.1 Definition Intelligenz

„Eine beneidenswerte Situation: Sie wissen nicht, was es ist. Aber sie können es messen“ [67, S. 242], so lautet ein gut formulierter aber inhaltlich unzutreffender Versuch, die Intelligenzforschung zu diskreditieren.

Die psychologische Forschung hat Definitionen und Modelle der Intelligenz veröffentlicht, die sich bewährt haben und die sich dennoch – und das ist ein gutes Zeichen – fortwährend weiterentwickeln. 1994 erschien ein Artikel, der die Definition der Intelligenz wieder gab, auf den sich die führenden Intelligenzforscher/-innen geeinigt hatten. Dass der Artikel zuerst im Wall Street Journal publiziert wurde, zeigt, wie wichtig das Personenmerkmal Intelligenz für die Wirtschaft ist. Später wurde dieser Konsens (Kasten 6.1) von der Initiatorin Linda S. Gottfredson [26] in der Fachzeitschrift „Intelligence“ publiziert.

Kasten 6.1 Konsens führender Intelligenzforscher über die Definition von Intelligenz [26]

Intelligenz ist eine sehr allgemeine geistige Kapazität, die u. a.

- die Fähigkeit zum schlussfolgernden Denken,
- zum Planen,
- zum Problemlösen,
- zum abstrakten Denken,
- zum Verstehen komplexer Ideen,
- zum raschen Lernen,
- zum Lernen aus Erfahrung

umfasst.

Intelligenz meint nicht das Lernen aus Büchern, bezieht sich nicht auf begrenzte akademische Begabung oder die Fähigkeit, Tests zu bearbeiten. Vielmehr reflektiert Intelligenz ein breiteres und tieferes Vermögen,

- unsere Umwelt zu verstehen,
- zu durchblicken,
- Sinn in Dingen zu erkennen,
- herauszubekommen, was zu tun ist.

([26, S. 13] eigene Übersetzung)

## 6.2 Zur Treffsicherheit intelligenztestbasierter Aussagen

Intelligenz ist der zentrale Erfolgsfaktor für Erfolg in Ausbildung und Beruf. Weder andere Fähigkeiten noch Erfahrungen sind bedeutsamer – so lautet sinngemäß das Fazit, mit dem Schmidt und Hunter [89, S. 162] in ihrer Überblicksarbeit den Stand der Forschung zum Thema Intelligenz und Arbeitswelt zusammenfassen. Die besonders hohe

Treffsicherheit von Intelligenztestbasierten Aussagen bei der Vorhersage des Erfolgs bei beruflichen Ausbildungen, Trainings sowie im Beruf konnte sowohl in Meta-Studien mit überwiegend nordamerikanischen Daten [90] als auch in entsprechenden Meta-Studien mit europäischen [86] und deutschen Daten [40, 53] aufgezeigt werden. Je nachdem, welches Kriterium vorhergesagt wird, werden Koeffizienten bis zu ,62 erreicht.

Dabei erlauben Intelligenztests Rückschlüsse, die für alle Berufe relevant sind [87, 91]. Die gängige Praxis, Intelligenztests bei der Azubi-Auswahl einzusetzen und bei der Beurteilung der Eignung von (potenziellen) Mitarbeiter/-innen für „höhere Positionen“ außen vor zu lassen, steht dem Stand der Forschung diametral gegenüber. Denn, Intelligenz ist umso erfolgsrelevanter, umso komplexer die Arbeit ist [60, 87, 89]. Auch eine andere weit verbreitete Behauptung ist unzutreffend: Die Prognosekraft der aus Intelligenztests abgeleiteten Aussagen ist nicht, wie gern behauptet wird, kurzfristig, sondern entfaltet sich nach Hossiep [36] sowie Schmidt und Hunter [89] mittelfristig. Auf der Basis von Intelligenztestergebnissen können selbst langfristige Karrierewege vergleichsweise gut vorhergesagt werden [45].

Die Frage, ob aus Intelligenztests abgeleitete Aussagen eine treffsichere Vorhersage der beruflichen Zukunft erlauben, ist längst beantwortet. Dass sie es tun, ist ein Nebenergebnis einer kaum noch überschaubaren Anzahl von Studien. Wie sie es tun, welche Intelligenzkomponenten prognostisch bedeutsam sind, ist aber nach wie vor eine Frage, die Forscher/-innen umtreibt. Weiter unten gehen wir im Kontext der Erläuterung von Intelligenzmodellen kurz auf die Debatte darüber ein, ob es sinnvoller ist, nur einen (übergeordneten) Intelligenzfaktor (z. B. die allgemeine Intelligenz, auch g-Faktor genannt oder die fluide Intelligenz) zu interpretieren oder aber von mehreren Intelligenzdimensionen auszugehen. Bezüglich der Vorhersagekraft von Intelligenztests herrschte lange die Auffassung vor, es wäre ausreichend, von einem Intelligenzwert auszugehen. Studien jüngerer Zeit weisen aber darauf hin, dass die Interpretation einzelner Intelligenzdimensionen auch unter dem Gesichtspunkt der Kriteriumsvalidität durchaus sinnvoll ist [24, 60, 86].

Die Grenzen der Intelligenztestbasierten Eignungsdiagnostik sind erreicht, wenn andere Kriterien als die „klassische“ Berufsleistung vorhergesagt werden soll. So lässt sich der berufliche Status beispielsweise besser durch Intelligenztestergebnisse prognostizieren, als das von vielen weiteren Faktoren abhängige Einkommen [105]. Insbesondere berufliche Zufriedenheit steht in einem engeren statistischen Zusammenhang mit „weichen“ Faktoren als mit Intelligenz. Insgesamt gilt es zu beachten, dass Intelligenz nicht der einzige Prädiktor für Berufserfolg ist. Persönlichkeitsmerkmale im engeren Sinne (siehe dazu das Kapitel von Hossiep und Weiß in diesem Band), insbesondere die Dimensionen Gewissenhaftigkeit, Extraversion und emotionale Stabilität des Fünf-Faktoren-Modells der Persönlichkeit spielen ebenfalls eine wichtige Rolle (die beiden zuletzt genannten Dimensionen in unterschiedlichem Ausmaß für verschiedene Berufe) [6]. Eine Kombination von Messungen dieser Persönlichkeitseigenschaften einerseits sowie Intelligenzmessungen andererseits könnte zu einer verbesserten Vorhersage beruflicher Leistungen führen.



Die hohe Treffsicherheit intelligenztestbasierter Aussagen lässt sich sowohl inhaltlich als auch methodisch erklären. Zum einen sind Fähigkeiten, wie etwa aus Erfahrungen lernen zu können oder in logischen Aufgaben die richtigen Schlüsse zu ziehen, zentral für den Berufserfolg. Zum anderen sind Intelligenztests aber auch deshalb erfolgreich, weil sie aufgrund ihrer „schlichten“ Vorgehensweise bei der Testung („Kreuzen Sie die richtige Antwort an“) und Auswertung (Auszählung der „richtigen“ Antworten) eine sehr hohe Objektivität erzielen. Gerade dieser Aspekt, dass lediglich „Kreuzchen“ gezählt werden, wird an Intelligenztests sowohl von Praktiker/-innen als auch von Forscher/-innen häufig kritisch gesehen. Laien sind überzeugt, dass ein aufwendiges Assessment-Center validere Informationen liefern würde als ein schlichter Intelligenztest. Dabei garantiert gerade diese „Schlichkeit“ der Durchführung und Auswertung von Intelligenztests, dass das Testergebnis objektiv ermittelt wird. Für das Testergebnis spielt es weder eine Rolle, wer den Test durchführt, auswertet und interpretiert, noch wie attraktiv die getestete Person ist usw. Intelligenztests sind folglich wesentlich objektiver als Interviews oder Verfahren zur Verhaltensbeobachtung und -beurteilung wie Gruppendiskussionen oder Rollenspiele. Dieser methodische Aspekt ist von hoher Bedeutung, da die Objektivität eine notwendige Voraussetzung für Reliabilität und Validität ist.

---

### 6.3 Intelligenzmodelle

Die Geschichte der Entwicklung von Intelligenzmodellen lässt sich nach Ansicht von Sternberg und Powell [103] in drei Stufen einteilen. Zu Beginn der empirischen Intelligenzforschung (erste Stufe) standen sich, vereinfacht dargestellt, zwei Auffassungen gegenüber. Einerseits gab es die Vorstellung einer einheitlichen, allgemeinen Intelligenz, dem sogenannten g-Faktor (Spearman), andererseits die Idee, es gäbe mehrere voneinander unabhängige Intelligenzen (Thomson). Diese beiden Konzepte wurden auf der zweiten Stufe weiterentwickelt, wobei die Kernelemente um jeweils ein Element ergänzt wurden. Einerseits gingen Vertreter/-innen der hierarchischen Sichtweise (z. B. Cattell) zwar weiterhin von einer einheitlichen Intelligenz aus, akzeptierten aber gleichzeitig, dass sich – auf einer hierarchisch untergeordneten Ebene – mehrere Fähigkeiten unterscheiden lassen. Andererseits wurde die Idee mehrerer Fähigkeiten u. a. von Thurstone dahin gehend weiterentwickelt, dass diese Fähigkeiten (Primärfaktoren, Kasten 6.4) als überlappend und somit abhängig voneinander angesehen wurden. Aktuelle Intelligenzmodelle befinden sich auf der dritten Stufe nach Sternberg und Powell [103]. Diese dritte Evolutionsstufe ist dadurch gekennzeichnet, dass intellektuelle Fähigkeiten sowohl überlappend als auch als hierarchisch geordnet konzeptualisiert werden können.

Als Beispiele für aktuelle Intelligenzmodelle möchten wir hier drei Modelle aus dem deutschsprachigen Bereich nennen: das Berliner Intelligenzstrukturmodell (BIS) [41, 42], das hierarchische Protomodell der Intelligenzstrukturforschung (HPI) [64] sowie das modifizierte Modell der Primary Mental Abilities (MPPMA) [50]. Wir wählen diese Modelle, da auf dieser Basis deutschsprachige Intelligenztests entwickelt wurden. Im

internationalen Bereich dominieren aktuell die Three-Stratum-Theorie von Carroll [14] sowie vor allem die Cattell-Horn-Carroll-Theorie (CHC-Theorie) von McGrew [69, 70].

### 6.3.1 Das Berliner Intelligenzstrukturmodell

Ausführlicher darstellen möchten wir das Berliner Intelligenzstrukturmodell (BIS) von Jäger [41, 42], das explizit als integratives Modell entwickelt wurde. Ausgangspunkt der Modellentwicklung war eine Inventarisierung aller möglichen Typen von Intelligenz- und Kreativitätsaufgaben. Dieses Inventar im Umfang von ca. 2000 Aufgabentypen wurde im Anschluss auf 191 Aufgaben (98 Aufgabentypen) reduziert, die die Unterschiedlichkeit der ursprünglich inventarisierten Aufgabentypen repräsentieren sollten. Lediglich auf die Aufgaben zur praktischen und sozialen Intelligenz sowie auf Aufgaben, die keine Leistungen, sondern Selbsteinschätzungen erfordern, wurde aus pragmatischen Gründen verzichtet. Letztendlich wurden 191 Aufgaben für die empirischen Untersuchungen zur Modellentwicklung genutzt.

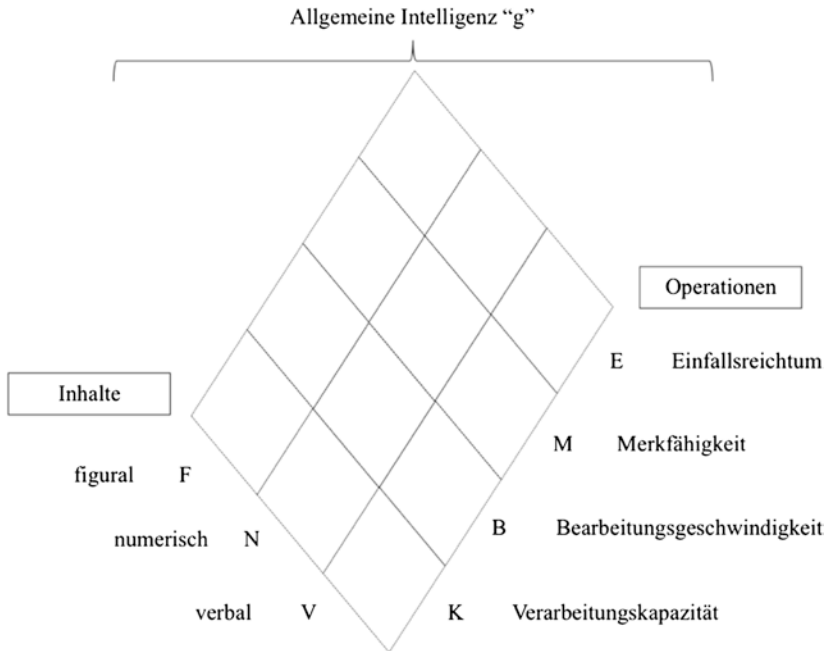
Aus den Analysen der Daten wurde das BIS-Modell abgeleitet, das durch drei Annahmen geprägt ist (Kasten 6.2).

#### Kasten 6.2 Die drei zentralen Annahmen des BIS-Modells

- An jeder Intelligenzleistung sind alle intellektuellen Fähigkeiten – zu unterschiedlichen Anteilen – beteiligt.
- Intelligenzleistungen und Fähigkeitskonstrukte lassen sich unter verschiedenen Modalitäten genannten Aspekten klassifizieren.
- Fähigkeitskonstrukte sind hierarchisch strukturiert, d. h. sie lassen sich unterschiedlichen Generalitätsebenen zuordnen.

(nach [43, S. 4])

Auf der generellen Ebene des BIS-Modells ist die Allgemeine Intelligenz als Integral aller Fähigkeiten angesiedelt. Auf der nächsten Ebene stehen vier Operationen des Denkens (Bearbeitungsgeschwindigkeit [B], Merkfähigkeit [M], Einfallsreichtum [E], und Verarbeitungskapazität [K]) und drei Inhaltsfaktoren des Denkens (sprachgebundenes [V], zahlengebundenes [N] und anschauungsgebundenes, figural-bildhaftes Denken [F]). Die Kreuzklassifikation der Operationen und Inhalte ergibt die zwölf Zellen auf der dritten Modellebene, Abb. 6.1.



**Abb. 6.1** Das Berliner Intelligenzstrukturmodell nach Jäger 1998 [42]. (Quelle: [https://upload.wikimedia.org/wikipedia/commons/b/ba/Berliner\\_Intelligenzstrukturmodell.png](https://upload.wikimedia.org/wikipedia/commons/b/ba/Berliner_Intelligenzstrukturmodell.png))

Die vier operativen und drei inhaltlichen Intelligenz-Fähigkeiten wurden von Jäger et al. [43, S. 6/7] wie folgt beschrieben (Kasten 6.3):

#### Kasten 6.3 Die Fähigkeiten des BIS-Modells [43, S. 6/7]

##### Operative Fähigkeiten

- Verarbeitungskapazität (K): Verarbeitung komplexer Informationen bei Aufgaben, die nicht auf Anhieb zu lösen sind, sondern Heranziehen, vielfältiges Beziehungsstiften, formal-logisch exaktes Denken und sachgerechtes Beurteilen von Informationen erfordern.
- Einfallsreichtum (E): Flexible Ideenproduktion, die Verfügbarkeit vielfältiger Informationen, Reichtum an Vorstellungen und das Sehen vieler verschiedener Seiten, Varianten, Gründe und Möglichkeiten von Gegenständen und Problemen voraussetzt, wobei es um problemorientierte Lösungen geht, nicht um ein ungesteuertes Luxurieren der Fantasie.
- Merkfähigkeit (M): Aktives Einprägen und kurzfristiges Wiedererkennen oder Reproduzieren von verschiedenartigem Material. (...)

- Bearbeitungsgeschwindigkeit (B): Arbeitstempo, Auffassungsleichtigkeit und Konzentrationskraft beim Lösen einfach strukturierter Aufgaben von geringem Schwierigkeitsniveau.

#### Inhaltsgebundene Fähigkeiten

- Sprachgebundenes Denken (V): Grad der Aneignung und der Verfügbarkeit des Beziehungssystems Sprache.
- Zahlengebundenes Denken (N): Grad der Aneignung und der Verfügbarkeit des Beziehungssystems Zahlen.
- Anschauungsgebundenes, figural-bildhaftes Denken (F): Einheitsstiftendes Merkmal scheint hier die Eigenart des Aufgabenmaterials zu sein, dessen Bearbeitung figural-bildhaftes und/oder räumliches Vorstellen erfordert.

Das BIS-Modell eignet sich als inhaltlicher Ordnungsrahmen für Intelligenztests. Es ist daher sehr nützlich für Personen, die darüber entscheiden, welche der zahlreichen Intelligenztests sie bei einer bestimmten Fragestellung einsetzen. Bei dieser Entscheidung sollte man sich nicht allein auf die Angaben verlassen, die von den jeweiligen Testautor/-innen getroffen werden, sondern man muss den Test in einen theoretischen Rahmen einordnen. Süß und Beauducel [106] weisen beispielsweise darauf hin, dass sowohl die Raven Matrizen als auch der Zahlenverbindungstest (ZVT) als Tests zur allgemeinen Intelligenz ausgewiesen sind. Vor dem Hintergrund des BIS-Modells wird deutlich, dass die beiden Tests allerdings sehr unterschiedliche Anforderungen stellen und z. B. mit figural-bildhaften (Raven) Material einerseits und numerischen Material (ZVT) andererseits arbeiten. Die Leistungen im Raven-Matrizen-Test sind, wie Süß und Schweickert [107] zeigen konnten, vor allem ein Indikator für die „figurale Verarbeitungskapazität“ im Sinne des BIS-Modells und nicht für die allgemeine Intelligenz.

Ein theoretisches Modell wie das BIS ist von hohem Wert für die Praxis, da es auf der Konstruktebene eine differenzierte Einordnung von Intelligenztests erlaubt. Für die Auswahl von Intelligenztests für die interne und externe Personalauswahl klassifiziert man idealerweise sowohl die auf das intellektuelle Vermögen bezogene Ergebnis der Anforderungsanalyse (siehe dazu das Kapitel von Nerdinger in diesem Band) als auch die möglichen Testverfahren nach dem BIS-Modell. In einem weiteren Schritt muss dann die psychometrische Qualität des Verfahrens berücksichtigt werden, darauf gehen wir in Abschn. 6.6 ein.

### 6.3.2 Das hierarchische Rahmen- bzw. Protomodell der Intelligenzstrukturforschung

Während das BIS-Modell auf der Basis einer Inventarisierung und empirischen Testung von Aufgaben formuliert wurde, wurde das „hierarchische Rahmen- bzw. Protomodell

der Intelligenzstrukturforschung (HPI)“ [64] aus den invarianten Merkmalen unterschiedlicher Intelligenztheorien abgeleitet. D. h. die Autoren haben auf der theoretischen Ebene analysiert, welche Annahmen bezüglich der Struktur, des Inhalts und der Hierarchie in mehreren Intelligenztheorien vorkommen. Aspekte, die mehreren Intelligenztheorien gemeinsam sind, wurden von den Autoren als modellübergreifend bedeutsam gewertet. Deutliche inhaltliche Konvergenzen verschiedener Intelligenzmodelle können nach Ansicht der Autoren festgestellt werden für die Fähigkeit zum schlussfolgernden Denken („reasoning“), verbale, numerische und räumlich-figurale Fähigkeiten, Kreativität und, mit Einschränkungen, für die Merkfähigkeit und wahrnehmungsbezogene Fähigkeiten. Das HPI ist ein hierarchisches Modell: Die Dimensionen verbale, numerische und figurale Intelligenz werden auf der übergeordneten Ebene zum schlussfolgernden Denken bzw. zur fluiden Intelligenz zusammengefasst, die Kenntnisse in verschiedenen Wissensdomänen werden auf der übergeordneten Ebene zum Wissen bzw. zur kristallinen Intelligenz gebündelt.

### 6.3.3 Modifiziertes Modell der „Primary Mental Abilities“

Das modifizierte Modell der „Primary Mental Abilities“ (MMPMA) ist eine direkte Weiterentwicklung des Modells von Thurstone, dem Modell der „Primary Mental Abilities“ (PMA) [110]. Nach dem ursprünglichen PMA-Modell lassen sich intellektuelle Leistungen im Wesentlichen auf die in Kasten 6.4 genannten sieben Primärfähigkeiten zurückführen.

Kasten 6.4 Die sieben Fähigkeiten nach Thurstones Modell der „Primary Mental Abilities“, Thurstone und Thurstone [110]

- R „Reasoning“: schlussfolgerndes Denken
- S „Space“: räumliches Denken
- N „Number“: rechnerisches Denken, Beherrschung von (relativ einfachen) Rechenoperationen
- V „Verbal comprehension“: sprachliches Denken
- M „Memory“: Merkfähigkeit (kurzfristiges Behalten relativ einfacher Sachverhalte)
- W „Word fluency“: Flüssigkeit sprachlicher Einfälle
- P „Perceptual speed“: rasches Erkennen von Details

Diese Fähigkeiten werden im MMPMA auf einer „Stratum II“ genannten Ebene angesiedelt, wobei das schlussfolgernde Denken als Zusammenfassung des sprachlichen, rechnerischen und räumlichen Denkens hierarchisch übergeordnet ist. Zusätzlich werden zwei weitere Hierarchie-Ebenen formuliert. An der Spitze der Hierarchie stehen auf der sogenannten Stratum-III-Ebene fluide und kristalline Intelligenz, auf der untersten Ebene (Stratum I) spezifische Fähigkeiten wie Kenntnisse in bestimmten Wissensdomänen.

### 6.3.4 Gemeinsamkeiten aktueller Intelligenzmodelle

Die drei Modelle BIS, HPI und MMPMA weisen u. a. die folgenden beiden Gemeinsamkeiten auf. Alle drei Modelle nehmen eine hierarchische Ordnung der Fähigkeiten vor, mit der allgemeine Intelligenz (BIS) oder aber der fluiden und kristallinen Intelligenz (HPI, MMPMA) an der Spitze der Hierarchie. Darüber hinaus realisieren die drei Modelle den facettentheoretischen Ansatz. Im BIS-Modell sind auf der Ebene der 12 Zellen die Inhalte und Operationen durchgängig bimodal konzipiert, im HPI-Modell sind die Fähigkeiten auf der übergeordneten Ebene (schlussfolgerndes Denken, Wissen, fluide und kristalline Intelligenz) mit den Inhaltsfaktoren (verbal, numerisch und figural) gekreuzklassifiziert und im MMPMA ist das schlussfolgernde Denken bezüglich der Inhalte ausbalanciert.

---

## 6.4 Ausgewählte Intelligenztests

Es gibt zahlreiche Intelligenztests, von denen wir hier nur einige Verfahren nennen, wobei wir uns auf deutschsprachige Tests konzentrieren, die im Kontext der internen und externen Personalauswahl eingesetzt werden können.

Zu den Verfahren, die der Ableitung nur eines Indikators für nur eine globale Fähigkeit dienen, zählen die figuralen Matrizenaufgaben von Raven [82], die die fluide Intelligenz indizieren sollen oder der verbale Mehrfachwahl-Wortschatz-Test (MWT) von Lehrl [61], der die kristallisierte Intelligenz messen soll. Dabei wird die globale Fähigkeit mit nur einem Aufgabentyp (gleichwohl mit vielen Items) erfasst. Weitere Matrizen-tests sind der adaptive Matrizen-test [35] und der Bochumer Matrizen-test (BOMAT) [38]. Zu den Wissenstests mit nur einem Aufgabentyp (aber vielen Items) zählt der Bochumer Wissenstest BOWIT [37].

Andere Tests erfassen eine Fähigkeit mit unterschiedlichen Aufgabentypen, so wird mithilfe des CFT-20 R [116] die fluide Intelligenz beispielsweise mit vier unterschiedlichen Aufgabentypen gemessen, wobei alle Aufgabentypen die Bearbeitung figuralen Materials erfordern.

Daneben gibt es Tests zur Messung mehrerer Intelligenzdimensionen. Bei einigen dieser Tests kann auf einer übergeordneten Ebene zusätzlich ein globales Intelligenzmaß bestimmt werden. Ein prominentes Beispiel hierfür ist die „Wechsler Adult Intelligence Scale IV“ WAIS-IV [81], das Nachfolge-Verfahren zum Hamburg-Wechsler Intelligenztest für Erwachsene (HAWIE-R). Während bei der Interpretation der Wechsler-Tests die Leistungen früher in einen Verbal- und in einen Handlungsteil differenziert wurden, sieht die Auswertung aktuell vier sogenannte „Indexwerte“ vor, die als Sprachverständnis, wahrnehmungsgebundenes logisches Denken, Arbeitsgedächtnis und Verarbeitungsgeschwindigkeit interpretiert werden. Darüber hinaus kann ein Gesamt-IQ bestimmt werden. Ungeachtet der hohen Bedeutung der Wechsler-Tests für die Geschichte der Intelligenztests spielt der Wechsler-Test in der Praxis der Eignungsdiagnostik kaum eine Rolle, da die Durchführung aufwendige Einzeltestungen vorsieht.

Aus Aufwandsgründen kommt auch der Test zum BIS, der BIS-4-Test [43] in der Praxis kaum zum Einsatz. Zwar können mit dem BIS-4 Gruppentestungen durchgeführt werden, der Test sieht aber keine Antwortbogen vor, sodass die Auswertung einen erheblichen Aufwand erfordert. Eine praktikabel durchführbare Variante des BIS-4 (mit elektronisch einlesbaren Antwortbogen) haben Kersting und Beauducel [51] mit dem BIS-r-DGP Test entwickelt, dieser Test steht als „confidential test“ aber nicht zur allgemeinen Verfügung. Daher verzichten wir hier auf die Darstellung des BIS-Tests, obwohl diesem Test zweifelsohne eine Leitfunktion für Intelligenztests zukommt und sich der Einsatz insbesondere für Studien zur Konstruktvalidität von Intelligenztests empfiehlt.

Tests zu den anderen beiden vorgestellten Intelligenzmodellen, dem HPI und dem MMPMA sind der I-S-T 2000 R [64] sowie der WIT-2 [50]. In der Tab. 6.1 werden einige Aspekte der beiden Verfahren gegenübergestellt.

Durch die konsequente Ausrichtung der Testkonstruktion des I-S-T 2000 R auf eine explizite theoretische Grundlage wurde ein konstruktvalides Instrument geschaffen, das vor allem mit der Option der Gewinnung von Indikatoren für die fluide und kristalline Intelligenz überzeugt. Dabei wurde ein konzeptionelles Problem anderer Tests überwunden. Bei vielen anderen Tests sind die Operationen des Denkens mit den Inhalten konfundiert. So wird behauptet, Matrizenests wie der Raven-Test würden fluide Intelligenz messen – dabei gehören auch die Inhaltsdimensionen verbal und numerisch, die in diesem Test nicht vorkommen, zur fluiden Intelligenz. Auf der anderen Seite wird die kristallisierte Intelligenz mit der Inhaltsdimension „verbal“ konfundiert, wenn man lediglich verbale Wissenstests einsetzt. Im I-S-T 2000 R werden auch bei den Wissensfragen die drei Inhaltsbereiche verbal, numerisch und figural konsequent berücksichtigt. Damit ist es anhand des I-S-T 2000 R im deutschen Sprachraum erstmalig möglich, Diagnosen der fluiden und kristallisierten Intelligenz im Sinne der Investmenttheorie von Horn und Cattell [34] weitgehend unabhängig von Inhaltskomponenten zu erstellen. Der I-S-T 2000 R überzeugt über alle Testgütekriterien hinweg, lediglich der Nachweis der Retest-Reliabilität fehlt und die Studien zur Kriteriumsvalidität, insbesondere zur Kriteriumsvalidität der Faktorwerte für fluide und kristallisierte Intelligenz, könnten ausgebaut werden. Unpraktisch für die Interpretation der Testwerte ist, dass die Normbasis für die einzelnen Module jeweils unterschiedlich ausfällt.

Liegt die Stärke des I-S-T 2000 R in der konsequenten Umsetzung des theoretischen Modells und der Konstruktvalidität, profiliert sich der WIT-2 vor allem durch den Berufsbezug und die Kriteriumsvalidität. Es werden gezielt relevante berufliche Schlüsselqualifikationen erfasst, beispielsweise nicht nur das schlussfolgernde Denken, sondern auch die Arbeitseffizienz. Aufgrund einer vollständigen Modularisierung ist der Testeinsatz des WIT-2 sehr flexibel möglich. Die Auswertung kann auf jeden individuellen Anwendungsfall hin anforderungsspezifisch gewichtet erfolgen, dazu findet sich ein konkreter Vorschlag im Testmanual. Der Merkfähigkeitstest und der Arbeitseffizienztest sind explizit berufsbezogen eingekleidet, die beiden Wissenstests zielen auf Wissensdomänen, die in vielen Berufen relevant sind, ab. Dies dürfte sich positiv auf die Augenscheinvalidität und somit auf die Akzeptanz auswirken. Allerdings sind die beiden Wissenstests



**Tab. 6.1** Merkmale des I-S-T 2000 R und des WIT-2

	I-S-T 2000 R erweitert	WIT-2
Sprachliches Denken	•	•
Rechnerisches Denken	•	•
Räumliches Denken	•	•
Schlussfolgerndes Denken	•	•
Merkfähigkeit	• Mit verbalem und figuralem Material	• Mit Störaufgabe zw. Einprägen und Abruf, mit verbalem, numerischem und figuralem Material, ausgewertet wird ein Gesamtwert
Arbeitseffizienz	–	•
Wissen	• 73 Items, Geografie/Geschichte, Kunst/Kultur, Naturwissenschaften, Mathematik, Alltag, verbal, numerisch und figural	• 40 Items, Informationstechnologie und Wirtschaft
Fluide Intelligenz	•	–
Kristallisierte Intelligenz	•	–
Faktorwerte	•	–
Objektivität	•	• Separates Anweisungsheft, Übungsheft für Teilnehmer/-innen
Reliabilität – Konsistenz	•	•
Reliabilität – Re-Test	–	•
Paralleltest	•	–
Konstruktvalidität	•	•
Kriteriumsvalidität	•	• Zahlreiche Studien
Normen	• N = 5800, keine Ernstfallerhebungen, hohe Altersheterogenität (15-60 Jahre)	• N = 2285 bis 10.024 pro Aufgabe, 15–42 Jahre, Normdaten „im Ernstfall“ erhoben
Akzeptanz	•	• Die Akzeptanz des WIT-2 wird bezüglich der Kontrollierbarkeit und im Gesamtwert besser eingeschätzt als die des I-S-T 2000 R (Kersting, 2008b)
Besonderheiten	• Sehr weit verbreitetes Verfahren	• Berufsorientierung

so spezifisch, dass keine Schlussfolgerungen auf die kristalline Intelligenz möglich sind. Diesbezüglich sollte der Test in Zukunft erweitert werden. Die umfassenden Normdaten zeichnen sich dadurch aus, dass sie im Kontext der Personalauswahl und nicht – wie sonst üblich – lediglich im Forschungskontext erhoben worden. Zur Kriteriumsvalidität wurden umfangreiche Studien vorgestellt, die – unter Einbezug der Daten des WIT-1-Tests – zu einer Metaanalyse zusammengefasst werden konnten [60]. Interessant ist, dass

mit dem Merkfähigkeitstest nicht nur die unmittelbare Behaltensleistung erfasst wird. Während bei den meisten Tests direkt nach der Einprägphase der Merkfähigkeitstest erfolgt, müssen die Teilnehmer/-innen beim WIT-2 nach dem Einprägen und vor dem Erinnerungstest eine Störaufgabe im Umfang von ca. 20 min bearbeiten. Der Test orientiert sich explizit an der DIN 33430, einer DIN-Norm zur Personalauswahl.

Ein weiterer „breiter“ deutschsprachiger Intelligenztest ist die Intelligenz-Struktur-Batterie (INSBAT) [3], die 14 Subtests umfasst und auf dem Cattell-Horn-Carroll-Modell (CHC) [69, 70] basiert. Neben der allgemeinen Intelligenz können auf einer weiteren Ebene fluide und kristalline Intelligenz, quantitatives Denken, visuelle Verarbeitung und das Gedächtnis interpretiert werden. Das CHC-Modell liegt auch dem SMART-Test [49] zugrunde, auf den wir weiter unten im Kontext der Akzeptanz von Intelligenztests näher eingehen.

Während die bislang besprochenen Verfahren sowohl als Papier-Bleistift-Test als auch als Computertest vorliegen, ist die INSBAT ein reiner Computertest. Es gibt verschiedene Anbieter, die Intelligenztests für die Eignungsdiagnostik ausschließlich online anbieten, etwa die Verfahren „numerical screening“ bzw. „verbal Screening“ von CEB/SHL, das Verfahren „scales“ (numerical/verbal) von „cut-e“, „Diagrammanalysen“ und weitere Verfahren von Eligo, das Verfahren „Kappa“ sowie weitere Verfahren von HR Diagnostics sowie die Verfahren der ITB-Consulting. Bei der Bearbeitung von Intelligenztests spielt es eine bedeutsame Rolle, ob die Verfahren mit oder ohne Aufsicht durchgeführt werden (siehe dazu das Kapitel von Ott, Ulfert und Kersting in diesem Band).

---

## 6.5 Zur Einsatzhäufigkeit und zur Akzeptanz von Intelligenztests

Intelligenztests werden im deutschsprachigen Raum vergleichsweise selten eingesetzt. Ryan et al. [85] analysierten die Personalauswahlpraxis von fast 1000 Organisationen aus 20 Industrieländern. Deutschland belegte in diesem internationalen Vergleich den drittletzten Platz, wenn es um die Häufigkeit des Einsatzes von Tests zur kognitiven Kompetenz ging. Zur Situation in Deutschland in Bezug auf die externe Personalauswahl liegen u. a. Daten für das Jahr 1993 [93] und für das Jahr 2003 vor [98]. Während die Einsatzhäufigkeit von Intelligenztests bei 34 % (im Jahr 1993) bzw. 30 % (2003) stagniert, kommen strukturierte Interviews auf eine Einsatzhäufigkeit von 70 % (1993) bzw. 82 % (im Jahr 2003), Arbeitsproben auf 44 % (1993) bzw. 45 % (2003) und Assessment-Center auf 39 % (1993) bzw. 58 % (2003). Zur internen Personalauswahl werden Intelligenztests noch seltener eingesetzt [32].

Die Einsatzhäufigkeit von Intelligenztests nimmt mit zunehmendem Alter, zunehmender (Berufs-)Erfahrung und zunehmendem hierarchischen Status der internen und externen Bewerber/-innen ab. Obwohl die Intelligenz umso bedeutsamer wird, umso komplexer die beruflichen Anforderungen sind [60], beschränkt sich der Einsatz von Intelligenztests häufig auf die Auswahl von Auszubildenden. Gründe hierfür sind verschiedene Annahmen und

Vorurteile von Personalverantwortlichen. Verbreitet ist beispielsweise die Annahme, auf beruflich „höherer Ebene“ könnte man sich die Intelligenzdiagnostik sparen. Schließlich seien auf dieser Ebene alle Personen intelligent und hätten dies auch durch (Hoch-)Schulabschlüsse, Berufserfolge und ggf. Promotionen nachgewiesen. Aber auch – und gerade – auf dieser Ebene macht die Intelligenz den Unterschied [57], sie lässt sich nicht ausreichend durch Interviewfragen schätzen oder aus dem Lebenslauf ablesen. Die Aussagekraft von Abschlüssen und Prädikaten im individuellen Lebenslauf sinkt in dem Maße, in dem die Anzahl von Abschlüssen und Prädikaten in der Gesellschaft insgesamt steigt. Die Zahl der Personen mit Hochschulzulassung und die Zahl der Studierenden sind ebenso immens gestiegen wie die Anzahl an „Bestnoten“. An die Stelle der aufwendigen Akademisierung junger Menschen ist die geschmeidige Produktion von akademischen Abschlüssen getreten. Es ist nicht davon auszugehen, dass alle Personen, die in gehobenen Positionen arbeiten, einen Intelligenzwert im obersten Bereich der Verteilung haben. Aber selbst wenn dies der Fall wäre, dürfte man auf die Diagnostik der Intelligenz innerhalb dieser homogen leistungsstarken Gruppe nicht verzichten, denn selbst innerhalb der Gruppe der sehr intelligenten Personen führen Intelligenzunterschiede noch zu Leistungsunterschieden [59, 66, 113].

Noch ein weiterer Grund spricht dafür, nicht nur auf die Biografie, sondern auch auf die Potenziale von internen und externen Bewerber/-innen zu achten. Die Idee, die Vergangenheit als Prädiktor für die Zukunft zu nutzen (das biografische Prinzip) stößt grundsätzlich an seine Grenzen, wenn es um die Prognose der Bewährung in einer neuen Situation geht (z. B. die Bewährung in einer anderen Unternehmens- und Führungskultur, in einer anderen Unternehmens- und Marktsituation usw.). Hier kommt es weniger auf die Erfahrung und mehr auf das Potenzial an – mit anderen Worten: auf die Intelligenz.

Häufig verhindert das Vorurteil, Intelligenztests seien erwachsenen Menschen schlichtweg nicht zumutbar, den Einsatz von Intelligenztests. Selbst Personen, die die Treffsicherheit intelligenztestbasierter Aussagen nicht bezweifeln, schrecken oft aus Sorge um die Akzeptanz vor einer Intelligenzdiagnostik zurück [56]. Grundsätzlich ist es wichtig, eine hohe Akzeptanz der zur Eignungsdiagnostik eingesetzten Verfahren und Prozesse anzustreben. Das Akzeptanzurteil korreliert, wie Hausknecht et al. [30] metaanalytisch zeigen konnten, substanzial mit der wahrgenommenen Attraktivität der Organisation, mit der Absicht, die Organisation weiterzuempfehlen und der Absicht, ein potenzielles Angebot der Organisation anzunehmen. Die Sorge, dass Intelligenztests grundsätzlich keine Akzeptanz finden, ist jedoch unbegründet. Anderson et al. [2] führten eine Metaanalyse u. a. zum Thema Akzeptanz durch, dabei belegten Intelligenztests einen mittleren Platz, zwar hinter Interviews, aber (nominell) vor Persönlichkeitsfragebogen. In Bezug auf die von Kandidat/-innen gewünschte Wahrung der Privatsphäre gingen Intelligenztests als „Sieger“ hervor.

Kersting [48] führte eine Akzeptanzbefragung u. a. zu den Verfahren BIS-4, d2, I-S-T 2000 R, Raven und WIT-2 durch. Alle Tests wurden im positiven Bereich beurteilt. Die Personen, die den Test bearbeitet haben, schätzen insbesondere, dass die Testung eine sehr gut kontrollierbare Situation darstellt. Alle Tests, insbesondere aber die Tests mit nur einem Aufgabentyp (d2, Raven), wurden allerdings unter dem Gesichtspunkt der

Augenscheinvalidität relativ kritisch bewertet. Die Augenscheinvalidität ist für das generelle Akzeptanzurteil jedoch entscheidend.

Einige Tests bemühen sich um eine hohe Augenscheinvalidität, indem die Aufgaben semantisch so eingekleidet sind, dass sie zur Berufswelt passen. Dazu gehören die oben genannten Online-Tests „numerical screening“, „verbal screening“, „scales“ (numerical/verbal), Diagrammanalysen, „Kappa“ und SMART [49] aber auch Tests, die sowohl in computerisierter Form als auch in Papierform vorliegen, beispielsweise das Verfahren Analyse des schlussfolgernden und kreativen Denkens (ASK) von Schuler und Hell [97] oder die Arbeitsprobe zur berufsbezogenen Intelligenz (AZUBI-BK) von Schuler und Klingner [99].

Beim SMART-Test [49] sind alle Aufgaben in eine Rahmengeschichte (Overarching Scenario) eingebunden. Die Teilnehmer/-innen werden in die Rolle eines Mitarbeiters eines Online-Versandhandels (der fiktiven Organisation „inversagi“) versetzt. Die realitätsnahen Aufgaben, die die Person zu lösen hat, ergeben sich aus der Rahmengeschichte und es entsteht der Eindruck, dass die einzelnen Aufgaben des Tests miteinander verbunden sind. Tatsächlich sind die Aufgaben aber, das ist für die Messung wichtig, unabhängig voneinander. Anders als bei den meisten herkömmlichen Intelligenztests können die Teilnehmer/-innen – wie in der Realität – selbst entscheiden, in welcher Reihenfolge sie die Aufgaben bearbeiten und wie sie die insgesamt zur Verfügung stehende Zeit auf die einzelnen Aufgaben aufteilen. Für Rechenaufgaben steht ihnen – ebenfalls wie im Alltag – ein Taschenrechner zur Verfügung. In empirischen Studien [49] konnte gezeigt werden, dass der SMART trotz dieser Besonderheiten Facetten der Intelligenz (konzipiert nach dem CHC-Modell) konstruktvalid erfasst und – wie intendiert – eine hohe Akzeptanz findet.

---

## 6.6 Kriterien der Testauswahl

Es ist sehr anspruchsvoll, unter der Vielzahl der verfügbaren Intelligenztests den für den jeweiligen Einsatzzweck geeigneten Test zu finden. Eine Hilfe bieten zunächst die weiter oben beschriebenen Intelligenzmodelle, insbesondere das BIS. Mithilfe der Modelle ist es möglich, die – unbedingt erforderlichen – Ergebnisse der Anforderungsanalyse (siehe Kapitel von Nerdinger im vorliegenden Band) in psychologische Eignungsmerkmale zu übersetzen. Bezüglich der kognitiven Kompetenzen sollte man sich dabei an einem theoretischen Modell, z. B. dem BIS, orientieren. Um sich innerhalb des Angebots an denjenigen Tests, die das anvisierte Merkmal messen, für einen konkreten Test zu entscheiden, muss man im nächsten Schritt Informationen über die Tests einholen und bewerten. Zu der Frage, welche Informationen für die Qualitätsbeurteilung relevant sind und wie diese zu bewerten sind, hat das Diagnostik- und Testkuratorium Richtlinien herausgebracht [108]. Diese Richtlinien sind Bestandteil des Testbeurteilungssystems des Testkuratoriums (TBS-TK). In Tab. 6.2 sind die Ergebnisse der bislang verfügbaren TBS-TK-Rezensionen zu Intelligenztests abgebildet.

**Tab. 6.2** Beurteilung der Intelligenz- und Konzentrationstests für Erwachsene, für die eine Rezension nach dem TBS-TK-System [108] vorliegt. (Stand: April 2016)

Testname/Autoren	Allg. Info	Objektivität	Zuverlässigkeit	Validität
I-S-T 2000 R [64, 92]	++	++	+	+
FAKT-II [73, 115]	+	++	+	+
INSBAT [3, 23]	+	+	++	+
CFT-20-R [116, 27]	+	+	+	+
WIE [4, 71]	++	+	+	–
d2-R [12, 20]	+	+	+	–
MSCEIT [100, 13]	+	+	+	–

Die Anzahl der bislang nach dem TBS-TK-System rezensierten Tests (siehe [www.zpid.de/tk](http://www.zpid.de/tk)) bleibt weit hinter der Anzahl der verfügbaren Tests zurück. Die Richtlinien können aber gedanklich auf alle Tests angewendet werden. So kann man jene Tests, die potenziell infrage kommen, weil sie die im Rahmen der Anforderungsanalyse identifizierten relevanten Fähigkeiten erfassen, nach dem Schema der TBS-TK-Rezensionen beurteilen. Zunächst wird dazu geprüft, ob zu dem Test überhaupt die Informationen vorliegen, die notwendig sind, um den Test beurteilen zu können. Welche Informationen das sind, wurde im „DIN Screen“ von Kersting [47] definiert. Diese Checkliste dient als „Standard zur Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens des Testkuratoriums der Föderation Deutscher Psychologenvereinigungen“. Aktuelle Testverfahren wie z. B. die oben genannten Verfahren d2-R [12] und WIT-2 [50] veröffentlichen bereits im Testhandbuch eine Übersicht, an welcher Stelle des Handbuchs die geforderten Informationen zu finden sind. Nur wenn diese Informationen überhaupt vorhanden sind, kann die Qualität eines Tests beurteilt werden. Wie die Beurteilung erfolgt, regeln die Beurteilungsrichtlinien [108]. Zum Zeitpunkt der Publikation des vorliegenden Beitrags wurde die überarbeitete Form der DIN 33430 publiziert. Das TBS-TK-System bezieht sich indirekt auf diese Norm. Daher wird in Folge der Modifikation der DIN auch das Testbeurteilungssystem modifiziert, die hier geschilderten Prinzipien bleiben aber unverändert. Wir empfehlen dem Leser, die jeweils aktuellste Version des TBS-TK-Systems heranzuziehen.

## 6.7 Kreativität

Kreativität wird häufig als geheimnisvolles Phänomen, als Komplex, als Syndrom oder als Symphonie beschrieben [75, 84, 114]. Nach über 70 Jahren intensiver Kreativitätsforschung ist bereits eine Vielzahl an Merkmalen zur Definition, Beschreibung und Erfassung von Kreativität bekannt (vgl. zum Verständnis von Kreativität am Arbeitsplatz [119]). Traditionell wird innerhalb der Kreativitätsforschung zwischen den folgenden vier Zugängen zur Kreativität unterschieden: das kreative Produkt („product“), die

kreative Person („person“), der kreative Prozess („process“) und die kreative Umwelt („press“) [72, 83]. Im Folgenden werden die „4p“ kurz vorgestellt.

- **Produkt:** Die Kreativitätsforschung bezeichnet all jene Produkte als innovativ, die zwei Attribute erfüllen: Sie sind neuartig und nützlich. Neuartig bedeutet: Ungewöhnlich, einzigartig, aus neuen Blickwinkeln betrachtet, variiert, originell, von bisherigen Mustern abweichend, adaptiert. Und nützlich umfasst: Wertvoll, effektiv, effizient und einen Beitrag für die Gemeinschaft leistend [79].
- **Person:** Die Personalpsychologie richtet ihren Fokus vornehmlich auf die Identifizierung jener personalen Eigenschaften, die kreatives Potenzial begründen bzw. effektiv abrufen helfen. Auf individueller Ebene wird kreatives Potenzial durch eine Kombination aus kognitiven (Intelligenz, Expertise), nicht-kognitiven (Persönlichkeit) und motivationalen („need for creativity“, Leistungsmotivation, Interessen) Faktoren definiert [78].
- **Prozess:** Gerade in hochgradig erfahrungs- bzw. wissensbasierten Domänen entstehen kreative Ideen nicht über Nacht. Die Forschung zum Kreativitätsprozess befasst sich mit den Herausforderungen, denen Kreative auf ihrem teils langen und mühevollen Weg von der initialen Problementdeckung bzw. Bedarfsidentifikation hin zu einer nachhaltigen und akzeptierten kreativen Lösung begegnen und beschreibt die relevanten Fähigkeiten und Verhaltensweisen in den jeweiligen Prozessphasen (für einen Überblick sei auf folgende Quellen verwiesen: [39, 78]). Ein grundlegendes Prozessverständnis unterscheidet vier Prozessstufen: Vorbereitung, Inkubation und Illumination, Verifikation und Implementierung [65].
- **Press (Umwelt):** Unter Umwelt werden die situativen und sozialen Kontextfaktoren zusammengefasst, die kreative Erfolge ermöglichen (oder verhindern), wie etwa der Handlungsspielraum, der Führungsstil und die wahrgenommene Unterstützung bzw. Bereitstellung von Ressourcen. Für einen Überblick auf Einflussvariablen im Unternehmenskontext sei auf folgende Beiträge verwiesen: [55, 77, 96] oder [118].

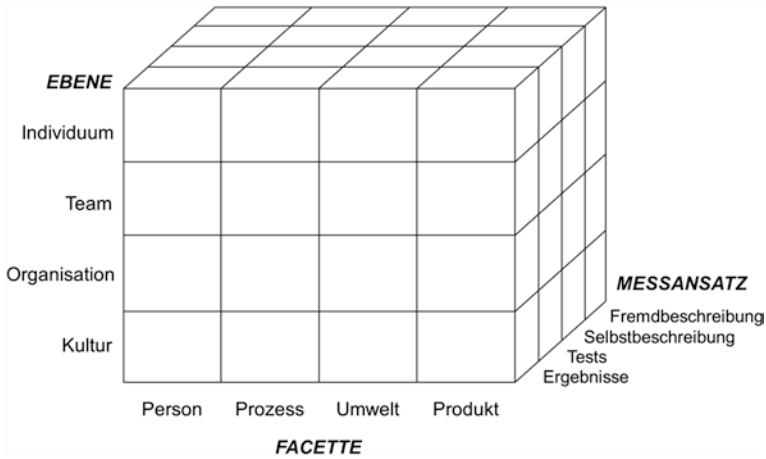
Damit lässt sich Kreativität wie folgt definieren:

- ▶ Kreativität ist die Interaktion zwischen Intelligenz, Expertise, Persönlichkeitsmerkmalen und Motivation als individuelle Dispositionen mit prozessstufen-spezifischen Anforderungen und Umgebungsbedingungen durch welche ein Individuum oder eine Gruppe ein reales Produkt schafft, das innerhalb eines sozialen Kontexts sowohl als neuartig als auch nützlich erachtet wird [78].

---

## 6.8 Kreativität und Intelligenz

Um die Vielzahl an Theorien, Modellen und Operationalisierungen von Kreativität besser einordnen zu können, empfiehlt sich eine Erweiterung des bewährten 4p-Modells (Produkt, Person, Prozess und Press/Umwelt). So können zum einen die verschiedenen sozialen Betrachtungsebenen und zum anderen unterschiedliche Messverfahren



**Abb. 6.2** Heuristisches Rahmenmodell der Kreativität [7]

berücksichtigt werden. Abb. 6.2 zeigt das in Anlehnung an Batey [7] konzipierte heuristische Rahmenmodell der Kreativität, wie es bei Palmer [78] beschrieben ist.

Im Rahmen dieses Beitrags interessiert vornehmlich die Beschreibung und Erfassung kreativen Potenzials mittels (kognitiver) Tests auf individueller Ebene; im Rahmenmodell also die Kombination aus Personenfacette, Individual-Ebene und Testverfahren. Aus der Vielzahl an Modellen und Tests, die sich in dieser Zelle befinden, kann aus Platzgründen nur ein Ausschnitt besprochen werden. Hierbei soll insbesondere auf die kognitiven Aspekte von Kreativität fokussiert werden.

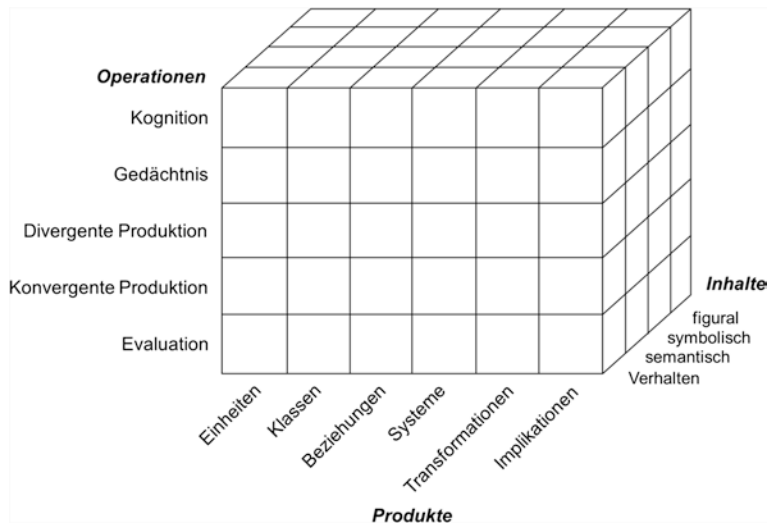
Die Relevanz von Kreativität für das Intelligenzkonzept zeigt Hell [31] auf. Im Modell von Thurstone [109] ist kreatives Denken durch „word fluency“ repräsentiert, im Modell von Vernon [112] durch „creative abilities“. Im Modell von Cattell und Horn [18] wird mit „fluency“ ein Faktor integriert, der ebenfalls mit kreativem Denken assoziiert ist. Auch Carroll [14] nimmt Kreativität mit „expressional fluency/orginality/creativity“ in sein Modell kognitiver Fähigkeiten auf. Und wie bereits beschrieben, findet Kreativität unter der Bezeichnung „Einfallsreichtum“ auch Eingang in das Berliner Intelligenzstrukturmodell (BIS) [42]. Die unterschiedlichen Bezeichnungen der Kreativitätselemente deuten darauf hin, dass Kreativität in Intelligenztests unterschiedlich operationalisiert wird.

### 6.8.1 Divergentes Denken

Von J. P. Guilford stammt eines der früheren und mit Sicherheit eines der umfangreichsten Modelle zur Intelligenzstruktur: das Strukturmodell der Intelligenz (SOI), das in Abb. 6.3 dargestellt ist.

Nach Guilford lässt sich jede kognitive Fähigkeiten durch drei Parameter beschreiben: Operationen, Inhalte und Produkte [29]. Unter „Operationen“ werden die fünf Hauptarten intellektueller Aktivitäten oder informationsverarbeitender Prozesse angeordnet:





**Abb. 6.3** Strukturmodell der Intelligenz nach Guilford [29]

Kognition, Gedächtnis, Divergente Produktion, Konvergente Produktion und Evaluation. Die zweite Dimension umfasst die „Inhalte“ bzw. die Bereiche der verarbeiteten Informationen: figural, symbolisch, semantisch und Verhalten. Letztlich werden in einer dritten Dimension die „Produkte“ informationsverarbeitender Aktivitäten klassifiziert. Guilford unterscheidet hier anhand der formalen Art der Informationen nach Einheiten, Klassen, Beziehungen, Systeme, Transformationen und Implikationen.

Durch die Operation „divergente Produktion“ sind Aspekte von Kreativität in das Strukturmodell der Intelligenz integriert. „Divergente Produktion“ ist definiert als die „Entwicklung logischer Alternativen aus gegebener Information, wobei die Betonung auf der Verschiedenheit, der Menge und der Bedeutung der Ergebnisse aus der gleichen Quelle liegt. Beinhaltet wahrscheinlich auch die Erinnerung an Transfer (ausgelöst durch neue Hinweise)“ [29, S. 34].

Krampen sieht eine gemeinsame Varianz von Intelligenz und Divergentem Denken von 25 % [54]. Neuere Forschungsergebnisse zum Zusammenhang von Divergentem Denken und Intelligenz deuten darauf hin, dass die gemeinsame Varianz vermutlich besonders durch die „Informationsverarbeitungsgeschwindigkeit“ („mental speed“) erklärt werden kann [78].

Zwar ermöglicht divergentes Denken das Generieren vielfältiger Lösungen in einem offenen Problemraum [22] und stellt damit eine Grundvoraussetzung zur Erbringung kreativer Leistungen dar [33, 76], aber bereits Guilford [28] lehnte eine Gleichsetzung von Kreativität mit „divergenter Produktion“ ab und wies darauf hin, dass an der Entstehung kreativer Leistungen weitere kognitive Fähigkeiten beteiligt sind. In der Meta-Analyse von Kim [52] ergab sich eine mittlere Korrelation von  $r = ,22$  für Tests des Divergenten Denkens mit kreativen Leistungen.

## 6.8.2 Kreativität im Berliner Intelligenzstrukturmodell

Jäger, Süß und Beauducel sehen die Integration von Kreativität in ein Strukturmodell der Intelligenz als zwingend an, da Problemlöseprozesse nur „im Wechsel von kreativ divergenten mit konvergenten Phasen“ erfolgreich gemeistert werden können [43, S. 7]. Im zu ihrem Berliner Intelligenzstrukturmodell zugehörigen Test (aktuelle Auflage: BIS-4) wird Kreativität unter dem Begriff Einfallsreichtum erfasst. Die Definition von Einfallsreichtum ist in Kasten 6.3 oben wiedergegeben. Die Antworten in den insgesamt 12 Einfallsreichtum-Aufgaben im BIS-4 werden nach Ideenflüssigkeit und Ideenflexibilität ausgewertet. Die Autoren erachten Einfallsreichtum jedoch nicht als Substitut von Kreativität, sondern als „eine ihrer zentralen Komponenten“ (S. 7) neben beispielsweise Originalität.

Dass Einfallsreichtum tatsächlich einen wichtigen Zusammenhang zu Allgemeiner Intelligenz aufweist, zeigt sich in der hohen absoluten Korrelation des Einfallsreichtums-Faktorwerts mit Allgemeiner Intelligenz ( $.56$ ) [43]. Zur obersten Fähigkeitsebene, der Allgemeinen Intelligenz, weist Einfallsreichtum nach Verarbeitungskapazität die zweithöchste Korrelation auf und trägt damit stärker zu diesem generellen Intelligenzmaß bei als Merkfähigkeit oder Bearbeitungsgeschwindigkeit.

## 6.8.3 Kreativität und das Modell der fluiden und kristallinen Intelligenz nach Cattell (1943)

Cattell [17] verzichtet in seinem Intelligenz-Modell auf ein übergeordnetes Maß allgemeiner Intelligenz und teilt das Intelligenzkonzept in zwei Facetten auf: fluide und kristalline Intelligenzleistungen [18]. Unter „fluider Intelligenz“ ( $g_f$ ) wird die generelle Fähigkeit gefasst, zwischen Merkmalen – ganz gleich ob bereits bekannt oder noch unbekannt – unterscheiden zu können und Beziehungen zwischen ihnen zu erkennen [17].

„Kristalline Intelligenz“ ( $g_c$ ) hingegen stellt die Fähigkeit dar, erlerntes Wissen, das über die Zeit und in Abhängigkeit kulturspezifischer „Interessen“ durch formelle oder informelle Bildung bzw. generelle Lebensereignisse erworben wurde, zur Problemlösung einzusetzen [1, 70]. Unter kristalliner Intelligenz wird vornehmlich sprachbasiertes deklaratives („wissen was“) und prozedurales („wissen wie“) Wissen subsumiert [70].

Kreative Leistungen kommen gemäß Cattell vornehmlich durch fluide Intelligenz zustande; den Einfluss von kristalliner Intelligenz und Persönlichkeitsfaktoren sah er als nachgeordnet an [102]. Die Studien von Batey et al. [10] und Batey et al. [8] scheinen diese Annahme zu belegen.

In einer anderen Studie berichten Batey et al. [9] hingegen einen Zusammenhang von  $g_c$  – und nicht  $g_f$  – mit Kreativität. Beauducel und Kersting [11] prüften die Beziehung des „Modells der fluiden und kristallinen Intelligenz“ zum „Berliner Intelligenzstruktur-Modell“. Einfallsreichtum (gemessen über den BIS-r-DGP) [43] erwies sich zwar in einer explorativen Faktorenanalyse als mit  $g_f$  assoziiert, in der anschließenden konfirmatorischen Faktorenanalyse war Kreativität hingegen besser durch  $g_c$  repräsentiert. Durch

einen zusätzlichen Pfad zwischen „fluency“ und  $g_f$  konnte ein noch passenderes Modell modelliert werden. Beauducel und Kersting (ebd.) weisen darauf hin, dass die Befunde ihrer Studie keine Verallgemeinerung für das globale Kreativitätskonstrukt erlauben. So haben sie Einfallsreichtum allein über Ideenflüssigkeit erfasst und nur die Quantität, nicht aber die Qualität der aufgabenbezogenen Antworten berücksichtigt, ebenso wurde auf Aufgaben zum numerischen Einfallsreichtum verzichtet.

#### 6.8.4 Cattell-Horn-Carroll-Modell

Das Cattell'sche Intelligenzmodell wurde später mit dem Three-Stratum-Modell der Intelligenz von Carroll (...) zum bereits in Abschnitt 6.3 angesprochenen Cattell-Horn-Carroll Modell (CHC) integriert. Der Fähigkeitsbereich Langzeitgedächtnis und -abfrage (long-term storage and retrieval; Glr; Stratum II) umfasst 13 darunterliegende Fähigkeiten. Einige dieser zugeordneten Fähigkeiten sind aus der Kreativitätsforschung bekannt [46]. So finden sich hier u. a. „associative memory“ (MA), „ideational fluency“ (FI), „figural flexibility“ (FX) und „originality/creativity“ (FO). Inwiefern kreative Leistungen auf Basis des CHC-Modells allerdings tatsächlich erklärt werden können, muss erst noch empirisch überprüft werden.

#### 6.8.5 Eine prozessbasierte Sicht auf Kreativität

Bereits früh haben Kreativitätsforscher/-innen darauf hingewiesen, zur Erklärung kreativer Leistungen aus kognitionspsychologischer Sicht nicht nur das divergente Denken, sondern auch schlussfolgerndes Denken sowie Expertise heranzuziehen. Eine prozessbasierte Sicht auf das Erbringen kreativer Leistungen verdeutlicht die unterschiedlichen Anforderungen, denen sich Kreative gegenübersehen.

Das Prozessmodell der Kreativität nach Schuler und Görlich [96] integriert bestehende Kreativitätsprozessmodelle nach empirischen Gesichtspunkten und kommt der Forderung nach einer stärkeren Anpassung des Kreativitätsprozessmodells an den beruflichen Kontext nach. In diesem Modell werden die folgenden acht Phasen unterschieden:

1. Problementdeckung,
2. Informationssuche, -aufnahme und -bewertung,
3. Konzeptkombination,
4. Ideenfindung,
5. Ausarbeitung und Entwicklung des Lösungsansatzes,
6. Ideenbewertung,
7. Anpassung und Umsetzung,
8. Implementierung.

Auch wenn der empirische Beleg für die Trennung in diese acht Prozessstufen noch aussteht, deutet sich an, dass sich über die Stufen hinweg differenzielle Korrelate klassischer konvergenter Intelligenzmaße (z. B. Matrizentests aus dem CFT 20-R) zu kreativen Leistungen (gemessen über die DBK-TE, Abschn. 6.3) ergeben [78]. Diese Zusammenhänge scheinen zudem geringer auszufallen als jene zwischen Maßen divergenten Denkens und Kreativität.

---

## 6.9 Kreativitätstests

Zur Erfassung kreativen Potenzials liegen ganz unterschiedliche Verfahrensarten vor: Selbsteinschätzungen, Adjektivchecklisten, Verhaltensbeobachtungen, biografische Inventare und die eigenschaftsbasierte Messung von Kreativität via Tests. In letztgenannter Verfahrensguppe lassen sich drei Verfahrensgruppen hinsichtlich ihres (hauptsächlich) erfassten Konstruktraums unterscheiden:

- Verfahren, welche die kognitiven Voraussetzungen überprüfen,
- Instrumente, die auf die Ausprägung relevanter Persönlichkeitsmerkmale abheben und
- Verfahren, welche die Motivstruktur messen.

Dieses Kapitel beschränkt sich auf kognitive Kreativitätstests, wobei aus Platzgründen lediglich einige wenige Kreativitätstests vorgestellt und nur ausgewählte Aspekte diskutiert werden können.

Lange Zeit war das Repertoire an kognitiven Kreativitätstests auf Tests zum Divergenten Denken beschränkt. Traditionelle Verfahren sind hier Guilfords Unusual-uses-Tests [29], die „Torrance Tests of Creative Thinking“ (TTCT) [111] oder die „Barron-Welsh Art Scale“ [117]. In einem modernen Verständnis der kognitiven Aspekte von Kreativität ergeben sich kreative Leistungen jedoch nicht allein durch divergente Denkleistungen, sondern auch durch schlussfolgerndes und bewertendes Denken [19]. Als Beispiele für deutschsprachige Verfahren, die sowohl divergentes als auch konvergentes Denken erfassen, sind der BIS-4 [43] oder die „Analyse schlussfolgernden und kreativen Denkens“ (ASK) [97] anzuführen.

Mit dem Verfahren „Diagnose berufsbezogener Kreativität“ ist es möglich, die berufsbezogene Kreativität entlang des kreativen Prozesses zu erfassen. Das Verfahren liegt bislang in zwei Varianten vor, die sich an unterschiedliche Zielgruppen richten: Je ein Test wurde für die breiten Tätigkeitsfelder „Planung und Gestaltung“ (DBK-PG) [95] und „Technik und Entwicklung“ (DBK-TK) [80] konzipiert.

Aus der Vielzahl an Validitätsbefunden zu Kreativitätstests werden hier ausgewählt Zusammenhänge zu Außenkriterien der Kreativität im beruflichen Umfeld vorgestellt. Es finden sich Korrelationen zwischen den Ergebnissen in Kreativitätstests und der Anzahl eingereichter Verbesserungsvorschläge [78]. Bei Entrepreneuren korreliert das kreative Potenzial mit Größen unternehmerischen Erfolgs, wie der Anzahl angemeldeter Patente,

der Anzahl an Mitarbeiter/-innen und dem Umsatz [79]. Schuler et al. [94] berichten einen Zusammenhang zwischen kognitiven Leistungstests der Kreativität und Vorgesetztenurteilen der Innovativität bei einer Stichprobe von Wissenschaftlern und Ingenieuren aus dem Forschungs- und Entwicklungsbereich ( $r = ,25$ ). Auch hier bestehen positive Zusammenhänge zur Anzahl eingereicherter und erlangter Patente – während die Ergebnisse von Intelligenztests keine Prädiktionskraft haben. Der Einsatz von Kreativitätstests weist folglich einen zusätzlichen Erklärungsbeitrag über die Prädiktionskraft von Maßen allgemeiner Intelligenz hinaus auf. Eine inkrementelle Validität von Kreativität über Intelligenz hinaus findet sich auch bei der Vorhersage von Schulnoten [78]. Dass eine Messung kreativen Potenzials auch einen wesentlichen Beitrag zur Vorhersage generellen beruflichen Erfolgs leisten kann, zeigen die Zusammenhänge von Kreativitätstests wie der DBK-TE mit Gehalt und dem Alter bei Übernahme der ersten Führungsposition [78].

---

## 6.10 Erweiterungen des Intelligenzbegriffs

Seit Beginn der Intelligenzforschung werden „alternative“ Intelligenzkonzepte propagiert. Das Konstrukt Intelligenz im Allgemeinen und die herkömmlichen Intelligenztests im Besonderen sind bei einigen Personen unbeliebt. Intelligenztests gehören zur „dunklen Seite“ der Diagnostik, vor der die andere, die bessere Diagnostik umso heller und leuchtender aufscheint. Hoffnungsträger waren auf der Seite der Messung u. a. die Kompetenzdiagnostik („testing for competence rather than for >intelligence<“ war der Titel eines Artikels von McClelland) [68], die „authentische“ Diagnostik [5] sowie die Diagnostik mit Problemlöseszenarien [21]. Auf der Seite der Konstrukte wurden u. a. die Lernfähigkeit, die operative, praktische, soziale und emotionale Intelligenz gefeiert. Dabei verliert ein Begriff jede wissenschaftliche Bedeutung, wenn man sämtliche Erscheinungen unter diesen Begriff subsumiert und ihn auf alles und jedes anwendet.

### 6.10.1 Praktisch-mechanisch-technisches Verständnis

Einen festen Platz in der Berufseignungsdiagnostik hat die Messung der praktischen Intelligenz im Sinne des praktisch-mechanisch-technischen Verständnisses. Vor langer Zeit konstruierte Tests wie der Mechanisch Technische Verständnistest (MTVT) [62] oder die Drahtbiegeprobe [63] werden noch immer zur Eignungsdiagnostik bei der Personalauswahl für Berufe mit einem handwerklichen Anteil eingesetzt. Die Kriteriumsvalidität der aufgrund solcher Tests getroffenen Aussagen ist – sofern die Fähigkeiten wirklich einschlägig für den Beruf sind – zufriedenstellend [74]. Probleme können sich daraus ergeben, dass Männer bei derartigen Aufgaben im Durchschnitt deutlich bessere Werte erzielen als Frauen. Dies würde bei einer Vergabe von Ausbildungsplätzen oder Jobs allein aufgrund der Testwerte zu einer Überrepräsentation von Männern führen, die nach bestimmten Fairnessauffassungen („fair share“) unerwünscht ist.

### 6.10.2 Praktische Intelligenz und Erfolgsintelligenz

Der Begriff „praktische Intelligenz“ wird allerdings auch für ein so genanntes „stilles Wissen“ („tacit knowledge“) genutzt, das – neben anderen Fähigkeiten und Erfahrungen – laut Sternberg und Wagner [104] zur Bewältigung von Alltagsanforderungen benötigt wird und mit herkömmlichen Intelligenztests nicht genügend erfasst wird. Später hat Sternberg [101] die praktische Intelligenz als eine der „Intelligenzen“ bezeichnet, die neben der analytischen und der kreativen Intelligenz die „Erfolgsintelligenz“ ausmachen. Zur Diagnostik der praktischen Intelligenz wurden „Situational Judgment Tests“ (SJT) entwickelt. Bei diesen Aufgaben sollen die Testteilnehmer/-innen verschiedene Reaktionen auf Situationen aus dem Ausbildungs- oder Arbeitsalltag dahin gehend bewerten, ob diese Reaktionen wirksam sind und/oder ihren eigenen Verhaltenspräferenzen entsprechen. Aktuelle Studien [58] wecken allerdings Zweifel daran, ob mit SJT wirklich das gewünschte Konstrukt erfasst wird, da die Aufgaben auch dann gelöst werden können, wenn man auf die Darstellung der Situation verzichtet. Dies spricht eher dafür, dass mit SJT explizites Wissen über sozial erwünschtes Verhalten gemessen wird.

### 6.10.3 Emotionale Intelligenz

Fast tumultartig war die Reaktion auf das Konzept der „emotionalen Intelligenz“. Den Anfang machten Salovey und Mayer [88], die den Begriff einführten, wobei weitgehend unklar blieb, wie sich der Begriff von dem bereits etablierten Konstrukt der sozialen Intelligenz abgrenzte. Wirklich „berühmt“ wurde das Konzept der emotionalen Intelligenz aber erst durch die Agitation von Goleman [25]. Mit dem „Mayer-Salovey-Caruso-Test zur emotionalen Intelligenz“ (MSCEIT), der seit 2011 auch in einer deutschen Fassung vorliegt [100], wird die emotionale Intelligenz erfasst. Es handelt sich nicht um einen Selbsteinschätzungsfragebogen, sondern um einen Leistungstest. Das Konzept der emotionalen Intelligenz hat die geweckten Hoffnungen nicht erfüllt. In der Metaanalyse von Joseph und Newman [44] zeigten sich substantielle Korrelationen zwischen Indikatoren der emotionalen Intelligenz und etablierten Konstrukten wie Gewissenhaftigkeit und akademischer Intelligenz, emotionale Intelligenz ist somit von etablierten Konstrukten schwer abzugrenzen. Die zusätzliche Berücksichtigung der emotionalen Intelligenz durch entsprechende Tests im Rahmen der Personalauswahl verspricht keinen wesentlichen zusätzlichen Erkenntnisgewinn.

---

## 6.11 Fazit

Intelligenz ist ein entscheidender Erfolgsfaktor in diversen Berufen, Organisationen und Branchen. Dennoch ist die Intelligenz alleine kein Garant für Berufserfolg. Sie ist eine notwendige, keine hinreichende Bedingung. Notwendig bedeutet: ohne Intelligenz

keine emotionale Intelligenz, keine Intuition und keine Kreativität. Auch und gerade das soziale Miteinander setzt Intelligenz voraus. Der Fetisch des Teams, die heiter propagierte Schwarmintelligenz, alles verkommt zur unproduktiven Kolchose, wenn es dem Einzelnen, insbesondere der Führungskraft, an Intelligenz mangelt. „Nicht hinreichend“ bedeutet: Intelligenz ist nicht alles, weitere Erfolgsfaktoren sind beispielsweise Gewissenhaftigkeit und Extraversion.

In der Frage des Einsatzes von Intelligenztests klaffen Theorie und Praxis bisher weit auseinander. Anders formuliert: Diesbezüglich besteht in der Praxis die Möglichkeit, mit nur kleinen Änderungen große Fortschritte in der Personalbeurteilung zu erzielen. Für die Auswahl eines konkreten Intelligenztests ist die Frage relevant, welche Intelligenzdimension mithilfe des Tests gemessen wird. Um dies zu klären, sollte man sich an etablierten Intelligenzmodellen wie z. B. dem BIS orientieren. Darüber hinaus ist es wichtig, dass der Test bestimmte Qualitätskriterien erfüllt. Diese Qualitätskriterien werden im TBS-TK-System formuliert. Die Wissenschaft ist gefordert, den Stand der Forschung zu kommunizieren und zugleich die Intelligenztests so weiter zu entwickeln, dass sie eine noch höhere Akzeptanz finden.

---

### Fragen

1. Was zeichnet aktuelle Intelligenzmodelle aus? Beschreiben Sie deren Inhalt.
2. Wieso sind theoretische Intelligenzmodelle wie das BIS hilfreich bei der praktischen Frage, welchen Intelligenztest man im Rahmen der Personalauswahl einsetzen will?
3. Wieso werden Intelligenztests trotz ihrer hohen Aussagekraft so selten eingesetzt – und wie lässt sich das ändern?
4. Bisher liegen nur wenige Testbeurteilungen nach dem TBS-TK-System vor – wieso ist das System dennoch für die Auswahl eines Testverfahrens nützlich?
5. Sollte man in der Praxis der Personalauswahl einen herkömmlichen Intelligenztest einsetzen, einen Test zur emotionalen Intelligenz oder beide Verfahren? Begründen Sie Ihre Aussagen.

---

### Literatur

1. Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, 22, 229–259.
2. Anderson, N., 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.
3. Arendasy, M., Hornke, L. F., Sommer, M., et al. (2012). *Intelligenz-Struktur-Batterie – INS-BAT*. Mödling: Schuhfried.
4. Aster, M., Neubauer, A., & Horn, R. (2006). *Wechsler Intelligenztest für Erwachsene WIE. Deutschsprachige Bearbeitung und Adaptation des WAIS-III von David Wechsler* (2., korrigierte Aufl.). Frankfurt: Pearson Assessment.



5. Baker, E. L., O'Neil, H. F., Jr., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist*, *48*, 1210–1218.
6. 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.
7. Batey, M. (2012). The measurement of creativity: From definitional consensus to the introduction of a new heuristic framework. *Creativity Research Journal*, *24*, 55–65.
8. Batey, M., Chamorro-Premuzic, T., & Furnham, A. (2010). Individual differences in ideational behavior: Can the big five and psychometric intelligence predict creativity scores? *Creativity Research Journal*, *22*, 90–97.
9. Batey, M., Chamorro-Premuzic, T., & Furnham, A. (2009). Intelligence and personality as predictors of divergent thinking: The role of general, fluid and crystallised intelligence. *Thinking Skills and Creativity*, *4*, 60–69.
10. Batey, M., Furnham, A., & Safiullina, X. (2010). Intelligence, general knowledge and personality as predictors of creativity. *Learning and Individual Differences*, *20*, 532–535.
11. Beauducel, A., & Kersting, M. (2002). Fluid and crystallized intelligence and the Berlin model of intelligence structure (BIS). *European Journal of Psychological Assessment*, *18*, 97–112.
12. Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2010). *d2-R. Test d2 – Revision. Aufmerksamkeits- und Konzentrationstest*. Göttingen: Hogrefe.
13. Burk, C. L., & Amelang, M. (2015). TBS-TK Rezension: „MSCEIT – Mayer-Salovey-CarusoTest zur Emotionalen Intelligenz. Deutschsprachige Adaptation des Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT)“. *Zeitschrift für Arbeits- und Organisationspsychologie*, *59*, 155–157.
14. Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
15. Carroll, J. B. (1996). A three-stratum theory of intelligence: Spearman's contribution. *Human abilities: Their nature and measurement, 1996*, 1–17.
16. Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues* (S. 122–130). New York: Guilford Press.
17. Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*, 153–193.
18. Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement*, *15*, 139–164.
19. Cropley, A. J. (2006). In praise of convergent thinking. *Creativity Research Journal*, *18*, 391–404.
20. Daseking, M., & Putz, D. (2015). TBS-TK Rezension: „Test d2-R. Aufmerksamkeits- und Konzentrationstest“. *Psychologische Rundschau*, *66*, 265–267.
21. Dörner, D. (1986). Diagnostik der operativen Intelligenz. *Diagnostica*, *32*, 290–308.
22. Ford, C. M. (1996). Theory of individual creative action in multiple social domains. *Academy of Management Review*, *21*, 1112–1142.
23. Frey, A., & Moshagen, M. (2015). TBS-TK Rezension: „Intelligenz-Struktur-Batterie (INS-BAT)“. *Psychologische Rundschau*, *66*, 86–88.
24. 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.

25. Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. New York: Bantam Books.
26. Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24, 13–23.
27. Gruber, N., & Tausch, A. (2015). TBS-TK Rezension: „CFT 20-R mit WS/ZF-R. Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest – Revision (WS/ZF-R)“. *Report Psychologie*, 10, 303–403.
28. Guilford, J. P. (1971). Some misconceptions regarding measurement of creative talents. *The Journal of Creative Behavior*, 5, 77–87.
29. Guilford, J. P., & Hoepfner, R. (1976). *Analyse der Intelligenz*. Weinheim: Beltz.
30. 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.
31. Hell, B. (2003). *Kognitive Leistungsfähigkeit in der Berufseignungsdiagnostik: Forschungsüberblick und Entwicklung eines neuen Instruments zur Erfassung Schlussfolgernden und Kreativen Denkens*. Berlin: dissertation.de.
32. Hell, B., Schuler, H., Boramir, I., et al. (2006). Verwendung und Einschätzung von Verfahren der internen Personalauswahl und Personalentwicklung im 10-Jahres-Vergleich. *Zeitschrift für Personalforschung*, 20, 58–78.
33. Hocevar, D. (1980). Intelligence, divergent thinking, and creativity. *Intelligence*, 4, 25–40.
34. Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253–270.
35. Hornke, L. F., Etzel, S., & Rettig, K. (2006). *Adaptiver Matrizen Test*. Mödling: Schuhfried.
36. Hossiep, R. (1995). *Berufseignungsdiagnostische Entscheidungen. Zur Bewährung eignungsdiagnostischer Ansätze*. Göttingen: Hogrefe.
37. Hossiep, R., & Schulte, M. (2008). *BOWIT – Bochumer Wissenstest*. Göttingen: Hogrefe.
38. Hossiep, R., Turck, D., & Hasella, M. (1999). *BOMAT – advanced. Bochumer Matrizenstest*. Göttingen: Hogrefe.
39. Howard, T. J., Culley, S. J., & Dekoninck, E. (2008). Describing the creative design process by the integration of engineering design and cognitive psychology literature. *Design Studies*, 29, 160–180.
40. 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.
41. Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195–226.
42. Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35, 21–35.
43. Jäger, A. O., Süß, H. M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test (BIS, Form 4)*. Göttingen: Hogrefe.
44. Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, 95, 54–78.
45. Judge, T. A., Higgins, C. A., Thoresen, C. J., et al. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52, 621–652.
46. Kaufman, J. C., Kaufman, S. C., & Lichtenberger, E. (2011). Finding creative potential on intelligence tests via divergent production. *Canadian Journal of School Psychology*, 26(2), 83–106.

47. Kersting, M. (in Druck). Zur Information über und Dokumentation von Instrumenten zur Erfassung menschl. Erlebens u. Verhaltens – Die DIN SCREEN Checkliste 1, Version 3. In: Diagnostik- und Testkuratorium (Hrsgb.). *Personalauswahl kompetent gestalten: Grundlagen u. Praxis der Eignungsdiagnostik nach DIN 33430*. Heidelberg: Springer.
48. Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33, 420–433.
49. Kersting, M. (2014). *Testmanual zum SMART – berufsbezogener Test zur kognitiven Kompetenz. Testmanual/Verfahrenshinweise*. Mödling: Schuhfried.
50. Kersting, M., Althoff, K., & Jäger, A. O. (2008). *Testmanual/Verfahrenshinweise zum WIT-2 (Wilde-Intelligenztest)*. Göttingen: Hogrefe.
51. Kersting, M., & Beauducel, A. (2004). BIS-r-DGP (sowie die Kurzformen A-1 und K-1). Berliner Intelligenzstruktur-Test der Deutschen Gesellschaft für Personalwesen e. V. In W. Sarges & H. Wottawa (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren* (S. 149–157). Lengerich: Pabst Science Publishers.
52. Kim, K. H. (2008). Meta-analyses of the relationship of creative achievement to both IQ and divergent thinking test scores. *Journal of Creative Behavior*, 42, 106–130.
53. Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland: Vertiefende und weiterführende Metaanalysen. *Psychologische Rundschau*, 60, 82–98.
54. Krampen, G. (1993). Diagnostik der Kreativität. In G. Trost, K. Ingenkamp, & R. S. Jäger (Hrsg.), *Tests und Trends 10. Jahrbuch der Pädagogischen Diagnostik* (S. 11–39). Weinheim: Beltz.
55. Krause, D. E. (2013). *Kreativität, Innovation und Entrepreneurship*. Wiesbaden: Springer.
56. Krause, D. E., & Gebert, D. (2003). A comparison of Assessment-Center practices in organizations in German-speaking regions and the United States. *International Journal of Selection and Assessment*, 11, 297–312.
57. Krause, D. E., Kersting, M., Heggestad, E. D., et al. (2006). Incremental validity of Assessment-Center ratings over cognitive ability tests. A study at the executive management level. *International Journal of Selection and Assessment*, 14, 360–371.
58. Krumm, S., Lievens, F., Hüffmeier, J., et al. (2015). How „situational“ is judgment in situational judgment tests? *Journal of Applied Psychology*, 100, 399–416.
59. Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science*, 19, 339–345.
60. Lang, J. W. B., Kersting, M., Hülsheger, U. R., et al. (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.
61. Lehrl, S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B*. Balingen: Spitta.
62. Lienert, G. A. (1958). *MTVT – Mechanisch-Technischer Verständnistest*. Göttingen: Hogrefe.
63. Lienert, G. A. (1967). *Drahtbiegeprobe (DBP)*. Göttingen: Hogrefe.
64. Liepmann, D., Beauducel, A., Brocke, B., et al. (2007). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)* (2., erweiterte und überarbeitete Aufl.). Göttingen: Hogrefe.
65. Lubart, T. L. (2001). Models of the creative process: Past, present and future. *Creativity Research Journal*, 13, 295–308.
66. Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspect Psychol Sci*, 1, 316–343.
67. Luczak, H. (1998). *Arbeitswissenschaft*. Berlin: Springer.
68. McClelland, D. C. (1973). Testing for competence rather than for intelligence. *American Psychologist*, 28, 1–14.
69. McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues* (S. 136–181). New York: Guilford Press.

70. 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.
71. Molz, G., Schulze, R., Schröder, U., et al. (2010). TBS-TK Rezension: „Wechsler Intelligenztest für Erwachsene WIE. Deutschsprachige Bearbeitung und Adaptation des WAIS-III von David Wechsler“. *Psychologische Rundschau*, 61, 229–230.
72. Mooney, R. L. (1963). A conceptual model for integrating four approaches to the identification of creative talent. In C. W. Taylor & F. Barron (Hrsg.), *Scientific creativity: Its recognition and development*. New York: Wiley.
73. Moosbrugger, H., & Goldhammer, F. (2007). *FAKT-II. Frankfurter Adaptiver Konzentrationsleistungs-Test* (2. Aufl.). Bern: Huber.
74. Muchinsky, P. M. (1993). Validation of intelligence and mechanical aptitude tests in selecting employees for manufacturing jobs. *The Journal of Business and Psychology*, 7, 373–382.
75. Mumford, M. D., & Gustafson, S. B. (1988). Creativity syndrome: Integration, application, and innovation. *Psychological Bulletin*, 103, 27–43.
76. Muñoz-Doyague, M. F., González-Álvarez, N., & Nieto, M. (2008). An examination of individual factors and employees' creativity: The case of Spain. *Creativity Research Journal*, 20, 21–33.
77. Oldham, G. R., & Cummings, A. (1996). Employee creativity: Personal and contextual factors at work. *Academy of Management Journal*, 39, 607–634.
78. Palmer, C. (2016). *Berufsbezogene Kreativitätsdiagnostik. Beschreibung und Messung der personalen Voraussetzungen von Innovationen*. Springer Fachmedien: Wiesbaden.
79. Palmer, C., Cesinger, B., Gelléri, P., et al. (2015). Psychometrical testing of entrepreneurial creativity. *International Journal of Entrepreneurial Venturing*, 7, 194–210.
80. Palmer, C., & Schuler, H. (in Vorbereitung) Diagnose berufsbezogener Kreativität – Technik und Entwicklung (DBK-TE). Manual. Göttingen: Hogrefe.
81. Petermann, F. (Hrsg.). (2015). Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV, dt. Version, 2., korr. Aufl.). Frankfurt: Pearson Assessment.
82. Raven, J. C., Raven, J., & Court, J. H. (1998). *Advanced progressive matrices (APM)*. Frankfurt: Pearson Assessment.
83. Rhodes, M. (1961). An analysis of creativity. *Phi Delta Kappa*, 42, 305–310.
84. Runco, M. A. (2006). Introduction to the special issue: Divergent thinking. *Creativity Research Journal*, 18, 249–250.
85. Ryan, A. M., McFarland, L., Baron, H., et al. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, 52, 359–391.
86. Salgado, J. F., Anderson, N., Moscoso, S., et al. (2003a). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, 56, 573–605.
87. Salgado, J. F., Anderson, N., Moscoso, S., et al. (2003b). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology*, 88, 1068–1081.
88. Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, 9, 185–211.
89. Schmidt, F. L., & Hunter, J. (2004). General mental ability in the World of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86, 162–173.
90. 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.

91. Schmidt, F. L., Law, K., Hunter, J. E., et al. (1993). Refinements in validity generalization methods: Implication for the situational specificity hypothesis. *Journal of Applied Psychology*, 78, 3–12.
92. Schmidt-Atzert, L., Rauch, W. (2008). TBS-TK Rezension: „Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). 2., erweiterte und überarbeitete Auflage von Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007)“. *Report Psychologie*, 33, 303–304.
93. Schuler, H., Frier, D., & Kaufmann, M. (1993). *Personalauswahl im europäischen Vergleich*. Göttingen: Verlag für Angewandte Psychologie.
94. Schuler, H., Funke, U., Moser, K., et al. (1995). *Personalauswahl in Forschung und Entwicklung. Eignung und Leistung von Wissenschaftlern und Ingenieuren*. Göttingen: Hogrefe.
95. Schuler, H., Gelléri, P., Winzen, J., et al. (2013). *Diagnose berufsbezogener Kreativität – Planung und Gestaltung (DBK-PG). Manual*. Göttingen: Hogrefe.
96. Schuler, H., & Görlich, Y. (2007). *Kreativität. Ursachen, Messung, Förderung und Umsetzung in Innovation*. Göttingen: Hogrefe.
97. Schuler, H., & Hell, B. (2005). *ASK – Analyse des Schlussfolgernden und Kreativen Denkens*. Bern: Hans Huber.
98. Schuler, H., Hell, B., Trapmann, S., et al. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. *Zeitschrift für Personalpsychologie*, 6, 60–70.
99. Schuler, H., & Klingner, Y. (2005). *Arbeitsprobe zur berufsbezogenen Intelligenz (AZUBI-BK)*. Göttingen: Hogrefe.
100. Steinmayr, R., Schütz, A., Hertel, J., et al. (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: Verlag Hans Huber.
101. Sternberg, R. J. (1999). The theory of successful intelligence. *Review of General Psychology*, 3, 292–316.
102. Sternberg, R. J., & O’Hara, L. A. (1999). Creativity and intelligence. In R. J. Sternberg (Hrsg.), *Handbook of creativity* (S. 251–272). New York: Cambridge University Press.
103. Sternberg, R. J., & Powell, J. S. (1982). Theories of intelligence. In R. J. Sternberg (Hrsg.), *Handbook of human intelligence* (S. 975–1005). New York: Cambridge University Press.
104. Sternberg, R. J., & Wagner, R. K. (1986). *Practical intelligence: Nature and origins of competence in the everyday world*. New York: Cambridge University Press.
105. Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35, 401–426.
106. Stüß, 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.
107. Stüß, H.-M., & Schweickert, S. (2002). Was ist ein gutes Maß für die Allgemeine Intelligenz? Eine Studie zur Konstruktvalidität von Intelligenztests. Universität Mannheim, Lehrstuhl Psychologie II, Mannheim.
108. 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.
109. Thurstone, L. L. (1957). *Primary mental abilities*. Chicago: University of Chicago Press.
110. Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago Press.

111. Torrance, E. P. (1966). Torrance tests of creative thinking. Directions manual, and scoring guide. Princetown: Personnel Press.
112. Vernon, P. E. (1961). *The structure of human abilities*. London: Methuen.
113. Wai, J. (2013). Investigating America's elite: Cognitive ability, education, and sex differences. *Intelligence*, 41, 203–211.
114. Ward, T. B. (2004). Cognition, creativity, and entrepreneurship. *The Journal of Business Venturing*, 19, 173–188.
115. Weis, S., & Nuerk, H.-C. (2010). TBS-TK Rezension: „FAKT-II. Frankfurter Adaptiver Konzentrationsleistungs-Test“. *Report Psychologie*, 36, 219–221.
116. Weiß, R. H. (2006). CFT 20-R mit WS/ZF-R. Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest – Revision (WS/ZF-R). Göttingen: Hogrefe.
117. Welsh, G. S. (1975). *Creativity and intelligence: A personality approach*. Chapel Hill: Institute for Research in Social Science.
118. Zhou, J., & Shalley, C. E. (2003). Research on employee creativity: A critical review and directions for future research. *Research in Personnel and Human Resources Management*, 22, 165–217.
119. Zhou, J., & Shalley, C. E. (2013). Zum Verständnis von Kreativität am Arbeitsplatz: Ein Überblick zu verschiedenen Ansätzen der Kreativitätsforschung. In D. E. Krause (Hrsg.), *Kreativität, Innovation und Entrepreneurship* (S. 1–41). Wiesbaden: Springer.

---

# Testverfahren II: Persönlichkeit und personenbezogene Attribute

# 7

Rüdiger Hossiep und Sabine Weiß

---

## Zusammenfassung

Persönlichkeitsmerkmale spielen bei Personalentscheidungen stets eine erhebliche Rolle – fraglich ist nur, ob ein mehr oder weniger intuitives Vorgehen zu ihrer Erfassung dominiert oder ob standardisierte Testverfahren zur Unterstützung der Entscheidungsfindung eingesetzt werden, die den Prozess objektivieren und systematisieren können. In diesem Kapitel werden verschiedene Arten persönlichkeitsorientierter Instrumente vorgestellt. Die Einsatzhäufigkeit persönlichkeitsorientierter Testverfahren belegt, dass das Potenzial persönlichkeitsorientierter Verfahren zur Optimierung der Qualität von Personalauswahlprozessen in Deutschland bisher noch unterschätzt wird. Bei Berücksichtigung der rechtlichen Rahmenbedingungen und sinnvoller Einbettung in den eignungsdiagnostischen Gesamtprozess können persönlichkeitsorientierte Testverfahren im Zusammenspiel mit etablierten Methoden der Personalauswahl wie dem klassischen Interview einen wertvollen inkrementellen Beitrag liefern.

---

R. Hossiep (✉) · S. Weiß  
Fakultät für Psychologie, Ruhr-Universität Bochum,  
Universitätsstr. 150, 44780  
Bochum, Deutschland  
E-Mail: ruediger.hossiep@rub.de

S. Weiß  
E-Mail: s.weiss@rub.de



## Inhaltsverzeichnis

7.1	Zur Relevanz persönlichkeitsorientierter Verfahren im eignungsdiagnostischen Kontext .....	160
7.2	Definitorische Überlegungen zu persönlichkeitsorientierten Verfahren im berufsbezogenen Kontext .....	161
7.3	Kriteriumsvalidität persönlichkeitsorientierter Testverfahren .....	162
7.4	Arten persönlichkeitsorientierter Verfahren .....	164
7.4.1	Persönlichkeits-Struktur-Tests .....	164
7.4.2	Typen-Tests .....	165
7.4.3	Objektive Persönlichkeitstests .....	167
7.4.4	Freie Verfahren .....	168
7.5	Einsatzhäufigkeit persönlichkeitsorientierter Verfahren in Deutschland .....	168
7.6	Zur Verfälschbarkeit persönlichkeitsorientierter Testverfahren .....	170
7.7	Rechtliche Rahmenbedingungen für den Einsatz persönlichkeitsorientierter Testverfahren .....	171
7.8	Einbettung persönlichkeitsorientierter Verfahren in den Personalauswahlprozess und darauf aufbauende Personalentwicklungsprozesse .....	175
	Literatur .....	178

---

## 7.1 Zur Relevanz persönlichkeitsorientierter Verfahren im eignungsdiagnostischen Kontext

Testverfahren, die im Rahmen eignungsdiagnostischer Maßnahmen Anwendung finden, können unterteilt werden in kognitive Fähigkeitstests (vgl. Kapitel von Kersting und Palmer in diesem Band) zur Messung kognitiver Fähigkeiten bzw. der intellektuellen Leistungsfähigkeit und persönlichkeitsorientierte Verfahren zur Erfassung von Persönlichkeitsmerkmalen und personenbezogenen Attributen [58]. Die intellektuelle Leistungsfähigkeit gilt aus wissenschaftlicher Sicht als der wesentliche Prädiktor beruflicher Leistung [52]. Inzwischen wird jedoch auch Persönlichkeitseigenschaften hinsichtlich beruflicher Eignung und Leistung vonseiten der Wissenschaft ein höherer Stellenwert zugesprochen [8], als dies noch vor einigen Jahren der Fall war.

Die Hauptursachen für das Scheitern sogenannter „High Potentials“ werden von Personalverantwortlichen häufig nicht auf fachlicher Ebene, sondern vielmehr in der mangelnden Ausprägung erfolgsrelevanter Soft Skills gesehen [33]. Fachliche Qualifikationen sind somit lediglich als Grundvoraussetzung für den Berufserfolg anzusehen. Um eine Position erfolgreich auszufüllen, gelten insbesondere persönlichkeitsbezogene Kompetenzen als Schlüsselfaktoren, um dem Trend „hired for ability, fired for personality“ entgegen zu wirken [21]. Heutzutage kommt kaum eine Stellenausschreibung ohne die Formulierung von Anforderungen aus, die die Persönlichkeit von Bewerbern betreffen. Dies unterstreicht noch einmal aus der Praxisperspektive die Bedeutung persönlichkeitsorientierter Merkmale im eignungsdiagnostischen Prozess.

Persönlichkeitsmerkmale sind also in jedem Personalprozess relevant [27]. Fraglich ist, ob zur Unterstützung der Entscheidungsfindung auch objektive Testinstrumente involviert werden oder ob hauptsächlich subjektive Eindrücke im Vordergrund stehen.

---

## 7.2 Definitivische Überlegungen zu persönlichkeitsorientierten Verfahren im berufsbezogenen Kontext

In der Persönlichkeitsforschung kommt dem Eigenschaftsparadigma eine zentrale Rolle zu. Ziel ist es dabei, die individuellen Besonderheiten von Menschen durch ihre Eigenschaften zu beschreiben, wobei Persönlichkeit als geordnete Gesamtheit dieser Eigenschaften verstanden wird [3]. Als Persönlichkeitseigenschaften werden relativ breite und zeitlich stabile Dispositionen zu bestimmten Verhaltensweisen bezeichnet, die konsistent in verschiedenen Situationen auftreten [63]. Auf Basis des lexikalischen Ansatzes, indem Listen von Adjektiven, welche bedeutsame Unterschiede zwischen Personen beschreiben, faktorenanalytisch zusammengefasst wurden, entstanden verschiedene Persönlichkeitsmodelle mit unterschiedlicher Faktorenanzahl. Das Fünf-Faktoren-Modell der Persönlichkeit („Big Five“) ist das weltweit verbreitetste allgemeine Persönlichkeitsmodell und postuliert fünf Grunddimensionen der Persönlichkeit: Neurotizismus (Gegenpol von emotionaler Stabilität), Extraversion, Offenheit für neue Erfahrungen, Verträglichkeit und Gewissenhaftigkeit [7]. Die meisten persönlichkeitsorientierten Verfahren, die derzeit im beruflichen Kontext angewendet werden, beruhen auf dem Eigenschaftsparadigma und erfassen unterschiedliche Anzahlen von Persönlichkeitseigenschaften, wengleich einige auch andere Modelle wie z. B. psychoanalytische Konzepte oder spezifische Theorien zugrunde legen (vgl. Abschn. 7.4.2).

Bis in die 1990er Jahre wurde der Schwerpunkt persönlichkeitsorientierter Verfahren im Bereich Forschung und im klinisch-psychologischen Setting gelegt. Instrumente wurden daher zunächst ausschließlich für diesen Zweck entwickelt. Die zunehmende Relevanz von Persönlichkeitsdiagnostik im beruflichen Kontext hat dazu geführt, dass vielfach auch Testverfahren in Organisationen zum Einsatz kommen, die nicht originär für die Anwendung im berufsbezogenen Setting konstruiert wurden und Fragen enthalten, die im beruflichen Kontext zu Irritationen führen.

Ein berufsbezogener Persönlichkeitstest ermöglicht auf Basis einer Selbsteinschätzung eine mehrdimensionale Persönlichkeitsbeschreibung in Bezug auf berufsbezogene Merkmale [24]. Um eine hinreichende Akzeptanz eines solchen Verfahrens im beruflichen Kontext sicherzustellen sollten die einzelnen Fragen Erlebens- und Verhaltensbereiche aus dem Berufsleben thematisieren, einen klaren Anforderungsbezug aufweisen und keinesfalls die Intimsphäre oder klinische Krankheitssymptome betreffen. Im Gegensatz zu Leistungstests erfolgt die Auswertung dieser Tests nicht im Sinne „richtiger“ oder „falscher“ Antworten, sondern bezogen auf geringer oder stärker ausgeprägte Persönlichkeitszüge. Die aus einem solchen Test resultierenden Profile ermöglichen eine qualitative

und quantitative Bewertung, z. B. in Hinblick auf die Passung der Person zu bestimmten überfachlichen Anforderungen. Die Mehrdimensionalität von Persönlichkeitstests sorgt dabei für eine entsprechende Breite und auch Tiefe der Ergebnisse. Sie liefern die Basis für gezielte Gespräche mit den jeweiligen Kandidaten, um deren Selbsteinschätzung zu hinterfragen und mit den spezifischen Anforderungen einer vakanten Position abgleichen zu können (vgl. Abschn. 7.8).

---

### 7.3 Kriteriumsvalidität persönlichkeitsorientierter Testverfahren

Der Einsatz persönlichkeitsorientierter Testverfahren im beruflichen Kontext ist grundsätzlich nur dann zu rechtfertigen, wenn bedeutsame Zusammenhänge der Ergebnisse dieser Verfahren mit beruflich relevanten Kriterien vorliegen [1]. Auch wenn die Anwendung von Persönlichkeitstests im eignungsdiagnostischen Bereich nicht ohne Kritik geblieben ist (z. B. [17, 42]), belegen eine Vielzahl von Studien die Sinnhaftigkeit des Einsatzes dieser Instrumente im beruflichen Kontext (für einen Überblick vgl. [29, 30, 46]).

Schmidt und Hunter [52] haben die Validität verschiedener Auswahlmethoden hinsichtlich allgemeiner beruflicher Leistung verglichen und zeigen, dass neben Intelligenztests ( $r = ,51$ ), strukturierten Vorstellungsgesprächen ( $r = ,51$ ) und Assessment-Centern ( $r = ,37$ ) auch persönlichkeitsorientierte Verfahren, insbesondere das Merkmal Gewissenhaftigkeit ( $r = ,31$ ), einen Zusammenhang mit beruflichem Erfolg aufweisen.

Ones et al. [44] fassen die metaanalytisch gewonnenen Validitätsbelege für die Big-Five-Merkmale als Gesamtkomplex zusammen und ermitteln für den Zusammenhang mit allgemeiner beruflicher Leistung einen korrigierten multiplen Validitätskoeffizienten in Höhe von ,27; mit Arbeitszufriedenheit von ,36 und mit Zufriedenheit mit der beruflichen Laufbahn in Höhe von ,39. Um die Einzelbeiträge der Big-Five-Merkmale zu beleuchten, werden exemplarisch im Folgenden die Ergebnisse für die Kriterien allgemeiner beruflicher Leistung auf Basis der Metaanalyse von Barrick et al. [4] sowie die Resultate für Arbeitszufriedenheit auf Basis der Metaanalyse von Judge et al. [31] und die Befunde für beruflichen Erfolg auf Basis der Metaanalyse von Ng et al. [43] berichtet (vgl. Tab. 7.1).

Ebenso wie bei den Big-Five-Merkmalen als Gesamtkomplex [44] ist auch bei der Betrachtung der einzelnen Faktoren ersichtlich, dass subjektive Kriterien durch Persönlichkeitsmerkmale besser vorhergesagt werden können als objektive Kriterien. Darüber hinaus wird deutlich, dass je nach Kriterium die prognostische Qualität der einzelnen Persönlichkeitsfaktoren schwankt.

Shaffer und Postlethwaite [60] konnten metaanalytisch belegen, dass Testverfahren mit berufsbezogen formulierten Items allgemein formulierten Persönlichkeitstests in Hinblick auf berufliche Validitätskriterien deutlich überlegen sind.

Für das explizit berufsbezogen konstruierte Verfahren BIP wurden anhand von mehr als 4800 Teilnehmern multiple Validitätskoeffizienten in Bezug auf die

**Tab. 7.1** Validität der Big-Five-Persönlichkeitsfaktoren für die Kriterien berufliche Leistung, Arbeitszufriedenheit und Berufserfolg [64, S. 43]

	Korrelation (r)					N Min	N Max
	N	E	O	V	G		
<b>Berufliche Leistung<sup>a</sup></b>							
Leistung (insgesamt)	-,13	,15	,07	,13	,27	23.225	48.100
Spezifische Kriterien							
Vorgesetztenbeurteilungen	-,13	,13	,07	,13	,31	18.535	33.312
Objektive Leistungskriterien	-,10	,13	,03	,17	,23	4401	7101
Trainingsleistung	-,09	,28	,33	,14	,27	3177	4100
Teamleistung	-,22	,16	,16	,34	,27	1820	3719
<b>Berufserfolg</b>							
Objektive Kriterien							
Entgelt <sup>b</sup>	-,12	,10	,04	-,10	,07	6286	6800
Berufl. Aufstieg <sup>b</sup>	-,11	,18	,01	-,05	,06	4428	4942
Subjektive Kriterien							
Zufriedenheit mit berufl. Laufbahn <sup>b</sup>	-,36	,27	,12	,11	,14	4634	10.556
Arbeitszufriedenheit <sup>c</sup>	-,29	,25	,02	,17	,26	11.856	24.527

Anm.: N = Neurotizismus (Gegenpol von emotionaler Stabilität), E = Extraversion, O = Offenheit für Erfahrungen, V = Verträglichkeit, G = Gewissenhaftigkeit

<sup>a</sup>Ergebnisse von Barrick et al. [4]

<sup>b</sup>Ergebnisse von Ng et al. [43]

<sup>c</sup>Ergebnisse von Judge et al. [31]

Berufserfolgskriterien Entgelt in Höhe von  $r = ,39$ , hierarchische Position in Höhe von  $r = ,41$ , subjektive Berufserfolgseinschätzung in Höhe von  $r = ,49$  und Arbeitszufriedenheit in Höhe von  $r = ,41$  ermittelt [25]. Für das BIP-6F werden auf Basis von Stichproben mit mehr als 4500 Teilnehmern multiple Korrelationskoeffizienten der BIP-6F-Faktoren mit Entgelt in Höhe von  $r = ,35$ , mit der hierarchischen Position in Höhe von  $r = ,35$ , mit der subjektiven Erfolgseinschätzung in Höhe von  $r = ,46$  und mit Arbeitszufriedenheit in Höhe von  $r = ,45$  empirisch aufgezeigt [23].

Neben der isolierten Validität bestimmter Persönlichkeitseigenschaften ist mit Blick auf den Einsatz von Persönlichkeitstests im berufsbezogenen Kontext vor allem auch ihre inkrementelle Validität entscheidend, d. h. das Ausmaß, in dem durch die Integration zusätzlicher Parameter im Rahmen einer Eignungsdiagnose die Vorhersage relevanter Validitätskriterien wie etwa berufliche Leistung oder Arbeitszufriedenheit verbessert werden kann [30].

Persönlichkeitseigenschaften korrelieren dabei vergleichsweise gering mit kognitiver Leistungsfähigkeit, welche als Hauptprädiktor für beruflichen Erfolg gilt [52]. Auch in Kombination mit anderen Variablen, wie beispielsweise Interviewdaten, können Persönlichkeitstests einen substantziellen Beitrag zur Erhöhung der Validität liefern [12].

---

## 7.4 Arten persönlichkeitsorientierter Verfahren

Hossiep und Mühlhaus [24] legen eine Unterscheidung in Persönlichkeits-Struktur-Tests, Typentests, objektive Persönlichkeitstests und freie Verfahren vor.

### 7.4.1 Persönlichkeits-Struktur-Tests

Das Hauptanliegen eines Persönlichkeits-Struktur-Tests ist die differenzierte Analyse des normierten Selbstbildes in Bezug auf mehrere Persönlichkeitsdimensionen, wobei je nach Test die Anzahl der Dimensionen zwischen 2 (z. B. Eysenck-Persönlichkeits-Inventar, E-P-I, [14]) und 16 (16-Persönlichkeits-Faktoren-Test, 16 PF, [54]) variiert [24].

Exemplarisch werden als prominente Vertreter dieser Verfahrensgruppe das NEO-Persönlichkeitsinventar (NEO-PI-R, [45]) und das NEO-Fünf-Faktoren-Inventar (NEO-FFI, [7]) sowie die explizit berufsbezogen formulierten Fragebogenverfahren „Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung“ (BIP, [25]) und „Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung – 6 Faktoren“ (BIP-6F, [23]) kurz vorgestellt.

Das NEO-PI-R und das NEO-FFI haben einen forschungsbezogenen Entwicklungshintergrund und ermöglichen die Erfassung unabhängiger, grundlegender Persönlichkeitsbereiche. Während das NEO-PI-R über 240 Items auch die den Grundfaktoren zugeordneten Teilfacetten konkretisiert (vgl. Tab. 7.2), fokussiert die Kurzfassung NEO-FFI mit lediglich 60 Testfragen auf die fünf Grunddimensionen Neurotizismus (bzw. dessen Gegenpol emotionale Stabilität), Extraversion, Offenheit für Erfahrungen, Verträglichkeit und Gewissenhaftigkeit. Für die Items wurde eine fünfstufige Likert-Skala von „starke Ablehnung“ bis „starke Zustimmung“ gewählt.

Die allgemeine, nicht kontextbezogene Abfassung des Big-Five-Persönlichkeitsmodells, der sowohl das NEO-FFI als auch das NEO-PI-R folgen, ist als wesentlicher Vorteil für einen breiten Einsatz in der wissenschaftlichen Forschung im Rahmen unterschiedlichster Fragestellungen anzusehen. Dies erweist sich jedoch gerade im berufsbezogenen Kontext als nachteilig. Einige Fragen, z. B. „Ich mag meine Zeit nicht mit Tagträumereien verschwenden“ (ein negativ gepoltes Item für den Faktor „Offenheit für Erfahrung“), sind vom Berufskontext relativ weit entfernt und für bestimmte Berufsgruppen sicher auch hinsichtlich der Polung fragwürdig, sodass der hohe Abstraktionsgrad bisweilen zu einer fehlenden Adäquatheit und einer gewissen Sperrigkeit des Modells in Bezug auf personalpsychologische Fragestellungen in Organisationen führt [21].

**Tab. 7.2** Zuordnung der Facetten zu Faktoren des NEO-PI-R [45]

Neurotizismus	Extraversion	Offenheit für Erfahrungen	Verträglichkeit	Gewissenhaftigkeit
Ängstlichkeit	Herzlichkeit	Offenheit für Fantasie	Vertrauen	Kompetenz
Reizbarkeit	Geselligkeit	Offenheit für Ästhetik	Freimütigkeit	Ordnungsliebe
Depression	Durchsetzungsfähigkeit	Offenheit für Gefühle	Altruismus	Pflichtbewusstsein
Soziale Befangtheit	Aktivität	Offenheit für Handlungen	Entgegenkommen	Leistungsstreben
Impulsivität	Erlebnishunger	Offenheit für Ideen	Bescheidenheit	Selbstdisziplin
Verletzlichkeit	Frohsinn	Offenheit des Werte- und Normensystems	Gutherzigkeit	Besonnenheit

Die Bochumer Inventare BIP und BIP-6F wurden explizit für den Einsatz im beruflichen Kontext konzipiert. Das BIP ermöglicht anhand von 210 Items die standardisierte Erfassung des Selbstbildes einer Person in Bezug auf 14 beruflich relevante Dimensionen, die den Bereichen berufliche Orientierung, Arbeitsverhalten, soziale Kompetenzen und psychische Konstitution zugeordnet sind (vgl. Tab. 7.3).

Das BIP-6F wurde faktorenanalytisch auf Basis des „Mutter“-Verfahrens BIP mit dem Ziel konstruiert, ein ökonomisch besonders effizientes Verfahren zur berufsbezogenen Diagnostik von breiten Persönlichkeitsfaktoren bereitzustellen. Es erfasst mit insgesamt lediglich 48 Items die sechs übergeordneten berufsbezogenen Persönlichkeitsfaktoren Engagement, Disziplin, Sozialkompetenz, Kooperation, Dominanz und Stabilität (vgl. Abb. 7.1). Die Items von BIP und BIP-6F, wie z. B. „Ich bin ausgesprochen ehrgeizig“, werden auf einer sechsstufigen Likert-Skala von „trifft voll zu“ bis „trifft überhaupt nicht zu“ beantwortet.

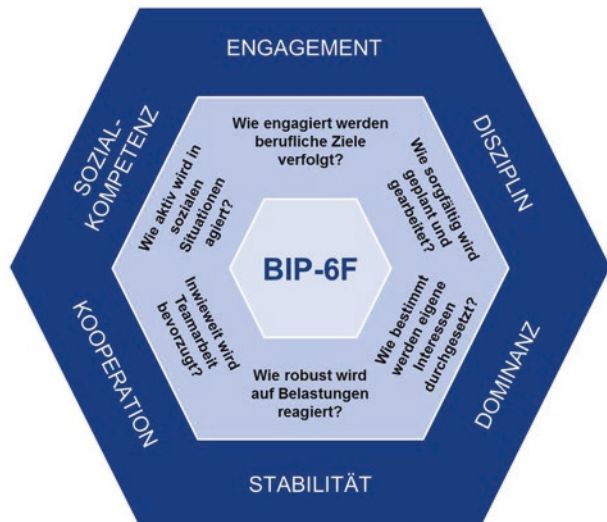
## 7.4.2 Typen-Tests

Typen-Tests beinhalten meist wenige Skalen mit zwei gegensätzlichen Polen, beispielsweise Extraversion – Introversion. Im Gegensatz zu Persönlichkeits-Struktur-Tests, die meist für beide Dimensionen die jeweilige Ausprägung ausweisen, wird die Informationsmenge reduziert, indem angegeben wird, welchem Pol der Kandidat auf Basis seiner Testantworten am ähnlichsten ist [24].

**Tab. 7.3** Die Skalen des BIP mit Leitfrage und Zuordnung zu Bereichen [25]

Bereich	Skala	Leitfrage
Berufliche Orientierung	Leistungsmotivation	Inwieweit stelle ich hohe Leistungsanforderungen an mich?
	Gestaltungsmotivation	Wie wirke ich auf Prozesse ein?
	Führungsmotivation	Wie wirke ich auf andere Personen ein?
Arbeitsverhalten	Gewissenhaftigkeit	Wie wichtig sind für mich Detailorientierung und Perfektionismus?
	Flexibilität	In welchem Ausmaß bin ich willens, mich immer wieder umzustellen?
	Handlungsorientierung	Wie zielgerichtet setze ich getroffene Entscheidungen in Handlungen um?
Soziale Kompetenzen	Sensitivität	Wie sicher erspüre ich Gefühle anderer?
	Kontaktfähigkeit	In welchem Umfang verhalte ich mich sozial offensiv?
	Soziabilität	Wie wichtig ist mir ein harmonisches Miteinander?
	Teamorientierung	Wie stark bevorzuge ich Teamarbeit?
	Durchsetzungsstärke	Mit welcher Vehemenz verfolge ich anderen gegenüber meine Ziele?
Psychische Konstitution	Emotionale Stabilität	In welchem Ausmaß bin ich emotional robust?
	Belastbarkeit	Wie viel will und kann ich mir an Belastung zumuten?
	Selbstbewusstsein	Wie überzeugt bin ich von mir als Person?

**Abb. 7.1** Hexagon-Modell des BIP-6F [23]





Exemplarisch sei hier das persolog Persönlichkeits-Profil genannt [16], besser bekannt unter dem Namen DISG, welches auf dem Modell menschlicher Reaktionen von Marston [39] fußt und aus zwei Grunddimensionen die Ausprägung der vier Verhaltenstendenzen Dominanz (D), Initiative (I), Stetigkeit (S) und Gewissenhaftigkeit (G) bestimmt (vgl. Tab. 7.4). Die Teilnehmer sollen im Test beispielsweise beurteilen, welcher Begriff widerspiegelt, wie sie sich im gewählten Umfeld „am wenigsten“ empfinden: gesellig, selbstsicher, geduldig oder ruhig.

Das DISG-Modell erfreut sich ebenso wie viele andere Typen-Tests in Deutschland trotz anhaltender deutlicher wissenschaftlicher Kritik [32, 35] großer Beliebtheit (vgl. Abschn. 7.5), welche sie möglicherweise auch der (vermeintlichen) Reduktion von Komplexität zu verdanken haben. Die Hauptkritikpunkte betreffen fehlende Nachweise zur Kriteriumsvalidität der Testergebnisse, die Orientierung an überholten Persönlichkeitsmodellen, die Verwendung eingeschränkter Antwortformate, welche zu einer überzeichneten Prägnanz in den Ergebnisprofilen führen sowie spekulative Aussagen in den Ergebnisberichten, die nicht durch belegbare Aussagen in den Testbeantwortungen fundiert werden können [24].

### 7.4.3 Objektive Persönlichkeitstests

Objektive Persönlichkeitstests sind dadurch gekennzeichnet, dass sie keine direkte Selbsteinschätzung von Kandidaten erfordern, sondern sich auf Verhaltenspräferenzen bei der Bearbeitung von Aufgaben stützen, anhand derer Teilnehmer das tatsächlich interessierende Persönlichkeitsmerkmal möglichst nicht selbst erkennen.

Beispielhaft sei hier die Testbatterie Arbeitshaltungen von Kubinger und Ebenhöf [36] erwähnt. Anhand der Aufgabe, einen Größenvergleich der Flächen verschiedener geometrischer Figuren vorzunehmen, wird auf den kognitiven Stil „Reflexivität vs. Impulsivität“ geschlossen.

**Tab. 7.4** Vier Typen menschlicher Reaktionen im persolog-Persönlichkeitsprofil [24, S. 104]

Wahrnehmung des Umfeldes als ...	Wahrnehmung der eigenen Person als ...	Reaktionen bzw. Verhaltenstendenzen
unangenehm/stressig	bestimmt	Dominanz (D)
	(stärker als das Umfeld)	(aktiv und entschlossen)
	zurückhaltend	Gewissenhaftigkeit (G)
	(schwächer als das Umfeld)	(diszipliniert und besorgt)
angenehm/nicht stressig	bestimmt	Initiative (I)
	(stärker als das Umfeld)	(gesprächig und offen)
	zurückhaltend	Stetigkeit (S)
	(schwächer als das Umfeld)	(verlässlich und kooperativ)

Die Undurchschaubarkeit eines objektiven Testinstruments sorgt für die Attraktivität solcher Verfahren. Gleichwohl steht immer auch die Frage der Zumutbarkeit eines intransparenten Vorgehens im Raum, was unter anderem im Kontext betrieblicher Mitbestimmung als kritisch angesehen werden kann, da das im Betriebsverfassungsgesetz verankerte Recht auf informationelle Selbstbestimmung verletzt wird, wenn dem Testanden nicht vor Augen steht, zu welchem Persönlichkeitsmerkmal er Auskunft gibt [24].

#### **7.4.4 Freie Verfahren**

Bei freien Verfahren ist die Gefahr groß, dass die Bewertung der Ergebnisse stark subjektiv geprägt ist [51]. Ihnen werden projektive Verfahren (auch Persönlichkeits-Entfaltungs-Verfahren genannt) zugeordnet. Bei Formdeutungsverfahren wie etwa dem Rorschach-Test [41] wird mehrdeutiges tintenklecksartiges Stimulusmaterial vorgelegt, das Projektionen des Teilnehmers abrufen, welche Rückschlüsse auf dessen Persönlichkeit erlauben sollen. Auch Handschriftanalyseverfahren gehören grundsätzlich zur Kategorie der freien Verfahren. Nicht zuletzt die fehlende Standardisierung und Auswertungsobjektivität führt nur zu vereinzeltem Einsatz solcher Verfahren im deutschsprachigen Raum.

---

### **7.5 Einsatzhäufigkeit persönlichkeitsorientierter Verfahren in Deutschland**

Von der ehemaligen Hohenheimer Arbeitsgruppe um Prof. Schuler wurden in Deutschland drei umfassende Untersuchungen zum Einsatz von Verfahren zur externen Personalauswahl, d. h. der Besetzung von Vakanzen durch noch nicht im Unternehmen tätige Personen, durchgeführt [56, 57, 59]. In der aktuellsten Studie wurden die umsatzstärksten 550 deutschen Unternehmen befragt und es konnten Daten von 125 Organisationen, die sich an der Untersuchung beteiligten, ausgewertet werden. Für Persönlichkeitstests wurde eine allgemeine Einsatzquote von 20 % ermittelt. Kognitive Fähigkeitstests, unterteilt in Intelligenztests mit einer Quote von 30,4 % und sonstige Leistungstests mit einer Quote von 40,8 % wurden demgegenüber deutlich häufiger in den Unternehmen durchgeführt.

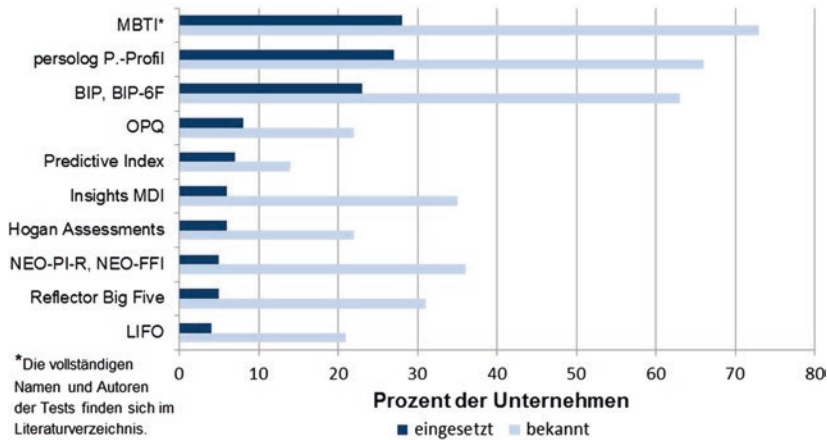
Die Einsatzhäufigkeit von Testverfahren variiert je nach betrachteter Bewerbergruppe. Insgesamt ist bei Persönlichkeitstests eine relativ niedrige Einsatzhäufigkeit in allen Bewerbergruppen zu konstatieren, wobei sich eine U-Funktion in Bezug auf die Höhe der hierarchischen Ebene abzeichnet: Bei der Auswahl von Auszubildenden und Trainees wenden 6 % der Unternehmen Persönlichkeitstests an, im Facharbeiter- und Angestelltenbereich kommen nahezu gar keine persönlichkeitsorientierten Verfahren zum Einsatz. Für verschiedene Führungskräfteebenen ist wieder eine ansteigende Tendenz erkennbar (vgl. erste Spalte Tab. 7.5).

**Tab. 7.5** Einsatzhäufigkeit von Persönlichkeitstests je Bewerbergruppe in Rahmen externer und interner Personalauswahl (Angaben in Prozent derjenigen Unternehmen, welche die jeweilige Berufsgruppe auch beschäftigten – maximal N = 125)

Zielgruppe	Persönlichkeitstests im Rahmen externer Personalauswahl [57] (%)	Persönlichkeitstests im Rahmen interner Personalauswahl [18] (%)
Technische Auszubildende	6,4	
Kaufmännische Auszubildende	6,3	
Trainees	6,3	2,7
Facharbeiter	1,1	0,0
Angestellte ohne Führungsaufgaben	0,0	1,7
Führungskräfte untere Ebene	4,1	5,0
Führungskräfte mittlere Ebene	6,6	5,2
Führungskräfte obere Ebene	7,6	5,7

Bei internen Personalauswahlprozessen, definiert als Auswahl von bereits in der Organisation tätigen Personen im Rahmen von angestrebtem Arbeitsplatzwechsel oder Personalentwicklungsmaßnahmen, zeigt sich ein anderes Bild. Im Vergleich zur externen Personalauswahl sinkt der prozentuale Anteil an Organisationen, die – über alle Zielgruppen insgesamt betrachtet – Testverfahren einsetzen, deutlich auf 3,2 % für Intelligenztests, 5,6 % für Leistungstests und 7,2 % für Persönlichkeitstests. Letztere stellen nunmehr – wengleich auf niedrigem Niveau – den stärksten Einsatzanteil aller Testverfahren [18]. Dies ist jedoch im Wesentlichen auf die geringe Einsatzfrequenz von Intelligenz- und Leistungstests zurückzuführen. Auch Persönlichkeitstests werden im Rahmen interner Personalauswahl im Vergleich zur externen Personalauswahl seltener durchgeführt (20 vs. 7,2 %). Betrachtet man die Zahlen je Bewerbergruppe für persönlichkeitsorientierte Verfahren, reduziert sich über alle Gruppen hinweg gesehen die Einsatzhäufigkeit zwar leicht, befindet sich aber auf einem ähnlich niedrigen Gesamtniveau (vgl. zweite Spalte Tab. 7.5).

Im Gegensatz zu den Studien der ehemaligen Hohenheimer Arbeitsgruppe um Prof. Schuler, welche die Einsatzhäufigkeit zahlreicher unterschiedlicher Verfahren im Rahmen von Personalauswahlmaßnahmen untersucht haben, z. B. auch Interviews, Assessment-Center und andere Methoden, fokussiert die Studie von Schecke [50] mit einer Befragung unter den 580 größten deutschen Unternehmen speziell den Einsatz persönlichkeitsorientierter Testverfahren. Von den 120 Rücksendungen gaben 80 Personalverantwortliche (67 %) an, dass in der Organisation, für die sie tätig sind, Persönlichkeitstests (unabhängig, ob zu Zwecken der externen oder internen Personalauswahl oder Personalentwicklung) eingesetzt werden, 40 hingegen (33 %) haben dies verneint [28, 50]. Verglichen mit den Untersuchungen von Schuler et al. [57], die für die externe Personalauswahl eine Quote von 20 % ermittelt haben, sowie von Hell et al. [18], die für die interne Personalauswahl eine Quote von 7,2 % ausweisen, handelt es sich um eine deutliche Steigerung



**Abb. 7.2** Einsatzhäufigkeit und Bekanntheit der zehn verbreitetsten Verfahren (N = 120, davon 80 Anwender; [28])

der Einsatzquote – auch wenn berücksichtigt werden muss, dass Schecke [50] im Gegensatz zu Schuler et al. [57] in seiner Studie ausschließlich persönlichkeitsorientierte Verfahren thematisiert und somit bei den Rücksendungen mit einem höheren „Drop-out“ von Unternehmen, die gar keine Persönlichkeitstests einsetzen, kalkuliert werden muss.

Bei der Frage nach dem Bekanntheitsgrad und den Einsatzhäufigkeiten einzelner Testverfahren dominieren Typentests (vgl. Abb. 7.2) wie der Myers-Briggs-Typenindikator (MBTI, [5]) sowie das persolog-Persönlichkeits-Profil [16]. Auf dem dritten Platz folgen als Vertreter von Persönlichkeits-Struktur-Tests die Bochumer Inventare zur berufsbezogenen Persönlichkeitsbeschreibung BIP [25] und BIP-6F [23]. Mit berichteten Einsatzquoten von 4–8 % belegen die Verfahren OPQ [61], Predictive Index [11], Insights MDI [37], Hogan Assessments [19], NEO-PI-R [45] bzw. NEO-FFI [7], Reflector Big Five [47] und LIFO [10] die hinteren Plätze unter den am häufigsten eingesetzten Verfahren.

Die Tatsache, dass die ersten beiden Rangplätze von Typen bildenden Verfahren eingenommen werden, deren Aussagekraft aus wissenschaftlicher Sicht stark angezweifelt wird [15, 32, 35], zeigt, dass es offenbar noch deutlicherer Anstrengungen bedarf, die Kluft zwischen Wissenschaft und Praxis zu schließen, um die Vorteile wissenschaftlich beforschter und empirisch abgesicherter Verfahren so herauszustellen, dass diese häufiger den Vorzug vor unwissenschaftlichen Methoden erhalten [28].

## 7.6 Zur Verfälschbarkeit persönlichkeitsorientierter Testverfahren

Ein gelegentlich vorgetragenes Argument gegen den Einsatz persönlichkeitsorientierter Fragebogenverfahren ist die Befürchtung, Kandidaten könnten bei der Beantwortung zu stark dem Phänomen der sozialen Erwünschtheit unterliegen. Dies bezeichnet die Tendenz,

eigene Einstellungen oder Verhaltensweisen an vorgestellte oder tatsächliche soziale Normen anzupassen, d. h. im Auswahlprozess könnten Kandidaten Antworten verfälschen, um einen bestimmten Eindruck zu hinterlassen.

Hierbei gilt es zu berücksichtigen, dass dieses Phänomen jegliche Form menschlicher Interaktion betrifft und damit auch im gesamten eignungsdiagnostischen Prozess wirksam wird – also beispielsweise auch im Einstellungsinterview oder im Assessment-Center [24].

Grundsätzlich bestehen bei vorliegenden Persönlichkeitsprofilen, die den Verdacht auf sozial erwünschtes Antwortverhalten nahe legen, drei Möglichkeiten (z. B. durch unwahrscheinliche Häufung von Extremwerten, unplausible Skalenkombinationen, andersartige Fremdbildeindrücke):

1. Der Kandidat unterliegt einer verzerrten Selbstwahrnehmung im Sinne einer Selbsttäuschung und ist davon überzeugt, dass die gegebenen Antworten ihn angemessen beschreiben.
2. Ein Versuch der Fremdtäuschung im Sinne von „Impression Management“ wird unternommen.
3. Eine Mischform von Selbst- und Fremdtäuschung liegt vor.

Zur Klärung der Effekte kann der in Abb. 7.3 veranschaulichte Prozess dienen.

Hinsichtlich der Auswirkung sozial erwünschten Antwortverhaltens auf die Höhe der Validität werden sowohl dämpfende, steigernde als auch neutrale Effekte beschrieben [38]. Hogan et al. [20] kommen allerdings auf Basis einer großen Untersuchung mit mehr als 5000 Mitarbeitern zu dem Schluss, dass Effekten von „Faking“ in realen Auswahl-situationen kaum Relevanz beizumessen ist.

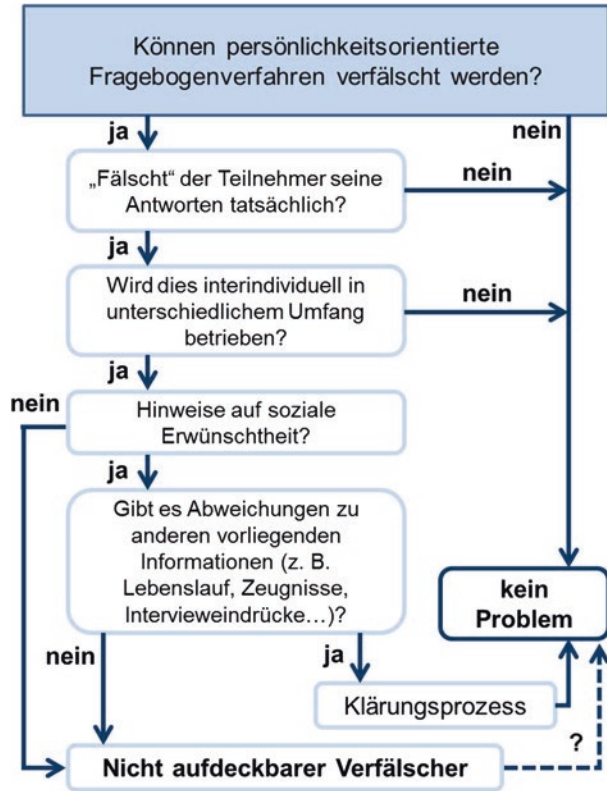
Letztlich kann die Problematik sozial erwünschten Verhaltens durch kein diagnostisches Instrument vollends befriedigend gelöst werden. Die etwaigen „unerwünschten“ Effekte bei der Nutzung persönlichkeitsorientierter Testverfahren können vom kundigen und ausreichend gesprächsführungskompetenten Anwender im Abgleich mit anderen Informationsquellen und Fremdwahrnehmungen jedoch in aller Regel aufgedeckt und im persönlichen Gespräch mit den Kandidaten hinterfragt und im weiteren Entscheidungsprozess eingebunden werden.

---

## 7.7 Rechtliche Rahmenbedingungen für den Einsatz persönlichkeitsorientierter Testverfahren

Insbesondere im Personalauswahlkontext hat die juristische Einwandfreiheit der eignungsdiagnostischen Maßnahmen besondere Relevanz. Seit der Einführung des Allgemeinen Gleichbehandlungsgesetzes (AGG) im Jahr 2006 hat auch die mediale Beachtung dieser Thematik stark zugenommen [64]. Gerade bei persönlichkeitsorientierten Verfahren scheut der ein oder andere Personalexperte möglicherweise den Einsatz aufgrund rechtsbezogener Befürchtungen und Unklarheiten, sodass an dieser Stelle der

**Abb. 7.3** Flussdiagramm zur Verfälschbarkeit persönlichkeitsorientierter Fragebogenverfahren [24, S. 104]



Besprechung der relevanten Aspekte ein etwas breiterer Raum gegeben werden soll und insbesondere auch erste empirische Ergebnisse im Kontext des AGG, welches in seinen Folgen für die Nutzung persönlichkeitsorientierter Instrumente bisher in der Literatur kaum diskutiert wurde, zusammengefasst werden.

Für alle abhängig Beschäftigten gilt das Arbeitsrecht, in welchem eine Reihe von Sonderregelungen zum Schutz der Arbeitnehmer implementiert sind, beispielsweise das Kündigungsschutzgesetz, Regelungen zum Schutz besonderer Beschäftigungsgruppen oder zum Datenschutz [48].

Nach § 32 Abs. 1 des Bundesdatenschutzgesetzes ist die Erhebung und Nutzung personenbezogener Daten eines Beschäftigten nur dann zulässig, wenn dies für die Entscheidung über die Begründung eines Beschäftigungsverhältnisses oder nach Begründung eines solchen für dessen Durchführung oder Beendigung erforderlich ist. Daher muss gewährleistet sein, dass mithilfe des Persönlichkeitstests ausschließlich anforderungsrelevante Merkmale zur Prüfung der Eignung einer Person in Bezug auf die Anforderungen in der (zukünftigen) Arbeitstätigkeit erhoben werden [40].

Mit dem Einsatz eines persönlichkeitsorientierten Fragebogenverfahrens darf zudem keine Verletzung des allgemeinen Persönlichkeitsrechts von Bewerbern einhergehen, zu

dessen Schutz der Arbeitgeber grundsätzlich verpflichtet ist. Das allgemeine Persönlichkeitsrecht ist insbesondere aus Artikel 1 Absatz 1 des Grundgesetzes zur Unantastbarkeit der Würde des Menschen sowie aus Artikel 2 Absatz 1 zum Recht auf freie Entfaltung der eigenen Persönlichkeit abgeleitet. Es wird als sogenanntes Richterrecht wirksam, welches gewohnheitsrechtlich anerkannt ist, d. h. jeder Fall wird auf Basis bereits ergangener Urteile abgewogen.

Aus der Notwendigkeit der Arbeitsplatzbezogenheit der erfassten Merkmale sowie aus dem Schutz des Persönlichkeitsrechts ist abzuleiten, dass beispielsweise Fragen, die die Privat- oder Intimsphäre des Kandidaten berühren, nicht gestellt werden dürfen [24].

Als eine weitere grundlegende Voraussetzung für die Zulässigkeit des Einsatzes eines Persönlichkeitstests ist die Einwilligung des Kandidaten anzusehen, die nach Aufklärung über Art und Umfang des Tests erteilt wurde [2]. Idealerweise wird diese Zustimmung schriftlich gegeben [48], in der Praxis erfolgt dies häufig jedoch auch mündlich bzw. meist durch sogenanntes schlüssiges Verhalten, indem der Kandidat durch Ausfüllen des Fragebogens sein Einverständnis signalisiert.

Auch das Mitbestimmungsrecht des Betriebsrats (nach § 5 Abs. 3 BetrVG allerdings nicht für leitende Angestellte, [2]) zielt darauf ab, das Persönlichkeitsrecht von Bewerbern zu schützen, indem unzulässige Methoden der Informationsgewinnung verhindert werden sollen. Personalfragebogen sind nach § 94 Abs. 1 BetrVG zustimmungspflichtig. Unter dem Begriff Personalfragebogen werden jegliche schriftliche oder elektronische Abfragen arbeitnehmerbezogener Daten zusammen gefasst, daher liegt die Mitbestimmungspflicht auch für psychologische Testverfahren nahe [48]. Für eine Zusammenfassung möglicher Vorbehalte von Arbeitnehmervertretungen in Bezug auf den Einsatz persönlichkeitsorientierter Fragebogen und entsprechende Gegenargumente wird auf Hossiep und Mühlhaus [24] verwiesen.

Ergebnisse persönlichkeitsorientierter Fragebogen sind sensible personenbezogene Daten, hinsichtlich deren Aufbewahrung auch die Persönlichkeitsrechte von Kandidaten zu wahren sind. Eine Archivierung wird nur dann als datenschutzrechtlich zulässig angesehen, wenn ein besonderes rechtliches Interesse vorliegt wie etwa bei Einstellung des Bewerbers oder wenn rechtliche Auseinandersetzungen hinsichtlich der Eignungsbeurteilung abzusehen sind. Bei erfolgloser Bewerbung ist somit im Regelfall mit Abschluss des konkreten Bewerbungsprozesses das Recht zur Nutzung der Daten beendet mit der Folge, dass diese zu vernichten sind, es sei denn, der Kandidat hat ihrer weiteren Aufbewahrung z. B. für einen Bewerberpool ausdrücklich zugestimmt [48].

Die DIN 33430 „Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen“ legt als Prozessnorm Qualitätsstandards fest [13]. Es handelt sich dabei jedoch nicht um rechtlich verbindliche Normen. Dennoch kann die DIN 33430 juristisch bedeutsam sein, z. B. wenn im Rahmen eines Rechtsstreits die Rechtmäßigkeit einer Personalauswahlentscheidung angezweifelt wird. Kann der Arbeitgeber in einem solchen Fall den Nachweis erbringen, bei der eignungsdiagnostischen Beurteilung den Anforderungen der DIN entsprochen zu haben, wird sich dies vermutlich zu seinen Gunsten auswirken [48].



Für den Einsatz von Testverfahren aller Art sind zudem die Bestimmungen des AGG zu berücksichtigen, auf die im Folgenden ein besonderer Fokus gelegt wird.

Ziel des AGG ist die Verhinderung bzw. Beseitigung von Benachteiligungen aufgrund des Geschlechts, des Alters, einer Behinderung, der sexuellen Identität, der Religion oder Weltanschauung sowie hinsichtlich Rasse oder ethnischer Herkunft (§ 1 AGG). Für die Anwendung psychologischer Instrumente und damit auch persönlichkeitsorientierter eignungsdiagnostischer Verfahren ist insbesondere von Bedeutung, dass Benachteiligungen in Bezug auf Bedingungen, einschließlich Auswahlkriterien und Einstellungsbedingungen, für den Zugang zu Erwerbstätigkeit und beruflichem Aufstieg unzulässig sind (§ 2 Abs. 1 AGG). Die in § 1 AGG genannten personenbezogenen Attribute dürfen also beim Einstellungsprozess keinerlei Berücksichtigung finden. Ausnahmen lässt das Gesetz nur zu, wenn unverzichtbare berufliche Anforderungen einer Gleichbehandlung entgegenstehen (§ 8 bis 10 sowie § 20 AGG). Des Weiteren sind auch positive Maßnahmen zugelassen (§ 5 AGG), um aktuell bestehende Nachteile wegen eines in § 1 genannten Grundes durch eine unterschiedliche Behandlung zu verhindern oder auszugleichen. Unter positiven Maßnahmen werden Aktivitäten zusammengefasst, die zur Herstellung von vollständiger und effektiver Chancengleichheit von Personengruppen beitragen, die benachteiligt sind oder anderweitig die Folgen vergangener oder gegenwärtiger Diskriminierung zu erleiden haben [34].

Das AGG unterscheidet zwischen unmittelbarer und mittelbarer Benachteiligung (§ 3 AGG): Erstere wäre gegeben, wenn eine Person wegen eines in § 1 genannten Grundes eine weniger günstige Behandlung als eine andere Person in einer vergleichbaren Situation erfährt, beispielsweise aufgrund des Geschlechts als nicht für die Vakanz infrage kommend betrachtet wird. Von mittelbarer Benachteiligung wird gesprochen, wenn scheinbar neutrale Vorschriften, Kriterien oder Verfahren Personen wegen eines in § 1 genannten Grundes gegenüber anderen Personen in besonderer Weise benachteiligen können. Die Anwendung eines persönlichkeitsorientierten Fragebogens, welcher zu einer Benachteiligung aufgrund eines in § 1 AGG genannten Merkmals führt, würde demnach eine mittelbare Benachteiligung darstellen und könnte eine Klagemöglichkeit für Kandidaten begründen.

Der dem AGG zugrunde liegende Fairnessbegriff folgt dem Regressionskonzept von Cleary [9], wonach Testverfahren als fair bezeichnet werden, wenn keine systematische Über- oder Unterschätzung von Kriteriumswerten für bestimmte Subgruppen vorliegt:

A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup. In other words, the test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroup [9, S. 115].

Die „Society for Industrial and Organizational Psychology“ (SIOP) [62] verwendet dafür die Termini prädiktiver Bias und differenzielle Prädiktion.

Wenn für alle Subgruppen die Beziehung zwischen einem Testwert und einem Kriterium durch eine gemeinsame Regressionsgerade adäquat beschrieben werden kann, wird die Bestenauswahl auf Basis eben jener Testwerte – ohne Berücksichtigung der Gruppenzugehörigkeit von Kandidaten und etwaiger Mittelwertunterschiede zwischen den Gruppen – als fair erachtet [53]. Während im Bereich von Testverfahren zur kognitiven Leistungsfähigkeit diese Thematik schon intensiv beforscht wurde (z. B. [6, 65]) und in der Regel kein nachteiliger prädiktiver Bias in Bezug auf die Validität für klassischerweise als benachteiligt angesehene Personengruppen (bzgl. Rasse bzw. Hautfarbe für die nicht-weiße Bevölkerung bzw. bezogen auf das Geschlecht für Frauen) konstatiert werden konnte, ist die Forschung im Bereich persönlichkeitsorientierter Verfahren noch unterrepräsentiert [62]. Eine erstmalige Untersuchung für den deutschsprachigen Raum mit mehr als 14.500 berufstätigen Personen legt Weiß [64] vor. Sie zeigt exemplarisch für die sechs übergeordneten Persönlichkeitsfaktoren Engagement, Disziplin, Sozialkompetenz, Kooperation, Dominanz und Stabilität des BIP-6F, dass bei Anwendung eines persönlichkeitsorientierten Fragebogens keine Benachteiligung in Bezug auf verschiedene Altersgruppen oder Geschlechter im Sinne des AGG entsteht.

---

## **7.8 Einbettung persönlichkeitsorientierter Verfahren in den Personalauswahlprozess und darauf aufbauende Personalentwicklungsprozesse**

Ein Personalauswahlprozess wird in der Regel sequenziell gestaltet [49].

Zunächst werden die wesentlichen Anforderungsmerkmale bestimmt (vgl. Kap. 4 in diesem Band). Von einer differenzierten Erfassung der Persönlichkeitsausprägungen derzeitiger und zukünftiger Mitarbeiter können Organisationen vor allem dann profitieren, wenn zuvor auch die überfachlichen Anforderungen systematisiert wurden. Eine schlagwortartige Liste relevanter Persönlichkeitseigenschaften ist jedoch nicht ausreichend, vielmehr sollte auch für die Anforderungsseite ein Soll-Profil oder Soll-Korridor mit den gewünschten Ausprägungsgraden definiert werden [22]. Das „Je mehr desto besser“-Prinzip, das aus dem oftmals einen unteren „Cut-off“ definierenden Bereich kognitiver Leistungsanforderungen bekannt ist, greift in Hinblick auf Persönlichkeitseigenschaften nicht. Je nach den Anforderungen einer Vakanz kann etwa eine hohe Ausprägung von Gewissenhaftigkeit anforderungsgerecht sein (z. B. für einen Sachbearbeiter im Controlling), während für eine Führungsposition im Vertrieb eine besonders hohe Ausprägung dieser Skala kontraproduktiv und eher eine mittlere bis leicht unterdurchschnittliche Ausprägung förderlich sein dürfte. Auch mit Blick auf die Forderungen der DIN-Norm (vgl. von Melchers in diesem Band) und die rechtlichen Rahmenbedingungen zum Einsatz von Testverfahren (vgl. Abschn. 7.7) ist gerade auch für den Bereich Persönlichkeit eine sorgfältige Anforderungsanalyse notwendig. Einen Ansatz zur Systematisierung

überfachlicher Anforderungen korrespondierend zur anschließenden Erhebung des Selbstbildes liefert beispielsweise das Anforderungsmodul zum BIP [22, 26].

Nach der Anforderungsanalyse erfolgt die Beschaffung potenziell geeigneter Bewerber, deren Anzahl durch Vorauswahlmaßnahmen wie eine Negativauslese auf Basis der Bewerbungsunterlagen und gegebenenfalls zusätzlich telefonischer Interviews oder Vorab-Durchführung von Testverfahren reduziert wird. Auch wenn traditionell persönlichkeitsorientierte Testverfahren erst später zum Einsatz kommen, wird ihr Einsatz zunehmend im Rahmen der Vorauswahl diskutiert. Da ein Persönlichkeitsfragebogen stets nur das Selbstbild eines Kandidaten erfassen kann, ist zur Entscheidungsfindung zumindest ein anschließendes persönliches Gespräch erforderlich, sodass vom Einsatz eines solchen Fragebogens im Rahmen der Vorauswahl ohne persönlichen Kontakt abzuraten ist. Ein persönlichkeitsorientiertes Fragebogenverfahren sollte daher niemals – gleiches gilt übrigens für andere Verfahren auch – als alleinige Entscheidungsquelle dienen und kommt klassischerweise meist während der Hauptauswahl im Rahmen späterer Selektionsstufen des eignungsdiagnostischen Prozesses zum Einsatz, z. B. begleitend zum Interview (vgl. Kapitel von Nerdinger in diesem Band). Nicht zuletzt auch aus Gründen des Personalmarketings ist es sinnvoll, Möglichkeiten zur Besprechung der Ergebnisse mit den Kandidaten zu schaffen [24].

Gerade der Umgang mit Ergebnissen persönlichkeitsorientierter Testverfahren erfordert einen umsichtigen und erfahrenen Personalexperten, der für das eingesetzte Verfahren ausreichend geschult ist. Das Ergebnis eines fundierten berufsbezogenen Persönlichkeitsfragebogens liefert eine systematische Zusammenfassung des Selbstbildes eines Kandidaten in Hinblick auf bestimmte beruflich relevante Persönlichkeitseigenschaften. Es bietet Anlass für ein vertiefendes Gespräch zur Klärung der überfachlichen Passung von Kandidaten für eine bestimmte berufliche Position. Besonderheiten, Widersprüche und Auffälligkeiten im Selbstbild werden für den kundigen Anwender durch die Systematik des Persönlichkeitsfragebogens erkennbar. Zur Komplettierung und Verifizierung des Bildes über die Persönlichkeit eines Bewerbers können und müssen diese Daten mit anderen Informationsquellen (z. B. dem Lebenslauf) und Fremdbildwahrnehmungen (z. B. aus dem Interview oder auch verhaltens- bzw. simulationsorientierten Verfahren wie dem Assessment-Center) abgeglichen werden.

Um die Erfassung von Persönlichkeitsmerkmalen im eignungsdiagnostischen Prozess einzubetten, ist generell ein multimethodales Vorgehen empfehlenswert [55]: Neben persönlichkeitsbezogenen Fragen im Interview oder dem Rückschluss auf bestimmte Persönlichkeitseigenschaften von Verhaltensbeobachtungen im Assessment-Center können persönlichkeitsorientierte Fragebogenverfahren einen wertvollen Beitrag zur Erfassung überfachlicher Kompetenzen derzeitiger oder zukünftiger Mitarbeiter liefern. Sie sind hoch ökonomisch hinsichtlich Kosten und Zeitaufwand, da – bei sachgemäßem Umgang mit den Resultaten – mit überschaubarem Aufwand eine hohe Informationsdichte zum Aspekt der Persönlichkeit, dessen Einbindung zur Verbesserung der Prognosequalität von Auswahlentscheidungen beiträgt, erzielt werden kann.

Zusätzlich zur Unterstützung der Entscheidungsfindung im Personalauswahlprozess können die Ergebnisse von Persönlichkeitsfragebogen zudem zahlreiche Ansatzpunkte für flankierende Personalentwicklungsmaßnahmen liefern, die die erfolgreiche Ausübung der jeweiligen beruflichen Position bzw. auch die Weiterentwicklung mit Blick auf weitere Karriereschritte unterstützen können.

Der folgende Kasten fasst die Argumente für den Einsatz seriöser persönlichkeitsorientierter Verfahren zusammen [24, S. 182].

#### Argumente für den Einsatz persönlichkeitsorientierter Verfahren:

- Liefern auf ökonomische Weise ein differenziertes Persönlichkeitsbild.
- Bieten Ansatzpunkte für den unkomplizierten Abgleich von Selbst- und Fremdbildern.
- Können die Laufbahn- und Karriereentwicklung durch wiederholten Einsatz langfristig begleiten und fördern.
- Ermöglichen Benchmarking durch hinterlegte Referenz- bzw. Vergleichsgruppen.
- Vermitteln eine anschauliche Struktur zur Reflexion und Kommunikation über Persönlichkeit.
- Können seitens erfahrener Anwender problemlos Bezüge zu den überfachlichen Anforderungen einer Position aufzeigen und zu deren Objektivierung dienen.
- Können nach Einsatz im Auswahlprozess später anhand der generierten Persönlichkeitsprofile einen sinnvollen Beitrag zu Personalentwicklungsmaßnahmen leisten.
- Ermöglichen über die Abbildung in Kompetenzmodellen hinaus anhand von Persönlichkeitsprofilen eine Erkennung von Verhaltensursachen und liefern somit Hinweise für Interventionen.
- Zeigen durch die resultierenden Profile ggf. Widersprüche innerhalb der Persönlichkeitsstruktur auf, die sonst meist nicht berücksichtigt werden.
- Liefern Ansatzpunkte für die weitere Exploration beruflich relevanter Persönlichkeitszüge.
- Bieten durch die Ergebnisprofile ein Fundament, um wesenswidrige und nicht persönlichkeitsadäquate Personalentwicklungsmaßnahmen zu unterlassen.

#### Fragen

1. Warum ist der Einsatz persönlichkeitsorientierter Verfahren im Personalauswahlbereich sinnvoll?
2. Welche Arten von Persönlichkeitstests sind zu unterscheiden? Nennen Sie jeweils ein Beispiel.
3. Welche rechtlichen Rahmenbedingungen sind beim Einsatz persönlichkeitsorientierter Verfahren zu berücksichtigen?

## Literatur

1. Aguinis, H., & Smith, M. A. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *Personnel Psychology, 60*, 165–199.
2. Arnold, C. (2013). Juristische Aspekte der Management-Diagnostik. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 50–63). Göttingen: Hogrefe.
3. Asendorpf, J. B., & Neyer, F. J. (2012). *Psychologie der Persönlichkeit*. Berlin: Springer.
4. 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.
5. Bents, R., & Blank, R. (2003). *Der M.B.T.I*. München: Claudius.
6. Berry, C. M., Clark, M. A., & McClure, T. K. (2011). Racial/ethnic differences in the criterion-related validity of cognitive ability tests: A qualitative and quantitative review. *Journal of Applied Psychology, 96*, 881–906.
7. Borkenau, P., & Ostendorf, F. (2008). *NEO-Fünf-Faktoren-Inventar nach Costa und McCrae*. Göttingen: Hogrefe.
8. Borkenau, P., Egloff, B., Eid, M., Hennig, J., Kersting, M., Neubauer, A. C., & Spinath, F. M. (2005). Persönlichkeitspsychologie: Stand und Perspektiven. *Psychologische Rundschau, 56*, 271–290.
9. Cleary, T. A. (1968). Test bias: Prediction of grades for Negro and White students in integrated colleges. *Journal of Educational Measurement, 5*, 115–124.
10. Czichos, R. (2001). *Profis managen sich selbst. Die LIFO-Methode für ihr persönliches Stärkenmanagement*. München: Reinhardt.
11. Daniels, A. (2004). The Predictive Index System. In W. Sarges (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren. Band I: Personalpsychologische Instrumente* (S. 637–642). Lengerich: Pabst Science Publishers.
12. DeGroot, T., & Kluemper, D. (2007). Evidence of predictive and incremental validity of personality factors, vocal attractiveness and the situational interview. *International Journal of Selection and Assessment, 15*, 30–39.
13. DIN. (2002). *DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth.
14. Eggert, D. (1983). *Eysenck-Persönlichkeits-Inventar (E-P-I)*. Göttingen: Hogrefe.
15. Fachgruppe Diagnostik des SDBB (2011, 14. September). Zertifizierung von Testverfahren MBTI und GPOP. [http://www.diagnostik.sdbb.ch/sites/testraum.localhost/files/Zertifizierung\\_von\\_Testverfahren\\_0.pdf](http://www.diagnostik.sdbb.ch/sites/testraum.localhost/files/Zertifizierung_von_Testverfahren_0.pdf). Zugegriffen: 13. Mai 2015.
16. Gay, F. (2004). *Das persolog-Persönlichkeits-Profil. Persönliche Stärke ist kein Zufall. Mit Fragebogen zur Selbstausswertung*. Offenbach: Gabal Verlag & persolog GmbH.
17. Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*, 135–164.
18. 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.
19. Hogan, R., & Hogan, J. (1992). *Hogan Personality Inventory: Manual*. Tulsa: Hogan Assessment Systems.
20. Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employee selection. *Journal of Applied Psychology, 92*, 1270–1285.
21. Hossiep, R. (2013). Persönlichkeits- und Motivationsinventare. Explizite Maße. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 592–601). Göttingen: Hogrefe.
22. Hossiep, R., & Bräutigam, S. (2008). Ansätze zur systematischen Erfassung überfachlicher Positionsanforderungen. In W. Sarges & D. Scheffer (Hrsg.), *Innovative Ansätze für die Eignungsdiagnostik* (S. 275–285). Göttingen: Hogrefe.

23. Hossiep, R., & Krüger, C. (2012). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung – 6 Faktoren*. Göttingen: Hogrefe.
24. Hossiep, R., & Mühlhaus, O. (2015). *Personalauswahl und -entwicklung mit Persönlichkeits-tests*. Göttingen: Hogrefe.
25. Hossiep, R., & Paschen, M. (2003). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung – BIP*. Göttingen: Hogrefe.
26. Hossiep, R., & Weiß, S. (i. V.). *Das BIP-AM. Anforderungsmodul zum Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung*. Göttingen: Hogrefe.
27. Hossiep, R., Paschen, M., & Mühlhaus, O. (2000). *Persönlichkeitstests im Personalmanagement. Grundlagen, Instrumente und Anwendungen*. Göttingen: Verlag für Angewandte Psychologie.
28. Hossiep, R., Shecke, J., & Weiß, S. (2015). Zum Einsatz von persönlichkeitsorientierten Fragebogen. Eine Erhebung unter den 580 größten deutschen Unternehmen. *Psychologische Rundschau*, 2(15), 127–129.
29. Hough, L. M., & Connelly, B. S. (2013). Personality measurement and use in industrial and organizational psychology. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Hrsg.), *APA handbook of testing and assessment in psychology, Bd. 1: Test theory and testing and assessment in industrial and organizational psychology* (S. 501–531). Washington: American Psychological Association.
30. Hough, L. M., & Dilchert, S. (2010). Personality: Its measurement and validity for employee selection. In J. L. Farr & N. T. Tippins (Hrsg.), *Handbook of employee selection* (S. 299–319). New York: Routledge & Taylor & Francis Group.
31. 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.
32. Kersting, M. (2013). *Persönlichkeit ist keine Typenfrage*. *personalmagazin*, 12, 26–29.
33. Kienbaum Consultants International GmbH. (2012). *High Potentials 2011/2012*. Gummersbach: Kienbaum Consultants International GmbH.
34. Klose, A., & Merx, A. (2010). *Positive Maßnahmen zur Verhinderung oder zum Ausgleich bestehender Nachteile im Sinne des § 5 AGG. Expertise*. Berlin: Antidiskriminierungsstelle des Bundes.
35. König, C. J., & Marcus, B. (2013). TBS-TK Rezension: „Persolog Persönlichkeits-Profil“. *Psychologische Rundschau*, 64, 189–191.
36. Kubinger, K. D., & Ebenhöf, J. (1996). *Arbeitshaltungen – Kurze Testbatterie*. Frankfurt a. M.: Swets.
37. Lothian, A., & Lothian, A. (2004). Insights discovery profil. In W. Sarges (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren. Band I: Personalpsychologische Instrumente* (S. 413–419). Lengerich: Pabst Science Publishers.
38. Marcus, B. (2009). „Faking“ from the applicant’s perspective: A theory of self-presentation in personnel selection settings. *International Journal of Selection and Assessment*, 17, 417–430.
39. Marston, W. M. (1928). *Emotions of normal people*. London: Routledge.
40. Maties, M., & Wottawa, H. (2011). Nicht immer erlaubt: Wann Eignungstests künftig zulässig sind. *Wirtschaftspsychologie aktuell*, 4, 28–32.
41. Morgenthaler, W. (1992). *Rorschach-Psychodiagnostik*. Bern: Huber.
42. Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60, 683–729.
43. Ng, T., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success. *A meta-analysis*. *Personnel Psychology*, 58, 367–408.
44. Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, 60, 995–1027.



45. Ostendorf, F., & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae*. Göttingen: Hogrefe.
46. Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, 57, 401–421.
47. PI Company. (2014). Reflector Big Five Personality. <http://www.picompany.de/produktverzeichnis/reflector-big-five-personality/>. Zugegriffen: 20. Nov. 2015.
48. Püttner, I. (2014). Rechtliche Aspekte der Personalarbeit. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 1201–1226). Göttingen: Hogrefe.
49. Sarges, W. (2000). Personal – Auswahl, Beurteilung und Entwicklung. In J. Straub, A. Kochinka, & H. Werbik (Hrsg.), *Psychologie in der Praxis – Anwendungs- und Berufsfelder einer modernen Wissenschaft* (S. 487–522). München: dtv.
50. Shecke, J. (2014). *Zum Einsatz von Persönlichkeitsfragebögen – Eine Erhebung unter den 580 größten deutschen Unternehmen*. Unveröffentlichte Masterarbeit, Ruhr-Universität Bochum: Bochum.
51. Scheffer, D. (2013). Persönlichkeits- und Motivationsinventare: Implizite Maße. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 601–608). Göttingen: Hogrefe.
52. 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.
53. Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik*. Berlin: Springer.
54. Schneewind, K. A., & Graf, J. (1990). *Der 16-Persönlichkeitsfaktoren-Test. Revidierte Fassung (16 PF-R)*. Bern: Huber.
55. Schuler, H. (2014). *Psychologische Personalauswahl*. Göttingen: Hogrefe.
56. Schuler, H., Frier, D., & Kauffmann, M. (1993). *Personalauswahl im europäischen Vergleich*. Göttingen: Hogrefe.
57. Schuler, H., Hell, B., Trapmann, S., Schaar, H., & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie*, 6, 60–70.
58. Schuler, H., Höft, S., & Hell, B. (2014). Eigenschaftsorientierte Verfahren der Personalauswahl. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 149–213). Göttingen: Hogrefe.
59. Schulz, C., Schuler, H., & Stehle, W. (1985). Die Verwendung eignungsdiagnostischer Methoden in deutschen Unternehmen. In H. Schuler & W. Stehle (Hrsg.), *Organisationspsychologie und Unternehmenspraxis: Perspektiven der Kooperation* (S. 126–132). Göttingen: Hogrefe.
60. Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, 65, 445–494.
61. SHL Group Limited. (2007). *OPQ32 User Manual*. SHL Group Limited. Thames Ditton: SHL Group Limited.
62. Society for Industrial and Organizational Psychology (2003). *Principles for the validation and use of personnel selection procedures*. Bowling Green: Society for Industrial and Organizational Psychology.
63. Stemmler, G., Bartussek, D., Hagemann, D., & Amelang, M. (2011). *Differentielle Psychologie und Persönlichkeitsforschung*. Stuttgart: Kohlhammer.
64. Weiß, S. (2014). *Zum Einsatz persönlichkeitsorientierter Testverfahren im beruflichen Kontext: Eine Analyse vor dem Hintergrund des Allgemeinen Gleichbehandlungsgesetzes (AGG)*. Dissertation, Ruhr-Universität Bochum, Bochum.
65. Young J. W., & Koblin J. L. (2001). *Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis* (College Board Research Report No. 2001-6). New York: The College Board.



Arleta Franczukowska

## Zusammenfassung

Im folgenden Beitrag werden zwei simulationsorientierte Verfahren der Personalauswahl eingehend diskutiert – die Arbeitsprobe und der Situational-Judgment-Test (SJT). Es gilt, einen umfassenden Überblick über die genannten Instrumente zu generieren sowie deren Gemeinsamkeiten und Unterschiede darzulegen. Ausgehend von der Begriffsbestimmung und -abgrenzung werden die Ziele, Typen und der genaue Ablauf beider Verfahren erläutert sowie deren Vor- und Nachteile gegenübergestellt. Empirisch-gestützte Aussagen zu den Gütekriterien (Validität, Reliabilität), subgruppenspezifischen Leistungsunterschieden („adverse impact“) sowie zur Verfälschbarkeit der Arbeitsprobe und des SJT ergänzen die Inhalte.

## Inhaltsverzeichnis

8.1	Arbeitsprobe	182
8.1.1	Ziele und Typen von Arbeitsproben	182
8.1.2	Gütekriterien von Arbeitsproben	184
8.1.3	Ablauf der Konstruktion von Arbeitsproben	185
8.1.4	Vor- und Nachteile von Arbeitsproben und zukünftiger Forschungsbedarf	186
8.2	Situational-Judgment-Test	188
8.2.1	Ziele und Typen von Situational-Judgment-Tests	189

A. Franczukowska (✉)

Fakultät für Wirtschaftswissenschaften, Alpen-Adria-Universität Klagenfurt,  
Klagenfurt, Österreich

E-Mail: Arleta.Franczukowska@aau.at

8.2.2	Ablauf der Konstruktion von Situational-Judgment-Tests .....	190
8.2.3	Gütekriterien von Situational-Judgment-Tests .....	191
8.2.4	Vor- und Nachteile von Situational-Judgment-Tests und zukünftiger Forschungsbedarf .....	192
8.3	Fazit .....	194
	Literatur .....	195

## 8.1 Arbeitsprobe

Seit Jahrzehnten gelten Arbeitsproben („work sample tests“) als zuverlässige Prädiktoren des beruflichen Erfolges im Rahmen von Personalauswahlprozessen [25]. Arbeitsproben lassen sich wie folgt definieren:

► Arbeitsproben bezeichnen einen Test, bei dem ein Bewerber eine bestimmte Auswahl an standardisierten, physischen oder physiologischen Aufgaben, die inhaltlich valide und erkennbar äquivalente Stichproben des erfolgsrelevanten beruflichen Verhaltens darstellen, bewältigen muss [24, 29].

Guion und Highhouse [10] weisen in Bezug auf Arbeitsproben darauf hin, dass sowohl der Arbeitsprozess als auch dessen Ergebnisse oder beides kombiniert beobachtet und gewertet werden können.

### Beispiel

Bei einem hochtechnologischen Flugsimulationstest, welchen Piloten häufig absolvieren müssen, würde man sich vorrangig den Prozess ansehen, während man bei der Auswahl einer geeigneten Sekretärin primär das Ergebnis der Arbeitsprobe (z. B. abgetippte Textqualität) fokussieren würde [10, S. 242].

### 8.1.1 Ziele und Typen von Arbeitsproben

Arbeitsproben zielen prinzipiell darauf ab, bestimmte Aufgaben aus dem Berufsalltag möglichst realitätsgetreu zu simulieren, um vom beobachteten Verhalten der Bewerber Rückschlüsse auf deren tatsächliche Fähigkeiten und Fertigkeiten sowie zukünftiges Arbeits- und Leistungsverhalten zu ziehen [4, S. 248, 11, S. 453]. Im Gegensatz zu Intelligenz- oder Persönlichkeitstests, spiegeln Arbeitsproben das charakteristische Verhalten einer Person wider. Demnach bringen sie nicht das mögliche, sondern das tatsächliche Verhalten einer Person zum Vorschein [32, S. 372]. Als eignungsdiagnostisches Verfahren werden Arbeitsproben seit Anfang des 20. Jahrhunderts entwickelt und eingesetzt [29, S. 225]. Abgesehen von der Personalauswahl eignen sie sich auch zur Validierung von Trainingsprogrammen, zur individuellen Leistungsbeurteilung und Potenzialanalyse als auch zur Arbeits- und Aufgabenzertifizierung [4, S. 248, 11, S. 453].

Üblicherweise werden Arbeitsproben bei der Auswahl bereits qualifizierter Arbeitskräfte bzw. Personen mit Berufserfahrung eingesetzt [28, S. 267].

### Beispiel

Will man feststellen, ob ein Mechaniker ausreichend qualifiziert ist, könnte man ihn im Zuge des Auswahlprozesses vor die Aufgabe stellen, einen defekten Motor zu reparieren. Hierbei empfiehlt es sich, nicht nur das Endergebnis („Motor läuft wieder oder nicht“) zu werten, sondern auch den Zielerreichungsprozess bzw. die jeweilige Herangehensweise [14, S. 218, 28, S. 267].

Eine solche simulierte Aufgabenstellung spiegelt die am häufigsten eingesetzte, klassische Form der Arbeitsprobe wider, nämlich die sogenannte psychomotorische („hands-on“) Arbeitsprobe, bei der es um die physiologische Manipulation einer konkreten Aufgabe geht [4, S. 252, 25]. Asher und Sciarrino [1, S. 519 ff.] unterscheiden in ihrem Überblicksartikel zwischen motorischen und verbalen Arbeitsproben, wobei ersteres mit der psychomotorischen Arbeitsprobe gleichgesetzt werden kann und letzteres die sprachbasierte oder sozial-interaktive Bearbeitung einer bestimmten Problemstellung beinhaltet (z. B. Textkorrektur). Ferner zählen Callinan und Robertson weitere eignungsdiagnostische Messverfahren auf, die sie ebenfalls unter dem Begriff der Arbeitsprobe subsumieren:

- Trainierbarkeitstests („trainability tests“), welche keine fundierte Berufserfahrung aufgrund einer integrierten Lernphase voraussetzen;
- Mündliche und schriftliche Situationstests (Situational-Judgment-Tests);
- Fachkenntnistests („job knowledge tests“) zur Erfassung berufsrelevanten Wissens;
- Assessment-Center-Aufgaben wie Rollenspiele, Gruppendiskussionen, Postkorbübungen etc. [4, S. 252 ff., 14, S. 219].

Anhand dieser Auflistung wird deutlich, dass auch Situational-Judgment-Tests, welche nachstehend ausführlicher behandelt werden, Fachkenntnistests sowie bestimmte Assessment-Center-Aufgaben im weiteren Sinne als Arbeitsproben aufgefasst werden.

Das Prinzip der Arbeitsprobe mag einfach erscheinen, bei genauerer Betrachtung eröffnen sich allerdings zahlreiche Gestaltungsmöglichkeiten [14, S. 218 f.]. Nach Callinan und Robertson [4, S. 256 ff.] lassen sich Arbeitsproben hinsichtlich von sieben Dimensionen differenzieren:

- Bandbreite,
- Realitätsbezug,
- Aufgabenspezifität,
- Erfahrungsbezug,
- Aufgabentypus,
- Antwortmodus,
- Bewertungsmodus.

Erstes Kriterium, die Bandbreite, beschreibt, in welchem Umfang die Probe die alltäglichen Arbeitsaufgaben abdeckt. Der Realitätsbezug hingegen bezieht sich auf das Ausmaß, in welchem die Probe die Arbeitsrealität widerspiegelt. Während die Aufgabenspezifität angibt, ob eher allgemeine oder spezifische Fertigkeiten erfasst werden, legt der Erfahrungsbezug nahe, inwiefern die Probe Berufserfahrung voraussetzt. Der Aufgabentypus weist auf die Art der Aufgabe (motorisch, verbal, sozial, kognitiv) hin, wobei die geforderte Reaktion (Antwortmodus) verhaltensbezogen, mündlich oder schriftlich erfolgen kann. Der Bewertungsmodus dient der standardisierten Einschätzung des jeweiligen Leistungsniveaus via Ratingskalen, Checklisten etc.

### 8.1.2 Gütekriterien von Arbeitsproben

Laut einer viel zitierten Metaanalyse von Schmidt und Hunter [28, S. 265] weisen Arbeitsproben im Vergleich zu Personalauswahlverfahren – darunter Intelligenz- und Persönlichkeitstests, Einstellungsinterviews, Assessment-Center etc. – die höchste prädiktive Validität (.54) auf. Demnach eignen sie sich am besten, um den zukünftigen Berufserfolg von Bewerbern vorherzusagen. Allerdings kommen Roth et al. [26, S. 1020] in einer jüngeren Metaanalyse auf einen deutlich geringeren Wert (.33 bei einfacher Minderungskorrektur) und weisen auf methodische Mängel in der der Metaanalyse zugrunde liegenden Studie von Hunter und Hunter [12] hin. Insgesamt betrachtet stellt aber auch der ermittelte Wert von .33 im Hinblick auf die prädiktive Validität des Verfahrens ein weiterhin zufriedenstellendes Ergebnis dar. Ferner spricht auch die hohe soziale Validität bzw. die Bewerberakzeptanz für den Einsatz von Arbeitsproben. In dieser Hinsicht können beispielsweise Intelligenztests nicht mithalten. Diese werden zwar auch mit Werten zwischen .39 bis .51 als mittel bis hoch valide eingestuft [28, S. 265, 267, 26, S. 1027], leiden aber unter ihrer mangelnden Akzeptanz [29, S. 379 ff.]. In Bezug auf die Reliabilität von Arbeitsproben, konnten Jackson et al. [13, S. 31] anhand der Evaluation einer selbst entwickelten standardisierten Arbeitsprobe Werte von .87 bei der Inter-Rater-Reliabilität und .88 bei der Retest-Reliabilität ermitteln.

Gemäß der Metaanalyse von Roth et al. [26, S. 1025] korrelieren Arbeitsproben mit Intelligenztests zu .32 (unkorrigiert); nach Schmidt und Hunter [28, S. 1011 ff.] zu .38. Demzufolge messen Proben auch zu einem hohen Anteil die kognitive Leistungsfähigkeit. Die inkrementelle Validität von Arbeitsproben gegenüber Intelligenztests wurde mit .06 [26, S. 1027] bzw. .12 [28, S. 267] ermittelt. Letzterer Wert entspricht einem Validitätszuwachs von 24 % im Falle der Kombination beider Auswahlverfahren. Dieses Ergebnis stimmt mit der Erkenntnis von Schmidt und Hunter [28, S. 266] überein, dass die Höhe der inkrementellen Validität sowohl von der Validität der kombinierten Verfahren als auch von deren Korrelation abhängig ist. Je kleiner deren Zusammenhang ist, umso mehr nimmt die Gesamtvalidität zu. Im Gegensatz zur Arbeitsproben-Intelligenztest-Beziehung liegt der ermittelte Wert für den Zusammenhang von Arbeitsproben und

Situational-Judgment-Tests lediglich bei 13, wonach es sich noch mehr anbietet, zusätzlich zu einer Arbeitsprobe einen Situational-Judgment-Test durchzuführen [26, S. 1029].

### 8.1.3 Ablauf der Konstruktion von Arbeitsproben

Jeder Arbeitsproben-Entwicklung geht eine umfangreiche Anforderungs- und Tätigkeitsanalyse voraus, wobei nicht alle Aufgabenbereiche von der Probe erfasst werden müssen [9, S. 468, 10, S. 243]. Da sich die Abdeckung des gesamten Tätigkeitsspektrums mittels einer einzigen Arbeitsprobe zumeist als schwierig bis unmöglich erweist, gilt es, vorrangig jene Arbeitsinhalte abzubilden, welche die zukünftige Berufstätigkeit am besten repräsentieren [4, S. 251]. Üblicherweise erfassen Arbeitsproben Aufgabengebiete, die auch im beruflichen Alltag regelmäßig anfallen. Will man jedoch prüfen, ob ein Bewerber auch mit ungewöhnlichen Arbeitssituationen umgehen kann, so ist es ratsam, auch selten anfallende Aufgaben in die Arbeitsprobe aufzunehmen. Hinsichtlich der angewandten Instrumente empfehlen Guion und Highhouse [10, S. 243] nur jene Materialien und Werkzeuge in der Probe einzusetzen, die auch aktuell im Arbeitsalltag zum Einsatz kommen. Jedenfalls sollte im Vorfeld der Personalauswahl eine Pilotstudie durchgeführt werden, um die neu entwickelte Arbeitsprobe und ihre Komponenten (Instruktionen, Ratingskalen, Aufgabeninhalte etc.) auf Plausibilität, Validität und Reliabilität zu evaluieren [10, S. 243].

Auch Roth et al. [26, S. 1011] und Görlich [7, S. 468] weisen in Bezug auf Arbeitsproben auf die Wichtigkeit der anforderungsbezogenen Konstruktion und standardisierter Durchführung hin, um die Validität und Reliabilität des Verfahrens zu gewährleisten und einen objektiven Vergleich der Kandidaten zu ermöglichen. Bestenfalls sollten die simulierten Inhalte der Auswahlmethodik 1:1 mit den realen beruflichen Aufgaben korrespondieren (Inhaltsvalidität) und im gleichen Ausmaß mit der zu prognostizierenden beruflichen Leistung (Kriteriumsvalidität). Um die Inhaltsvalidität des Verfahrens sicherzustellen, ist es notwendig, arbeitsrelevante, erfolgskritische Fähigkeiten, Fertigkeiten und Verhaltensweisen von irrelevanten abzugrenzen und nur erstere zu erfassen. Letztendlich sollte es möglich sein, vom ermittelten Leistungsniveau eines Kandidaten gleichermaßen auf dessen zukünftigen beruflichen Erfolg schließen zu können. Die hohe Anforderungsbezogenheit des Verfahrens wirkt sich wiederum positiv auf die wahrgenommene Fairness und soziale Validität des Verfahrens aus [4, S. 251 f., 9, S. 468].

Im Hinblick auf die Leistungsevaluation können sowohl numerische als auch nicht-numerische Bewertungsskalen wie Checklisten, Globalurteile („zufrieden“ vs. „nicht zufrieden“) oder Punktevergabesysteme zum Einsatz kommen. Die ausgewählte Evaluationsmethode ist letztlich vom Untersuchungsdesign der jeweiligen Arbeitsprobe abhängig, wobei es sich hierbei unbedingt um geprüfte Skalen handeln sollte [9, S. 471]. Sofern möglich, sollte die Leistungseinschätzung von mehr als nur einer Person vorgenommen werden, um die Ergebnisse untereinander vergleichbar zu machen und klassische Beurteilungsfehler und Wahrnehmungsverzerrungen zu vermeiden [10, S. 244].

Zudem sollten die beurteilenden Personen hinsichtlich ihres Fachwissens ausreichend qualifiziert und mit den angewandten Evaluationsmethoden vertraut sein [14, S. 218]. Die Beachtung all dieser Regeln bei der Konstruktion von Arbeitsproben kann zu deutlich höheren Validitäten beitragen [9, S. 471].

#### **8.1.4 Vor- und Nachteile von Arbeitsproben und zukünftiger Forschungsbedarf**

Die Vorteile, welche sich durch den Einsatz von Arbeitsproben in Personalauswahlprozessen ergeben, lassen sich nicht nur auf der Unternehmensseite sondern auch auf der Bewerberseite finden. Zum einen profitieren Unternehmen nicht nur dadurch, dass sie Informationen über das konkrete Arbeits- und Leistungsverhalten ihrer Bewerber in arbeitsplatzrelevanten Situationen gewinnen, sondern auch durch die hohe prädiktive und soziale Validität dieser Verfahren [14, S. 215 f.]. Die hohe Akzeptanz resultiert nicht zuletzt aufgrund der wahrgenommenen Fairness (Information, Transparenz, Situationskontrolle, Feedback) und Arbeitsbezogenheit der Auswahlmethodik [25, S. 179 f., 29, S. 370 ff.]. Ferner wirken sich Arbeitsproben verglichen mit anderen Auswahlverfahren, wie z. B. Intelligenztests, kaum nachteilig auf Minoritäten aus („adverse impact“) bzw. unterscheidet sich die im Rahmen von Arbeitsproben erfasste Leistung innerhalb verschiedener ethnischer Gruppierungen nur marginal voneinander [4, S. 248, 25, S. 178 f.].

Des Weiteren weisen Arbeitsproben im Gegensatz zu anderen Personalauswahlverfahren den höchsten Realitätsbezug im Hinblick auf die Stimulus- und Response-Komponente auf. Während die Stimulus-Komponente auf das Untersuchungsmaterial, mit welchem ein Bewerber während des Auswahlprozesses konfrontiert wird, verweist, bezieht sich die Response-Komponente auf die damit einhergehenden Reaktionsmöglichkeiten des Bewerbers. Je stärker beide Komponenten den beruflichen Alltag widerspiegeln, desto höher ist deren Realitätsbezug [14, S. 216 f.].

##### **Beispiel**

Eine Arbeitsprobe beinhaltet die Anfertigung eines bestimmten Werkstückes. Eine solche Stimulus-Komponente spiegelt die Realität 1:1 wider. Gleichzeitig weist eine solche Arbeitsprobe einen hohen Realitätsbezug im Hinblick auf die Response-Komponente auf, da der Bewerber im Zuge der Anfertigung des Werkstückes zu 100 % reagieren und sein wahres handwerkliches Können vorführen kann. Der Realitätsbezug wäre weitaus geringer, würde es sich seitens des Stimulus lediglich um eine Fragestellung („Wie würden Sie das vor Ihnen liegende Werkstück anfertigen?“) mit daraus resultierender mündlicher Beschreibung oder der Wahl zwischen mehreren Antwortalternativen handeln [14, S. 216 f.].

Bewerber können von einer Arbeitsprobe aus mehreren Gründen profitieren: Abgesehen davon, dass sie einen ersten Einblick in ihr zukünftiges Arbeitsgebiet und den damit

verbundenen Aufgaben erhalten, können sie im Zuge dessen auch abschätzen, inwiefern ihnen die Aufgaben liegen bzw. mit ihren eigenen Fähigkeiten und Interessen korrespondieren. Dank des erhaltenen Feedbacks bekommen sie zusätzlich die Möglichkeit, ihre eigenen Stärken und Schwächen zu reflektieren, was wiederum ihrer beruflichen Orientierung und Weiterbildung zugutekommen kann. Demzufolge begünstigen Arbeitsproben auch die Selbstbeurteilung und Selbstselektion und tragen folglich zur Fluktuationsreduktion bei [8, 14, 29].

Trotz all ihrer Vorteile, bergen Arbeitsproben auch Nachteile. Schmidt und Hunter [28] heben zwar in ihrer Metaanalyse die hohe prädiktive Validität von Arbeitsproben hervor, weisen gleichzeitig aber auch darauf hin, dass sie – wie alle anderen Personalauswahlverfahren auch – idealerweise ergänzend zu Intelligenztests eingesetzt werden sollten. Als Begründung führen sie einige Limitationen an, die der Arbeitsprobe zugrunde liegen.

In den meisten Fällen stellen Arbeitsproben kein geeignetes Mittel dar, um die Leistung von Bewerbern ohne Berufserfahrung zu erfassen. Zudem kann deren Entwicklung und Durchführung ein sehr komplexes, kostspieliges und zeitaufwendiges Unterfangen darstellen. Bestenfalls sollten Arbeitsproben auf den jeweiligen Arbeitsplatz zugeschnitten sein, was wiederum mit einer spezifischen Entwicklung und Validierung einhergeht. Es gilt, Kosten und Nutzen sorgfältig gegeneinander abzuwägen [4, S. 249]. Hinsichtlich der Validität von Arbeitsproben raten Kanning und Schuler [14] davon ab, metaanalytische Resultate auf konkrete Arbeitsproben zu beziehen – nicht zuletzt aufgrund ihrer Heterogenität bzw. vielfältigen Gestaltungsmöglichkeiten. Vielmehr empfehlen sie, die Validität einer bestimmten Arbeitsprobe vor Ort empirisch zu überprüfen, was allerdings in der Praxis aufgrund begrenzter finanzieller und zeitlicher Mittel sowie zu kleiner Stichproben oftmals nicht möglich erscheint [14, S. 219].

Zur besseren Übersicht stellt nachstehende Tabelle (Tab. 8.1) die Vor- und Nachteile von Arbeitsproben gegenüber.

Obwohl Arbeitsproben zu den validesten Personalauswahlinstrumenten zählen, werden sie international noch wenig eingesetzt. Ein Hauptgrund hierfür liegt womöglich im hohen

**Tab. 8.1** Vor- und Nachteile von Arbeitsproben. (Quelle: eigene Darstellung)

Vorteile	Nachteile
Mittlere bis hohe prädiktive Validität	Hoher Konstruktions- und Entwicklungsaufwand
Hohe Inhaltsvalidität, belegte inkrementelle Validität	Hoher Durchführungsaufwand
Hoher Realitätsbezug	Hoher Heterogenitätsgrad
Hohe Bewerberakzeptanz (soziale Validität)	Erforderliche, aber oftmals schwer mögliche empirische Validierung einer konkreten Arbeitsprobe
Geringe Benachteiligung unterschiedlicher Gruppen und Minoritäten	Berufserfahrung als notwendige Voraussetzung der Anwendung
Unterstützung der Selbstbeurteilung und Selbstselektion von Bewerbern	



Entwicklungs-, Konstruktions- und Durchführungsaufwand [15, S. 34]. Aus wissenschaftlicher Sicht bleibt fraglich, welche Faktoren Arbeitsproben nun tatsächlich erfassen und auf welche Weise sie dies tun. Zwar wird angenommen, dass Arbeitsproben den Einfluss erfolgsabhängiger Faktoren wie Intelligenz, Berufserfahrung, Fachwissen und Motivation medieren, allerdings ist deren Beitrag von der jeweiligen Gestaltungsform der Arbeitsprobe abhängig und aufgrund dieser Komplexität nur schwer zu erfassen [4, S. 251]. Zusätzliche Studien würden helfen, Arbeitsproben verstärkt hinsichtlich ihrer multidimensionalen Konstrukte und prädiktiven Validität zu eruieren [26, S. 1031]. Davon könnten insbesondere metaanalytische Untersuchungen profitieren. Welchen Einfluss ein Minimum an beruflicher Qualifikation (z. B. sieben Jahre Berufserfahrung) auf die Ergebnisse einer Arbeitsprobe nehmen kann, gilt es ebenso genauer zu erforschen. Zugleich empfiehlt es sich, den zukünftigen Forschungsfokus verstärkt auf die Korrelation von Arbeitsproben mit anderen Prädiktoren der beruflichen Leistung zu legen. Ferner sollten Arbeitsproben hinsichtlich interkultureller Unterschiede analysiert werden [27, S. 656 f.].

---

## 8.2 Situational-Judgment-Test

Diese Tests stellen ein beliebtes Verfahren der Personalauswahl dar [18, S. 730, 33, S. 188]. McDaniel und Nguyen [18, S. 103] definieren Situational-Judgment-Tests (SJT) als „assessments designed to measure judgment in work settings“. Demzufolge verstehen die Autoren unter einem SJT einen Test zur Messung berufs- und arbeitsplatzrelevanter Situationsbeurteilungen. Ein SJT umfasst mehrere Items, die sich wiederum aus einer erfolgsrelevanten Situationsbeschreibung und mehreren vorgegebenen Antwortalternativen zusammensetzen. Die Darstellung der Items kann schriftlich oder digital (z. B. Videoclip) erfolgen. Die Situationsbeurteilung ist abhängig von der eingeschätzten Wirksamkeit oder der eigenen Verhaltenstendenz [2, S. 483, 10, S. 245, 14, S. 222 f.].

► Unter einem SJT versteht man ein Messinstrument zur Erfassung berufs- und arbeitsplatzrelevanter Situationsbeurteilungen [18, S. 103].

Die SJT stellen Mini-Szenarios des beruflichen Alltags dar, die einer mentalen Arbeitsprobe gleichkommen. Der größte Unterschied zur Arbeitsprobe besteht darin, dass die Situation nicht real ist, sondern lediglich in schriftlicher oder digitaler Form vorliegt. Demnach spiegeln die Antworten der Kandidaten nur Verhaltensintentionen bzw. Verhaltenssimulationen wider. Beiden Verfahren gemeinsam ist, dass sie auf der Annahme basieren, zukünftiges Verhalten ließe sich am besten über vergangenes prognostizieren [2, S. 485]. Häufig werden SJT auch mit ihrem verwandteren Konzept, dem situativen Interview, gleichgesetzt. Zwar zielen beide Verfahren auf die Messung unterschiedlicher Konstrukte ab, unterscheiden sich jedoch grundsätzlich in ihrer jeweiligen Vorgehensweise. Im Gegensatz zu SJT finden situative Interviews in mündlicher Form statt und stellen zudem keine Antwortalternativen zur Verfügung [14, S. 228, 33, S. 190].

### 8.2.1 Ziele und Typen von Situational-Judgment-Tests

In einem SJT werden Kandidaten mit mehreren berufs- und arbeitsplatzrelevanten Situationen konfrontiert, mit dem Ziel, deren Urteil im Hinblick auf jede einzelne beschriebene Situation zu erfassen [33, S. 188]. Ausgehend vom ermittelten Antwort-Score, der mit einer Referenzlösung verglichen wird, wird anschließend auf das zukünftige Leistungsverhalten des Bewerbers geschlossen [2, S. 483]. Im Gegensatz zu anderen Personalauswahlverfahren, wie beispielsweise Intelligenztests, zielen SJT auf die Messung mehrerer unterschiedlicher Konstrukte, wobei diese unterschiedlich gewichtet sein können. Die Debatte darüber, ob es sich bei einem SJT um ein Testverfahren für praktische Intelligenz oder um eine Methodik zur Erfassung multipler Konstrukte handelt, ist jedoch bis heute nicht abgeklungen [18, S. 106 f., 33, S. 190]. Nachstehend (Tab. 8.2) wird ein Item aus dem „Situational-Judgment-Test on Personal Initiative“ nach Bledow und Frese [3, S. 6] illustriert.

#### Beispiel

**Tab. 8.2** SJT-Item aus dem „SJT on Personal Initiative“ [3, S. 6]

Sie arbeiten in einem Großraumbüro. Die Arbeitsplätze sind sehr ungünstig organisiert. Es ist sehr eng, da Sie eine Menge auf Ihrem Schreibtisch unterbringen müssen. Außerdem müssen Sie lange Wege zurücklegen. In einigen Monaten werden Sie eine andere Stelle in Ihrem Unternehmen übernehmen. Damit wird sich das Problem für Sie von selbst lösen. Wie würden Sie sich verhalten?

Trifft am wenigsten zu		Trifft am ehesten zu
0	Ich verändere die Organisation auf meinem Schreibtisch, damit ich mehr Dinge auf meinem Schreibtisch unterbringen kann	0
1	Bis zu meinem Arbeitsplatzwechsel kann ich mit der gegenwärtigen Situation gut auskommen	-1
0	Ich mache meinen Chef darauf aufmerksam, dass etwas verändert werden sollte	0
-1	Ich stelle die Möbel mit meinen Kollegen um, damit ich mehr Platz zur Verfügung habe	1

Motowidlo et al. [20, S. 640] ordnen Situational-Judgment-Tests aufgrund ihres geringen Realitätsbezuges in die Kategorie der sogenannten „low fidelity simulations“ ein – anders als Arbeitsproben, die zu den „high fidelity simulations“ gezählt werden. Ferner lassen sich im Arbeitskontext entwickelte SJT-Items hinsichtlich ihrer fokussierten Inhalte differenzieren. Behrmann [2, S. 486] zählt folgende häufig vorliegende Themenschwerpunkte von SJT auf:

- Interaktionsprobleme mit Vorgesetzten, Kollegen, Mitarbeitern etc.,
- Herausfordernde Aufgaben (schwierige/unbeliebte Aufgaben, veränderte Aufgabenbereiche etc.),

- Kontextbezogenes Verhalten,
- Unvorhergesehene Ereignisse oder Entscheidungsdilemmata,
- Technische oder technologische Probleme.

Abgesehen von der inhaltlichen Differenzierung der Items können SJT auch hinsichtlich folgender Kriterien unterschieden werden [2, S. 486, 14, S. 223]:

- Anzahl, Heterogenitätsgrad und Interaktivität/Verknüpfungsgrad der Items,
- Umfang, Verständlichkeit und Art (schriftlich vs. digital) der Situationsbeschreibungen,
- Anzahl, Verständlichkeit und Art (schriftlich vs. digital) der Antworten,
- Antwortinstruktion (Wahl der wahrscheinlichsten vs. unwahrscheinlichsten Alternative; Wahl der besten vs. schlechtesten Alternative; Ranking/Rating der Alternativen; Would-do-Instruktion vs. Should-do-Instruktion etc.),
- Bewertungsschlüssel (Optimalantwort-Keying & Scoring-Strategien).

### **8.2.2 Ablauf der Konstruktion von Situational-Judgment-Tests**

Wie bei allen simulationsorientierten Verfahren setzt auch die Konstruktion eines SJT eine umfangreiche berufs- und arbeitsplatzspezifische Anforderungsanalyse (s. hierzu Kapitel von Nerding in diesem Band) voraus. Es gilt, erfolgskritische Schlüsselsituationen zu erfassen, die im Zusammenhang mit der angestrebten beruflichen Position stehen [17, S. 429, 20, S. 641]. Für diesen Vorgang bietet sich insbesondere die „Critical Incident Technique“ an, bei der Arbeitsplatzexperten erfolgsrelevante, berufsspezifische Situationen schildern sowie positive und negative Verhaltensmöglichkeiten aufzeigen. Die Entwicklung eines SJT kann sich aber auch auf neuere Ansätze wie rational-deduktive oder konstruktorientierte Methoden stützen. Ferner können Items auch interaktiv gestaltet werden, indem sie auf logische Weise miteinander verknüpft werden bzw. sich eine Situation im nächsten Item fortsetzt. Liegen die einzelnen Situationsbeschreibungen in schriftlicher oder digitaler Form vor, folgt die sorgfältige Herausarbeitung der Handlungs- und Antwortalternativen, wobei sich deren Anzahl erfahrungsgemäß auf maximal fünf beschränkt. Anzumerken ist, dass der Realitätsbezug des Verfahrens, jedoch auch dessen Konstruktionsaufwand, mit der Anzahl der Antwortalternativen steigt [2, S. 485 f., 14, S. 223 ff.]. Ferner sind die Testentwickler gefordert, sich Gedanken über die Antwortinstruktion zu machen, die auf zweierlei Art formuliert werden kann und die Einschätzung der vorgegebenen Antwortalternativen beeinflusst. Entweder fordert man die Kandidaten dazu auf, anzugeben, wie sie sich in einer bestimmten Situation idealerweise verhalten sollten („Should-do-Instruktion“) oder man instruiert sie, aufzuzeigen, wie sie sich hierbei selbst verhalten würden („Would-do-Instruktion“). Die Beantwortung dieser Fragestellung hängt letztlich davon ab, ob vorrangig kognitive Fähigkeiten oder Persönlichkeitsmerkmale erfasst werden sollen [5, S. 223, 10, S. 245, 14, S. 223 f.]. Untersuchungen zufolge [18, 21] korrelieren „Should-do-Instruktionen“ stärker mit der

Intelligenz von Testpersonen, während „Would-do-Instruktionen“ mehr mit Persönlichkeitsmerkmalen zusammenhängen und folglich aussagekräftiger hinsichtlich des zukünftigen Verhaltens sind, wobei sie aber auch eine erhöhte Manipulationsanfälligkeit aufweisen. Nachdem die Antwortinstruktion festgelegt worden ist, steht die Entscheidung über die Bearbeitung der Alternativen an. Am einfachsten gestaltet es sich für die Kandidaten, wenn sie sich lediglich für eine der vorgegebenen Antwortalternativen entscheiden oder die beste vs. schlechteste Antwort auswählen müssen. Schwieriger wird es für sie, wenn sie die Antworten reihen („ranking“) oder auf einer mehrstufigen Skala einschätzen („rating“) sollen. Die Auswahl der Antwortalternativen erfolgt nahezu immer schriftlich, obwohl auch verschiedene Videoclips oder Audiodateien als Antworten denkbar wären [5, S. 220, 14, S. 223 f.]. Zu guter Letzt folgt die Festlegung des Bewertungsschlüssels („scoring keys“), indem die vorgegebenen Verhaltensalternativen im Sinne der Ausarbeitung einer Referenzlösung hinsichtlich ihrer Antwortqualität (z. B. beste vs. schlechteste Antwort) beurteilt, eingestuft oder bewertet werden. Für die Auswertung eines SJT können einfache Punktschichten für richtig erachtete Antworten, Minuspunkte für falsch erachtete Antworten oder komplexe Berechnungsmethoden wie die Rangreihung der Alternativen durch Antwortprofilunterschiede herangezogen werden [2, S. 485 f.]. Wie erkennbar, bedarf es für die Leistungsbeurteilung keiner direkten Beobachtung durch geschulte Assessoren [17, S. 428]. Eine vollständige Testkonstruktion schließt allerdings empirische Voruntersuchungen ein, um die Aufgabenkomplexität und Reliabilität des Verfahrens zu überprüfen [14, S. 225].

### 8.2.3 Gütekriterien von Situational-Judgment-Tests

Dank ihres Arbeitsplatzbezuges wird einem SJT Inhaltsvalidität zugeschrieben [2, S. 487, 5, S. 222]. Ebenso zeichnen sich die Tests durch eine hohe soziale Validität aus bzw. werden von den Testpersonen positiv wahrgenommen [2, S. 488, 14, S. 227]. In puncto Reliabilität variieren die Werte sehr stark, nicht zuletzt aufgrund des hohen Heterogenitätsgrades der SJT-Konstrukte [5, S. 222, 14, S. 226]. In diesem Fall wird weniger die Berechnung des Cronbachs Alpha als vielmehr die der Retest-Reliabilität empfohlen [14, S. 226, 33, S. 190 f.]. Im Hinblick auf die kriteriumsbezogene (prädiktive) Validität von SJT bringen Untersuchungen durchgängig gute Ergebnisse hervor. Motowidlo et al. [20, S. 640] konnten in ihrer Studie, in der ein SJT zur Auswahl von externen Nachwuchsführungskräften in der Telekommunikationsbranche entwickelt und empirisch überprüft wurde, Korrelationen von SJT mit bestimmten Leistungsindikatoren im Bereich von ,28 bis ,37 ermitteln. Auf einen ähnlichen Wert (,34, korrigiert) kommen auch McDaniel et al. [19, S. 734] in ihrer Metaanalyse, in der die berufliche Leistung als Validitätskriterium herangezogen wurde. Ferner konnte die Forschergruppe um McDaniel über eine Moderatorenanalyse einen Validitätsanstieg im Falle des Vorliegens einer arbeitsplatzbezogenen Tätigkeitsanalyse ermitteln. Weitere Studien, wie jene von Weekly und Jones [30] oder Oswald et al. [22], bestätigen durchgehend den ermittelten Wertebereich der Validitätskoeffizienten von

SJT. In einer nachfolgenden Metaanalyse von Christian et al. [6, S. 105] wurde ferner aufgezeigt, dass mittels videobasierter SJT höhere Validitäten erreicht werden können als mittels herkömmlichen Papier-Bleistift-Tests. Zudem lassen sich mit SJT, welche die interpersonale Team- und Führungsfähigkeit erfassen, die höchsten Werte erzielen.

Dass SJT mit Faktoren wie Intelligenz und bestimmten Persönlichkeitsmerkmalen korrelieren (Konstruktvalidität), wurde bereits hervorgehoben. Die anfängliche Kritik einiger Wissenschaftler, SJT seien simple Intelligenztests, konnte dank empirischer Studien frühzeitig entkräftet werden [5, S. 221]. Zwar korrelieren SJT mit kognitiven Leistungstests – laut Weekly und Jones [30, S. 679] sogar zu ,45 (gewichteter Durchschnittswert), nahezu gleichermaßen (,46) auch nach McDaniel et al. [19, S. 736] – erfassen jedoch nicht vollständig dieselben Konstrukte. Außerdem nimmt der Zusammenhang ab, je weniger kognitiv-basierende Konstrukte (Planungs-, Organisations- und Problemlösungsfähigkeit etc.) und mehr verhaltensbezogene Konstrukte (Team- und Führungsfähigkeit etc.) erhoben werden [5, S. 221].

Was die Korrelation von SJT mit anderen Variablen wie Berufserfahrung oder Persönlichkeitsmerkmalen angeht, liegen folgende Befunde vor: Weekly und Jones [30, S. 678] ermitteln für den Zusammenhang von Berufserfahrung und SJT einen gewichteten Durchschnittswert von ,20; Oswald et al. [22, S. 194] errechnen SJT-Korrelationswerte von ,17 für Extraversion, ,38 für Verträglichkeit, ,28 für Gewissenhaftigkeit, ,17 für emotionale Stabilität und ,21 für Offenheit für neue Erfahrungen. Darüber hinaus werden SJT in der Kombination mit anderen Verfahren, wie kognitive Leistungs- oder Persönlichkeitstests, inkrementelle Validität zugeschrieben. Clevenger et al. [7, S. 413 f.] berichten von einem Validitätszuwachs im Wertebereich von ,016–,026, wenn zusätzlich zu anderen Auswahlverfahren – mitunter Intelligenztests – ein SJT durchgeführt wird [5, S. 221]. Selbiges bestätigt auch die Studie von Patterson et al. [23], die im Rahmen von medizinischen Auswahlprozessen durchgeführt wurde. Hervorzuheben ist, dass die inkrementelle Validität zunimmt, je mehr verhaltensbezogene Konstrukte durch den SJT erfasst werden [33, S. 194]. Was die Kombination von Arbeitsproben mit SJT betrifft, sei auf Abschn. 8.1.3 verwiesen.

#### **8.2.4 Vor- und Nachteile von Situational-Judgment-Tests und zukünftiger Forschungsbedarf**

Die Vorteile von SJT decken sich größtenteils mit denen von Arbeitsproben, wenn auch nicht im gleichen Ausmaß. Neben ihrer belegten Inhaltsvalidität, weisen SJT auch gute Werte in Bezug auf ihre prädiktive und soziale Validität auf. Zudem kann ihre Prognosekraft in Kombination mit anderen Auswahlverfahren wie kognitiven Fähigkeitstests oder Persönlichkeitstests zusätzlich gesteigert werden (inkrementelle Validität). Abgesehen von der guten Validität spricht auch die vergleichsweise geringe Verfälschbarkeit im Sinne sozial erwünschten Antwortverhaltens für den Einsatz von SJT. Wie Arbeitsproben unterstützen auch die SJT Bewerber bei der Selbstbeurteilung und Selbstselektion und wirken sich wenig diskriminierend auf unterschiedliche Gruppen und Minoritäten aus

[2, 14, 17]. Ein entscheidender Vorteil gegenüber Arbeitsproben liegt darin, dass die SJT komplizierte Tätigkeitsbereiche repräsentieren können, die sich mittels Arbeitsproben nicht so einfach simulieren lassen [10, S. 244 f.]. Außerdem sind die SJT den „high fidelity simulations“ aufgrund ihres geringeren Kosten- und Konstruktionsaufwandes überlegen, wenngleich sie auch mit einem bestimmten Aufwand verbunden sind [2, S. 487]. Ferner ermöglichen sie es, mehrere Kandidaten gleichzeitig zu testen [17, S. 434].

Im Gegensatz zu Arbeitsproben weisen SJT einen vergleichsweise geringeren Realitätsbezug hinsichtlich ihrer Stimulus- und Response-Komponente auf, sofern sie in schriftlicher Form absolviert werden [14, S. 216, 17, S. 427]. Was ihre Reliabilität angeht, können SJT aufgrund der Heterogenität ihrer Konstrukte wenig punkten. In den meisten Fällen setzen auch „low fidelity simulations“ eine umfassende arbeitsplatzbezogene Anforderungs- und Tätigkeitsanalyse voraus, die einige Zeit in Anspruch nehmen kann. Darüber hinaus kann die gesamte Konstruktion – vor allem die videobasierte – und die Validierung eines SJT mit einem bestimmten finanziellen Aufwand einhergehen. Zudem liegt die Vermutung nahe, dass SJT-Ergebnisse über spezielle Übung oder Vorbereitung (coaching) manipuliert werden können. Außerdem besteht in der Wissenschaft noch immer wenig Klarheit darüber, was SJT letztlich messen – Intelligenz, Persönlichkeit oder Wissen? [5, S. 222, 17, S. 429 ff.].

Nachstehende Übersicht (Tab. 8.3) stellt die einzelnen Vor- und Nachteile von SJT gegenüber.

Wenngleich Situational-Judgment-Tests gute Prädiktoren der beruflichen Leistung mit aktuellem Trend zum vermehrten Einsatz darstellen, besteht noch in vielerlei Hinsicht erheblicher Forschungsbedarf. Um die Qualität und Aussagekraft von SJT zukünftig zu steigern, ist es notwendig, deren theoretische Fundierung auszuweiten und sie im Hinblick auf ihre Konstrukte, Validität und Reliabilität eingehender zu erforschen.

**Tab. 8.3** Vor- und Nachteile von SJT. (Quelle: eigene Darstellung)

Vorteile	Nachteile
Mittelhohe prädiktive Validität	Geringe Reliabilität
Belegte Inhaltsvalidität und inkrementelle Validität	Hoher Heterogenitätsgrad
Hohe Bewerberakzeptanz (soziale Validität)	Geringer Realitätsbezug im Falle schriftlicher SJT
Geringe Benachteiligung unterschiedlicher Gruppen und Minoritäten	Verfälschbarkeit der Testergebnisse über spezielle Vorbereitungsmaßnahmen (coaching)
Unterstützung der Selbstbeurteilung und -selektion von Bewerbern	Unklarheit bezüglich der konkreten Messziele von SJT (Wissen, Intelligenz, Persönlichkeit?)
Geringer Kosten- und Konstruktionsaufwand im Falle schriftlicher SJT	Erhöhter Kosten- und Konstruktionsaufwand im Falle digitaler SJT
Möglichkeit der Repräsentation komplizierter Tätigkeitsbereiche und der simultanen Überprüfung mehrerer Probanden	

Insbesondere gilt es, die Frage zu beantworten, welche multidimensionalen Konstrukte die SJT letztlich messen und wie man es schafft, nur diejenigen zu erfassen, die von Interesse sind. Voraussetzung hierfür ist die korrekte Unterscheidung zwischen Messmethodik (z. B. Arbeitsprobe, SJT etc.) und Konstrukt (z. B. Wissen, Persönlichkeitsmerkmale etc.). Zudem sollten die SJT auf moderierende Einflussfaktoren genauer untersucht werden [2, S. 488, 6, S. 107]. Bislang konnten die Antwortinstruktion („Should-do-Instruktion“ vs. „Would-do-Instruktion“) und Verfahrensweise (schriftlich vs. digital) als Determinanten von SJT identifiziert werden. Zukünftig gilt es, nach weiteren potenziellen Einflussfaktoren zu forschen – wie beispielsweise die Länge und Spezifität der Items, Scoring-Strategien, Motive, Berufserfahrung der Probanden [5, S. 239, 17, S. 437]. Weekly und Ployhart [31, S. 100] weisen weiter darauf hin, dass es zukünftig wichtig sei, die Korrelationen von SJT über verschiedene Berufe und Organisationen hinweg zu untersuchen, um Variationen in der kognitiven Ladung von SJT besser zu erklären. Laut Krumm et al. [16, S. 413] sollte auch der Einfluss kontext-ausgerichteter vs. nicht kontext-ausgerichteter SJT auf Ergebnisse, Bewerberwahrnehmung und Validität genauer analysiert werden. Ferner sollte die Forschung ihren Schwerpunkt auf die Entwicklung und Validierung neuer SJT-Formate legen.

---

**Beispiel**

Ein innovatives Antwort-Format könnte beispielsweise so aussehen, dass Kandidaten nach der Präsentation einer video-basierten Situation ihre Antwort selbst vorführen müssen, während sie gefilmt werden [17, S. 437].

Weiterer Forschungsbedarf besteht im Hinblick auf die Analyse von SJT hinsichtlich interkultureller Unterschiede und subgruppenspezifischen Leistungsunterschieden [18, S. 437]. Auch sollte vermehrt nach Möglichkeiten geforscht werden, um sozial erwünschtem Antwortverhalten bzw. der Manipulierbarkeit von SJT entgegenzuwirken [5, S. 232 ff., 33, S. 199].

---

**8.3 Fazit**

Sowohl Arbeitsproben als auch SJT stellen valide, gut akzeptierte und faire Verfahren der Personalauswahl dar, welche zur Selbstselektion und Fluktuationsreduktion beitragen können. Beide Instrumente zielen auf die Vorhersage des zukünftigen Leistungsverhaltens von Kandidaten über die Simulation zukünftiger berufsbezogener Tätigkeitsbereiche ab, wobei deren Prognosegüte in Kombination miteinander oder mit anderen Verfahren zusätzlich gesteigert werden kann. Im Gegensatz zu Arbeitsproben setzen SJT eine weniger aufwendige, anforderungsbezogene Konstruktion, Durchführung und Evaluation voraus, leiden aber oftmals unter ihrem geringeren Realitätsbezug und der Manipulierbarkeit ihrer Ergebnisse. Ob und wie viel Berufserfahrung als notwendige Voraussetzung zur erfolgreichen Durchführung beider Verfahren erforderlich ist, muss im Speziellen eruiert werden. Im Hinblick auf den zukünftigen Forschungsbedarf lässt sich erkennen,



dass weder die Arbeitsprobe noch der SJT hinreichend hinsichtlich ihrer Konstrukte analysiert sind. Zudem werden beide Verfahren im Gegensatz zu anderen Personalauswahlverfahren noch verhältnismäßig wenig eingesetzt, allerdings wird ein deutliches Wachstum prognostiziert.

### Fragen

1. Was versteht man unter einer Arbeitsprobe und was unter einem Situational-Judgment-Test?
2. Welche Funktion erfüllen beide Verfahren bei der Auswahl von Mitarbeitern und wie werden sie entwickelt?
3. Worin bestehen Gemeinsamkeiten und Unterschiede zwischen Arbeitsproben und SJT?
4. Wo liegen deren jeweilige Vor- und Nachteile im Hinblick auf Testgütekriterien, subgruppenspezifische Leistungsunterschiede und Verfälschbarkeit der Ergebnisse?

### Literatur

1. Asher, J. J., & Sciarrino, J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology, 27*, 519–533.
2. Behrmann, M. (2007). Situational-Judgment-Tests. In H. Schuler & K. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 484–489). Göttingen: Hogrefe.
3. Bledow, R., & Frese, M. (2005). *Situational-Judgment-Test on personal initiative*. Gießen: Justus-Liebig-Universität.
4. Callinan, M., & Robertson, I. T. (2000). Work sample testing. *International Journal of Selection and Assessment, 8*(4), 248–260.
5. Chan, D., & Schmitt, N. (2005). Situational-judgment-test. In A. Evers, N. Anderson, & O. Smit-Voskuil (Hrsg.), *Handbook of personnel selection* (S. 219–246). London: Blackwell.
6. Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational-Judgment-Tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63*, 83–117.
7. Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Schmidt Harvey, V. (2001). Incremental validity of Situational-Judgment-Tests. *Journal of Applied Psychology, 86*(3), 410–417.
8. Downs, S., Farr, R. M., & Colbeck, L. (1978). Self-appraisal: A convergence of selection and guidance. *Journal of Occupational Psychology, 51*, 271–278.
9. Görlich, Y. (2007). Arbeitsproben. In H. Schuler & K. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 468–474). Göttingen: Hogrefe.
10. Guion, R. M., & Highhouse, S. (2006). *Essentials of personnel assessment and selection*. New York: Lawrence Erlbaum Associates.
11. Hedge, J. W., & Teachout, M. S. (1992). An interview approach to work sample criterion measurement. *Journal of Applied Psychology, 77*(4), 453–461.
12. Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72–98.
13. Jackson, D. N., Harris, W. G., Ashton, M. C., McCarthy, J. M., & Tremblay, P. F. (2000). How useful are work samples in validation studies? *International Journal of Selection and Assessment, 8*(1), 29–33.

14. Kanning, U. P., & Schuler, H. (2014). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 215–256). Göttingen: Hogrefe.
15. Krause, D. E. (2011). *Trends in der internationalen Personalauswahl. Praxis der Personalpsychologie*. Göttingen: Hogrefe.
16. Krumm, S., Hüffmeier, J., Lievens, F., Lipnevich, A. A., Bendels, H., & Hertel, G. (2015). How „situational“ is judgment in Situational Judgment Tests? *Journal of Applied Psychology, 100*(2), 399–416.
17. Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational Judgment Tests: A review of recent research. *Personnel Review, 37*(4), 426–441.
18. McDaniel, M. A., & Nguyen, N. T. (2001). Situational Judgment Tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*(1/2), 103–113.
19. McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of Situational Judgment Test to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*(4), 730–740.
20. Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*(6), 640–647.
21. Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a Situational Judgment Test. *International Journal of Selection and Assessment, 13*(4), 250–260.
22. Oswald, F. L., Schmitt, N., Kim, B. H., Gillespie, M. A., & Ramsay, L. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89*(2), 187–207.
23. Patterson, F., Lievens, F., Kerrin, M., Zibarras, L., & Carette, B. (2012). Designing selection systems for medicine: The importance of balancing predictive and political validity in high-stakes selection contexts. *International Journal of Selection and Assessment, 20*(4), 486–496.
24. Ployhart, R. E., Schneider, B., & Schmitt, N. (2006). *Staffing organizations: Contemporary practice and research*. New York: Lawrence Erlbaum Associates.
25. Robertson, I. T., & Kandola, R. S. (1982). Work sample tests: Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology, 55*, 171–183.
26. Roth, P. L., Bobko, P., & McFarland, L. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology, 58*, 1009–1037.
27. Roth, P. L., Bobko, P., McFarland, L., & Buster, M. (2008). Work sample tests in personnel selection: A meta-analysis of black-white differences in overall and exercise scores. *Personnel Psychology, 61*, 637–662.
28. Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262–274.
29. Schuler, H. (2014). *Psychologische Personalauswahl. Eignungsdiagnostik für Personalentscheidungen und Berufsberatung*. Göttingen: Hogrefe.
30. Weekly, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*, 679–700.
31. Weekly, J. A., & Ployhart, R. E. (2005). Situational judgment: Antecedents and relationships with performance. *Human Performance, 18*(1), 81–104.
32. Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*(5), 372–376.
33. Whetzel, D. L., & McDaniel, M. A. (2009). Situational-Judgment-Tests: An overview of current research. *Human Resource Management Review, 19*, 188–202.

---

## Zusammenfassung

Dieses Kapitel verfolgt das Ziel, eine zusammenfassende Übersicht zum AC zu vermitteln. Das AC wird definiert und Ziele (Personalauswahl, Personalentwicklung) sowie allgemeine Ablaufphasen (Analyse, Design, Ausführung, Evaluation) werden beschrieben. Ferner werden Anforderungsdimensionen, die häufig Anwendung finden, charakterisiert und geschildert, welche Probleme mit der Art und Anzahl der verwendeten Anforderungsdimensionen im AC verbunden sind. Ein weiterer Teil des Kapitels widmet sich den im AC eingesetzten Übungen und Simulationen (z. B. Präsentationen, Postkorb, Fallstudie, führerlose Gruppendiskussion) sowie der Frage, inwiefern die Kombination von AC mit weiteren Diagnosemethoden (z. B. psychometrischen Testverfahren) von Vorteil sein kann. Schließlich legt das Kapitel dar, wie sich der Beobachterpool im AC konstituiert. Es erfolgt eine Beschreibung der Beobachtungsmethoden und der Inhalte von Beobachtertrainings. Der abschließende Teil des Kapitels widmet sich ausgewählten psychometrischen Qualitäten des AC, nämlich metaanalytischen Befunden zur prädiktiven Validität und Maßnahmen ihrer Verbesserung sowie empirischen Ergebnissen zur inkrementellen Validität des AC.

---

D.E. Krause (✉)  
Berlin, Deutschland  
E-Mail: [diana.eva.krause@gmail.com](mailto:diana.eva.krause@gmail.com)

## Inhaltsverzeichnis

9.1 AC: Ziele, Definition, Beschreibung .....	198
9.2 Anforderungsdimensionen .....	199
9.3 Übungen, Simulationen und zusätzliche Diagnoseinstrumente .....	203
9.4 Beobachter und Beobachtertraining .....	206
9.5 Prädiktive und inkrementelle Validität .....	209
9.6 Fazit .....	211
Literatur .....	211

---

### 9.1 AC: Ziele, Definition, Beschreibung

In der letzten Dekade hat sich unsere Arbeitswelt in mehrfacher Hinsicht gewandelt. Dies betrifft z. B.

- die Zusammensetzung des Arbeitsmarktes in Bezug auf Diversity; ein gestiegener Anteil von Minoritäten im Arbeitsleben;
- die Arbeitsgestaltung;
- die Arbeitsanforderungen sowie die Anforderungen in Bezug auf kognitive Fähigkeiten und soziale Kompetenz von Arbeitnehmern;
- die Arbeitsintensität;
- Arbeitgeber-Arbeitnehmer-Beziehungen (z. B. Lebenszeitstellen sind vielfach durch kürzere Arbeitsverhältnisse ersetzt).

Diese Veränderungen erfordern ein entsprechendes Personalmanagement. AC sind eine Methode, um den veränderten Bedingungen in der modernen Arbeitswelt zu begegnen. AC dienen dazu, organisationsinterne und -externe Personen auszuwählen und zu entwickeln. Demzufolge verfolgen AC folgende Hauptzielsetzungen: a) Personalauswahl, b) Personalentwicklung und c) Kombination aus Personalauswahl und Personalentwicklung.

► Die „International Task Force of AC Operations“ [14] definiert AC wie folgt:

1. Ein AC ist ein Verfahren, in dem Informationen über den Kandidaten in Bezug auf vorab bestimmte, multiple verhaltensbezogene Anforderungen, sogenannten Anforderungsdimensionen, gesammelt werden.
2. Dazu werden mehrere Übungen und Simulationen in einem AC durchgeführt.
3. Mehrere Beobachter dokumentieren und beurteilen die Leistung des Kandidaten.
4. Es erfolgt eine Aggregation der Leistungen des Kandidaten, die übungsbasiert (Aggregation über alle Übungen hinweg), dimensionsbasiert (Aggregation über alle Anforderungsdimensionen hinweg) oder gesamturteilsbasiert (sogenanntes OAR, „Overall Assessment Rating“) sein kann.

5. Nach dem AC erhält jeder Kandidat von den Beobachtern mündliches oder schriftliches Feedback bezüglich der Leistung.

Die Funktionen des AC im Rahmen des Personalmanagements bestehen nach Thornton und Rupp [45] in:

- Rekrutierung
- Personalauswahl
- Leistungsbeurteilung
- Personalplanung
- Beförderung
- Versetzung
- Training und Personalentwicklung
- Organisationsentwicklung

Generell lassen sich vier Phasen des AC-Prozesses unterscheiden [2, 38]: 1) Analyse, 2) Design, 3) Ausführung und 4) Evaluation.

In der Analysephase wird eine Anforderungsanalyse durchgeführt (vgl. dazu Kapitel von Nerdinger in diesem Band), es werden die wesentlichsten arbeitsbezogenen Verhaltensweisen (z. B. Kommunikation) sowie jene KSAO (knowledge, skills, abilities, other characteristics) identifiziert, die diesen Verhaltensweisen zugrunde liegen. Ferner werden in dieser initialen Phase des AC Verhaltensdimensionen identifiziert, die mit den KSAO zusammenhängen. Während der Designphase werden Übungen und Simulationen zusammengestellt oder entwickelt, um die zuvor ermittelten Anforderungen zu messen. Darüber hinaus werden die Beobachter geschult und ein Pilottest des AC durchgeführt. Im Rahmen der Phase der Durchführung werden zunächst die Teilnehmer des AC informiert, das AC wird absolviert und auf der Basis der Leistungen der Kandidaten ein Rating erstellt, welches übungsbasiert, dimensionsbasiert oder gesamturteilsbasiert [35] sein kann. Schließlich werden die Daten integriert (Datenintegrationsprozess) und an die Kandidaten Feedback übermittelt (mündlich, schriftlich oder beides) (Details zur Datenintegration und zum Feedback s. [24]). In der Evaluationsphase wird das AC im Hinblick auf Reliabilität und Validität überprüft und in Abhängigkeit der Ergebnisse der Evaluation das ursprüngliche AC modifiziert und verbessert (zu spezifischen Varianten des AC s. [13, 15, 45]).

---

## 9.2 Anforderungsdimensionen

Welche Anforderungsdimensionen in deutschen Unternehmen zugrunde gelegt werden, variiert erheblich [25]. Ein Ranking der in Deutschland, Österreich und der Schweiz genutzten Anforderungsdimensionen zeigt Tab. 9.1 [26]. Demnach werden in fast allen

**Tab. 9.1** Genutzte Anforderungsdimensionen in AC im deutschsprachigen Raum [26]

Rang	Anforderungsdimension	% der Fälle in 144 Unternehmen
1	Kommunikationsfähigkeit	89
2	Durchsetzungskraft	86
3	Kooperationsfähigkeit	85
4	Führungskompetenz	76
5	Konfliktfähigkeit	74
6	Problemlösefähigkeit	70
7	Systematisches Denken	65
8	Zielorientierung	65
9	Entscheidungsfreude	62
10	Belastbarkeit	59
11	Engagement	58
12	Kreativität	46
13	Organisationstalent	45
14	Kundenorientierung	41
15	Offenheit	38
16	Authentizität	29
17	Delegationsfähigkeit	29
18	Fachwissen	14
19	Charisma	11
20	Unternehmerisches Denken	9
21	Ambiguitätstoleranz	7
22	Motivation	6
23	Lernbereitschaft	3
24	Einfühlungsvermögen	3
25	Argumentationsfähigkeit	3
26	Flexibilität	2
27	interkulturelle Kompetenz	1

Anmerkung: Mehrfachantworten. Angabe in % der deutschsprachigen Unternehmen, die das jeweilige Anforderungsmerkmal im AC diagnostizieren

Unternehmen Kommunikationsfähigkeit, Durchsetzungsfähigkeit und Kooperationsfähigkeit im AC geprüft. Mit der Art und Anzahl dieser Anforderungsdimensionen sind folgende Risiken verbunden:

- Die Anforderungsdimensionen unterscheiden sich im Grad ihrer Beobachtbarkeit. Bei schlechter Beobachtbarkeit einzelner Merkmale (z. B. Authentizität) stehen den Beobachtern zu wenige Informationen zur Verfügung, um zu einem validen Urteil zu gelangen. Da sie aufgrund des „Urteils-Zwangs“ dennoch entscheiden müssen, versuchen sie ihre Unsicherheit zu kompensieren, indem sie greifbare Verhaltensinformationen aus anderen, gut beobachtbaren Verhaltensweisen hinzuziehen.
- Verschiedene Personen (beispielsweise AC-Entwickler, Beobachter und Teilnehmer) verstehen unter ein und derselben Anforderung (z. B. Führungskompetenz) etwas Unterschiedliches (z. B. charismatisch führen, Mitarbeiter motivieren, Aufgaben delegieren, Zielerreichung überprüfen etc.). Besteht z. B. Kooperationsfähigkeit darin, dass jemand die Interessen seiner Kollegen und Vorgesetzten neben seinen eigenen Interessen berücksichtigt? Oder ist jemand dann kooperationsfähig, wenn er Gemeinschaftslösungen anstrebt, andere zu integrieren versucht oder Konsens in Arbeitsgruppen herzustellen vermag?
- Ferner bleibt zu fragen, ob das Standardrepertoire der aktuell beobachteten Merkmale der Anforderungswirklichkeit im Zeitalter der Internationalisierung noch hinreichend entspricht. Möglicherweise sind aufgrund sich wandelnder Bedingungen veränderte Anforderungsprofile notwendig, die stärker Konstrukte wie etwa interkulturelle Kompetenz, Mehrdeutigkeitstoleranz oder Charisma abbilden. Besonders die interkulturelle Kompetenz stellt einen erfolgskritischen Faktor dar, weil es heute darauf ankommt, ob jemand in der Lage ist, sich auf internationaler Ebene adäquat zu bewegen. Insofern wäre es nicht nur aus Gründen der Veränderungen in der Position, Organisation und Umwelt der Organisation wünschenswert, Aktualisierungen der Art der Anforderungsdimensionen vorzunehmen, sondern insbesondere aus Gründen der Gewährleistung kriterien- und konstruktvalider AC.
- Darüber hinaus ist die Anzahl der genutzten Anforderungsdimensionen entscheidend. In vielen Unternehmen werden schlicht zu viele Anforderungsdimensionen diagnostiziert. Dies stellt ein Problem für die Konstruktvalidität des AC dar, denn eine Vielzahl genutzter Anforderungsdimensionen überfordert die kognitive Verarbeitungskapazität der Beobachter, die dann nicht mehr hinreichend zwischen den Anforderungsdimensionen differenzieren können (zu kognitiven Prozessen der Beobachter s. [31]). Einzelstudien [11], Reviews [14, 32] und Metaanalysen [3, 5] zeigen übereinstimmend, dass eine geringe Anzahl an Anforderungsdimensionen einer hohen Anzahl an Anforderungsdimensionen vorzuziehen ist. In diesem Sinne zeigt sich metaanalytisch, dass eine geringe Anzahl an Anforderungsdimensionen mit einer höheren konvergenten Validität ( $r = ,37$ ) und eine hohe Anzahl an Anforderungsdimensionen mit einer geringeren konvergenten Validität ( $r = ,27$ ) korrespondiert [5]. Empfehlenswert erscheint die Verwendung von fünf bis maximal sieben Anforderungsdimensionen pro AC.

Die Frage, welche Anforderungsdimensionen dies sein könnten, wurde metaanalytisch untersucht [3, 5, 48]. Diese Metaanalysen zeigen, dass folgende sechs Anforderungsdimensionen konstrukt- und kriterienvalide sind: Kommunikation, Berücksichtigung anderer



Menschen, Antrieb, Beeinflussung anderer Menschen, Organisation und Planung sowie Problemlösen.

Nach Arthur und Kollegen [3] meint „Kommunikation“ das Ausmaß, in dem jemand mündliche und schriftliche Informationen zur Verfügung stellt und mit Fragen und Herausforderungen umgeht. „Berücksichtigung anderer Menschen“ definieren die Autoren als das Ausmaß, in dem individuelles Verhalten die Gefühle und Bedürfnisse anderer Menschen reflektiert. Ferner beinhaltet es, die Klarheit über die Auswirkungen und Implikationen von Entscheidungen anderer innerhalb und außerhalb der Organisation. Mit „Antrieb“ ist das Ausmaß gemeint, in dem eine Person ein hohes Aktivitätsniveau zeigt und aufrecht erhält, sich hohe Leistungsstandards setzt, persistent an der Zielerreichung arbeitet und einen Aufstiegswunsch hat. Mit „Beeinflussung anderer Menschen“ ist der Grad gemeint, in welchem ein Individuum andere dazu veranlasst, etwas zu tun oder eine Meinungsänderung bei anderen zu erreichen, um gewünschte Ergebnisse zu erzielen. Demgegenüber meint „Organisation und Planung“ den Grad, in dem eine Person die eigene Arbeit und Ressourcen systematisch arrangiert, um Aufgaben effizient zu absolvieren sowie das Ausmaß, in dem jemand zukunftsbezogene Ereignisse antizipiert und diese vorbereitet. Die Dimension „Problemlösen“ beinhaltet vielfältige Aspekte, nämlich den Grad, in dem eine Person Informationen einholt, relevante technische und professionelle Informationen versteht, Daten und Informationen effektiv analysiert, Optionen und Ideen zur Problemlösung generiert, Handlungspläne für Probleme und Situationen auswählt, verfügbare Ressourcen in neuer Art und Weise nutzt sowie imaginäre Handlungsergebnisse generiert.

In diversen Einzelstudien wurde u. a. untersucht, in welcher Häufigkeit Unternehmen in Westeuropa [19, 20, 21], in Nordamerika [9, 24, 44], in Südafrika [29] und in Indonesien [30] diese sechs Anforderungsdimensionen im AC diagnostizieren. Es zeigt sich das in Tab. 9.2 zusammengefasste Bild.

Es wird deutlich, dass zwei Drittel der Organisationen in Südafrika die Dimensionen Kommunikation, Organisation und Planung, Problemlösen und Beeinflussung anderer Menschen und ca. ein Drittel die Dimensionen Berücksichtigung anderer Menschen und Antrieb diagnostizieren. In Westeuropa nutzen fast alle Organisationen in dieser Stichprobe Kommunikation und Problemlösen und ca. zwei Drittel der Unternehmen legen die anderen vier in Tab. 9.2 aufgeführten Anforderungsdimensionen zugrunde. In Nordamerika zählen Kommunikation, Problemlösen und Organisation und Planung zu den am häufigsten genutzten Anforderungsdimensionen. In Indonesien nutzen ca. die Hälfte der untersuchten Unternehmen Kommunikation und Problemlösen, während ca. ein Drittel der Unternehmen die anderen vier Anforderungsdimensionen im AC zugrunde legen. Der Vergleich der absoluten Häufigkeiten der genutzten Anforderungsdimensionen macht deutlich, dass all die in Tab. 9.2 genannten Dimensionen in Südafrika und Indonesien seltener diagnostiziert werden als in Westeuropa und Nordamerika. Die Metaanalyse von Arthur et al. [3]. bestätigte, dass die besten anforderungsbezogenen Leistungsprädiktoren im AC Kommunikation, Beeinflussung anderer Menschen, Organisation und Planung sowie Problemlösen darstellen. Diese Anforderungsdimensionen

**Tab. 9.2** Diagnostizierte Anforderungsdimensionen in AC in Südafrika, Westeuropa, Nordamerika und Indonesien [18]

	Südafrika	Westeuropa	Nordamerika	Indonesien
	(N = 43)	(N = 45)	(N = 52)	(N = 47)
Kommunikation	63	98	90	47
Berücksichtigung anderer Menschen	33	80	49	34
Antrieb	33	70	37	26
Beeinflussung anderer Menschen	58	75	61	23
Organisation und Planung	67	73	83	34
Problemlösen	67	84	90	51

Anmerkung: Angabe in % der Unternehmen, die das jeweilige Anforderungsmerkmal im AC diagnostizieren. Südafrika 43 Organisation, Westeuropa 45 Organisationen, Nordamerika 52 Organisationen, Indonesien 47 Organisationen

erklärten in der erwähnten Studie 20 % der Varianz der Leistung. Insofern kann geschlussfolgert werden, dass jene Anforderungsdimensionen, die in Südafrika, Westeuropa und Nordamerika am häufigsten im AC gemessen werden auch jene mit der besten prädiktiven Validität sind.

### 9.3 Übungen, Simulationen und zusätzliche Diagnoseinstrumente

Der folgende Abschnitt widmet sich der Frage, durch welche Übungen und Simulationen die Anforderungsdimensionen im AC diagnostiziert werden. Zu den prominentesten Übungen und Simulationen in verschiedenen Ländern und Regionen gehören die Postkorbübung, Präsentationen, Interview über den Hintergrund der Person, situatives Interview, Rollenspiele (z. B. „one-to-one“), Fallstudien, Planungsübungen, Gruppendiskussionen, Faktenfindung, Planungsübungen und soziometrische Aufgabenstellungen [18]. Tab. 9.3 gibt ein Beispiel einer Anforderungs-Übungsmatrix [27], d. h. einen Überblick, welche Anforderungsdimensionen durch welche Übungen und Simulationen erfasst werden können. Ausgewählte Übungen und Simulationen sollen nachfolgend kurz beschrieben werden.

#### Präsentation

Die Kandidaten werden aufgefordert, formale Informationen und/oder neue Ideen einer Gruppe von Personen (zumeist den Beobachtern) zu präsentieren. Typische Beispiele sind Besprechungen der Geschäftsleitung, eine Pressekonferenz, Vorstellung des Unternehmens, eine Besprechung von Abteilungsleitern aus verschiedenen Regionen oder

**Tab. 9.3** Anforderungs-Übungsmatrix. (Quelle: [27])

Dimensionen	Übungen			
	Interview	Präsentation	Führerlose Gruppen- diskussion	Entscheidungsorien- tierte Gruppenaufgabe
Kommunikation	+	+	+	+
Soziale Kompetenz	+		+	+
Stresstoleranz	+	+	+	+
Faktenorientierte Argumentation	+	+	+	+
Aktivität	+		+	+
Vorstellungsvermögen		+	+	+
Führungskompetenz	+		+	+
Motivation	+			

eine Skype-Konferenz mit Unternehmensvertretern eines anderen Landes. Zumeist werden den Kandidaten wenig bis keine Hintergrundinformationen dargeboten, sie werden demgegenüber gebeten, sich diese Informationen zu beschaffen und die Präsentation vorzubereiten. Die Beobachter, die mitunter auch als Rollenspieler fungieren, haben standardisierte Fragen, die dazu dienen sollen, die Kandidaten herauszufordern und eine dynamischere Interaktion zu schaffen.

### **Führerlose Gruppendiskussion**

Während einer führerlosen Gruppendiskussion soll die Dynamik in einem Entscheidungsfindungsprozess in Kleingruppen stimuliert werden. Zumeist wird den Kandidaten ein Organisationsproblem zur Verfügung gestellt und Zeit gegeben, das Problem zu analysieren. Im Anschluss wird das Problem entweder in kleinen Gruppen bis zu drei Personen oder in Gruppen bis zu sieben Personen diskutiert, wobei eine Lösung herbeigeführt werden soll. Typisch ist, dass niemand in der Kleingruppe die Rolle einer Führungskraft innehat. Mitunter erhalten die Kandidaten in führerlosen Gruppendiskussionen vor Beginn der Diskussion formale Rollen zugeteilt. Jeder Kandidat wird dazu ermuntert, eine bestmögliche Lösung zu entwickeln und dabei den Nutzen für die Gruppe zu maximieren. Es gibt auch Fälle, in denen keine formalen Rollen zugewiesen werden (z. B. Ad-hoc-Meetings), wobei es hier darum geht, die individuellen Lösungsvorschläge in einen kollektiven Handlungsplan zu transformieren.

### **Rollenspiel („one-to-one“)**

Der Kandidat trifft eine andere Person (zumeist einen Beobachter, der über standardisierte Verhaltensinstruktionen verfügt oder einen professionellen Rollenspieler), um ein spezifisches Problem einer Arbeitseinheit oder einer Organisation zu diskutieren. Typische Beispiele sind: Interview mit einem Bewerber, Gespräch mit einem leistungsschwachen

Mitarbeiter, Meeting mit einem neuen Kunden oder Geschäftspartner, Gespräch mit einem Kollegen zur Informationsgewinnung oder Gespräch mit einem unzufriedenen Kunden. Der Kandidat ist aufgefordert, mit der jeweils anderen Person Informationen auszutauschen und eine Lösung für das Problem herbeizuführen oder eine Handlungsabsicht zu formulieren. Der Rollenspieler ist vorab instruiert, spezifische Emotionen zu simulieren, bestimmte Fragen zu stellen oder bestimmte Vorschläge zu unterbreiten. Dabei verfügen die Rollenspieler über ein Skript, das in standardisierter Weise vorgibt, wie sie auf Fragen oder Antworten des Kandidaten zu reagieren haben. Vor dem Rollenspiel wird den Kandidaten zumeist Zeit zur Verfügung gestellt, um sich auf das Rollenspiel vorzubereiten (z. B. weitere Informationen einholen). Im Anschluss an das Rollenspiel wird der Kandidat ersucht, einen Bericht bzw. einen Handlungsplan über das Gespräch zu erstellen.

### **Postkorbübung**

In einer Postkorbübung erhält der Kandidat eine große Menge an Informationen (z. B. E-Mails, Briefe, Erinnerungsschreiben, Berichte), die sich in seiner E-Mail-Box befinden. Der Kandidat erhält weiterhin Informationen über die Organisation, das Personal und einen Kalender mit Meetings, Terminen und Veranstaltungen. Die Kandidaten werden aufgefordert, Aktionen zu zeigen und dabei die Dringlichkeit, arbeitsbezogene Relevanz, Komplexität und Interdependenz der Einzelinformationen, Termine etc. zu berücksichtigen. Typischerweise erfolgt die Postkorbübung unter hohem Zeitdruck. Nach Absolvierung der Postkorbübung durch den Kandidaten wird der Kandidat von einem Beobachter interviewt, der beabsichtigt, mehr über die Schwerpunktsetzung bei den durchgeführten Entscheidungen und die Gründe für die erfolgten Aktionen von dem Kandidaten zu erfahren.

### **Fallstudie**

Bei einer Fallstudie analysiert der Kandidat Hintergrundinformationen über ein organisationales Problem (z. B. finanzielle Entscheidungen, Prozeduren, organisationsweite Strategiepläne) und erstellt individuell einen schriftlichen Aktionsplan, wobei er diesen Aktionsplan gegenüber den Beobachtern zu begründen hat.

Mit Blick auf die Erhöhung der Validität von AC ist es empfehlenswert, zusätzlich zu den beschriebenen verhaltensbezogenen Übungen und Simulationen weitere Diagnoseinstrumente einzusetzen [29]. Dazu bieten sich etwa Arbeitsproben, Persönlichkeitstests, Fähigkeitstests, Intelligenztests, biografische Fragebogen oder Wissenstests an [17, 24]. Empirisch hat sich gezeigt, dass der kombinierte Einsatz von AC mit diesen zusätzlichen diagnostischen Methoden validitätssteigernde Wirkung entfacht [29], d. h. die Adäquatheit der auf der Basis der AC Ergebnisse getroffenen Personalentscheidungen verbessert. In der Personalpraxis wird diesem Umstand bislang zu wenig Bedeutung geschenkt: Sowohl Organisationen in Europa als auch in den USA und Kanada setzen nur in Ausnahmefällen zusätzlich zu den verhaltensbezogenen Übungen und Simulationen auch Testverfahren im Rahmen des AC ein ([18, 41]; zu den Gründen hierzu vgl.

[19]). In diesem Aspekt besteht für die aktuelle AC-Anwendung eindeutig Verbesserungsbedarf.

---

## 9.4 Beobachter und Beobachtertraining

Der Beobachterpool im AC setzt sich aus unterschiedlichen Funktionsgruppen zusammen, zumeist Linienmanager und Personalexperten, wodurch die Beurteilung der Teilnehmer auf einer breiten Basis abgesichert wird. In Organisationen in Deutschland werden stärker als in anderen Ländern Linienmanager in den AC-Prozess integriert [23]. Diese stärkere Integration von Linienmanagern kann mit der spezifischen Arbeitsgesetzgebung im deutschsprachigen Raum erklärt werden. Sowohl Kündigungsschutzgesetze als auch die Rechtsprechung in deutschen Arbeitsgerichten verleihen der Frage der Personalauswahl und Personalbeförderung ein erhöhtes Gewicht: Aufgrund spezifischer deutscher Rechtsnormen und deutscher Rechtsprechung ist die Langfristigkeit der getroffenen Personalentscheidungen hoch und die Reversibilität dieser Entscheidungen gering [20]. Dies könnte ein Grund dafür sein, dass in der Hierarchie hoch angesiedelte Manager in den Personalauswahl- und Personalentwicklungsprozess stärker einbezogen werden als z. B. in den USA. Psychologen nehmen nur in Ausnahmefällen den Beobachterstatus ein. Allerdings fanden Thornton et al. [22], dass die prädiktive Validität des AC durch den Einsatz von Psychologen als Beobachter steigt. Aufgrund des Betriebsverfassungsgesetzes nehmen außerdem Vertreter des Betriebsrats und Frauenbeauftragte als Beobachter am AC teil.

Bei der Zusammenstellung des Beobachterpools sollten unterschiedliche Kriterien berücksichtigt werden wie etwa die hierarchische Ebene, der funktionale Arbeitskontext, das Geschlecht, die Herkunft, die Ethnizität, das Alter und der Ausbildungshintergrund der Beteiligten [24]. Ein Vergleich des Ausmaßes, in dem diese Kriterien in Organisationen in Westeuropa und Nordamerika berücksichtigt werden [24] zeigt folgendes Bild: Die organisationale Ebene der Beobachter wird häufiger in westeuropäischen Organisationen als Kriterien der Konstitution des Beobachterpools zugrunde gelegt als in nordamerikanischen Organisationen, wo hingegen der Herkunft der Beobachter ein größerer Stellenwert beigemessen wird. Insgesamt spielen in Organisationen in beiden Ländergruppen die Kriterien Geschlecht, Ethnizität, Ausbildungshintergrund und Alter der Beobachter eine untergeordnete Rolle bei der Zusammensetzung des Beobachterpools, was problematisch ist, da die meisten Beobachterpools unbalanciert in Bezug auf diese wichtigen Kriterien sind. Dieser Umstand ist insofern als problematisch anzusehen als die unausgewogene Zusammensetzung der Beobachter im Hinblick auf diese Kriterien die Exaktheit der auf Basis der AC-Ergebnisse getroffenen Entscheidungen beeinträchtigt (in Bezug auf geschlechtsbezogene Effekte von AC-Ratings siehe [4, 8, 10]).

Im Hinblick auf die konstrukt- sowie kriterienbezogene Validität des AC ist es wichtig, geeignete Beobachtungssysteme zu verwenden. Im Wesentlichen unterscheidet

man dabei zwischen Beobachtungsskalen, Verhaltenschecklisten und verhaltensnahen Beschreibungen. Bei Beobachtungsskalen werden konkrete Verhaltensindikatoren des Konstrukts auf einer meist 7-fach gestuften Likert-Skala eingeschätzt. Bei Verhaltenschecklisten entscheidet der Beobachter jeweils, ob der Teilnehmer die konkrete Operationalisierung des Konstrukts zeigt oder nicht zeigt. Bei verhaltensnahen Beschreibungen sind die Verhaltensweisen auf einer Skala einzustufen, deren Pole nicht numerisch, sondern verbal-inhaltlich definiert sind. In der Art der verwendeten Beobachtungssysteme unterscheiden sich Organisationen im deutschsprachigen Raum von US-amerikanischen Organisationen. Deutschsprachige Organisationen setzen häufiger Beobachtungsskalen ein als US-amerikanische Organisationen, welche umgekehrt verhaltensnahe Beschreibungen häufiger einsetzen [28]. Verhaltenschecklisten werden in diesen Ländern von ca. der Hälfte der Unternehmen eingesetzt. Reilly et al. [37] belegten empirisch, dass sich die Verwendung von Verhaltenschecklisten oder verhaltensnahen Beschreibungen positiv auf die Konstruktvalidität des AC auswirkt, weil die Beobachter eine eindeutigere Zuordnung des Teilnehmerverhaltens in die jeweiligen Anforderungsdimensionen vornehmen können. Durch Beobachtungsskalen werden die ohnehin nicht hinreichend trennscharfen Anforderungsdimensionen meist nur schlagwortartig reproduziert, wodurch die Heterogenität in der Bedeutungszuweisung zwischen den Beobachtern pro Merkmal steigt, was eine fehlerbehaftete Gesamtbeurteilung zur Folge hat.

Der Umgang mit den Beobachtungssystemen wird in der Regel in entsprechenden Beobachtertrainings geübt. Diesbezüglich zeigt sich, dass fast alle Unternehmen in Deutschland, Österreich, der Schweiz und den USA separate Beobachtungstrainings durchführen [24]. Ein Beobachtertraining dauert dabei in den USA signifikant länger ( $M = 3,84$  Tage,  $S = ,74$ ) als im deutschsprachigen Raum ( $M = 1,72$  Tage,  $SD = ,72$ ,  $t = -2,66$ ,  $p < ,001$ ). Von der Dauer der durchgeführten Beobachtertrainings lässt sich jedoch nicht auf die Qualität dieser Trainings rekurren [32]. Mit zunehmender Länge der Schulung nimmt also keineswegs automatisch die Qualität der vermittelten Inhalte zu. Analog zu dieser These konnte metaanalytisch belegt werden [11], dass die prädiktive Validität des AC nicht durch die Dauer des Beobachtertrainings determiniert wird. Um Beobachter hinsichtlich ihrer eigenen Wahrnehmungsverzerrungen zu sensibilisieren und ihnen ausreichende Sicherheit in der Methode der Verhaltensbeobachtung zu vermitteln, kommt es demnach weniger auf die Länge als vielmehr auf die Qualität des Trainings an.

Im Hinblick auf die Qualität des Beobachtertrainings kommt es zum einen auf die vermittelten Inhalte und zum anderen auf die angewandte Methodik an. Inhaltlich sollte ein qualitativ hochwertiges Beobachtertraining die Beobachter mit der Methode der Verhaltensbeobachtung sowie der adäquaten Anwendung der Beobachtungssysteme einschließlich der korrekten Skaleninterpretation vertraut machen. Dies schließt ein, dass die Beobachter Kenntnis über systematische Beurteilungsfehler bekommen – wie etwa den Halo-Effekt, Attributionsfehler, den Primacy-Effekt oder die Aktivierung impliziter Persönlichkeitstheorien –, um diese reduzieren zu können. Ferner sollten die Beobachter lernen,

sich pro Übung auf diejenigen Anforderungsdimension zu konzentrieren, die mit der Übung erfasst werden sollen, gleichzeitig aber zwischen den Anforderungsdimensionen zu differenzieren lernen sowie üben, die verschiedenen Anforderungsdimensionen unabhängig voneinander zu beobachten. Tab. 9.4 gibt einen Überblick über die in westeuropäischen und nordamerikanischen Organisationen vermittelten Inhalte des Beobachtertrainings.

Zur Erhöhung der Anschaulichkeit dieser zu vermittelnden Inhalte könnte es methodisch hilfreich sein, eine Videodemonstration vorzuführen, in der alle Beobachter die Aufgabe haben, dieselbe AC-Sequenz zu beobachten und anschließend den dargestellten

**Tab. 9.4** Inhalte des Beobachtertrainings in westeuropäischen und nordamerikanischen Organisationen

	Westeuropa (%)	Nordamerika (%)
<i>Inhalte des Beobachtertrainings</i>		
Wissen über die eingesetzten Übungen	95	80
Fähigkeit der Beobachtung, des Protokollierens und der Klassifikation des Teilnehmerverhaltens in Bezug auf die Anforderungsdimensionen	68	68
Professioneller Umgang mit den Teilnehmern während des AC	75	65
Wissen und Sensibilisierung für Beurteilungsfehler	93	63
Wissen über die Organisation	10	55
Wissen über die Zielposition	25	50
Methode der Verhaltensbeobachtung	83	58
Wissen über die Anforderungsdimensionen (Definition, Abgrenzungen)	60	48
Verwendung der Beobachtungssysteme inkl. korrekter Skaleninterpretation	70	58
Fähigkeit, akkurat mündliches und schriftliches Feedback zu vermitteln	65	58
Konsistenz im Rollenspiel	53	58
Relation zwischen den Anforderungsdimensionen und der Leistung	28	43
Fähigkeit, die Anforderungsdimensionen unabhängig voneinander zu beobachten	43	43
Fokussierung auf diejenigen Anforderungsdimensionen, für die die Übung konstruiert wurde	48	45
Wissen über den Datenintegrationsprozess	63	33
Methode der Urteilsbildung (statistisch, nicht-statistisch)	20	40
Grenzen der AC-Methode	55	33
Wissen über diagnostische Standards	33	18
Wissen, um zwischen den Anforderungsdimensionen zu unterscheiden	38	23
Formen der Beeinflussung in Gruppen während des Datenintegrationsprozesses	35	28

Anmerkung: Westeuropa: 45 Organisationen. Nordamerika: 52 Organisationen



Teilnehmer auf den entsprechenden Anforderungsdimensionen einzustufen. Dies kann ermöglichen, dass den Beobachtern die Unterschiede in der Beobachtung zwischen den verschiedenen Beobachtern bei der Beurteilung der gleichen Beobachtungssequenz verdeutlicht werden und sie außerdem für eigene Fehler in der Beurteilung sensibilisiert werden.

---

## 9.5 Prädiktive und inkrementelle Validität

### Evidenz für Korrelationen zwischen AC-Gesamturteilen und berufsbezogenen Kriterien

Eine Vielzahl von Studien dokumentiert, dass AC-Gesamturteile („overall assessment ratings“, OAR) verschiedene Kriterien – wie etwa Gehaltsverbesserungen, beruflichen Aufstieg, Ausbildungserfolg, Arbeitsleistung – für ein breites Spektrum an Berufen und Organisationen valide zu prognostizieren imstande sind [33, 42]. Die Ergebnisse der Metaanalysen für die Beziehung zwischen AC-Urteilen und berufsbezogenen Ergebnissen variieren erheblich, nämlich zwischen  $r = ,41$  [33],  $r = ,37$  [11],  $r = ,25$  bis  $r = ,39$  [3],  $r = ,22$  [1] und  $r = ,31$  [12]. Diese Unterschiede in den ermittelten prädiktiven Validitäten lassen sich durch die Berücksichtigung unterschiedlicher Einzelstudien pro Metaanalyse, Unterschiede in der Qualität des AC [19, 20, 27] und die Anwendung divergierender metaanalytischer Verfahren erklären. Lievens und Thornton [33] argumentieren, dass der viel zitierte Validitätskoeffizient ( $r = ,37$ ) nach Gaugler et al. [11] möglicherweise eine Unterschätzung der Validität des AC darstellt, da sie eine vergleichsweise konservative (also hohe) Schätzung der Reliabilität des Kriteriums ( $,86$ ) in ihren Korrekturen bezüglich der Unreliabilität des Kriteriums vorgenommen haben. Sofern die Reliabilität des Kriteriums auf  $,52$  geschätzt würde [47], würde die geschätzte Validität des AC von  $r = ,7$  auf  $r = ,47$  steigen.

Zusammenfassend kann festgehalten werden, dass sorgfältig entwickelte und professionell durchgeführte AC (s. dazu „International Task Force on Assessment Center Guidelines“ [14]) die berufsbezogene Leistung für unterschiedliche Berufsgruppen vorhersagen kann.

Will man die prädiktive Validität des AC weiterhin erhöhen, so bieten sich unterschiedliche Vorgehensweise an wie etwa die Kombination eines AC mit weiteren Diagnosemethoden [7, 27], die Erhöhung der Anzahl und Art der eingesetzten Übungen [45, 46], die Integration von Psychologen im Beobachterpool [46], die Verwendung von Peer-Rating [46] sowie die Anwendung intransparenter Anforderungsdimensionen [15].

Wenngleich Evidenz dafür vorliegt, dass AC zu einem statistisch signifikanten und praktisch relevanten Ausmaß mit Leistungskriterien korrelieren, fallen diese Korrelationen der AC-Ratings gewöhnlich niedriger aus als die Korrelationen zwischen kognitiven Fähigkeitstests und Leistungsergebnissen. Die metaanalytischen Ergebnisse von Schmidt

und Hunter [39] machen deutlichen, dass die Relation zwischen kognitiven Fähigkeitstests und der beruflichen Leistung ( $r = ,51$ ) stärker ist als die Relation zwischen AC-Ratings und beruflicher Leistung ( $r = ,37$ ). Auch Pynes und Bernadin [36] fanden, dass Ergebnisse kognitiver Fähigkeiten höher mit dem Trainingserfolg korrelieren ( $r = ,31$ ) als AC-Ergebnisse.

Angesichts der höheren kriterienbezogenen Validitäten von kognitiven Fähigkeitstests und der Tatsache, dass diese weniger kostenintensiv und leichter zu administrieren sind als AC, müssen AC also die Prädiktor-Kriterien-Beziehungen erhöhen, um ihre Anwendung zu rechtfertigen. Mit anderen Worten: AC müssen zusätzliche prädiktive Validität ausweisen, wenn sie in Kombination mit solchen psychometrischen Testverfahren eingesetzt werden.

### **Inkrementelle Validität des AC**

Schon vor ca. 30 Jahren wiesen Klimoski und Brickner [16] darauf hin, dass AC zwar funktionieren, dass sie jedoch wenig mehr messen als Intelligenz. Die Frage, ob AC-Ratings das Verständnis und die Prognose der Leistung zusätzlich zu kognitiven Fähigkeitstests verbessern, hängt ab von a) der Komplexität der Leistungsdomäne und b) der Relation zwischen AC und kognitiven Fähigkeitstests. Die Leistungsdomäne einer Vielzahl von Tätigkeit ist als komplex zu bezeichnen, wobei dies insbesondere für Führungsaufgaben gilt [43]. Aus diesem Grund ist davon auszugehen, dass unterschiedliche Messungen der verschiedenen Konstrukte einen jeweils eigenständigen Beitrag für die Prädiktion der Leistung aufweisen. So zeigt sich im Gegensatz zur erwähnten Metaanalyse von Schmidt und Hunter [39], dass AC-Ratings inkrementell valide sind, wenn sie zusätzlich zu kognitiven Fähigkeitstests zum Einsatz kommen. Dies wird etwa in einer Studie in der israelischen Polizei deutlich [7]. Die Autoren zeigen, dass AC-Gesamturteile signifikant mit dem Ausbildungserfolg korrelieren wie z. B. die finale Ausbildungsnote ( $r = ,34$ ), die gemittelten Peer-Evaluationen pro Dimension ( $r = ,14$ ), den zukünftigen Berufserfolg ( $r = ,43$ ), die Leistung on-the-job gemessen durch Vorgesetztenurteile ( $r = ,25$ ) sowie periodische Evaluationen durch Vorgesetzte ( $r = ,24$ ). Ferner wiesen in dieser Studie die AC-Gesamturteile inkrementelle Validität auf in Bezug auf die finale Ausbildungsnote, Peer-Evaluationen und den zukünftigen Berufserfolg, wenn zusätzlich zu Tests der allgemeinen Intelligenz Anwendung finden.

Bezüglich der Beziehung zwischen AC-Ratings und kognitiven Fähigkeiten schätzten Schmidt und Hunter [39] die korrigierte Korrelation zwischen AC und Intelligenztests auf  $r = ,50$ , während eine Metaanalyse von Scholz und Schuler [40] empirisch belegte, dass allgemeine Intelligenz mit der Leistung einer Person im AC in Höhe von  $r = ,33$  korreliert. Metaanalytische Resultate von Collins et al. [6] haben ebenfalls eine enge Beziehung zwischen AC-Gesamturteilen und kognitiven Fähigkeiten ( $r = ,67$ ) aufgezeigt. Diese Ergebnisse haben Schmidt und Hunter [39] zu dem Argument veranlasst, dass AC-Ergebnisse keine inkrementelle Validität ( $r = ,02$ ) aufweisen würden, wenn Informationen aus den Testverfahren der allgemeinen Intelligenz zur Verfügung ständen.

Aktuellere Studien zeigen im Gegensatz hierzu, dass AC-Ratings inkrementelle Validität gegenüber dem parallelen Einsatz von kognitiven Fähigkeitstests [27] und Persönlichkeitstests [34] aufweisen.

---

## 9.6 Fazit

Durch AC-Verfahren können verschiedene berufsbezogene Kriterien (z. B. Leistung, Gehaltsentwicklung, Karriereentwicklung, Ausbildungserfolg) in diversen Berufen, Organisationen und Branchen vorhergesagt werden. Dies gilt auch in diversen Ländern und Regionen, in denen die AC-Anwendung bereits untersucht worden ist. Die berichteten Ergebnisse legitimieren den Einsatz des AC als Instrument der Personalauswahl und der Personalentwicklung. Zukünftige Forschung im Bereich der Personalauswahl und der Personalentwicklung durch AC sollte stärker als dies bislang zu verzeichnen ist, auch interkulturelle Aspekte z. B. im AC-Design (Zusammenstellung der Übungen und Simulationen, Konstitution des Beobachterpools, Übermittlung von Feedback) zu berücksichtigen.

---

### Fragen

1. Beschreiben Sie die Merkmale eines AC sowie verwandte Anforderungsdimensionen und prominente Übungen.
2. Welche Gruppen von Personen sind aus welchen Gründen in deutschsprachigen Organisationen im Beobachterpool vertreten? Welche Inhalte sollte ein Beobachtertraining vermitteln?
3. Nennen Sie metaanalytische Ergebnisse für die prädiktive Validität eines AC und begründen Sie, durch welche Maßnahmen man die prädiktive Validität erhöhen kann.
4. Zeigen Sie die widersprüchliche Befundlage zur inkrementellen Validität des AC auf.

---

### Literatur

1. Aamodt, M. G. (2004). *Research in law enforcement selection*. Boca Raton: BrownWalker.
2. Arthur, W., Jr., & Day, E. (2011). AC. In *Blackwell handbook of personnel selection* (S. 205–229). Blackwell.
3. Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of AC dimensions. *Personnel Psychology, 56*, 125–154.
4. Anderson, N., Lievens, F., Van Dam, K., & Born, M. (2006). A construct-driven investigation of gender differences in a leadership-role AC. *Journal of Applied Psychology, 91*, 555–566.
5. Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on AC ratings. *Journal of Applied Psychology, 91*, 1114–1124.

6. Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of AC evaluations? *International Journal of Selection and Assessment*, *11*, 17–29.
7. Dayan, K., Kasten, R., & Fox, S. (2002). Entry-level police candidate AC: An efficient tool or a hammer to kill a fly? *Personnel Psychology*, *55*, 827–849.
8. Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in AC ratings: A meta-analysis. *Journal of Applied Psychology*, *93*, 685–691.
9. Eurich, T., Krause, D. E., Cigularov, K., & Thornton, G. C., III. (2009). The current use of AC programs in the U.S. *Journal of Business and Psychology*, *24*, 387–407.
10. Falk, A., & Fox, S. (2014). Gender and ethnic composition of AC and its relationship to participants' success. *Journal of Personnel Psychology*, *13*, 11–20.
11. Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III, & Benson, C. (1987). Meta-analysis of AC validity. *Journal of Applied Psychology*, *72*, 493–511.
12. Hardison, C. M., & Sackett, P. R. (2004). AC criterion related validity: A meta-analytic update. Paper presented at the 18<sup>th</sup> annual conference of the society for industrial and organizational psychology, Chicago, April.
13. Höft, S., & Funke, U. (2006). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 145–187). Göttingen: Hogrefe.
14. International Task Force on AC Guidelines. (2015). Guidelines and ethical considerations for AC operations. *Journal of Management*, *41*(4), 1244–1273.
15. Kleinmann, M. (2003). *AC*. Göttingen: Hogrefe.
16. Klimoski, R. J., & Brickner, M. (1987). Why do AC work? The puzzle of AC validity. *Personnel Psychology*, *40*, 243–260.
17. Krause, D. E. (2010). *Trends in der internationalen Personalauswahl. Praxis der Personalpsychologie*. Göttingen: Hogrefe.
18. Krause, D. E. (2015). A comparison of assessment centres application in South Africa, Indonesia, North America and Western Europe – Results from three survey studies. Keynote address presented at the 35<sup>th</sup> annual AC study group conferences. Somerset West, March 23–27, Somerset West, South Africa.
19. Krause, D. E., & Gebert, D. (2003a). A comparison of AC practices in organizations in German-speaking regions and the United States. *International Journal of Selection and Assessment*, *11*, 297–312.
20. Krause, D. E., & Gebert, D. (2003b). Unterschiede in der Durchführung zwischen internen und externen AC – Eine empirische Analyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, *47*, 87–94.
21. Krause, D. E., & Thornton, G. C., III. (2005). International research project: Cultural value orientation and AC practices in America, Europe, and Asia. *Journal of Personnel Psychology*, *4*, 206–208.
22. Krause, D. E., & Thornton, G. C., III. (2007). Kulturelle Vielfalt und AC: Handlungsempfehlungen für Theorie und Praxis. *Zeitschrift für Führung und Organisation*, *4*, 223–230.
23. Krause, D. E., & Thornton, G. C., III. (2008). International perspectives on current AC practices and future challenges. In G. Roodt & S. Schlebusch (Hrsg.), *AC* (S. 285–300). Johannesburg: Knowres Publishing.
24. Krause, D. E., & Thornton, G. C., III. (2009). A cross-cultural look at AC practices: A survey in Western Europe and Northern America. *Applied Psychology: An International Review*, *58*, 557–585.
25. Krause, D. E., Meyer zu Kniendorf, C., & Gebert, D. (2001a). Aktuelle Trends in der AC Anwendung. *Wirtschaftspsychologie*, *8*, 47–55.

26. Krause, D. E., Meyer zu Kniendorf, C., & Gebert, D. (2001b). Das AC in der deutschsprachigen Wirtschaft. *Personal – Zeitschrift für Human Resource Management*, 53, 638–642.
27. Krause, D. E., Kersting, M., Heggstad, E. D., & Thornton, G. C., III. (2006). Incremental validity of AC ratings over cognitive ability tests. A study at the executive management level. *International Journal of Selection and Assessment*, 14, 360–371.
28. Krause, D. E., Gebert, D., & Thornton, G. C., III. (2007). Interkulturelle Unterschiede in der AC Anwendung. In H. Schuler (Hrsg.), *AC zur Potenzialanalyse* (S. 359–378). Göttingen: Hogrefe.
29. Krause, D. E., Rossberger, R. J., Dowdeswell, K., Venter, N., & Joubert, T. (2011). State of the art of AC practices in South Africa. *International Journal of Selection and Assessment*, 19, 262–275.
30. Krause, D. E., Anderson, N., Rossberger, R. J., & Parastuty, Z. (2014). AC practices in Indonesia: An exploratory study. *International Journal of Selection and Assessment*, 22, 384–398.
31. Lance, C. E., Foster, M. R., Thorensen, J. D., & Gentry, W. A. (2004). Assessor cognitive processes in an operational AC. *Journal of Applied Psychology*, 89, 22–35.
32. Lievens, F. (1998). Factors which improve the construct validity of AC: A review. *International Journal of Selection and Assessment*, 6, 141–152.
33. Lievens, F., & Thornton, G. C., III. (2005). AC: Recent developments in practice and research. In A. Evers, N. Anderson, & O. Voskuilj (Hrsg.), *The Blackwell handbook of personnel selection*. Malden: Blackwell.
34. Meriac, J. P., Hoffman, B. J., Woehr, A. J., & Fleisher, M. S. (2008). Further evidence for the validity of AC dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, 93, 1042–1052.
35. Putka, D. J., & Hoffman, B. J. (2013). Clarifying the contribution of assessee-, dimension-, exercise-, and assessor-related effects to reliable and unreliable variance in AC ratings. *Journal of Applied Psychology*, 98, 114–133.
36. Pynes, J. E., & Bernadin, H. J. (1989). Predictive validity of an entry-level police officer AC. *Journal of Applied Psychology*, 74, 831–833.
37. Reilly, R. R., Henry, S., & Smithers, J. W. (1990). An examination of the effects of using behavior checklists on the construct validity of AC dimensions. *Personnel Psychology*, 43, 71–84.
38. Schlebusch, S., & Roodt, G. (2008). *AC*. Johannesburg: Knowres Publishing.
39. 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 finding. *Psychological Bulletin*, 124, 262–274.
40. Scholz, G., & Schuler, H. (1993). Das nomologische Netzwerk des AC: Eine Meta-Analyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 73–85.
41. Schuler, H. (2000). Personalauswahl im europäischen Vergleich. In E. Regnet & L. M. Hoffmann (Hrsg.), *Personalmanagement in Europa* (S. 129–139). Göttingen: Hogrefe.
42. Speer, A. B., Christiansen, N. D., Goffin, R. D., & Goff, M. (2014). Situational bandwidth and the criterion-related validity of AC ratings: IS cross-exercise convergence always desirable? *Journal of Applied Psychology*, 99, 282–295.
43. Tett, R. P., Guterman, H. A., Bleir, A., & Murphy, P. J. (2000). Development and content validation of a „hyperdimensional“ taxonomy of managerial competence. *Human Performance*, 13, 205–251.
44. Thornton, G. C., III, & Krause, D. E. (2009). Comparison of practices in selection vs. developmental AC: An international survey. *International Journal of Human Resource Management*, 20, 478–498.

45. Thornton, G. C., III, & Rupp, D. E. (2006). *AC in Human Resource Management*. London: Lawrence Erlbaum.
46. Thornton, G. C., III, Gaugler, B. B., Rosenthal, D., & Bentson, C. (1992). Die prädiktive Validität des AC – eine Metaanalyse. In H. Schuler & W. Stehle (Hrsg.), *AC als Methode der Personalentwicklung* (2. Aufl., S. 36–60). Göttingen: Hogrefe.
47. Viswesaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557–574.
48. Woehr, D. J., & Arthur, W., Jr. (2003). The construct-related validity of AC ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*, 231–258.

---

# „Online-Assessments“ und „Self-Assessments“ in der Eignungsdiagnostik

# 10

Michael Ott, Anna-Sophie Ulfert und Martin Kersting

---

## Zusammenfassung

„Online-Assessments“ finden zunehmende Verbreitung im Kontext der Eignungsdiagnostik. In diesem Beitrag werden sie zur Personalauswahl und Eignungsdiagnostik beschrieben. Dabei definieren wir „Online-Assessment“ als Oberbegriff und differenzieren anhand von drei Unterscheidungsmerkmalen relevante Varianten: Erstens anhand des primären Ziels (Selbst- oder Fremdselektion), zweitens bezüglich der Methode (Testen oder Informieren), drittens anhand des Ausmaßes an Kontrolle in der Durchführung (beaufsichtigt oder unbeaufsichtigt). Wir berichten über die Effekte der unbeaufsichtigten Testdurchführung. Auf der Basis dieser Differenzierungen definieren wir „Self-Assessments“ als diejenige Form des „Online-Assessments“, die mittels unbeaufsichtigter Testung auf die Selbstselektion von Bewerbern abzielen. Abschn. 10.2 dient der Erläuterung dieser drei Unterscheidungsmerkmale. Abschn. 10.3 thematisiert die Besonderheiten von „Online-Assessments“ (z. B. Usability, Datenschutz, Akzeptanz), die sich aus dem online-basierten Charakter ergeben. Diese Ausführungen bilden die Grundlage zur Qualitätsbeurteilung von „Online-Assessments“, wobei wir verdeutlichen, welche Konsequenzen sich aus den Besonderheiten für die praktische

---

M. Ott (✉) · A.-S. Ulfert · M. Kersting  
Institut für Psychologie und Sportwissenschaft, Justus-Liebig-Universität Giessen,  
Giessen, Deutschland  
E-Mail: Michael.Ott@psychol.uni-giessen.de

A.-S. Ulfert  
E-Mail: Anna-Sophie.Ulfert@psychol.uni-giessen.de

M. Kersting  
E-Mail: Martin.Kersting@psychol.uni-giessen.de



diagnostische Anwendung ableiten. Abschn. 10.4 stellt Kriterien vor, anhand derer die Qualität von entsprechenden Online-Tools bewertet bzw. gesichert werden kann.

## Inhaltsverzeichnis

10.1	Einleitung	216
10.2	Unterscheidungsdimensionen	219
10.2.1	Selbst- und Fremdselektion	220
10.2.2	Testen und Informieren	223
10.2.3	Durchführung mit und ohne „Aufsicht“ – Herausforderungen	224
10.2.4	Kombination der Gestaltungselemente	227
10.3	Spezifika der Durchführung als Folge des Online-Charakters	228
10.3.1	Usability und Attraktivität von Tools	229
10.3.2	Wahrgenommener Datenschutz und wahrgenommene Fairness	230
10.3.3	Interindividuelle Unterschiede bezüglich der Akzeptanz von Online-Testungen	230
10.4	Bewertung der Qualität von „Online-Assessments“ und „Self-Assessments“	231
10.5	Fazit	234
10.5.1	Was bei der Konstruktion und dem Einsatz von „Online-Assessments“ und „Self-Assessments“ zu beachten ist	234
10.5.2	Forschungsdesiderata	236
10.5.3	Schlusswort	238
	Literatur	238

## 10.1 Einleitung

Die Begriffe „Online-Assessment“, „online tools“, „Online-Tests“, „e-Assessment“, „Recruitment“, „Online-Self-Assessment“ oder auch nur „Self-Assessment“ werden im Kontext der Eignungsdiagnostik häufig genutzt. Da verschiedene diagnostische Praktiken und Ziele unter diese Begriffe subsumiert werden, verschwimmt mitunter die Prägnanz der Begriffe. Der folgende Beitrag verfolgt aus diesem Grund zwei Ziele: Zunächst nehmen wir begriffliche Setzungen vor. Wir schlagen vor, den Begriff „Online-Assessment“ als Oberbegriff zu nutzen, und definieren „Self-Assessments“ als Unterkategorie dieses Oberbegriffs. Nach welchen Gesichtspunkten „Online-Assessments“ differenziert werden können und wie „Self-Assessments“ hierin zu verorten sind, wird im ersten Teil des Beitrags verdeutlicht. Darüber hinaus berichten wir empirische Befunde zum Einsatz von „Online-Assessments“, wobei wir auch Befunde aus benachbarten Themenbereichen (z. B. Online-Bewerbungen) aufführen, soweit sie für die Thematik bedeutsam und für die Beurteilung, Sicherung und Optimierung der Qualität des (wachsenden) Angebots hilfreich sind.

Betrachten wir zunächst die bisherige Begriffsverwendung und beginnen mit dem Begriff „Online-Assessment“. Konradt und Sarges [36] nutzen diesen Begriff im Kontext

der Personalauswahl ganz allgemein für jede online durchgeführte Diagnostik, die das Ziel verfolgt, die berufliche Eignung von Bewerbern festzustellen. „Online-Assessments“ dienen somit der Selektion (oder Vorselektion) von geeigneten Bewerbern, wobei die diagnostischen Verfahren in der Regel nicht frei zugänglich sind, sondern vom Bewerber im Laufe des Auswahlprozesses erst dann durchgeführt werden können, wenn eine entsprechende Einladung durch die auswählende Organisation erfolgt [14]. Neben der Selektion kann das „Online-Assessment“ nach unserer Definition auch die Selbstselektion umfassen.

- ▶ „Online-Assessments“ bieten zahlreiche Vorteile:
  - Unabhängigkeit von Raum, Zeit und Ressourcen (Rund-um-die-Uhr-Angebot, von jedem Ort der Welt [mit Internetanschluss] erreichbar)
  - Anonyme Durchführung möglich (bei „Self-Assessments“)
  - Hohe Auswertungs- und Interpretationsobjektivität
  - Ökonomie
  - Passung zur Zielgruppe (das Internet ist das Medium der Zielgruppe, Angebote im Internet erreichen die Zielgruppe und tragen zu einem „modernen“ Image des Anbieters bei)

Der vorliegende Beitrag beschränkt sich auf die Nutzung von „Online-Assessments“ im Bereich der Eignungsdiagnostik. Es gibt aber weitere Anwendungsfelder von „Online-Assessments“ im Personalbereich, beispielsweise das Personalmarketing und die Personalentwicklung [25]. Kirbach und Wottawa [34] verweisen explizit darauf, dass „Online-Assessments“ (und „Self-Assessments“) nicht nur Potenziale für die Personalauswahl, sondern insbesondere auch für das Personalmarketing besitzen. In Bezug auf „Self-Assessments“ verweisen Kubinger et al. [39] mit Bezug auf den Hochschulsektor darauf, dass diese nicht nur der Selbstselektion, sondern auch der positiven Außendarstellung einer Organisation dienen. Dem breiten Anwendungsfeld entsprechend mannigfaltig sind die möglichen Zielgruppen des „Online-Assessments“, von Schülern und Studieninteressierten über Berufsanfänger bzw. -einsteiger bis hin zu Personen mit Berufserfahrung [25].

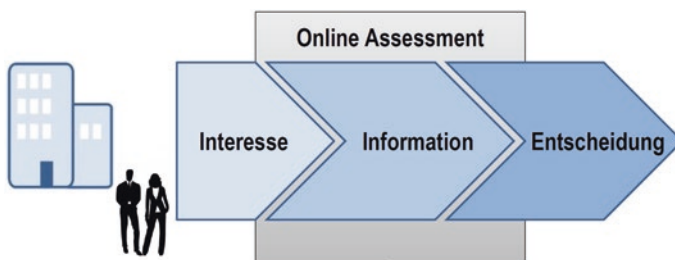
Der Begriff „Self-Assessment“ ist in der Personalauswahl relativ neu. In der wörtlichen Übersetzung einer „Selbsteinschätzung“ handelt es sich aber um nichts Neues. Neu hingegen ist, dass unter dem Begriff „Self-Assessment“ eine Reihe unterschiedlicher Konzepte aufkommen, die die Internet-Technologie nutzen und somit einen computergestützten, onlinebasierten diagnostischen Prozess ermöglichen.

Nach Kubinger [38] testet sich eine Person beim „Self-Assessment“ „... (zumeist über Internet) selbst und absolviert folglich den diagnostischen Prozess eigenverantwortlich und fachpsychologisch unkontrolliert“. Damit können „Self-Assessments“ sowohl im Sinne einer Selbsteinschätzung, aber auch als eine Form selbst (auch „an sich selbst“) durchgeführter Diagnostik verstanden werden (vgl. auch der Begriff des „Selbsttests“, z. B. Hell [24]).

Um eine Systematik zu etablieren, orientieren wir uns bei unseren Definitionen an dem eignungsdiagnostischen Prozess, den wir stark vereinfachend in drei Phasen untergliedern und am Beispiel der Selektion (Personalauswahl) erläutern (siehe Abb. 10.1). Zu Beginn der Selektion, in der ersten Phase, steht ein mehr oder minder konkretes Interesse der potenziellen Bewerber an einer Tätigkeit oder Position bzw. Organisation und/oder das Interesse der Organisation an bestimmten Bewerbern. In der zweiten Phase werden Informationen über die tatsächlichen Bewerber und/oder die Anforderungen generiert. Diese Informationen sind die Basis für die Entscheidung, der dritten Phase unseres vereinfachten Prozessmodells. Dabei kann die Entscheidung sowohl die Entscheidung des Bewerbers sein, die Bewerbung aufrecht zu erhalten oder ein Angebot anzunehmen als auch die Entscheidung der Organisation, einem Bewerber ein Angebot zu unterbreiten.

Der Begriff „Online-Assessment“ bezieht sich auf die zweite Phase des dargestellten Prozesses. „Online-Assessments“ können sowohl zur Personal(vor-)auswahl als auch zur Selbsterkenntnis bzw. Selbstselektion der Teilnehmer eingesetzt werden. Den Begriff „Self-Assessment“ sollte man hingegen, so unser Vorschlag, nur nutzen, wenn die Selbsterkenntnis des Teilnehmers und die darauf basierende Selbstselektion primäres Ziel sind. Eine Gemeinsamkeit aller hier unter „Online-Assessment“ subsumierten Verfahren ist die computergestützte und onlinebasierte Umsetzung. Um vorhandene Angebote sachgerecht zu beurteilen und neue Instrumente zielorientiert zu gestalten, schlagen wir drei zentrale Unterscheidungsdimensionen vor, anhand derer sich „Online-Assessments“ charakterisieren bzw. kategorisieren lassen:

- ▶ • Ziele: Selbst- und Fremdselektion
- Methoden: Testen und Informieren als Funktionsmechanismen zur Erreichung dieser Ziele
- Kontrolle: Unbeaufsichtigte und beaufsichtigte Testung als Eigenschaft der Darbietungsweise



**Abb. 10.1** Vereinfachtes Prozessmodell der Eignungsdiagnostik bei der Selbst- oder Fremdselektion

Neben diesen drei Unterscheidungskriterien existieren weitere Unterscheidungsdimensionen, auf die wir hier nicht eingehen. So könnte man auch

- das Ausmaß der Interaktivität (niedrig/hoch) sowie
- das Ausmaß der Einbindung spielerischer Elemente bei der Testung (niedrig/hoch) (Stichwort: „Recrutainment“, [13])

als Unterscheidungsdimensionen nutzen. Für den Hochschulsektor könnte man die von Hell [24] vorgeschlagenen Differenzierungen

- „hochschulübergreifend“ versus „hochschulspezifisch“,
- „fachübergreifend“ versus „fachspezifisch“

sowie weitere Einteilungsgesichtspunkte nutzen. Wir beschränken uns auf die zuerst genannten drei Dimensionen.

Nach einer näheren Beschreibung der drei oben genannten Unterscheidungsdimensionen thematisieren wir die Besonderheiten, die aus dem Online-Setting für den diagnostischen Prozess erwachsen. Dies betrifft die drei Hauptaspekte Usability, Datenschutz und Akzeptanz von „Online-Assessments“. Diese Ausführungen bilden die Grundlage zur Qualitätsbeurteilung von „Online-Assessments“, auf die wir im letzten Teil des Beitrags eingehen. Dabei nehmen wir eine Analyse von Erfolgs- sowie Risikofaktoren vor. Da „Online-Assessments“ häufig automatisiert erstellte Ergebnis- und (zum Teil) Entscheidungsrückmeldungen umfassen, gehen wir auch auf diese Form des Feedbacks ein und thematisieren die besonderen Anforderungen, die sich dabei stellen.

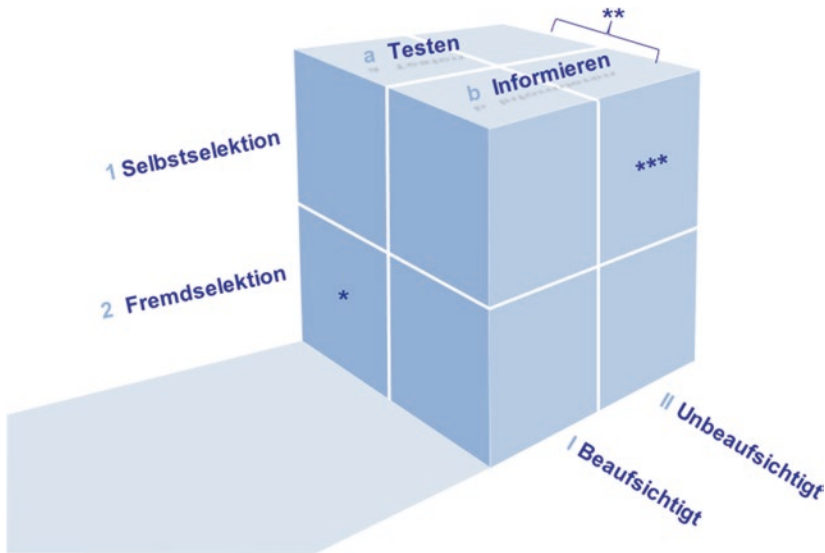
---

## 10.2 Unterscheidungsdimensionen

Zunächst differenzieren wir die unterschiedlichen Ziele der „Online-Assessments“ anhand des Entscheidungsträgers und unterscheiden zwischen Selbst- und Fremdselektion [48]. Die nächste Unterscheidungsdimension betrachtet zwei unterschiedliche Methoden, mit denen die Selbst- und Fremdselektion erreicht werden kann: Das Testen und das Informieren. Schließlich unterscheiden sich „Online-Assessments“ im Ausmaß der Kontrolle der diagnostischen Situation. Das „Online-Assessment“ kann beaufsichtigt oder unbeaufsichtigt erfolgen.

Diese drei Dimensionen dienen dazu, unterschiedliche „Online-Assessments“ zu charakterisieren. Alle drei Unterscheidungsdimensionen (siehe Abb. 10.2) stehen orthogonal zueinander, so dass theoretisch jede Kombinationsvariante der drei Dimensionen möglich ist.

Ein „Online-Assessment“ kann mit dem Ziel der Fremdselektion beaufsichtigt durchgeführt werden und informative Elemente umfassen oder aber auch mit dem Ziel der Selbstselektion unbeaufsichtigt administriert werden und sich auf das Testen beschränken. Ungeachtet dessen kommen in der Praxis bestimmte Kombinationen häufiger vor,

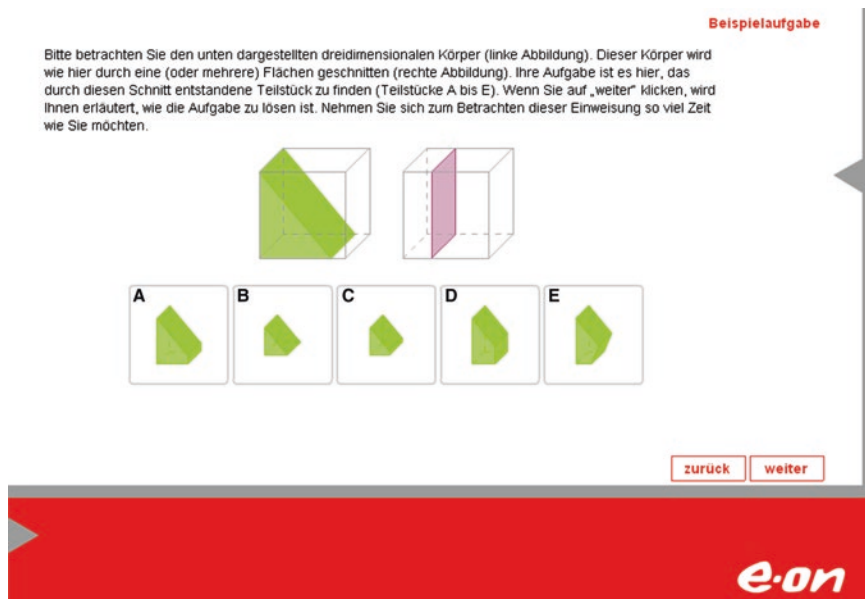


**Abb. 10.2** Unterscheidungsdimensionen von „Online-Assessments“. \*Fremdselektion (via beaufsichtigtem Testen); \*\*Selbstselektion (via Testen und Informieren, unbeaufsichtigt bzw. frei zugänglich); \*\*\*Selbstselektion (via Informieren, unbeaufsichtigt bzw. frei zugänglich)

beispielsweise die beaufsichtigte Fremdselektion und die Koppelung von Selbstselektion einerseits und Information andererseits. Konkrete Beispiele für gängige Kombinationen geben wir nach der Erläuterung der drei Dimensionen. Mit einem „Online-Assessment“ können auch verschiedene Elemente, die sich in der Taxonomie gegenüber stehen, gleichzeitig realisiert werden, beispielsweise kann ein „Online-Assessment“ sowohl informative als auch diagnostische (testende) Elemente enthalten.

### 10.2.1 Selbst- und Fremdselektion

Die Frage nach der Selbst- bzw. Fremdselektion betrifft die Frage, wer die mit dem Test erhobenen Daten interpretiert und eine Entscheidung trifft. Der klassische Fall der Personalauswahl sieht eine Interpretation der Daten durch die einstellende Organisation vor. Person X bewirbt sich bei Organisation Y, absolviert einen Test, danach treffen Entscheidungsträger aus der Organisation Y eine Entscheidung über den weiteren Fortgang der Bewerbung. In der praktischen Anwendung läuft es häufig so, dass Bewerber bei selektierenden „Online-Assessments“, nachdem sie in Form einer gelungenen Bewerbung den ersten Schritt des Auswahlprozesses absolviert haben, eine individuelle Einladung zur Teilnahme am „Online-Assessment“ erhalten. Solche „Online-Assessments“ finden Anwendung beispielsweise bei der Auszubildendenauswahl bei der Deutschen Post, E.ON oder der Targobank [68, 69, 70] (für ein Beispiel siehe Abb. 10.3).



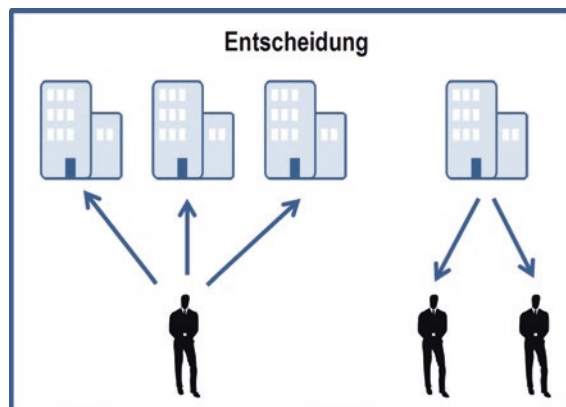
**Abb. 10.3** Beispiel für eine Aufgabe aus einem Online-Assessment. Beispielaufgabe aus dem „E.ON Phasenprüfer“. (Quelle: Cyquest GmbH)

Mittels „Self-Assessments“ können „Interessenten auch ohne eine Bewerbung etwas über die eigenen Stärken und Schwächen erfahren“ [31]. Bei der Selbstselektion kann eine Person X vor der eigentlichen Bewerbung in einem Test ausprobieren, ob ihr Eignungsprofil zu dem Anforderungsprofil passt (vgl. dazu Kapitel Palmer und Kersting in diesem Band). Die Testung kann dabei anonym erfolgen, d. h. dass nur Person X ihr Testergebnis kennt. Person X entscheidet dann über ihr weiteres Bewerbungsverhalten, z. B. über Bewerbung, keine Bewerbung, Bewerbung bei einer anderen Organisation oder auch das Annehmen eines Stellenangebots. „Self-Assessments“ (also der Selbstselektion dienende „Online-Assessments“) eröffnen einem Bewerber die Möglichkeit, Informationen über die eigene Passung zu einer Ausbildung, einem Beruf, einer bestimmten Stelle und/oder einer Organisation zu erhalten. Dies soll die „Selbstreflexion“ über die Passung von persönlicher Eignung und Anforderungen anregen und eine stimmige Selbstselektion hervorrufen. Die zugrunde liegende Annahme lautet, dass geeignete Bewerber sich für eine Bewerbung entscheiden, während weniger geeignete Bewerber entweder von der Bewerbung absehen oder aber weitere Informationen sowie ggf. Kompensationsangebote nachfragen. Bei denjenigen, die sich für eine Bewerbung entscheiden, kann das „Self-Assessment“ zu einer ersten Identifikation mit der Organisation und zur Steigerung des „commitment“ führen.

- Das „Self-Assessment“ soll die absolute Zahl der Bewerber (und somit den Aufwand) minimieren und den relativen Anteil an geeigneten Bewerbern in der Gesamtgruppe der Bewerber maximieren.

Da das „Self-Assessment“, wie die meisten „Online-Assessments“, kostenfrei, zeit- und ortsunabhängig durchgeführt werden kann, handelt es sich um ein niederschwelliges Angebot: Es ist für potenzielle Nutzer leicht zugänglich und mit einem überschaubaren Aufwand verbunden. Die Teilnahme an „Self-Assessment“ Angeboten erfolgt darüber hinaus – anders als „Online-Assessments“ zur Fremdselektion – in der Regel anonym. Ob eine geeignete Person sich nach Teilnahme an dem „Self-Assessment“ für die Bewerbung bei der Organisation entscheidet, hängt wesentlich davon ab, wie attraktiv die Organisation von den (potenziellen) Bewerbern wahrgenommen wird. In Zeiten des Personalmangels wird dieser Aspekt der Entscheidung des Kandidaten immer relevanter, so dass die Organisationen viel Wert auf eine „positive candidate experience“ legen. Die Attraktivität und Professionalität des „Self-Assessments“ hat eine hohe Bedeutung dafür, wie attraktiv die Organisation eingeschätzt wird. Die unterschiedlichen Entscheidungsperspektiven der Selbst- und Fremdselektion sind in Abb. 10.4 dargestellt (Abb. 10.4).

Cronbach und Gleser [12] grenzen die Personen-Selektion von der Bedingungs-Selektion ab. Bei der Personenselektion entscheidet sich eine Organisation für einen Bewerber. Bei der Bedingungsselektion wird eine Wahl zwischen verschiedenen Bedingungen getroffen. So könnte eine Organisation überlegen und entscheiden, in welchen Bereich der Organisation ein Bewerber am besten passt. Bei der Selbst-Bedingungsselektion werden beispielsweise den Testteilnehmern im Anschluss an die Teilnahme verschiedene, individuell für sie „passende“ Ausbildungsberufe empfohlen (oder einem Studieninteressierten werden passende Studiengänge empfohlen).



**Abb. 10.4** Entscheidungsträger im „Online-Assessment“



## 10.2.2 Testen und Informieren

Neben dem Einsatz von Tests kann ein „Online-Assessment“ auch dazu genutzt werden, Informationen zu vermitteln. Dies kann unabhängig von der Testung geschehen, indem z. B. Test- und Informationsblöcke sich abwechseln. Testung und Information können aber auch auf vielfältige Art und Weise miteinander verknüpft werden. So können z. B. über verschiedene Medien (Text, Audio, Video) dargebotene Informationen über eine Organisation oder Position anschließend in einer Art „Wissensquiz“ abgefragt werden. Oder ein „Online-Assessment“ startet direkt (ohne Vorabinformationen) mit Wissensfragen und die Teilnehmer erhalten eine informative Rückmeldung darüber, ob ihre Antworten richtig oder falsch waren. In diesem Sinne funktionieren beispielsweise so genannte „Erwartungschecks“, bei dem die Teilnehmer etwa Aussagen zu der infrage stehenden Berufsausbildung als „richtig“ oder „falsch“ einstufen sollen. Anschließend erhalten sie eine Erklärung zu dem Sachverhalt. Auf diese Art und Weise kann man verbreitete Fehlannahmen zu einer Ausbildung oder einem Beruf (z. B. „Lehrer haben ab mittags frei“) korrigieren. Durch eine derartige Testung wird die Informationsvermittlung interaktiv. Dadurch erhöht sich die Wahrscheinlichkeit, dass Informationen, die auf einer Website oder in einer Broschüre unbeachtet geblieben wären, zur Kenntnis genommen werden. Bei der Information kann es sich auch um eine Information über geeignete Lernwege und -methoden handeln. Im Idealfall beziehen sich z. B. Informationen zu Lernmodulen passgenau auf die vorher durch das Testen diagnostizierten individuellen Wissenslücken und Fähigkeits-/Fertigkeitsdefizite und ermöglichen so ein selbstgesteuertes Lernen.

Ein Vorteil von „Online-Assessments“ mit Informationen ist, dass hier alle für die Teilnehmer interessanten Informationen an einem (virtuellen) Ort gesammelt präsentiert werden. Die Organisation, die das „Online-Assessment“ gestaltet, kann entscheiden, welche Informationen der Teilnehmer erhält. Durch die Interaktivität und Kontrollfragen erhöht sich die Wahrscheinlichkeit, dass die Informationen – anders als Informationen in Broschüren oder auf Websites – auch zur Kenntnis genommen werden. Man kann sogar die Informationsaufnahme durch Testfragen kontrollieren. Auf diese Weise kann eine Organisation auch auf Berufsfelder oder Aspekte der Organisation hinweisen, die eventuell noch unbekannt sind und nach denen der Teilnehmer deshalb auch nicht aktiv suchen würde. Schließlich können auch die Testaufgaben selbst Informationen vermitteln, sodass „Testen“ und „Informieren“ keine Gegensätze sind. Wenn die Testaufgaben nämlich die Aufgaben simulieren, die auch im „wirklichen“ Leben (in der Ausbildung, im Studium, im Beruf) auf Menschen zukommen, so bieten diese Testaufgaben einen „realistic job preview“. Außerdem sind solche Aufgaben „augenscheinvalide“, was sich positiv auf die Akzeptanz auswirkt [30].

Die Unterscheidungsdimension „Testen/Informieren“ ist häufig mit der Unterscheidungsdimension Selbst- und Fremdelektion verbunden. Während sich Tools zur Selbst- und Fremdelektion (Abschn. 10.2.1) in Aufbau und verwendeter Methodik häufig

ähnlich sind, unterscheiden sie sich oft bezüglich des Anteils an Informationen, wobei die Selbstselektion mit einem höheren Informationsanteil einhergeht (Abb. 10.2). Wir benennen die Pole dieser Unterscheidungsdimension hier vereinfachend „Testen und Informieren“, korrekt müssten die Pole aber „ausschließliches Testen“ einerseits und „Kombination von Testen und Informieren“ andererseits genannt werden. Verfahren, die keine Testanteile umfassen und lediglich der Information dienen zählen nicht zur Diagnostik und sollten nicht „Online-Assessment“ genannt werden.

### 10.2.3 Durchführung mit und ohne „Aufsicht“ – Herausforderungen

Der Begriff des „online“ Testens wird häufig synonym verwendet zu einer Testung „ohne Aufsicht“. „Online“ bedeutet aber lediglich, dass ein Computer, auf dem die Testung stattfindet, sich in einem Netzwerk mit anderen Computern (z. B. via Internet oder Intranet) befindet. Teilnehmer, die einen Test online bearbeiten, können dabei sehr wohl beaufsichtigt werden. Der relevante Einteilungsgesichtspunkt besteht also nicht in der Frage „online“ oder „offline“, sondern in der Frage, ob die Testungen unter „Aufsicht“ oder „ohne Aufsicht“ durchgeführt werden. Die International Test Commission [62] unterscheidet vier Modi der Testadministration, die sich im Ausmaß an Aufsicht bzw. Kontrolle voneinander unterscheiden und auf einer übergeordneten Ebene den Kategorien „unbeaufsichtigt“ und „beaufsichtigt“ zugeordnet werden können (Tab. 10.1).

Im sogenannten „open mode“ liegt am wenigsten Kontrolle vor. Hierzu zählen online frei zugängliche Tools. Der „controlled mode“ sieht, ebenso wie der „open mode“, zwar keine menschlichen Testleiter vor, die Tools sind jedoch insofern zugangsbeschränkt, dass sie beispielsweise nur bestimmten Personen mit speziellen Zugangsdaten für einen bestimmten Zeitraum zugänglich sind. Eine teilweise Kontrolle bzw. starke Kontrolle durch menschliche Testleiter liegt im „supervised (proctored) mode“ bzw. im „managed mode“ vor. Der Unterschied beider Modi liegt darin, dass im „managed mode“ zusätzlich zur Anwesenheit eines menschlichen Testleiters Rahmenbedingungen wie der Ort der Testung von den Testdurchführenden festgelegt werden. Ein Beispiel für den „managed mode“ ist die beaufsichtigte Bearbeitung von computerbasierten Tests in den Räumen der auswählenden Organisation. Die ITC [62] subsumiert die ersten beiden Typen unter unbeaufsichtigte („unproctored“) und die letzten beiden Typen unter beaufsichtigte („proctored“) Testung.

**Tab. 10.1** Formen der Testbeaufsichtigung. (Quelle: [62])

Unbeaufsichtigte Testung („unproctored“)	<ul style="list-style-type: none"> <li>• Open mode</li> <li>• Controlled mode</li> </ul>
Beaufsichtigte Testung („proctored“)	<ul style="list-style-type: none"> <li>• Supervised mode</li> <li>• Managed mode</li> </ul>

Im Kontrast zu klassischen Auswahlverfahren finden „Online-Assessments“ und „Self-Assessments“ häufig in einer unbeaufsichtigten Umgebung (z. B. zu Hause) und ohne menschlichen Testleiter statt (d. h. im obigen Sinne „unproctored“). Daraus resultieren spezifische Herausforderungen für die Selbst- und Fremdselektion. Potenziell verzerrend auf die Ergebnisse kann sich auswirken, dass Teilnehmer ein „Online-Assessment“ unter ungünstigen oder, allgemeiner formuliert, unter ungleichen äußeren oder motivationalen Bedingungen bearbeiten. Hinzu kommt die Möglichkeit des Einsatzes unerlaubter Hilfsmittel, die nur bei beaufsichtigten Testungen vollständig kontrolliert werden kann. Im Folgenden berichten wir empirische Befunde, die auf die Äquivalenz (bzw. Divergenz) der Ergebnisse von beaufsichtigten und unbeaufsichtigten Testungen schließen lassen.

### **Effekte der unbeaufsichtigten Testdurchführung auf die Testergebnisse**

„Online-Assessments“ finden im Internet und häufig ohne Aufsicht statt. In der Literatur wird das unkontrollierte Testen bzw. das Testen ohne Aufsicht mit dem Begriff „unproctored Internet-based testing“ (UIT) bezeichnet [42]. Da die meisten empirischen Untersuchungen zur Validität von Personalauswahlverfahren sich auf das „klassische“, beaufsichtigte Setting beziehen, drängt sich die Frage auf, inwiefern UIT im Rahmen der Personalauswahl äquivalente Ergebnisse liefert. Durch die (relativ) unkontrollierte Testung im Rahmen eines „Online-Assessments“ resultiert eine spezifische Verfälschbarkeit, die die Validität ungünstig beeinflussen kann. In diesem Kontext ist allerdings zu beachten, dass die Intention, positive Ergebnisse vorzutäuschen („faking good“) vor allem bei der Fremdselektion, nicht aber bei der Selbstselektion zu erwarten ist. Eine weitere Verfälschungsmöglichkeit besteht in einem unkontrollierten Ausmaß an „Übung“, indem sich Bewerber unter verschiedenen Identitäten immer wieder Zugang zu einem „Online-Assessment“ verschaffen, bevor sie das Assessment unter dem eigenen Namen bearbeiten. Dies ist besonders problematisch, wenn die Bewerber in stark unterschiedlichem Maße Gelegenheit zur Übung hatten [65]. Natürlich könnten Bewerber auch andere (geeignete) Personen bitten, die Testaufgaben zu lösen.

Inkonsistente Testwerte zwischen UIT und beaufsichtigter Testung stellen einen Indikator für Effekte der Verfälschung dar. Das Ausmaß lässt sich beispielsweise durch wiederholtes Testen quantifizieren. Kantowitz und Dainis [28] analysierten die Daten eines Fähigkeitstests, den über 4000 Bewerber im Rahmen von Bewerbungen bei verschiedenen Organisationen zunächst unbeaufsichtigt und danach beaufsichtigt durchführten. Inkonsistente Testwerte fanden sich bei 6,4 % der Bewerber. Templer und Lange [59] konnten anhand einer Stichprobe potenzieller Bewerber für einen kognitiven Fähigkeitstest sowie mehrere berufsbezogene Persönlichkeitsverfahren keine systematischen Unterschiede zwischen UIT und beaufsichtigter Testung feststellen. Ployhart et al. [51] verglichen bei realen Bewerbern Papier-Bleistift-Testungen mit beaufsichtigten Online-Testungen (jeweils  $N > 2300$ ), wobei ein Persönlichkeitsfragebogen, ein biografischer

Fragebogen und ein Situational-Judgement-Test (SJT) zum Einsatz kamen. Zwar zeigten sich Divergenzen, jedoch sprachen diese aus statistischer und diagnostischer Sicht überwiegend für die Internet-basierte Testung (z. B. höhere Varianzen und interne Konsistenzen). Offen ist allerdings, inwiefern sich das auf „Online-Assessments“ zur Personalauswahl übertragen lässt.

Einen wichtigen Beitrag in diesem Bereich leisteten Beaty et al. [6]. In einer Metaanalyse verglichen sie Validitätskoeffizienten von Auswahlverfahren via UIT gegenüber beaufsichtigten Online-Testungen. Die Autoren stützten sich auf interne Daten eines Personaldienstleisters ( $k = 52$ ), der für Stellen in verschiedenen Branchen und Berufsfeldern Auswahlverfahren durchgeführt hat. Als Prädiktoren wurden die erfasste Gewissenhaftigkeit und Vertriebsbeignung sowie ein biografischer Fragebogen herangezogen, als Kriterien standen verschiedene Indikatoren des Berufserfolgs zur Verfügung. Die ermittelten Unterschiede in den Validitätskoeffizienten waren geringfügig und überwiegend nicht signifikant. Inwiefern sich diese Ergebnisse generalisieren lassen muss noch beantwortet werden, da der Großteil der Studien sich auf Stellen mit eher niedriger Komplexität bezog (z. B. Tätigkeit im Callcenter). Ebenso wurden nur nicht-kognitive Maße untersucht. Zu kognitiven Maßen liegt u. a. eine Studie von Wright et al. [66] vor. Die Autoren verglichen die Daten eines Fähigkeitstests, den Bewerber in beaufsichtigter und unbeaufsichtigter Testung bearbeiteten. In zwei großen Stichproben ( $N > 24.000$  bzw. 9000) zeigten sich zwar Testleistungsunterschiede (bezogen auf die theta-Werte) zwischen unbeaufsichtigter und beaufsichtigter Testung, die Effekte waren jedoch klein und nicht systematisch ( $d = ,20$  bzw.  $-,15$ ).

In der Literatur [49] wird ein Absinken von Testergebnissen von UIT zu einer nachfolgenden, beaufsichtigten Testung als Indiz für Verfälschung durch die Teilnehmer gewertet. Bedeutsame Leistungsverschlechterungen in diesem Sinne konnten jedoch Untersuchungen von Nye et al. [49] sowie Lievens und Burke [43] nicht bestätigen (stattdessen fanden sich teilweise Leistungsverbesserungen, womöglich aufgrund von Übungs- oder motivationalen Effekten).

Eine andere Frage, die sich bei der unbeaufsichtigten Testung stellt, ist die nach der Qualität des von der Testperson genutzten Gerätes zur Bearbeitung. Sofern keine Aufsicht über die Testung besteht, kann auch nicht kontrolliert werden, welches „Endgerät“ die Teilnehmer für die Testung nutzen (dies wäre nur im „managed mode“ möglich). Eine Herausforderung stellen diesbezüglich die mobilen Endgeräte dar. Arthur et al. [2] analysierten einen sehr großen Datensatz von über 3,5 Mio. Bewerbern, die über UIT Verfahren zur Erfassung der Persönlichkeit sowie kognitiver Fähigkeiten bearbeiteten. Eine Teilstichprobe von gut 69.000 Bewerber bearbeitete die Verfahren über Mobilgeräte. Für diese Gruppe konnte geprüft werden, ob die Testergebnisse in Abhängigkeit vom genutzten Gerät variieren. Die Autoren konnten in Bezug auf die Faktorstruktur der eingesetzten Skalen und die Skaleninterkorrelationen Hinweise auf Messinvarianz bzw. -äquivalenz finden. Bei den Skalenmittelwerten des Persönlichkeitsverfahrens gab es nur

geringfügige Unterschiede, beim Fähigkeitstest hingegen zeigten sich deutlich schlechtere Werte bei der mobilen gegenüber der nicht-mobilen Bearbeitung ( $d = ,90$ ). Morelli et al. [46] führten mit einem großen Datensatz von Bewerbern, die via UIT unter anderem einen kognitiven Fähigkeitstest (siehe dazu Kapitel von Kersting und Palmer in diesem Band), biografische Fragebogen (siehe dazu Kapitel von Strobel in diesem Band) und einen SJT (siehe dazu Kapitel von Franczukowska in diesem Band) bearbeiteten, eine ähnliche Untersuchung durch. Die Messinvarianz konnte für alle Verfahren bestätigt werden, im SJT zeigten sich jedoch schlechtere Leistungen bei der mobilen Durchführung. Arthur et al. [2] äußern die Vermutung, dass eine bessere Darstellung von Verfahren über Smartphone-Applikationen dazu beitragen könnte, derartige Unterschiede abzubauen. Hier ist sicherlich noch weitere Forschung und auch technische Entwicklungsarbeit nötig.

Zur Kontrolle von Verfälschung empfiehlt die ITC [62] im Falle von UIT bei „high-stakes“-Testungen (also z. B. im Auswahlprozess) eine nachfolgende Prüfung unter kontrollierten Bedingungen. Für die Durchführung eines beaufsichtigten Tests nach einem UIT stellen Guo und Drasgow [20] sowie Tendeiro et al. [60] statistische Verfahren vor, die eine systematische Identifikation inkonsistenter Testwerte erlauben.

Darüber hinaus variiert das Verfälschungsrisiko in Abhängigkeit von der Gestaltung des „Online-Assessments“. Die öffentliche Verbreitung von Aufgabenmaterial aus Leistungstests (vgl. hierzu die „Testknacker“-Ratgeberliteratur) oder auch der Einfluss von Testtrainings ist geringer, wenn Items randomisiert aus großen Item-Datenbanken gezogen werden. Allerdings lässt sich von bestimmten Aufgabentypen (z. B. Matrizenaufgaben) leichter eine große Item-Menge erzeugen, als von anderen Aufgabentypen (z. B. Items mit verbalem Inhalt). Es ist außerdem aus Sicht des Teilnehmers wesentlich leichter, sich eine Wissensfrage zu merken als eine Matrizenaufgabe. Dies kann dazu führen, dass bei „Online-Assessments“ und „Self-Assessments“ häufiger auf Aufgaben zum schlussfolgernden Denken mit figuralem Material zurückgegriffen wird. Die Auswahl der Aufgaben muss sich aber aus der Anforderungsanalyse ergeben und nicht aus pragmatischen Gesichtspunkten. Der Hinweis, dass bei Aufgaben zum schlussfolgernden Denken mit figuralem Material ein höherer Testschutz besteht ist unerheblich, wenn der gedankliche Umgang mit figuralem Aufgabenmaterial für die infrage stehende Ausbildung resp. den Beruf irrelevant ist.

#### 10.2.4 Kombination der Gestaltungselemente

Die hier behandelten drei Unterscheidungsdimensionen sind unabhängig voneinander. Diese prinzipielle Kombinierbarkeit von Gestaltungsmerkmalen stellt ein großes Potenzial von „Online-Assessments“ dar und ermöglicht deren Anpassung an den jeweiligen Bedarf (siehe die Anwendungsfelder nach Hiltmann [25]). Ungeachtet dessen sind einige Kombinationen sinnvoller und häufiger als andere. So sind die in Tab. 10.2 aufgeführten Kombinationen für bestimmte Anwendungen besonders gängig.

**Tab. 10.2** Gängige Kombinationen von Gestaltungselementen des „Online-Assessments“ für bestimmte Einsatzziele. (Quelle: eigene Darstellung)

	Gestaltung					
	Selektion		Funktionsmechanismus		Darbietungsweise	
Ziel des „Online-Assessment“	Selbst	Fremd	Testen	Informieren	Beaufsichtigt	Unbeaufsichtigt
Personal-Vorauswahl (Screening)		●	●			●
Personal-Endauswahl		●	●		●	
Self-Assessment	●		●	●		●

Aus der Bezeichnung „Online-Assessment“ allein lässt sich weder ableiten, um welche Art der Diagnostik es sich handelt, noch lassen sich Rückschlüsse auf eine hohe oder niedrige Qualität ziehen. Um qualitativ hochwertige „Online-Assessments“ zu erkennen oder zu gestalten, muss man einige Besonderheiten berücksichtigen, die sich aus dem englischen Wort „online“ in der Bezeichnung „Online-Assessment“ ergeben. Eine dieser Besonderheiten ist die Frage nach der Beaufsichtigung der Testung, die wir weiter oben bereits thematisiert haben. In den folgenden Abschnitten gehen wir auf weitere Besonderheiten des „Online-Assessments“ ein.

### 10.3 Spezifika der Durchführung als Folge des Online-Charakters

Die Gründe, die zur zunehmenden Anwendung von „Online-Assessments“ (und spezifischer von „Self-Assessments“) führen sind vielfältig. Zum einen liegt es an dem immer weiter zunehmenden Stellenwert, den das Internet bei den heutigen potenziellen Bewerbern hat. Bei der Zielgruppe solcher Rekrutierungsverfahren ist das Internet die erste und zumeist auch die einzige Informationsquelle im Berufswahl- oder Bewerbungsprozess, wodurch die Präsenz von Organisationen in diesem Bereich inzwischen unabdingbar ist. In einer Umfrage der Universität Erfurt gaben insgesamt 85 % der Schüler einer gymnasialen Oberstufe an, das Internet genutzt zu haben, um Informationen bezüglich des Studiums oder Berufs zu recherchieren [17]. Das Internet stellt nach dieser Studie bei Gymnasiasten in der Oberstufe die wichtigste Informationsquelle im Berufsorientierungsprozess dar. Es ist davon auszugehen, dass diese Zahl in den vergangenen Jahren noch weiter angestiegen ist.

Zum anderen spielt die Attraktivität eines Arbeitgebers eine wichtige Rolle wenn es darum geht, Bewerber auf die Organisation aufmerksam zu machen. Dabei geht es nicht nur um die Attraktivität der tatsächlichen Position oder Karriere, sondern auch darum, wie attraktiv sich eine Organisation präsentiert, beispielsweise durch ihren Online-Auftritt. Wie zuvor dargestellt, müssen viele Organisationen verstärkt um die Aufmerksamkeit qualifizierter Bewerber kämpfen, wobei ein attraktiver Auftritt eine wichtige Rolle spielt. Viele Organisationen können es sich bei dem momentanen Bewerbermangel nicht mehr leisten, geeignete Bewerber zu übersehen [32]. Die Erweiterung von Fremdselektionsverfahren vom klassischen Papier-Bleistift-Modus auf den Computer und (darüber hinaus) ins Internet sowie die Ergänzung um die Möglichkeit zur Selbstselektion stellen aktuell eine große Herausforderung für die Personalauswahl dar.

Mittlerweile gibt es umfangreiche Forschungsarbeiten zu den Besonderheiten computerbasierter Diagnostik. Dies zu vertiefen würde den Umfang dieses Beitrags übersteigen (siehe dazu z. B. [4, 35]). Stattdessen soll die Aufmerksamkeit auf drei Aspekte gerichtet werden, die sich spezifisch aus dem online und häufig ohne Aufsicht stattfindenden diagnostischen Prozess ergeben: 1) Usability und Attraktivität von Online-Tools, 2) der wahrgenommene Datenschutz und die wahrgenommene Fairness sowie 3) interindividuelle Unterschiede in der Akzeptanz von Online-Tools.

### 10.3.1 Usability und Attraktivität von Tools

Wer im Internet eine „gute Figur“ abgeben möchte, muss für einen professionellen Internetauftritt sorgen. Ein intuitiv verständliches und optisch attraktives Tool wird Akzeptanz erfahren, optisch unzeitgemäße oder fehleranfällige Tools werden zu Frustration und Ablehnung führen. Dass sich eine ansprechende und benutzerfreundliche Website auch auf Bewerberreaktionen auswirkt konnten Braddy et al. [10] zeigen. Die Wahrnehmung des Unternehmens durch die Teilnehmer war dabei maßgeblich durch die Gestaltung der Website beeinflusst. Während einfach zu navigierende, attraktive Seiten zu positiven Einschätzungen in Bezug auf das Unternehmen führten, hatte der Umstand, ob die einschätzenden Personen das Unternehmen kannten oder nicht nur wenig Einfluss auf die Einschätzung.

Die Gestaltung der Online-Tools spielt auch eine Rolle wenn es darum geht, möglichst viele Bewerber für eine Stelle anzuwerben. Deshalb sind „Online-Assessments“ ungeachtet ihrer diagnostischen Funktion immer auch Instrumente des Personalmarketings. In der Literatur [11, 58] wird die Annahme vertreten, dass Merkmale eines Online-Auswahlsystems, wie es beispielsweise ein „Online-Assessment“ ist, dem Bewerber als Informationsbasis für ein Urteil über die Organisation dienen. Dabei fließen gerade leicht verfügbare Informationen, wie z. B. die Benutzerfreundlichkeit eines entsprechenden Tools, die nicht direkt mit der zu besetzenden Stelle in Bezug stehen, in die Urteilsbildung ein. Sinar et al. [56]



befragten über 23.000 Bewerber, die sich online bei verschiedenen Unternehmen bewarben und dabei auf der Webseite Fragen beantworten mussten (z. B. einen biografischen Fragebogen), die zur Vorselektion genutzt wurden. Es zeigten sich deutliche positive Korrelationen zwischen der Einstellung gegenüber Organisationen, die das Internet zur Personalauswahl nutzen und der subjektiven Geschwindigkeit und Benutzerfreundlichkeit der jeweiligen Webseite ( $r = ,39$  bzw.  $,41$ ; partielle Korrelationen, kontrolliert um subjektiven Stellenbezug der Fragen). Braddy et al. [10] konnten ebenfalls positive Effekte der Benutzerfreundlichkeit von Rekrutierungswebseiten mehrerer großer US-Unternehmen auf die Einstellung potenzieller Bewerber gegenüber den jeweiligen Unternehmen demonstrieren.

### 10.3.2 Wahrgenommener Datenschutz und wahrgenommene Fairness

Die wahrgenommene vertrauliche Behandlung der Daten ist ein entscheidender Punkt für die Akzeptanz eines „Online-Assessments“. Dieser Aspekt ist umso wichtiger, je größer der Testanteil in einem Tool wird. Harris et al. [21] fanden in einer Untersuchung mit belgischen und US-amerikanischen Studierenden eher schwache Zusammenhänge zwischen Bedenken bezüglich Datenschutz und -sicherheit sowie der ablehnenden Haltung gegenüber onlinebasierten Auswahlprozessen. Interessanterweise fand sich aber (bei den US-amerikanischen Studierenden) eine deutliche Korrelation ( $r = ,50$ ) zwischen der ablehnenden Haltung gegenüber onlinebasierten Auswahlprozessen und der Sorge, man könne aufgrund technischer Probleme in einem Online-Auswahlverfahren benachteiligt werden. Bauer et al. [5] konnten sowohl anhand einer Labor- als auch anhand einer Feldstudie zeigen, dass bei Bewerbungen Bedenken bezüglich des Datenschutzes einen negativen Effekt auf die wahrgenommene Fairness des Auswahlprozesses (siehe dazu Kapitel von Kanning in diesem Band) haben, welche wiederum die wahrgenommenen Attraktivität der Organisation sowie die Zuversicht, bei guter Bewerbungsleistung die Stelle zu erhalten, prädiziert. Die Akzeptanz eines Online-Auswahlprozesses hängt nicht nur von äußeren Merkmalen der Webseite, wie der Geschwindigkeit des Online-Systems und dessen Benutzerfreundlichkeit, ab [58], sondern auch von Aspekten wie dem wahrgenommenen Datenschutz, der wahrgenommenen Fairness sowie der Berufsbezogenheit des Inhalts und der wertschätzenden Interaktion des Systems mit dem Bewerber (z. B. keine sittenwidrigen Fragen) (für die letzten beiden Aspekte siehe [37]).

### 10.3.3 Interindividuelle Unterschiede bezüglich der Akzeptanz von Online-Testungen

Hinsichtlich der Akzeptanz von „Online-Assessments“ ist zu beachten, dass die Reaktionen der Bewerber auch von interindividuell variierenden Merkmalen der Bewerber abhängen. Wiechmann und Ryan [64] führten eine Studie durch, in der Probanden in einer simulierten Auswahl-situation eine Postkorbübung bearbeiten mussten. Die selbst-

eingeschätzte, spezifische Erfahrung (z. B. Programmierkenntnisse) mit Computern hatte einen signifikant positiven Effekt auf die Leistung in der computerbasierten Darbietungsform ( $b = ,31$ ). Die Scheu vor dem Umgang mit Computern hingegen hatte deutliche negative Effekte auf die Akzeptanz des Verfahrens, so auf die erlebte Fairness ( $b = -,24$ ), die Augenscheinvalidität ( $b = -,40$ ) und darauf, wie angenehm der Test empfunden wurde ( $b = -,36$ ). Diese Aspekte sollten berücksichtigt werden, da sie einen Einfluss darauf haben, ob „Online-Assessments“ ihr Ziel erreichen oder nicht. So zeigte sich in der bereits zitierten Untersuchung von Harris et al. [21] – zumindest in der US-amerikanischen Teilstichprobe – ein negativer Zusammenhang zwischen dem selbstberichteten Wissen über das Internet und den Bedenken bezüglich des Datenschutzes bei onlinebasierten Auswahlverfahren ( $r = -,25$ ).

Es ist davon auszugehen, dass sich Bewerber in ihrem Wissen über das Internet und in ihren spezifischen Erfahrungen mit Computern unterscheiden. Daraus erwächst die Gefahr, dass, wenn nicht die Ergebnisse selbst, so doch zumindest die Akzeptanz von „Online-Assessments“ in Abhängigkeit von Merkmalen der Bewerber unterschiedlich ausfallen. Sofern es sich bei der Erfahrung im Umgang mit Computern (oder auch der Scheu davor) um ein eignungsirrelevantes Merkmal handelt, welches die diagnostische Entscheidung (oder Empfehlung) beeinflusst, handelt es sich um einen „bias“ [53], der das Eignungsurteil verfälscht. Es darf allerdings die optimistische Vermutung geäußert werden, dass die Erfahrung im Umgang mit dem Internet und Computern bei den sogenannten „digital natives“ [52], zu denen zumindest ein Großteil der heutigen Berufseinsteiger zählen dürfte, als weitgehend gegeben erachtet und somit vorausgesetzt werden kann. Inwieweit dies bei anderen (z. B. lebensälteren) Bewerbergruppen ebenfalls angenommen werden kann, bedarf weiterer Forschungsaktivität. Untersuchungswert sind auch geschlechtsspezifische Erfahrungsunterschiede mit dem Internet sowie Unterschiede in der Einstellung gegenüber Computern und ihre Auswirkungen auf die Akzeptanz.

---

#### 10.4 Bewertung der Qualität von „Online-Assessments“ und „Self-Assessments“

Neben den bereits behandelten Aspekten der Usability, des wahrgenommenen Datenschutzes und der wahrgenommenen Fairness kommt bei „Online-Assessments“ mit Testelementen auch und vor allem der Qualität dieser Testung eine hohe Bedeutung zu. Die Bewertung der Qualität der Testung kann im Sinne des Testbeurteilungssystems des Testkuratoriums erfolgen (TBS-TK, Testkuratorium der Föderation Deutscher Psychologenvereinigungen [61]). Das TBS-TK-System berücksichtigt bereits einige Qualitätsforderungen der DIN 33430, einer Norm zur berufsbezogenen Eignungsbeurteilung [16, 29]. Diese Norm formuliert aber auch Anforderungen an die Eignungsdiagnostik, die über die Qualität des Tests (siehe dazu Kapitel von Melchers in diesem Band) hinausgehen und bei der Bewertung von „Online-Assessments“ ebenfalls zu beachten sind. Hierzu zählt beispielsweise die Forderung nach einem eindeutigen Anforderungsbezug

oder nach einer regelgeleiteten Durchführung und Auswertung der Verfahren sowie einer regelgeleiteten Interpretation der Verfahrensergebnisse sowie Urteilsbildung.

Weitere Qualitätsstandards beziehen sich spezifisch auf computergestütztes Testen und thematisieren etwa den Datenschutz, die technischen Voraussetzungen und die Frage, wer eigentlich z. B. für „automatisch“ erstellte Diagnosen verantwortlich ist. Einschlägig sind hier beispielsweise die „International Guidelines on Computer-Based and Internet-Delivered Testing“ der International Test Commission [62] oder die „Guidelines for computer-based testing“, die von der Association of Test Publishers [3] herausgegeben wurden. Wir greifen hier aus der Vielzahl der Qualitätsaspekte für „Online-Assessments“ vier Teilbereiche heraus: die Validität, den Nutzen, die Akzeptanz sowie die Qualität der Rückmeldung, die die Teilnehmer erhalten.

### **Validität**

Die Validität von Entscheidungen, die aus diagnostischen Verfahren abgeleitet wurden, ist in Bezug auf die Vorhersage des beruflichen Erfolgs vielfach belegt worden [55]. Auch die Vorhersagekraft von „Self-Assessments“ im Rahmen der Studienwahl für Studiennoten und Studienzufriedenheit ist durch eine wachsende Zahl an Untersuchungen belegt (eine Übersicht liefert [22]). Bislang liegen allerdings für „Online-Assessments“ in der Personalauswahl – anders als für den traditionellen Einsatz von Personalauswahlverfahren – noch keine umfangreichen Validitätsdaten vor. Die Validität wurde bislang vor allem bezüglich einzelner Tests überprüft, kaum jedoch für gesamte „Online-Assessments“. Die weiter oben geschilderten Befunde zur Testung ohne Aufsicht erlauben einen vorsichtigen Optimismus und geben Anlass zu der Vermutung, dass die Validität weniger eine Frage des Darbietungsmediums (Papier oder Bildschirm) ist, sondern – nicht überraschend – eine Frage der Qualität der diagnostischen Verfahren, z. B. des Fragebogens und Tests, selbst.

### **Nutzen**

Den Nutzen von „Online-Assessments“ können zwei Praxisbeispiele aus dem Personalbereich veranschaulichen. In Bezug auf die „Online-Assessments“ der Deutschen Lufthansa AG berichtet Wiedmann [65], dass sich in einem mehrstufigen Selektionsprozess durch die Vorselektion aufgrund der „Online-Assessments“ eine bessere Basisrate bei späteren Stufen im Auswahlprozess erzielen lässt. Dies führt zu deutlichen Kosteneinsparungen, da sich der Anteil der zu beurteilenden Bewerber in späteren (meist teureren) Auswahlstufen (z. B. Assessment-Center) erheblich reduziert. Wurow [67] beschreibt ein „Online-Assessment“ der Conergy AG, das zur internen Stellenbesetzung eingesetzt wird. Auf Basis eines darin administrierten verhaltensbasierten Fragebogens werden von den Bewerbern Profile ermittelt. Diese Profile werden in späteren Auswahlritten, wie strukturierten Interviews, aufgegriffen. Es wird von einer erhöhten Validität und reduzierten Kosten berichtet.

Während die beiden Beispiele den Nutzen aus der Perspektive der Organisation betrachten, zeigt die folgende Studie des „Centre of Human Resources Information Systems“ (CHRIS) der Universitäten Frankfurt am Main und Bamberg den subjektiven Nutzen von

„Self-Assessments“ im Bewerbungsprozess aus der Sicht der Nutzer (sprich den Bewerbern selbst) auf. Für ein „Self-Assessment“ ist zu verlangen, dass die Person nach der Teilnahme mehr über ihre eigene Passung zu den Anforderungen weiß als zuvor und sich zu einer fundierten Entscheidung für oder gegen eine Bewerbung befähigt fühlt. Zwischen 2008 bis 2012 wurden für die CHRIS-Studie über 26.000 (potenzielle) Bewerber u. a. nach dem subjektiven Nutzen von „Self-Assessments“ befragt [18]. Das Urteil über „Self-Assessments“ fiel überwiegend positiv aus. So gaben (für das Befragungsjahr 2008) jeweils über 40 % der Befragten an, ein „Self-Assessment“ biete eine gute Rückmeldung über die eigene Eignung für eine Position und erhöhe die Effektivität und Qualität der Bewerbung. Immerhin 33 % gaben an, dass „Self-Assessments“ die Bewerbungsentscheidung erleichtern. Kritisch zu werten ist das mangelhafte Vertrauen in den Datenschutz sowie in die Aussagekraft. Nur etwa 30 % der Befragten waren überzeugt, dass die Daten nicht an Dritte weitergeleitet würden und nur etwa ein Fünftel äußerte Vertrauen in das Ergebnis eines „Self-Assessments“.

### **Soziale Akzeptanz**

Auf die hohe Bedeutung der sozialen Akzeptanz sind wir weiter oben bereits eingegangen. „Online-Assessments“ die unbeaufsichtigt dargeboten werden, sind aufgrund der allgemeinen Zugänglichkeit ein mediales Aushängeschild einer Organisation und beeinflussen somit nicht nur die Selbst- oder Fremdselektion, sondern auch das Ansehen jener Organisationen, welche sie einsetzen. Eine mangelnde Akzeptanz von Personalauswahlverfahren kann negative Konsequenzen für die auswählende Organisation und deren Reputation haben [23, 30]. Die Bewertung der Qualität von „Online-Assessments“ muss sich also nicht nur mit den Hauptgütekriterien, sondern auch mit dem Nebengütekriterium der sozialen Akzeptanz auseinandersetzen.

### **Rückmeldung**

Mit den Reaktionen der Bewerber ist ein weiteres, spezifisches Qualitätsmerkmal von „Online-Assessments“ assoziiert, nämlich die Qualität der automatisiert erstellten Rückmeldung. Ilgen et al. [26] stellten in einer Literaturübersicht ein Modell vor, das eine Prozesskette zwischen Feedback und der Reaktion von Personen auf das Feedback beschreibt. Große Bedeutung kommt hier der Akzeptanz des Feedbacks zu, also dem subjektiven Urteil über die Richtigkeit der im Feedback enthaltenen Aussagen über die eigene Person. Nach den von Ilgen et al. [26] zusammengetragenen Befunden neigen Menschen dazu, positives Feedback eher zu akzeptieren als negatives. Dies kann durch das Selbstaufwertungsmotiv („self-enhancement“) erklärt werden [7, 47]. Anseel und Lievens [1] demonstrierten dies qua manipulierter Passungsrückmeldungen auf Basis eines Persönlichkeitsfragebogens: Der (positive) Zusammenhang zwischen der Rückmeldung (niedrige vs. hohe Passung) und der Einstellung gegenüber der auswählenden Organisation wurde durch die Akzeptanz des Feedbacks mediiert. Beim Einsatz von Persönlichkeitsskalen ist zusätzlich zu beachten, dass Menschen dazu tendieren, allgemeingültige und unkonkrete Beschreibungen ihrer eigenen Person als zutreffend zu erachten (der sog. Barnum-Forer-

Effekt [19, 33, 45]). Auch das Selbstbestätigungsmotiv („self-verification“ [47]) kann einer wirksamen Selbstselektion entgegenstehen, beispielsweise wenn eine negative Passungsrückmeldung mit stark ausgeprägten Ausbildungs- oder Berufswünschen kollidiert. In einem Versuch, diese sich widersprechenden Motive zu integrieren, postulierten Swann und Schroeder [57], dass mit zunehmender Verarbeitungstiefe – welche wiederum abhängt von aktueller Motivation und verfügbaren kognitiven Ressourcen – verstärkt die Plausibilität von selbstbezogenen Informationen überprüft wird (Selbsteinschätzungsmotiv oder auch „self-assessment“<sup>1</sup> [47]). Feedbacks in einem „Self-Assessment“ sollten möglichst konkret und anforderungsbezogen sein und darüber hinaus die Bewerber zu einer sorgfältigen Verarbeitung der Rückmeldungsinformationen ermutigen [1].

Diese Verarbeitung ist stets risikobehaftet. Aufgrund der automatisierten Befundung besteht keine Möglichkeit, Missverständnisse oder Fehlinterpretationen aufzufangen. Dieser Tatsache ist durch eine überaus sorgfältige und zurückhaltende Gestaltung der Feedbacks Rechnung zu tragen. So sollte bei Testungen ohne Aufsicht beispielsweise darauf hingewiesen werden, dass die erzielten Ergebnisse auch davon abhängen, unter welchen äußeren und motivationalen Bedingungen die Tests bearbeitet wurden. Darüber hinaus sollte bei „Self-Assessments“ darauf hingewiesen werden, dass neben dem Verfahrensergebnis auch weitere Informationen, wie z. B. die Schulnoten sowie weitere Beratungen (z. B. durch Eltern und Freunde) bei der Entscheidung (d. h. der Selbstselektion) berücksichtigt werden sollten. Sinnvoll ist auch der Hinweis auf konkrete externe Beratungsangebote. Die Teilnehmer sollten bei der Auseinandersetzung mit den Feedbacks in ihrer Reflexion geleitet und dazu angehalten werden, die Konsistenz oder Inkonsistenz verschiedener Informationen (Ergebnis im „Self-Assessment“, Schulnoten, Fremdeinschätzungen von z. B. Eltern, Freunden usw.) zu berücksichtigen. Insgesamt ist die Wirkung der aus „Online-Assessments“ abgeleiteten automatischen Feedbacks bislang zu wenig erforscht.

---

## 10.5 Fazit

### 10.5.1 Was bei der Konstruktion und dem Einsatz von „Online-Assessments“ und „Self-Assessments“ zu beachten ist

„Online-Assessments“ und „Self-Assessments“ bieten ein großes Potenzial für die Personalauswahl. Sie können, wie eingangs bereits erwähnt, orts- und zeitunabhängig bearbeitet werden, ermöglichen automatisierte und damit in hohem Maße objektive Testauswertungen und -rückmeldungen und bieten einen großen Spielraum in der Umsetzung von Testformaten [54]. Wie in Abschn. 1.2 dargelegt wurde, lassen sich Elemente von Selbst- und Fremdselektion kombinieren, sodass „Online-Assessments“ sowohl

---

<sup>1</sup>Der Begriff ist in dieser Verwendung nicht zu verwechseln mit dem in diesem Beitrag beschriebenen Online-Tool.

in Funktion als auch bezüglich des Inhalts (Testen vs. Informieren) an die jeweiligen Bedürfnisse der Situation angepasst werden können. Angesichts der weiter wachsenden Bedeutung des Internets für alle Lebensbereiche ist davon auszugehen, dass der Einsatz von „Online-Assessments“ für Personalauswahl und -marketing noch zunehmen wird. Sowohl Bewerber als auch Organisationen können davon profitieren. „Online-Assessments“ und speziell „Self-Assessments“ können jedoch nur dann ihre Ziele erreichen, wenn bestimmte Anforderungen erfüllt sind. Diese lassen sich in drei Felder gliedern: Anforderungen an die eingesetzten Verfahren, an die Technik und den Datenschutz sowie an die sinnvolle Integration in den diagnostischen Prozess.

Die in „Online-Assessments“ und „Self-Assessments“ eingesetzten diagnostischen Verfahren müssen den Anforderungen der Norm DIN 33430 für die berufsbezogene Eignungsbeurteilung [16, 29] sowie des TBS-TK [61] und den weiter oben genannten computerspezifischen Qualitätsstandards genügen [3, 62]. Hervorzuheben ist die Forderung nach einem eindeutigen Anforderungsbezug sowie die Forderung nach möglichst hoher Objektivität, Reliabilität und Validität.

Ebenso wichtig ist die professionelle und bedarfsgerechte technische Umsetzung des „Online-Assessments“. Laier [41] führt als Maßstab für die technische Qualität von Online-Tests und damit auch von „Online-Assessments“ und „Self-Assessments“ die Norm DIN EN ISO 9241-110 [15] zu Grundsätzen der Dialoggestaltung auf und stellt vier zentrale Aspekte heraus, die erfüllt werden müssen: 1) die Aufgabenangemessenheit, also die effektive und effiziente Bedienbarkeit durch den Benutzer, 2) die Steuerbarkeit des Prozesses, 3) die Erwartungskonformität, also eine erkennbare logische Struktur des Tools, und schließlich 4) die Fehlertoleranz. Wie bereits erwähnt spielt der Datenschutz auch eine zentrale Rolle für die Akzeptanz der Bewerber. Organisationen sollten hier in jedem Fall die jeweiligen gesetzlichen Vorgaben einhalten (in Deutschland beispielsweise das Bundesdatenschutzgesetz [BDSG]).

Schließlich sind die Besonderheiten der onlinebasierten Diagnostik zu berücksichtigen. Spezielle Formen des „Online-Assessments“ finden in einem weitgehend unkontrollierten Setting statt. Da es hierbei keinen Testleiter gibt, der prüfen kann, ob ein Bewerber Hilfsmittel nutzt oder sich auf andere Weise Vorteile verschafft, bestehen verschiedene Risiken. Hierzu zählen die Verfälschung, Übungseffekte, die Verbreitung von Aufgabenmaterial und -lösungen sowie das Risiko, dass die Assessments mit ungeeigneten Endgeräten bearbeitet werden. In der Regel sollten „Online-Assessments“ ohne Aufsicht daher nur zur Vorselektion von Bewerbern eingesetzt werden (siehe hierzu [40]) (Tab. 10.3).

Ansätze, die zuerst genannten drei Problematiken zu entschärfen sind beispielsweise die randomisierte Darbietung von Items aus großen Itemdatenbanken [54], die Testwiederholung unter kontrollierten Bedingungen bzw. die entsprechende Ankündigung der selbigen sowie die Aufklärung über Konsequenzen von Verfälschung [62, 65].

- ▶ Falls die Teilnehmer die Testung unbeaufsichtigt absolviert haben, ist dies bei der Gestaltung des Feedbacks zu berücksichtigen. Es ist denkbar, dass die Teilnehmer das „Online-Assessment“ unter ungünstigen äußeren oder motivationalen

**Tab. 10.3** Checkliste zu ausgewählten Anforderungen an „Online-Assessments“. (Quelle: eigene Darstellung sowie [41])

Sind die Anforderungen an diagnostische Verfahren gemäß TBS-TK und DIN 33430 erfüllt?	Insbesondere <ul style="list-style-type: none"> <li>• Anforderungsbezug</li> <li>• Objektivität, Reliabilität und Validität</li> <li>• Prüfung der Gültigkeit der Normwerte</li> <li>• Akzeptanz</li> </ul>
Ist das „Online-Assessment“ technisch bedarfsgerecht umgesetzt gemäß DIN EN ISO 9241-110?	Insbesondere <ul style="list-style-type: none"> <li>• Bedienbarkeit</li> <li>• Steuerbarkeit</li> <li>• Erwartungskonformität</li> <li>• Fehlertoleranz</li> </ul>
Ist der Datenschutz gewährleistet nach dem Bundesdatenschutzgesetz (BDSG)? Sind folgende Datenschutzerfordernungen erfüllt?	<ul style="list-style-type: none"> <li>• Datenvermeidung und -sparsamkeit</li> <li>• Erfassung nur von erforderlichen Daten</li> <li>• Zweckbindung der erfassten Daten</li> <li>• Transparenz bezüglich der Verwendung der Daten</li> </ul>
Wie ist das „Online-Assessment“ in den diagnostischen Prozess integriert?	<ul style="list-style-type: none"> <li>• Fokus auf Selbst- vs. Fremdselektion</li> <li>• Wird das „Online-(Self-)Assessment“ zur Vorauswahl genutzt?</li> <li>• Ist die Rückmeldung konkret, weist sie Handlungsempfehlungen auf?</li> <li>• Vorkehrungen gegen und Kontrolle von Verfälschungstendenzen</li> </ul>

Bedingungen bearbeiten, was die Ergebnisse verzerrt. Auf diesen Aspekt ist bei der Ergebnismeldung hinzuweisen. Außerdem besteht die Gefahr, dass Teilnehmer die „Diagnose“ überbewerten. Diesem Risiko ist durch eine sorgfältige und zurückhaltende Gestaltung der Feedbacks Rechnung zu tragen.

### 10.5.2 Forschungsdesiderata

Zukünftige Forschungsarbeiten sollten sich u. a. damit beschäftigen, unter welchen Bedingungen eine erfolgreiche Selbstselektion stattfinden kann. Dies betrifft die grundlegende Frage, ob Bewerber sich in Online-Tools zur Selbstselektion anders verhalten als in jenen zur Fremdselektion. Bewerber als „sich Bewerbende“ möchten naturgemäß einen positiven Eindruck bei einem potenziellen Arbeitgeber hinterlassen. Eine „erfolgreiche“ Selbstselektion, also eine die eigene Passung angemessen reflektierende Bewerbungsentscheidung, stützt sich indes auf eine zutreffende Selbsteinschätzung.

Ausgehend von Paulhus [50] ist sozial erwünschtes Antwortverhalten in zwei verschiedene Faktoren differenzierbar: Selbsttäuschung („self deception“) betrifft Selbstbeschreibungen, die unangemessen positiv sind, wobei die Person jedoch von der Richtigkeit der Selbsteinschätzung überzeugt ist. „Impression management“ bedeutet demgegenüber das



(gezielte) Bestreben, ein positives Bild von der eigenen Person nach außen zu transportieren. Während bei der Fremdselektion insbesondere mit „impression management“ seitens der Bewerber zu rechnen ist, spielt bei der Selbstselektion die Selbsttäuschung eine größere Rolle.

Zum Phänomen des „impression management“, insbesondere zur aktiven Verfälschung („faking“) sowie möglicher Gegenmaßnahmen liegt inzwischen umfangreiche Forschungsarbeit vor. Metaanalytisch zeigen sich beispielsweise unterschiedliche Mittelwerte in Skalen von Persönlichkeitsfragebogen, sowohl unter experimenteller Instruktion von Verfälschung („fake good“ bzw. „fake bad“ [63]) als auch beim Vergleich von Bewerberstichproben gegenüber Nicht-Bewerberstichproben [8]. „Impression management“ kann bei der Personalauswahl problematisch sein [27] und stellt zweifellos eine Herausforderung für die Fremdselektion dar.

Unklar ist aber, in welchem Ausmaß „impression management“ – und insbesondere auch „self deception“ – bei der Selbstselektion wirksam ist. Marcus [44] schlägt ein Prozessmodell vor, in dem die positive Selbstdarstellung in der Personalauswahl als Funktion u. a. von der Attraktivität einer Position, der Fähigkeit zur Selbstdarstellung und der subjektiven Instrumentalität der Selbstdarstellung verstanden wird. Dieser Prozess beginnt nach der Bewerbungsentscheidung – die Selbstselektion ist an dieser Stelle bereits abgeschlossen. Es ist möglich, dass „impression management“ und „self-deception“ insofern die Selbstselektion beeinträchtigen, indem sie eine valide Selbsteinschätzung behindern. Eine Beurteilung der eigenen Passung enthält Beurteilungen über die eigene Person oder, allgemeiner gesprochen, die eigene Identität oder das eigene Selbstkonzept. Bozeman und Kacmar [9] beschreiben ein Modell zur Erklärung von „impression management“ in Organisationen. Eine Implikation des Modells ist, dass beim Vergleich der aktuellen und der erwünschten sozialen Identität ein Bedarf besteht, die eigene Identität anzupassen, zu schützen oder auch aufrecht zu erhalten. Dieser Bedarf lässt sich durch „impression management“ decken. In dem Maße wie diese Prozesse am Werk sind wird sich die Selbstselektion zu einem großen Teil nur auf eine verzerrte Selbsteinschätzung stützen können.

Während zum Thema „faking“ und möglicher Gegenmaßnahmen in der Fremdselektion zahlreiche empirische Untersuchungen vorliegen, besteht bezüglich des Selbstkonzepts einerseits und der Selbstselektion andererseits noch Forschungsbedarf. Damit zusammen hängend ist zu klären, ob Bewerbern auch bei der Selbstselektion dieselbe grundlegende Motivation zur positiven Selbstdarstellung unterstellt werden kann wie bei der Fremdselektion – oder ob letzteres Phänomen erst dann auftritt, wenn der Rubikon überschritten wurde – also ein Bewerber die Bewerbungsentscheidung getroffen hat.

Die Wirkung von „Self-Assessments“ wird – wie oben dargestellt – wesentlich von der Gestaltung des Feedbacks beeinflusst – über entsprechende Erfolgsfaktoren und Risiken bei der Gestaltung ist aber nur wenig bekannt. Eine Forschungsfrage ist schließlich auch, wer freiwillige „Self-Assessment“ Angebote nutzt. Es ist zu befürchten, dass insbesondere solche Personen das Angebot nutzen, die ohnehin dazu neigen, ihre Entscheidung gut abzusichern (Matthäus-Effekt). Diesbezüglich müssten Strategien zur Steigerung der Anwendungshäufigkeit formuliert und überprüft werden.

### 10.5.3 Schlusswort

„Online-Assessments“ bergen große Potenziale, weisen aber auch Risiken auf. Das gleiche gilt für „Self-Assessments“ als Sonderform des „Online-Assessments“. Eine erfolgreiche, onlinebasierte Selbstselektion kann die Personalauswahl verbessern und gleichzeitig zu einer positiven Darstellung der Organisation und der zu besetzenden Position und deren Anforderungen beitragen. Die weitere Forschung hierzu kann und sollte von der wachsenden Erfahrung des Praxiseinsatzes von „Online-Assessments“ und „Self-Assessments“ sowohl in der Personalauswahl als auch in Berufs- und Studienorientierung profitieren (Tab. 10.3).

#### Fragen

1. Worin unterscheiden sich „Self-Assessments“ von „Online-Assessments“?
2. Nach welchen Unterscheidungsdimensionen lassen sich „Online-Assessments“ beschreiben?
3. Welche Vor- und Nachteile bietet der Einsatz von „Online-Assessments“?
4. Welche Möglichkeiten gibt es, die Gefahr von Verfälschung bei „Online-Assessments“ bzw. „Self-Assessments“ zu begrenzen?
5. Welche Prozesse können einer erwünschten Selbstselektion entgegenwirken?
6. Was ist bei der Gestaltung automatisierter Rückmeldungen in „Online-Assessments“ und speziell in „Self-Assessments“ zu beachten?

#### Literatur

1. Anseel, F., & Lievens, F. (2009). The mediating role of feedback acceptance in the relationship between feedback and attitudinal and performance outcomes. *International Journal of Selection & Assessment*, 17(4), 362–376.
2. Arthur, W., Doverspike, D., Muñoz, G. J., et al. (2014). The use of mobile devices in high-stakes remotely delivered assessments and testing. *International Journal of Selection & Assessment*, 22(2), 113–123.
3. Association of Test Publishers (ATP). (2002). *Guidelines for computer-based testing*. Washington: ATP.
4. Bartram, D., & Hambleton, R. K. (Hrsg.). (2005). *Computer-based testing and the internet*. West Sussex: Wiley.
5. Bauer, T. N., Truxillo, D. M., Tucker, J. S., et al. (2006). Selection in the information age: The impact of privacy concerns and computer experience on applicant reactions. *Journal of Management*, 32(5), 601–621.
6. Beaty, J. C., Nye, C. D., Borneman, M. J., et al. (2011). Proctored versus unproctored internet tests: Are unproctored noncognitive tests as predictive of job performance? *International Journal of Selection & Assessment*, 19(1), 1–10.
7. Bell, S. T., & Arthur, W. (2008). Feedback acceptance in developmental assessment-centers: The role of feedback message, participant personality, and affective response to the feedback session. *Journal of Organizational Behavior*, 29(5), 681–703.

8. Birkeland, S. A., Manson, T. M., Kisamore, J. L., et al. (2006). A Meta-Analytic investigation of job applicant faking on personality measures. *International Journal of Selection & Assessment*, 14(4), 317–335.
9. Bozeman, D. P., & Kacmar, K. (1997). A cybernetic model of impression management processes in organizations. *Organizational Behavior and Human Decision Processes*, 69(1), 9–30.
10. Braddy, P. W., Meade, A. W., & Kroustalis, C. M. (2008). Online recruiting: The effects of organizational familiarity, website usability, and website attractiveness on viewers' impressions of organizations. *Computers in Human Behavior*, 24(6), 2992–3001.
11. Cober, R. T., Brown, D. J., Levy, P. E., et al. (2003). Organizational web sites: Web site content and style as determinants of organizational attraction. *International Journal of Selection & Assessment*, 11(2–3), 158–169.
12. Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions* (2. Aufl.). Urbana: University of Illinois Press.
13. Diercks, J., & Kupka, K. (Hrsg.). (2013a). *Recrutainment*. Wiesbaden: Springer Fachmedien.
14. Diercks, J., & Kupka, K. (2013b). Recrutainment – Bedeutung, Einflussfaktoren und Begriffsbestimmung. In J. Diercks & K. Kupka (Hrsg.), *Recrutainment* (S. 1–18). Wiesbaden: Springer Fachmedien.
15. DIN EN ISO 9241-110. (2006). *Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung*. Berlin: Beuth.
16. DIN 33430. (2016). *Anforderungen an berufsbezogene Eignungsdiagnostik*. Berlin: Beuth.
17. Driesel-Lange, K., & Hany, E. (2005). Berufsorientierung am Ende des Gymnasiums: Die Qual der Wahl. In B. Kracke & E. Hany (Hrsg.), *Schriften zur Berufsorientierungsforschung* (Bd. 1). Erfurt: Universität Erfurt.
18. Eckhardt, A., Laumer, S., & Vornewald, K. (2013). Bewertung von Self- und E-Assessments durch Kandidaten und Unternehmen. In J. Diercks & K. Kupka (Hrsg.), *Recrutainment* (S. 19–32). Wiesbaden: Springer Fachmedien.
19. Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *The Journal of Abnormal and Social Psychology*, 44(1), 118–123.
20. Guo, J., & Drasgow, F. (2010). Identifying cheating on unproctored internet tests: The z-test and the likelihood ratio test. *International Journal of Selection & Assessment*, 18(4), 351–364.
21. Harris, M. M., Hoye, G. van, & Lievens, F. (2003). Privacy and attitudes towards internet-based selection systems: A cross-cultural comparison. *International Journal of Selection & Assessment*, 11(2–3), 230–236.
22. Hasenberg, S. (2012). *Zur prädiktiven Validität von Self-Assessments für die Studienzufriedenheit*. Unveröffentlichte Dissertation, Philipps-Universität Marburg, Marburg.
23. Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57(3), 639–683.
24. Hell, B. (2009). Selbsttests zur Studienorientierung: nützliche Vielfalt oder unnützer Wildwuchs? In G. Rudinger (Hrsg.), *Self-Assessment an Hochschulen. Von der Studienfachwahl zur Profilbildung* (S. 9–20). Bonn: Bonn University Press.
25. Hiltmann, M. (2013). Online-Self-Assessments: Ein Impuls zur persönlichen und beruflichen Weiterentwicklung. *Wirtschaftspsychologie*, 1, 72–80.
26. Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64(4), 349–371.
27. Kanning, U. P. (2003). Sieben Anmerkungen zum Problem der Selbstdarstellung in der Personalauswahl. *Zeitschrift für Personalpsychologie*, 2(4), 193–195.
28. Kantrowitz, T. M., & Dainis, A. M. (2014). How secure are unproctored pre-employment tests? Analysis of inconsistent test scores. *Journal of Business and Psychology*, 29(4), 605–616.

29. Kersting, M. (2008a). *Qualität in der Diagnostik und Personalauswahl – der DIN-Ansatz*. Göttingen: Hogrefe.
30. Kersting, M. (2008b). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33(9), 420–433.
31. Kersting, M. (2011a). *Personalgewinnung und Personalauswahl im öffentlichen Dienst: Potpourri der Methoden*, 2011(1), 10–12.
32. Kersting, M. (2011b). Schlaraffenland ist abgebrannt – warum sich die Bewerberauswahl ändern muss. *Wirtschaftspsychologie aktuell*, 4, 23–26.
33. Kersting, M. (2014). Zirkustricks entlarven. Qualität von Persönlichkeitsfragebogen. *Personalmagazin*, 1(14), 28–32.
34. Kirbach, C., & Wottawa, H. (2009). Online-Tools zur Gewinnung passender Mitarbeiter. In H. Steiner (Hrsg.), *Online-Assessments. Grundlagen und Anwendung von Online-Tests in der Unternehmenspraxis* (S. 68–82). Heidelberg: Springer.
35. Klinck, D. (2002). *Computergestützte Diagnostik. Beeinflusst das Medium der Testbearbeitung die Testcharakteristika, die Testfairness oder das Erleben der Testsituation?*. Göttingen: Hogrefe.
36. Konradt, U., & Sarges, W. (2003). *E-Recruitment und E-Assessment: Rekrutierung, Auswahl und Beurteilung von Personal im Inter- und Intranet*. Göttingen: Hogrefe.
37. Konradt, U., Warszta, T., & Ellwart, T. (2013). Fairness perceptions in web-based selection: Impact on applicants' pursuit intentions, recommendation intentions, and intentions to reapply. *International Journal of Selection & Assessment*, 21(2), 155–169.
38. Kubinger, K. D. (2009). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens* (2., überarb. und erw. Aufl.). Göttingen: Hogrefe.
39. Kubinger, K. D., Frebort, M., Khorramdel, et al. (2013). Prinzipien und Verfahren der Self-Assessments vom „Wiener Autorenkollektiv Studienberatungstests“. *Wirtschaftspsychologie*, 1(1), 17–24.
40. Kubinger, K. D. (2013). Self-Assessment im Rahmen der Personalauswahl: Möglichkeiten und Grenzen. *Wirtschaftspsychologie*, 1, 81–87.
41. Laier, W. (2009). Einsatz von Online-Tests aus technischer Sicht. In H. Steiner (Hrsg.), *Online-Assessments. Grundlagen und Anwendung von Online-Tests in der Unternehmenspraxis* (S. 37–54). Heidelberg: Springer.
42. Landers, R. N., & Sackett, P. R. (2012). Offsetting performance losses due to cheating in unproctored internet-based testing by increasing the applicant pool. *International Journal of Selection & Assessment*, 20(2), 220–228.
43. Lievens, F., & Burke, E. (2011). Dealing with the threats inherent in unproctored internet testing of cognitive ability: Results from a large-scale operational test program. *Journal of Occupational and Organizational Psychology*, 84(4), 817–824.
44. Marcus, B. (2009). 'Faking' from the applicant's perspective: A theory of self-presentation in personnel selection settings. *International Journal of Selection & Assessment*, 17(4), 417–430.
45. Meehl, P. E. (1956). Wanted – a good cook-book. *American Psychologist*, 11(6), 263–272.
46. Morelli, N. A., Mahan, R. P., & Illingworth, A. J. (2014). Establishing the measurement equivalence of online selection assessments delivered on mobile versus nonmobile devices. *International Journal of Selection & Assessment*, 22(2), 124–138.
47. Morf, C. C., & Koole, S. L. (2014). Das Selbst. In K. Jonas, W. Stroebe, & M. Hewstone (Hrsg.), *Sozialpsychologie. Eine Einführung* (6., komplett neu verfasste und aktual. Aufl., S. 141–195). Berlin: Springer.
48. Nerdinger, F. W. (2014). Gravitation und organisationale Sozialisation. In F. W. Nerdinger, G. Blickle, & N. Schaper (Hrsg.), *Arbeits- und Organisationspsychologie* (S. 71–82). Berlin: Springer.

49. Nye, C. D., Do, B., Drasgow, F., et al. (2008). Two-step testing in employee selection: Is score inflation a problem? *International Journal of Selection & Assessment*, 16(2), 112–120.
50. Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598–609.
51. Ployhart, R. E., Weekley, J. A., Holtz, B. C., et al. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgement tests comparable? *Personnel Psychology*, 56(3), 733–752.
52. Prensky, M. (2001). Digital natives, Digital immigrants Part 1. *On the Horizon*, 9(5), 1–6.
53. Rudolph, C. W., Wells, C. L., Weller, M. D., et al. (2009). A meta-analysis of empirical studies of weight-based bias in the workplace. *Journal of Vocational Behavior*, 74(1), 1–10.
54. Schaper, N. (2009). Online-Tests aus diagnostisch-methodischer Sicht. In H. Steiner (Hrsg.), *Online-Assessments. Grundlagen und Anwendung von Online-Tests in der Unternehmenspraxis* (S. 17–36). Heidelberg: Springer.
55. Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274.
56. Sinar, E. F., Reynolds, D. H., & Paquet, S. L. (2003). Nothing but 'net? Corporate image and web-based testing. *International Journal of Selection & Assessment*, 11(2–3), 150–157.
57. Swann, W. B., & Schroeder, D. G. (1995). The search for beauty and truth: A framework for understanding reactions to evaluations. *Personality and Social Psychology Bulletin*, 21(12), 1307–1318.
58. Sylva, H., & Mol, S. T. (2009). E-Recruitment: A study into applicant perceptions of an online application system. *International Journal of Selection & Assessment*, 17(3), 311–323.
59. Templer, K. J., & Lange, S. R. (2008). Internet testing: Equivalence between proctored lab and unproctored field conditions. *Computers in Human Behavior*, 24(3), 1216–1228.
60. Tendeiro, J. N., Meijer, R. R., Schakel, L., et al. (2012). Using cumulative sum statistics to detect inconsistencies in unproctored internet testing. *Educational and Psychological Measurement*, 73(1), 143–161.
61. Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen. (2009). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 09. September 2009. *Report Psychologie*, 34, 470–478.
62. The International Test Commission. (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, 6(2), 143–171.
63. Viswesvaran, C., & Ones, D. S. (1999). Meta-Analyses of faking estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59(2), 197–210.
64. Wiechmann, D., & Ryan, A. M. (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection & Assessment*, 11(2–3), 215–229.
65. Wiedmann, J. (2009). Mehrstufiges Auswahlverfahren mit Online-Assessments bei der Luftwaffe. In H. Steiner (Hrsg.), *Online-Assessments. Grundlagen und Anwendung von Online-Tests in der Unternehmenspraxis* (S. 105–126). Heidelberg: Springer.
66. Wright, N. A., Meade, A. W., & Gutierrez, S. L. (2014). Using invariance to examine cheating in unproctored ability tests. *International Journal of Selection & Assessment*, 22(1), 12–22.
67. Wurow, A. (2009). Interne Stellenbesetzung innerhalb der Conergy Gruppe mit dem verhaltensbasierten Fragebogen »Predictive Index«. In H. Steiner (Hrsg.), *Online-Assessments. Grundlagen und Anwendung von Online-Tests in der Unternehmenspraxis* (S. 127–140). Heidelberg: Springer.

## Internetverweise

68. Information zum E-Assessment der Deutschen Post AG. <http://www.eligo.de/de/files/2012/07/Personalmarketing-Azubis-eligo.pdf>. Zugegriffen: 13. März 2015.
69. Information zum E-Assessment der Targobank. <http://www.cyquest.net/projektliste/webbasiertes-eassessment-zur-rekrutierung-von-auszubildenden-bei-der-targobank/>. Zugegriffen: 13. März 2015.
70. Self-Assessment von E.ON. <http://www.cyquest.net/projektliste/e-on-testet-ausbildungsplatzbewerber-online-mit-dem-e-on-phasenprufer/>. Zugegriffen: 13. März 2015.

---

# Anwendung der Personalauswahlverfahren im internationalen Vergleich

Diana Eva Krause

---

## Zusammenfassung

Das Ziel dieses Kapitels ist, Gemeinsamkeiten und Unterschiede im Einsatz von Personalauswahlverfahren in verschiedenen Ländern darzustellen. Dabei stehen die Analyse der Bewerbungsunterlagen und Referenzen, das Einstellungsinterview, Arbeitsproben, Testverfahren, der biografische Fragebogen, Assessment-Center und graphologische Gutachten im Zentrum der Betrachtung. Pro Verfahren der Personalauswahl wird die gegenwärtige Anwendung in diversen Ländern verdeutlicht und durch Rekurs auf Validitätsaspekte auf Divergenzen zwischen wissenschaftlichem Kenntnisstand und praktischer Umsetzung aufmerksam gemacht. Daraus werden Empfehlungen für eine Verbesserung der Personalauswahlpraxis generiert.

## Inhaltsverzeichnis

11.1	Einleitung	244
11.2	Vergleich der Anwendung der Verfahren der Personalauswahl im internationalen Kontext	245
11.2.1	Analyse der Bewerbungsunterlagen und Referenzen im internationalen Vergleich	245
11.2.2	Anwendung des Einstellungsinterviews im internationalen Vergleich	247
11.2.3	Anwendung von Arbeitsproben im internationalen Vergleich	249
11.2.4	Anwendung von Testverfahren im internationalen Vergleich	250
11.2.5	Anwendung des biografischen Fragebogens im internationalen Vergleich	255
11.2.6	Anwendung des Assessment-Center im internationalen Vergleich	257

---

D.E. Krause (✉)  
Berlin, Deutschland  
E-Mail: [diana.eva.krause@gmail.com](mailto:diana.eva.krause@gmail.com)



11.2.7 Anwendung der Graphologie und sonstige Maßnahmen im internationalen Vergleich .....	263
11.3 Fazit .....	265
Literatur .....	266

---

## 11.1 Einleitung

Der Trend von Unternehmen, das Geschäft jenseits nationaler Grenzen zu betreiben, setzt sich fort [10]. Dieser Trend spiegelt sich im Export von Produkten und Dienstleistungen, Joint Ventures, Strategischen Allianzen und in Mergers & Acquisitions im Ausland wider. Daraus ergibt sich für Unternehmen u. a. die Konsequenz, den Arbeitsmarkt aus einem internationalen Blickwinkel heraus zu betrachten. Dies hat weitreichende Implikationen für alle Facetten der gegenwärtigen und zukünftigen Personalarbeit in international und multinational agierenden Unternehmen. Eine Herausforderung, die sich damit verbindet, besteht in der Auswahl von Personal im internationalen Kontext. Traditionelle Methoden und Verfahren der Personalauswahl stoßen angesichts aktueller Veränderungen auf Grenzen [34]. Insofern sind Methoden in der Personalauswahl notwendig, die internationale Kontextbedingungen von Unternehmen berücksichtigen. So haben Personalverantwortliche heute beispielsweise dem Bedarf nach solchen Human-Ressource-Systemen Rechnung zu tragen, die in verschiedenen Ländern Anwendung finden können, dabei aber gleichzeitig den spezifischen Gegebenheiten in einem bestimmten Land oder einer bestimmten Kultur gerecht werden [25].

Eine grundsätzliche Frage, die sich international agierende Unternehmen in Bezug auf Personalauswahlentscheidungen stellen muss, ist die Frage nach der Balance zwischen dem Bedarf nach kulturspezifischer Ausgestaltung der Personalauswahlverfahren und dem gleichzeitigen Bedarf nach kulturübergreifender Standardisierung der Verfahren [20]. Der Bedarf nach kulturspezifischer Ausgestaltung von Personalauswahlverfahren betrifft dabei etwa die Auswahl der Kriterien der Leistung und des Berufserfolgs in der jeweiligen Landeskultur, die Frage nach erfolgskritischen Anforderungsdimensionen, die Auswahl und Adaptation der einzusetzenden Methoden, die Frage kulturspezifischer Inhalte des Auswahlverfahrens oder die Frage des auf die jeweilige Landeskultur zugeschnittenen Feedbacks an die Bewerber (synonym hier: Kandidaten).

---

### Beispiel

Ein Beispiel soll diesen Sachverhalt verdeutlichen: Anforderungsanalysen für ein und dieselbe Zielposition (z. B. Bereichsleiter Finanzierung) mögen zu identischen Verhaltensdimensionen für verschiedene Länder geführt haben (z. B. Fähigkeit zum Konfliktmanagement als kritische, erfolgsrelevante Dimension); allerdings mag die Akzeptanz der entsprechenden Verhaltensindikatoren interkulturell variieren (z. B. akzeptables Konfliktverhalten in Deutschland vs. akzeptables Konfliktverhalten in Indien).

Angesichts der Notwendigkeit nach kultureller Anpassung von Personalauswahlverfahren kann bereits an dieser Stelle nicht deutlich genug hervorgehoben werden, dass eine Kulturanpassung nicht beliebig erfolgen kann, sondern unter der Perspektive der Validität und Reliabilität zu erfolgen hat. Daher ist die Realisierung einer kulturübergreifenden Standardisierung der jeweils eingesetzten Verfahren der Personalauswahl erforderlich, sodass eine Vergleichbarkeit der Resultate der Personalauswahlverfahren gegeben ist und international valide Leistungsprognosen aus dem Personalauswahlsystem ableitbar sind. Deshalb kann z. B. die Auswahl der Kriterien der Leistung und des Berufserfolgs oder die Auswahl der einzelnen Methoden des internationalen Personalauswahlsystems nicht beliebig relativiert werden. Im Gegensatz zu den gewandelten und sich fortwährend wandelnden Bedingungen, denen die Personalarbeit Rechnung zu tragen hat, gibt es also eine konkurrierende Bedingung, die Personalverantwortliche ebenfalls zu berücksichtigen haben: Stabilität. Stabilität, die als partiell widersprüchlich zum Wandel interpretiert werden kann, manifestiert sich im Rahmen von internationalen Personalauswahlssystemen im dargelegten Bedarf nach Standardisierung. In diesem Zusammenhang verdeutlichen Huo et al. [14, S. 31]: „Whether a specific personnel selection practice should be universally adopted remains an unresolved issue.“

Damit ist bereits das Anliegen dieses Kapitels umrissen: Wir werden uns der Anwendung von Personalauswahlverfahren in verschiedenen Ländern zuwenden. Mitunter sehen sich Personalverantwortliche im internationalen Kontext mit der Rekrutierung von Personal, Personalmarketing und der Auswahl von Expatriates konfrontiert. Diese Themen zählen nicht (!) zum Gegenstand dieses Kapitels. Leser, die an diesen Fragestellungen interessiert sind, seien auf folgende Literatur verwiesen: [1, 7, 60, 66].

---

## **11.2 Vergleich der Anwendung der Verfahren der Personalauswahl im internationalen Kontext**

### **11.2.1 Analyse der Bewerbungsunterlagen und Referenzen im internationalen Vergleich**

Widmen wir uns zunächst der Analyse der Bewerbungsunterlagen im internationalen Kontext. In einer Studie von Schuler et al. [54] wurde festgestellt, dass in den umsatzstärksten Unternehmen die Sichtung der Bewerbungsunterlagen zu den gebräuchlichsten Verfahren der Personalauswahl in Deutschland gehört. Wie Tab. 11.1 zeigt, variiert die Häufigkeit der Sichtung und Analyse der Bewerbungsunterlagen (Anschreiben, Curriculum Vitae, Spezialkenntnisse, Zeugnisse, Referenzen etc.) zwischen den untersuchten Ländern erheblich [45]. Während in einigen Ländern die Sichtung der Bewerbungsunterlagen häufig Anwendung findet (Japan, Malaysia, Hongkong, Frankreich, Italien, Singapur, Südafrika, USA), finden sie in anderen Ländern kaum Beachtung (Griechenland, Schweden). Insgesamt gesehen werden Bewerbungsunterlagen in allen Ländern vergleichsweise häufig (Mittelwert 3,53 auf einer 5-stufigen Skala) im ersten Schritt

**Tab. 11.1** Sichtung der Bewerbungsunterlagen, Einholung von persönlichen Referenzen sowie Referenzen früherer Arbeitgeber und Anwendung des Einzelinterviews sowie des gruppenbasierten Panel-Interviews und Arbeitsproben im internationalen Vergleich. (Quelle [58], modifizierte Darstellung)

Land	Sichtung der Bewerbungsunterlagen $M$	Einholung von Referenzen früherer Arbeitgeber $M$	Einholung von persönlichen Referenzen $M$	Einzelinterview $M$	Gruppenbasiertes Panel-Interview $M$	Arbeitsproben $M$
Australien	3,59	4,32	2,88	3,59	4,08	1,89
Belgien	3,94	2,64	2,74	4,70	2,75	1,40
Deutschland	3,65	2,03	3,06	4,65	1,88	1,50
Frankreich	4,09	3,32	2,79	4,85	2,06	1,50
Griechenland	2,92	2,30	3,44	4,92	2,71	1,79
Großbritannien	4,26	4,37	3,51	3,88	3,82	1,59
Hongkong	4,75	3,75	2,75	3,38	3,63	1,83
Irland	3,46	4,53	3,46	3,34	4,00	1,39
Italien	4,19	2,69	3,04	4,93	1,50	1,00
Japan	5,00	3,33	3,33	4,00	4,00	1,32
Kanada	3,29	4,05	3,00	4,35	3,57	1,80
Malaysia	5,00	5,00	4,00	3,00	4,00	1,49
Neuseeland	3,64	4,40	3,40	4,05	3,71	1,69
Niederlande	3,55	2,72	2,47	3,78	4,30	1,67
Portugal	3,40	3,14	3,14	4,77	3,29	1,71
Schweden	1,19	4,49	3,94	4,84	2,82	1,40
Singapur	4,67	3,13	3,07	4,13	3,47	2,15
Spanien	3,22	2,43	2,91	4,72	3,63	1,22
Südafrika	4,20	4,09	3,63	4,70	2,45	1,84
USA	4,12	4,02	3,18	4,78	3,27	1,89
Alle Länder	3,53	3,77	3,22	4,30	3,36	1,40

Anmerkung:  $N = 959$  Organisationen aus 20 Ländern.  $M =$  Mittelwert. Antwortskala: 1 = niemals, 2 = selten (1–20 %), 3 = manchmal (21–50 %), 4 = häufig (51–80 %), 5 = sehr häufig oder immer (81–100 %)

des Personalauswahlprozesses berücksichtigt. Vergleicht man die Art der Bewerbungsunterlagen in verschiedenen Ländern, dann zeigt sich, dass die Bewerbungsunterlagen in Großbritannien stark auf die spezifische Organisation sowie die spezifische Tätigkeit

zugeschnitten sind, während in anderen Ländern häufiger Standardunterlagen genutzt werden [56].

In manchen Fällen sind Referenzen Bestandteil von Bewerbungsunterlagen. Referenzen stellen in der Regel die Einschätzung des Kandidaten durch eine dritte Partei dar. Dies kann in schriftlicher oder mündlicher Form erfolgen. Zu unterscheiden sind hier persönliche Referenzen und Referenzen früherer Arbeitgeber. Wie aus Tab. 11.1 deutlich wird, werden länderübergreifend häufiger Referenzen früherer Arbeitgeber anstelle persönlicher Referenzen eingeholt. Länder, in denen Referenzen früherer Arbeitgeber einen dominanten Einfluss auf die Personalauswahlentscheidung haben, sind Australien, Großbritannien, Irland, Kanada, Malaysia, Neuseeland, Schweden, Südafrika und die USA. Relativ selten werden Referenzen früherer Arbeitgeber in Belgien, Deutschland, Griechenland, Italien, Niederlande und Spanien eingeholt.

### 11.2.2 Anwendung des Einstellungsinterviews im internationalen Vergleich

Interviews sind weltweit das am häufigsten genutzte Personalauswahlverfahren (s. Tab. 11.1). Dieser Befund ist konsistent mit früheren Untersuchungen in verschiedenen Einzelländern [8, 33, 54], in Europa [48], aber auch vergleichenden Sammelreferaten [5, 30, 39, 42]. Die durchgeführten Interviews unterscheiden sich in verschiedenen Ländern in ihrem Format und in ihrer Struktur.

Einzelinterviews finden im globalen Rahmen mit Abstand am häufigsten Einsatz. Dabei ist die Frage interessant, ob diese Interviews in strukturierter oder unstrukturierter Form durchgeführt werden. Denn der Strukturierungsgrad eines Interviews hat sich als eine entscheidende Determinante der prädiktiven Validität des Interviews erwiesen. Strukturiert ist ein Interview dann, wenn jedem Kandidaten die gleichen Fragen in der gleichen Reihenfolge gestellt werden.

- ▶ Metaanalysen zur Validität von strukturierten vs. unstrukturierten Interviews belegen übereinstimmend, dass strukturierte Interviews eine hohe prädiktive Validität aufweisen ( $r = ,40$  bis  $,56$ ), während unstrukturierte Interviews eine relativ geringe prädiktive Validität ( $r = ,13$  bis  $,20$ ) haben [13].

In Australien werden strukturierte Interviews häufiger eingesetzt als unstrukturierte Interviews [8]. Auch in Deutschland ist eine häufigere Anwendung strukturierter Interviews zu verzeichnen [54]. Eine Untersuchung von Baron et al. [3] zeigt, dass Organisationen in Nordamerika (Kanada und USA) und der Schweiz strukturierte Interviews bevorzugen, während Organisationen in Frankreich, Griechenland und Italien – als Beispiel für partikularistische Kulturen – unstrukturierte Interviews häufiger zur Personalauswahl einsetzen als strukturierte Interviews. Aus der Validitätsperspektive muss man daher für Frankreich, Griechenland und Italien Verbesserungsbedarf bezüglich der genutzten Form

des Interviews ableiten. Eine Studie von Zee et al. [67] macht in diesem Zusammenhang darauf aufmerksam, dass Personalverantwortliche generell intuitive, unstrukturierte Interviews gegenüber strukturierten Interviews bevorzugen und dass diese Bevorzugung auch mit den in der jeweiligen Organisationen genutzten Formen des Interviews korrespondiert. Als Gründe für die Bevorzugung unstrukturierter Interviews gegenüber der valideren Interviewform führen die Autoren generelle Einstellungen, subjektive Normen sowie „glaubensbasierte“ Einstellungen der Personalverantwortlichen an, die maßgeblichen Einfluss auf ihre Intentionen zur Nutzung dieser oder jener Interviewform haben. Fraglich bleibt, ob der „Glauben“ der Personalverantwortlichen auch zu validen Ergebnissen im Auswahlprozess führt.

- Konkrete Maßnahmen, durch die man ein Interview verbessern kann, sind nach Schuler [52]:
- die anforderungsbezogene Gestaltung,
  - die Beschränkung auf diejenigen berufsrelevanten Merkmale, die nicht durch andere Personalauswahlverfahren zuverlässiger ermittelt werden können,
  - die Durchführung in strukturierter Form,
  - die Verwendung geprüfter und verankerter Skalen,
  - die empirische Überprüfung von Einzelfragen,
  - der Einsatz mehrerer Interviewer,
  - die Integration von Verfahrenskomponenten aus dem Assessment-Center,
  - die Trennung von Information und Entscheidung,
  - die Standardisierung der Gewichtung- und Entscheidungsprozedur,
  - die Vorbereitung der Interviewer durch ein Interviewertraining.

Eine weitere Form des Interviews ist das gruppenbasierte Panel-Interview (s. Tab. 11.1). Bei einem Panel-Interview wird ein Kandidat von mehreren Interviewern sequentiell befragt. Insgesamt gesehen sind auch Panel-Interviews weit verbreitet, doch werden sie relativ betrachtet seltener eingesetzt als Einzelinterviews. Länder, in denen Panel-Interviews häufig Anwendung finden, sind Australien, Irland, Japan, Malaysia und die Niederlande. Demgegenüber spielen sie in Frankreich, Italien und Schweden eine untergeordnete Rolle.

Weitere Formen des Interviews, die die eignungsdiagnostischen Defizite unstrukturierter Interviews zu überwinden suchen, sind das situative Interview und das multimodale Interview. Bei einem situativen Interview wird der Kandidat gebeten anzugeben, wie er sich in einer konkreten Situation verhalten würde (Beispiel: Sie sind Hauptabteilungsleiter der Firma X. Der gestiegene Wettbewerbsdruck zwingt Ihr Unternehmen, Kürzungsmaßnahmen zu beschließen. Was würden Sie in dieser Situation tun?). Die Bewertung der Antwort des Kandidaten erfolgt hier durch die Anwendung einer verhaltensverankerten Skala. Das situative Interview ist insbesondere in den USA und Großbritannien verbreitet [56]. Das multimodale Interview kombiniert demgegenüber mehrere

Fragetypen miteinander (z. B. situative Fragen, biografieorientierte Fragen, Fragen zu Berufswahl, beruflichen Interessen, Organisationswahl), was zu einer hohen prädiktiven Validität sowie einer hohen wahrgenommenen Fairness dieser Interviewform durch die Kandidaten führt (vgl. [29, S. 865]). Wenngleich sich daher aus akademischer Sicht der Einsatz dieser Interviewform empfiehlt, muss die Frage, inwieweit sie bereits in die internationale Praxis diffundiert ist, aufgrund mangelnder empirischer Resultate offen bleiben.

### 11.2.3 Anwendung von Arbeitsproben im internationalen Vergleich

Arbeitsproben („work sample tests“) sind kleine Simulationen der zukünftigen Berufstätigkeit des Kandidaten [43]. Arbeitsproben können inhaltlich stark variieren: So besteht etwa eine Arbeitsprobe für die Tätigkeit einer Sekretärin in der Aufgabe des Tippens, für einen Piloten werden dagegen Arbeitsproben in hochtechnologischen Flugsimulatoren durchgeführt. Generell kann die prädiktive Validität von Arbeitsproben mit  $r = ,54$  als hoch bezeichnet werden [50]. Darüber hinaus sind Arbeitsproben insofern von Vorteil als sie relativ frei von Diskriminierung bestimmter ethnischer Gruppen sind, den Kandidaten einen realistischen Überblick über ihre Tätigkeit vermitteln, es ermöglichen, dass der Kandidat selbst einschätzen kann, wie gut er zu der Position passt und schließlich als faires Auswahlinstrument von den Kandidaten wahrgenommen (zu Aspekten der Fairness s. das Kapitel von Kanning in diesem Band) werden.

Werfen wir nun einen Blick in die Empirie, so wird deutlich, dass Arbeitsproben weltweit eine untergeordnete Rolle im Personalauswahlgeschehen spielen (s. Tab. 11.1). In 19 von 20 Ländern werden Arbeitsproben nie bis selten zu Personalauswahlzwecken eingesetzt. Eine Ausnahme stellt Singapur dar, wo die Arbeitsprobe einen höheren Stellenwert genießt als in anderen Ländern. Nach aktuellen Befunden von Schuler et al. [54] wird die Arbeitsprobe in deutschen Großunternehmen von 45 % der Organisationen im Rahmen der Personalauswahl genutzt. Am Beispiel der Arbeitsprobe zeigt sich die Kluft zwischen Theorie und Praxis: Aufgrund der hohen prädiktiven Validität der Arbeitsprobe empfiehlt sich ihr Einsatz (vgl. dazu Kapitel von Anderson in diesem Band). Allerdings kommt die Arbeitsprobe im internationalen Feld kaum zum Einsatz. Ein Grund hierfür mag darin bestehen, dass der Entwicklungs-, Konstruktions- und Durchführungsaufwand von Arbeitsproben nicht zu unterschätzen ist [12]. Ferner ist zu bedenken, dass es für manche Tätigkeiten, insbesondere Managementtätigkeiten im internationalen Rahmen, schwer ist, eine tätigkeitsäquivalente Arbeitsprobe zu konstruieren. Das Ziel einer Arbeitsprobe besteht darin, die zukünftige Berufstätigkeit in all ihren Facetten umfassend und repräsentativ abzubilden.

#### Beispiel

Wenn wir nun z. B. in China eine Führungskraft für den Bereich Personal auswählen wollen und diese Führungskraft zudem noch personelle und finanzielle Verantwortung in anderen Ländern (z. B. Indonesien, Sri Lanka, Australien) tragen soll, dann wird

an diesem Beispiel schnell klar, dass es zum einen schwierig ist, das Spektrum der erfolgskritischen Situationen zu definieren (da diese partiell unvorhersehbar sind) und zum anderen das Tätigkeitsspektrum repräsentativ und erschöpfend in einer Arbeitsprobe abzubilden.

#### 11.2.4 Anwendung von Testverfahren im internationalen Vergleich

Gegenüber der Arbeitsprobe basieren Testverfahren auf einem anderen eignungsdiagnostischen Hintergrund. Es handelt sich um konstruktionsorientierte Verfahren der Personalauswahl, die dem eigenschaftsorientierten Ansatz der Berufseignungsdiagnostik zuzurechnen sind [15]. Durch den Einsatz von Testverfahren werden personenbezogene Merkmale gemessen, die als stabile Konstrukte („traits“) gelten. So wird etwa durch den Einsatz eines Intelligenztests das Merkmal Intelligenz (Vergleich das Kapitel von Kersting und Palmer in diesem Band) erfasst. Inzwischen liegen ca. 5000 standardisierte Testverfahren zur Erfassung von personenbezogenen Merkmalen vor [40, S. 161], wobei die meisten in Nordamerika und Westeuropa konstruiert und evaluiert worden sind.

Tabelle 11.2 zeigt, dass Testverfahren insgesamt gesehen in mittlerem Ausmaß für Personalauswahlzwecke angewandt werden ( $M = 3,09$ ), wobei eine erhebliche Streuung zwischen den Ländern zu beobachten ist. Länder, in denen Testverfahren vergleichsweise häufig zum Einsatz kommen, sind Malaysia, Belgien, Neuseeland, Schweden, Südafrika und Portugal. Eine Studie von Taylor et al. [61] belegt, dass Testverfahren in Neuseeland gegenwärtig häufiger für Personalauswahlzwecke eingesetzt werden als dies vor 10 Jahren der Fall war. Relativ selten werden Testverfahren eingesetzt in Deutschland, Italien, Japan und den USA. Schuler et al. [54] stellen folgende Anwendungshäufigkeiten von Testverfahren in Deutschland fest: 20 % der Unternehmen nutzen Persönlichkeitstests, 40 % wenden Leistungstests an und 30 % setzen Intelligenztests ein. Fragen wir an dieser Stelle, warum Testverfahren in Deutschland insgesamt gesehen eher selten zur Personalauswahl genutzt werden, so liegt folgende Erklärung nahe.

- ▶ Der Einsatz von Testverfahren stößt in Deutschland auf Akzeptanzbarrieren und reflektiert spezifische Idiosynkrasien: Aus ideologischen Gründen ist der Einsatz von Testverfahren in Deutschland unpopulär. Personalverantwortliche in Deutschland vertreten häufig die Auffassung, durch den Einsatz von Testverfahren würden Eignungsunterschiede im Sinne unveränderbarer „traits“ festgeschrieben und daher soziale Ungleichheiten in den beruflichen Chancen der Kandidaten reproduziert [18]. Insbesondere im Zuge der Anlage-Umwelt-Debatte in den 1960er Jahren wird es verständlich, dass Tests von Personalverantwortlichen in Deutschland als repressiv angesehen wurden und sich mit Skeptizismus verbanden – ein Umstand, unter dem die heutige Personalauswahlpraxis noch immer leidet. Um die soziale Akzeptanz des Personalauswahlsystems zu sichern, werden in Deutschland demgegenüber eher solche Verfahren eingesetzt, die den Charakter der Situation betonen, die dem



**Tab. 11.2** Anwendung verschiedener Testverfahren im internationalen Vergleich. (Quelle [45], modifizierte Darstellung)

Land	Testverfahren (gesamt) $M$	Kognitive Fähigkeits-tests $M$	Integrity-Tests $M$	Berufsbezogene Interessentests $M$	Persönlichkeitstests $M$	Situational-Judgement-Tests $M$
Australien	2,75	2,39	1,16	1,51	2,56	1,35
Belgien	3,73	3,85	1,60	1,98	3,75	2,16
Deutschland	1,74	1,90	1,00	1,40	1,70	1,50
Frankreich	2,63	2,29	1,00	2,33	3,42	1,33
Griechenland	2,65	2,54	1,85	2,29	3,14	1,15
Großbritannien	3,41	2,98	1,12	1,33	3,21	1,85
Hongkong	2,88	1,83	1,33	1,50	2,50	3,00
Irland	2,98	2,79	1,04	1,07	3,17	1,31
Italien	1,58	1,33	1,00	1,00	1,86	1,00
Japan	2,33	3,76	0,00	0,00	3,29	2,30
Kanada	2,87	2,59	1,52	1,76	2,78	–
Malaysia	4,00	3,37	0,00	0,00	3,59	–
Neuseeland	3,45	2,83	1,20	1,54	2,67	1,57
Niederlande	3,03	3,27	1,69	1,65	3,00	1,82
Portugal	3,27	3,25	1,92	2,62	3,66	2,23
Schweden	3,74	2,09	1,11	1,53	1,62	1,64
Singapur	2,64	3,75	1,67	1,33	4,43	1,67
Spanien	3,09	3,08	2,21	–	3,46	2,33
Südafrika	3,44	2,86	1,62	1,73	3,68	1,94
USA	2,51	2,39	1,09	1,15	2,56	1,71
Alle Länder	3,09	3,85	1,33	1,61	3,75	1,78

Anmerkung:  $N = 959$  Organisationen aus 20 Ländern.  $M =$  Mittelwert. Antwortskala: 1 = niemals, 2 = selten (1–20 %), 3 = manchmal (21–50 %), 4 = häufig (51–80 %), 5 = sehr häufig oder immer (81–100 %)

Kandidaten die Möglichkeit geben, mit komplexem Material umzugehen, und die aktives Handeln von den Kandidaten verlangen. Derartige Verfahren der Personalauswahl werden als tätigkeitsspezifischer wahrgenommen, sodass ihnen eine höhere Augenscheinvalidität zugeschrieben wird als anderen Verfahren.

Auch in den USA und Kanada spielen Testverfahren in der Personalauswahl keineswegs eine dominante Rolle. Tests werden in den USA und in Kanada lediglich als Screening der Kandidaten in frühen Phasen des Personalauswahlprozesses eingesetzt [26].

Hintergrund dessen ist, dass durch den Einsatz von Testverfahren (insbesondere kognitiver Fähigkeitstests) die Gefahr steigt, Mitgliedern der Majorität einen Vorteil einzuräumen gegenüber Mitgliedern einer Minoritätengruppe. Zusätzlich sind in Nordamerika rechtliche Besonderheiten zu berücksichtigen, die entscheidenden Einfluss auf die Zuschreibung von Legalität der Verfahren der Personalauswahl haben.

Wenden wir uns nun der Anwendung von spezifischen Testverfahren im internationalen Vergleich zu: kognitiven Fähigkeitstests (z. B. Intelligenztests, Tests allgemeiner kognitiver Fähigkeiten, GMA), Persönlichkeitstests, Integrity-Tests und berufsbezogene Interessentests.

Dabei haben wir zunächst die prädiktive Validität dieser spezifischen Testverfahren zur Kenntnis zu nehmen, die nach den metaanalytischen Befunden von Schmidt und Hunter [50] wie folgt beziffert werden:

- kognitive Fähigkeitstests  $r = ,51$ ,
- Integrity-Tests  $r = ,41$ ,
- Interessentests  $r = ,10$ ,
- Gewissenshaftigkeitstests  $r = ,31$ .

Damit wird deutlich, dass kognitive Fähigkeitstests (z. B. der I-S-T 2000 R [31]) sowie Integrity-Tests (z. B. IBES [37]) eine hohe prädiktive Validität aufweisen und in ihrem prognostischen Gehalt lediglich von der Arbeitsprobe übertroffen werden. Für den europäischen Bereich nennen Salgado et al. [49] sogar eine Kriteriumsvalidität für kognitive Fähigkeitstests in Höhe von  $r = ,62$ . Für die höhere Managementebene berichten Krause et al. [21] eine prädiktive Validität von kognitiven Fähigkeitstests in Höhe von  $r = ,53$ . Nun zeigt Tab. 11.2, dass die Anwendung von kognitiven Fähigkeitstests länderspezifisch erheblich variiert.

- Bringt man die Länder bezüglich ihres Einsatzes kognitiver Fähigkeitstests in eine Rangfolge (s. dazu für den europäischen Raum [48]), dann lautet diese Rangfolge (beginnend mit dem Land der häufigsten Nutzung in absteigender Reihenfolge) folgendermaßen: 1) Belgien, 2) Singapur, 3) Japan, 4) Malaysia, 5) Niederlande, 6) Portugal, 7) Spanien, 8) Großbritannien, 9) Südafrika, 10) Neuseeland, 11) Irland, 12) Kanada, 13) Griechenland, 14) USA, 15) Australien, 16) Frankreich, 17) Schweden, 18) Deutschland, 19) Hongkong, 20) Italien.

Im internationalen Vergleich belegt Deutschland damit Platz 18 von 20 insgesamt zu vergebenden Rangplätzen in der Anwendung kognitiver Fähigkeitstests. Und dies ist ein Problem: Kognitive Fähigkeitstests sind sehr gute Prädiktoren verschiedener Leistungskriterien im beruflichen Kontext [21, 28, 35, 38, 41, 49].

Allerdings werden sie in manchen Ländern kaum zur Personalauswahl genutzt. So wird auch im Hinblick auf den Einsatz von kognitiven Fähigkeitstests die Divergenz zwischen

wissenschaftlichem Kenntnisstand und praktischer Realität deutlich (zur Diskrepanz zwischen Theorie und Praxis s. Kapitel von Anderson in diesem Band). Insofern wäre es für die zukünftige Personalpraxis in manchen Ländern (z. B. Deutschland, Frankreich, Hongkong, Italien) empfehlenswert, mehr als bislang kognitive Fähigkeitstests einzusetzen.

Andere Testverfahren, die ebenfalls eine hohe prädiktive Validität aufweisen, sind Integrity-Tests. Mitunter werden sie auch als Ehrlichkeitstests bezeichnet. Sie messen die Integrität eines Kandidaten und sind insbesondere zur Vorhersage kontraproduktiven Verhaltens in Organisationen geeignet (z. B. Diebstahl am Arbeitsplatz, Absentismus, aggressives Verhalten am Arbeitsplatz). Ein Blick in Tab. 11.2 macht deutlich, dass Integrity-Tests trotz ihrer hohen prädiktiven Validität in keinem Land häufig zu Personalauswahlzwecken genutzt werden. Die Befragten aus allen 20 Ländern gaben übereinstimmend an, Integrity-Tests niemals oder selten (in 1 bis 20 % der Fälle) einzusetzen. Auch der relative Vergleich der angewandten Testverfahren im internationalen Rahmen zeigt, dass Integrity-Tests am seltensten genutzt werden im Vergleich zu kognitiven Fähigkeitstests, berufsbezogenen Interessentests und Persönlichkeitstests. Insofern wird am Beispiel der Integrity-Tests die Kluft zwischen Theorie und Praxis abermals deutlich. Aus der Perspektive der Validität wäre es vor diesem Hintergrund zu empfehlen, Integrity-Tests häufiger anzuwenden als dies bislang der Fall ist.

Für eine weitere Kategorie von Testverfahren, nämlich berufsbezogene Interessentests, zeigt Tab. 11.2, dass sie von nur marginaler Bedeutung für die Personalauswahl in den dargestellten 20 Ländern sind. Interessentests messen in der Regel die relativen Interessen bzw. Berufspräferenzen eines Kandidaten ohne eine Aussage über die absolute Höhe der Interessen zu ermöglichen. Allerdings werden sie insgesamt gesehen häufiger angewandt als Integrity-Tests. Dies kann einen unerfreulichen Ist-Zustand zum Ausdruck bringen, denn die metaanalytisch nachgewiesene prädiktive Validität von Interessentests ( $r = ,10$ ) liegt mit großem Abstand weit unter der metaanalytisch belegten prädiktiven Validität von Integrity-Tests ( $r = ,41$ ). Die weltweit geringe Einsatzhäufigkeit von Interessentests kann nach Kanning und Schuler [15] mitunter auch dadurch bedingt sein, dass Interessentests für beabsichtigte Verfälschungen durch den Kandidaten noch anfälliger sind als andere Testverfahren. Schlussfolgernd bleibt für die organisationale Praxis festzuhalten, dass die Unterrepräsentanz von berufsbezogenen Interessentests im Rahmen der Verfahren der Personalauswahl mit ihrer geringen Prognosekraft konsistent ist, weshalb hier kein Veränderungsbedarf bezüglich des Einsatzes von Interessentests abzuleiten ist.

Länder unterscheiden sich auch in der Anwendung von Persönlichkeitstests zur Erfassung der „Big Five“ (Extraversion, emotionale Stabilität, Verträglichkeit, Gewissenhaftigkeit, Offenheit für neue Erfahrungen). Wie bereits erwähnt, zählen Persönlichkeitsmerkmale zu den Verfahren mit hoher prädiktiver Validität (s. dazu das Sammelreferat [11]). In diesem Zusammenhang zeigt eine Untersuchung von Salgado [46] für den europäischen Raum, dass emotionale Stabilität und Gewissenhaftigkeit valide Prädiktoren für verschiedene berufsbezogene Variablen und Berufsgruppen darstellen, während Extraversion, Verträglichkeit und Offenheit für Erfahrungen nur für bestimmte berufsbezogene Kriterien (z. B. Ausbildungserfolg) und Berufsgruppen valide Prädiktoren sind.

Wie deutlich wird (s. Tab. 11.2), werden Persönlichkeitstests im Vergleich zu anderen Testverfahren relativ häufig eingesetzt in Ländern wie Singapur, Belgien, Südafrika, Frankreich, Spanien, Japan, Großbritannien und Griechenland. Vergleichsweise selten finden sie dagegen Anwendung in Italien, Deutschland und Schweden. Darüber hinaus zeigt sich, dass in einigen Ländern (z. B. Australien, Frankreich, Griechenland, Großbritannien, Hongkong, Irland, Kanada, Portugal, Singapur, Spanien, Südafrika) Persönlichkeitstests häufiger zur Personalauswahl angewandt werden als kognitive Fähigkeitstests. Di Milia [8] zeigt für Australien einen gegenteiligen Befund: Dieser Studie zufolge werden Persönlichkeitstests seltener zur Personalauswahl herangezogen als kognitive Fähigkeitstests. In anderen Ländern (s. Tab. 11.2) werden wiederum kognitive Fähigkeitstests häufiger eingesetzt als Persönlichkeitstests (z. B. Belgien, Japan, Neuseeland, Niederlande, Schweden). Es soll nicht unerwähnt bleiben, dass es in den untersuchten Ländern eine deutliche Positivbeziehung zwischen dem Einsatz von Persönlichkeitstests und kognitiven Fähigkeitstests gibt. D. h. in Ländern, in denen eine Art dieser Tests angewandt wird, ist es sehr wahrscheinlich, dass auch die andere Testart Anwendung findet. Diese Beziehung stimmt überein mit den Ergebnissen, die Shackleton und Newell [56] schon vor 15 Jahren fanden.

Außerdem sei nicht nur an die bereits erwähnte prädiktive Validität von Persönlichkeitstests erinnert, sondern auch an eine Metaanalyse 2. Ordnung [4], in die 15 unabhängig durchgeführte Metaanalysen mit mehr als 184.000 Arbeitnehmern eingingen. Es zeigten sich bedeutsame Zusammenhänge zwischen den Persönlichkeitskonstrukten und der Arbeitsleistung (gemessen anhand objektiver Indikatoren), dem Ausbildungserfolg und der Teamarbeit (s. Tab. 11.3). Dabei ist ferner interessant, dass die Arbeitsleistung insbesondere durch Gewissenhaftigkeit und Verträglichkeit prädiktiert wird. Für den Ausbildungserfolg sind demgegenüber Extraversion und Gewissenhaftigkeit am wichtigsten. Mit dem Kriterium Teamarbeit zeigen Verträglichkeit und Gewissenhaftigkeit sowie emotionale Stabilität die stärksten positiven Zusammenhänge. Auf der Maßnahmenebene empfiehlt es sich daher insbesondere für jene Länder, die Persönlichkeitstests

**Tab. 11.3** Ausgewählte Ergebnisse der Metaanalyse 2. Ordnung zum Zusammenhang von Persönlichkeitsmerkmalen und organisationalen Ergebnisvariablen. (Quelle: [4])

Persönlichkeitsmerkmal	Objektiv gemessene Arbeitsleistung	Ausbildungserfolg	Teamarbeit
Extraversion	,13	,28	,16
Emotionale Stabilität	,10	,09	,22
Verträglichkeit	,17	,14	,34
Gewissenhaftigkeit	,23	,27	,27
Offenheit für Erfahrungen	,03	,33	,16

Anmerkung.  $N = 15$  unabhängige Metaanalysen gingen in diese Metaanalyse zweiter Ordnung ein.  $N > 184,376$  Mitarbeiter

Dargestellt sind die korrigierten Korrelationen auf der Konstruktebene

bisher nur gelegentlich für Personalauswahlzwecke verwenden, sie zukünftig stärker einzusetzen.

Eine weitere Form von Testverfahren, die im Rahmen der internationalen Personalauswahl bedeutsam sind, sind Tests zur Beurteilung von Situationen (Situational-Judgment-Tests, siehe dazu das Kapitel von Franczukowska in diesem Band). Hier geht es darum, dass man versucht, neben kognitiven Fähigkeiten weitere berufsrelevante Kriterien zu messen. Konkret heißt das, dass man Personen Situationen in schriftlicher Form vorlegt und die Personen dann angeben, welche Verhaltensstrategie aus ihrer Sicht die effektivste in der Situation X ist [32]. Salgado [47] summiert Tests zur Beurteilung von Situationen unter der Kategorie Wissenstests. Da sie aber anders konstruiert sind als Wissenstests, sollen sie hier gesondert behandelt werden.

#### Beispiel

Ein Beispiel mag zum besseren Verständnis dienen: „Ihr Sportclub plant einen Ausflug nach Berlin, um sich das Fußballspiel Deutschland/Italien anzusehen, das in zwei Wochen stattfindet. Sie sind für die Organisation des Ausflugs verantwortlich. Was beabsichtigen Sie zu tun?“ (modifiziert nach [32], übersetzt d. Verf.).

Diese Form der Tests hat eine hohe Prognosegüte [32]. Dennoch zeigt die internationale Personalauswahlpraxis, dass sie bislang kaum zum Einsatz kommen (s. Tab. 11.2). Ausnahmen von diesem Trend sind Belgien, Hongkong und Japan, wo Funktion und Effizienz von Tests zur Situationsbeurteilung für Personalauswahlzwecke schon in die Praxis diffundiert ist. Für die anderen Länder gilt, dass sie diese Testform häufiger als bislang einsetzen sollten.

### 11.2.5 Anwendung des biografischen Fragebogens im internationalen Vergleich

Im Vergleich zu Testverfahren basiert der biografische Fragebogen auf einem anderen eignungsdiagnostischen Ansatz, nämlich dem biografieorientierten Ansatz [53]. Hintergrund ist, dass der Eignungsdiagnostiker bestimmte Attribute zusammenstellt und in Form von Fragen zusammenfasst, die Vergangenheitsbezug haben (Beispielfrage: Wie wichtig war Unabhängigkeit als Grundlage Ihrer Berufswahl?, Frage aus [37, 55]). Anschließend werden diese Fragen, die nicht notwendigerweise in erkennbarem Bezug zu der angestrebten Position bzw. zur Tätigkeit stehen, mit Erfolgskriterien korreliert, sodass deutlich wird, welche Fragen den höchsten Zusammenhang mit diesen Erfolgskriterien aufweisen. Diese Korrelationen bilden dann die Grundlage für Item-Selektionen, wobei jene Fragen mit den schwächsten Korrelationen aus dem Biobogen eliminiert werden. Mitunter nutzt man hier Gruppenvergleiche, d. h. vorab werden zwei Gruppen festgelegt, die Gruppe der Erfolgreichen und die Gruppe der Erfolglosen und man analysiert im Anschluss, wie das Antwortverhalten der Erfolgreichen ausfällt und sich von der zweiten Gruppe unterscheidet.

Wenngleich bis heute unklar ist, warum der biografische Fragebogen als Verfahren der Personalauswahl funktioniert – festzuhalten ist, dass er funktioniert. So wurden meta-analytisch prädiktive Validitäten in Höhe von  $r = ,20$  bis  $r = ,50$  für den biografischen Fragebogen ermittelt [51].

- ▶ Allerdings bleibt einschränkend zu bedenken, dass der Biobogen keine homogenen Konstrukte misst [37, 55] und auch der Berufsbezug unklar ist, so dass letztlich unbekannt ist, was durch dieses eignungsdiagnostische Instrument erfasst wird. Wenden wir uns nun der Frage zu, in welchem Ausmaß der biografische Fragebogen in verschiedenen Ländern zum Einsatz kommt. Dies ist in Tab. 11.4 dargestellt.

**Tab. 11.4** Anwendung des biograischen Fragebogens im internationalen Vergleich. (Quelle [45], modifizierte Darstellung)

Land	Biografischer Fragebogen <i>M</i>
Australien	1,23
Belgien	1,52
Deutschland	2,77
Frankreich	1,20
Griechenland	3,87
Großbritannien	1,23
Hongkong	1,62
Irland	1,18
Italien	1,92
Japan	1,33
Kanada	1,19
Malaysia	1,00
Neuseeland	1,35
Niederlande	1,53
Portugal	2,29
Schweden	1,59
Singapur	1,80
Spanien	1,68
Südafrika	1,41
USA	1,21
Alle Länder	1,52

Anmerkung:  $N = 959$  Organisationen aus 20 Ländern.  $M =$  Mittelwert

Antwortskala 1 = niemals, 2 = selten (1–20 %), 3 = manchmal (21–50 %), 4 = häufig (51–80 %), 5 = sehr häufig oder immer (81–100 %)

Insgesamt gesehen, spielt der biografische Fragenbogen in der Mehrzahl der Länder eine untergeordnete Rolle. Die meisten Organisationen in 18 Ländern gaben an, ihn selten bis nie einzusetzen. Ausnahmen von dieser Regel sind Deutschland, Griechenland und Portugal, wo der Biobogen deutlich häufiger für Personalauswahlzwecke genutzt wird als in anderen Ländern. Schuler et al. [54] belegen in einer aktuellen Studie, dass der biografische Fragebogen in Deutschland in geringerem Maße Anwendung findet als dies noch vor 15 Jahren der Fall war. Wie bereits erwähnt, ist die Frage, warum durch den Biobogen valide Personalentscheidungen getroffen werden, nach wie vor ungeklärt. Ferner steigt durch den Einsatz des Biobogens, der auf der Historie des Kandidaten basiert, die Gefahr der systematischen Benachteiligung bestimmter Bewerbergruppen. Aus diesem Grund wird für die internationale Personalauswahl auch nicht empfohlen ihn stärker als bislang einzusetzen. Ein Verfahren, das gegenüber dem biografischen Fragebogen sehr häufig genutzt wird, ist das Assessment-Center.

### 11.2.6 Anwendung des Assessment-Center im internationalen Vergleich

Das AC ist nicht nur im deutschsprachigen Raum, sondern auch im internationalen Raum ein wichtiges Instrument der Personalauswahl und der Personalentwicklung [25]. AC sind „multiple diagnostische Verfahren, welche systematisch Verhaltensleistungen bzw. Verhaltensdefizite von Personen erfassen“. Hierbei schätzen mehrere Beobachter gleichzeitig für einen oder mehrere Teilnehmer seine/ihre Leistungen nach festgelegten Regeln in Bezug auf vorab definierte Anforderungsdimensionen ein [16, S. 1]. Das AC findet zunehmend stärkere Anwendung [20, 42, 54]. Es bestehen jedoch große nationale Unterschiede in der Häufigkeit, in der AC zum Einsatz kommen. Sie sind weit verbreitet in großen Organisationen in den USA, Großbritannien, Deutschland, Belgien und Dänemark, aber relativ selten in Australien, Frankreich, Spanien und Italien [56]. AC werden heute nicht nur in Industrieländern angewandt, sondern auch in Ländern wie Indonesien, Kuwait, Korea, Südafrika und den Philippinen [26, 36, 65]. Generell gilt, dass die AC-Anwendung nicht nur in Abhängigkeit von der primär verfolgten organisationalen Zielsetzung innerhalb und zwischen Organisationen eines Landes variiert [19, 22], sondern auch eine erhebliche Anwendungsvielfalt zu beobachten ist, je nachdem, in welchem Land das AC zum Einsatz kommt [18, 24]. AC werden in verschiedenen Ländern also verschieden durchgeführt, weil die situativen Randbedingungen des Personalmanagements international heterogen sind.

- Für die internationale Personalauswahl ist es wichtig, solche AC zu konstruieren und zu implementieren, die von einem Unternehmen interkulturell und international einsetzbar sind und in dem jeweiligen Land sowohl von Personalverantwortlichen als auch von AC-Teilnehmern akzeptiert werden. Eine Voraussetzung für interkulturell und international anwendbare und akzeptierte AC, deren Ergebnisse dann auch verglichen werden können, ist jedoch die Kenntnis des länderspezifischen Einsatzes des AC.



Zu dieser Frage liegen bislang wenige Untersuchungen vor von Spsychalski et al. [59], die die AC-Anwendung in den USA untersuchte, von Kudisch et al. [27], der erstmals die AC-Anwendung im internationalen Raum untersuchte, jedoch die Daten für alle Länder aggregierte, sodass mögliche Unterschiede zwischen den Ländern systematisch verdeckt werden, von Krause und Gebert [18], die die AC-Anwendung in Deutschland, Österreich und der Schweiz untersuchten, von Krause und Thornton [26], die die AC-Anwendung in Nordamerika und Westeuropa fokussieren, von Krause et al. [23], die die AC-Anwendung in Südafrika untersuchten und von Krause et al. [17], die sich mit der AC-Anwendung in Indonesien beschäftigten. Nachfolgend widmen wir uns ausgewählten Ergebnissen der Untersuchung von Krause und Thornton [26], da sie Unterschiede in der AC-Anwendung zwischen nordamerikanischen und westeuropäischen Organisationen aufzeigt. An dieser Untersuchung nahmen 45 Organisationen aus Westeuropa (Deutschland, Großbritannien, Schweiz, Österreich, Schweden, Frankreich, Niederlande) und 52 Organisationen aus Nordamerika (Kanada und USA) teil.

In beiden Regionen wird eine bedarfsorientierte Personalplanung vorgenommen und die Richtlinien zur AC-Anwendung [62] in fast allen Organisationen berücksichtigt (s. Tab. 11.5). Verglichen mit der früheren AC-Anwendung [18, 27, 59] kann dies als positiver Trend benannt werden. Die Richtlinien geben an, dass ein AC durchgeführt werden soll, um spezifische Ziele zu erreichen. Wie dargestellt (s. Tab. 11.5), setzen die meisten Organisationen AC ein, um Beförderungs- und Personalentwicklungsentscheidungen zu treffen. In etwa einem Drittel der Fälle werden AC zur Personalauswahl angewandt. Dieses Ergebnis reflektiert einen generellen Trend, Potentialentwicklungs-AC (s. [65, S. 57–76]), in denen das Lernen und die Entwicklung des Kandidaten eine dominante Rolle spielen, heute häufiger zu nutzen als in der Vergangenheit.

Verglichen mit früheren Studien von Spsychalski et al. [59] ( $M = 2,48$  Tage) und von Krause und Gebert [18] ( $M = 3,20$  Tage) reflektiert dieses Ergebnis, dass Assessment-Programme heute „schlanker“ sind als noch vor fünf Jahren. Der Bedarf nach „lean assessments“ (z. B. Personalentscheidungen von hoher Qualität in möglichst kurzer Zeit zu treffen) macht diesen Befund verständlich. Allerdings bleibt zu berücksichtigen, dass diese Sparmaßnahmen die Akkuratheit und Effektivität des AC beeinflussen [36]. Die Richtlinien zur AC-Anwendung geben darüber hinaus an, dass Organisationen eine Anforderungsanalyse durchzuführen haben. Wie deutlich wird, gaben fast alle Organisationen in Nordamerika und Westeuropa an, dies zu tun — ein Befund, der mit früheren Untersuchungen übereinstimmt.

Ein weiterer positiver Trend ist die Tatsache, dass in beiden Ländergruppen ein breites Spektrum an Anforderungsanalysen zum Einsatz kommt (s. Tab. 11.5). Dies scheint angemessen, denn eine einzelne Methode der Anforderungsanalyse ist nicht hinreichend [65]. Vergleicht man die Häufigkeiten, mit denen einzelne Methoden der Anforderungsanalyse angewandt werden, mit den Ergebnissen früherer Untersuchungen [18, 59], so wird deutlich, dass sich die Einsatzhäufigkeiten sowohl in Westeuropa als auch in Nordamerika verändert haben. In Westeuropa werden am häufigsten Interviews mit dem Vorgesetzten, neue Kompetenzmodelle und Stellenbeschreibung eingesetzt. In Nordamerika

**Tab. 11.5** Praxis der AC-Entwicklung und Durchführung in Westeuropa und Nordamerika. (Quelle: [26])

AC-Maßnahme	Praxis in Westeuropa (%)	Praxis in Nordamerika (%)
Berücksichtigung von Richtlinien zur AC-Durchführung	89	93
Hauptziele des AC		
Personalauswahl externer Bewerber	34	20
Beförderung (nächste Hierarchieebene)	25	59
Personalentwicklung	41	21
AC-Dauer		
< 0.5 Tage	–	25
1 Tag	48	38
2 Tage	39	18
3 Tage	9	13
4 Tage	2	2
> 4 Tage	2	4
Durchführung einer Anforderungsanalyse	89	95
Art durchgeführter Anforderungsanalysen		
Tätigkeitsbeschreibung	41	76
Interview mit dem Stelleninhaber	39	56
Interview mit dem Vorgesetzten	59	54
Fragebogen Stelleninhaber	14	41
Beobachtung des Stelleninhabers	7	31
Critical-incident-Technik	36	27
Fragebogen des Vorgesetzten	16	29
Neues Kompetenzmodell	43	20
Vorhandenes Kompetenzmodell	39	20
Durchführung eines Workshops in Teamarbeit	36	15
Art diagnostizierter Anforderungsdimensionen		
Kommunikation	98	90
Problemlösen	84	90
Organisation und Planung	73	83
Beeinflussung anderer	75	61
Wahrnehmung anderer/Berücksichtigung anderer	80	49
Antrieb	70	37
Art der angewandten Übungen im AC		
Rollenspiel	88	78
Präsentation***	92	58
Postkorb	35	55
Planungsübungen	40	48
Situatives Interview	48	48
Gruppendiskussion***	90	45
Fallstudie***	78	38
Interview über bisherigen Werdegang	48	40
Soziometrische Elemente	5	5

(Fortsetzung)

**Tab. 11.5** (Fortsetzung)

AC-Maßnahme	Praxis in Westeuropa (%)	Praxis in Nordamerika (%)
Anwendung zusätzlicher Diagnosemethoden (Gesamt)	38	10
Art zusätzlich angewandter Diagnosemethoden		
Wissenstests	15	30
Test von Fertigkeiten und Fähigkeiten	20	30
Persönlichkeitstests	25	20
Biografischer Fragebogen (Biodaten)	23	10
Arbeitsproben	10	15
Intelligenztest (GMA)	20	10
Graphologie	3	–

Anmerkungen: Angegeben ist die Prozentzahl der Organisationen, die die entsprechende AC-Maßnahme anwenden

Westeuropa  $N = 45$  Organisationen, Nordamerika  $N = 52$  Organisationen

\*\*\* $p < ,001$

sind dagegen Stellenbeschreibungen, Interview mit dem Positionsinhaber und Interview mit dem Vorgesetzten verbreiteter als andere Methoden der Anforderungsanalyse.

Im Hinblick auf die Art der Anforderungsdimensionen, die im AC diagnostiziert werden, nutzten Krause und Thornton [26] die Ergebnisse zweier Metaanalysen [2, 6]. Diese Metaanalysen bestätigten, dass folgende Anforderungsdimensionen konstrukt- und kriterienvalide sind: Kommunikation, Berücksichtigung der Bedürfnisse anderer Menschen, Antrieb, Beeinflussung anderer Menschen, Organisation und Planung, Problemlösen.

Die Studie von Krause und Thornton [26] zeigt, dass fast alle Organisationen Kommunikationsfähigkeiten und Problemlösen im AC diagnostizieren und zwei Drittel der Organisationen auch Beeinflussung anderer sowie Organisation und Planung als Anforderungsdimensionen zu Grunde legen (s. Tab. 11.5). Diese vier Dimensionen sind auch jene, die Kudisch in seiner Studie [27] als die am häufigsten diagnostizierten Anforderungsdimensionen bestätigte. Es bleibt darauf hinzuweisen, dass metaanalytisch gezeigt wurde, dass diese vier Dimensionen 20 % der Varianz der Arbeitsleistung erklären [2]. Daher können wir schlussfolgern, dass die am häufigsten im AC diagnostizierten Anforderungsdimensionen auch jene sind, die die höchste prädiktive Validität aufweisen. Dies gilt sowohl für Organisationen in Westeuropa als auch für Organisationen in Nordamerika. In Westeuropa werden zusätzlich deutlich häufiger als in Nordamerika „Berücksichtigung der Bedürfnisse anderer Menschen“ sowie „Antrieb“ im AC diagnostiziert. Allerdings lässt sich durch diese zwei Dimensionen die zukünftige Arbeitsleistung wesentlich schlechter prognostizieren als durch die anderen vier Dimensionen [2]. Daher besteht für westeuropäische Organisationen Verbesserungsbedarf, was die Festlegung der Anforderungsdimensionen anbelangt.

- Organisationen in beiden Kontinenten nutzen eine Vielfalt unterschiedlicher Übungen im AC. Auch dieser Tatbestand kann als positives Zeichen in der heutigen AC-Anwendung interpretiert werden, da eine positive Beziehung besteht zwischen der Anzahl angewandter AC-Übungen und der prädiktiven Validität des AC [9]. Die am häufigsten angewandten Übungen in Westeuropa sind Präsentation, Gruppendiskussion und Rollenspiel. Es gibt eine geringe Streuung in Westeuropa in Bezug auf die Anwendung dieser Übungen. Diese Befunde sind vergleichbar mit den Ergebnissen von Krause und Gebert [18], die für Deutschland, Österreich und die Schweiz belegten, dass Gruppendiskussionen, Präsentation und Interview zu den gebräuchlichsten AC-Übungen in diesen Ländern zählen.

Die Anwendung von Präsentationen und Rollenspielen ist konsistent mit dem Argument von Thornton und Rupp [64], dass situative Übungen nach wie vor den Mainstream des AC ausmachen. Viele Organisationen in Westeuropa setzen auch Gruppendiskussionen im AC ein. Dieses Ergebnis stimmt mit früheren Untersuchungen überein [18, 59]. Im Vergleich zu Westeuropa sind die am häufigsten genutzten Übungen in Nordamerika Rollenspiel, Präsentation und Postkorb. Manche dieser Übungen unterscheiden sich von den populärsten Übungen früherer Untersuchungen. So fanden Spsychalski et al. [59] beispielsweise, dass Gruppendiskussionen und Postkorb in den USA die populärsten Übungen im AC sind. Im Vergleich zeigten Krause und Thornton [26] interkulturelle Unterschiede im Hinblick auf die Anwendung bestimmter Übungen im AC: Präsentation, Gruppendiskussion und Fallstudien sind in Westeuropa verbreiteter als in Nordamerika. Die interkulturellen Unterschiede können durch die soziale Akzeptanz des AC erklärt werden. Beispielsweise bevorzugen Organisationen in Westeuropa solche Übungsarten, die die Fähigkeit des Kandidaten erfassen, mit komplexem Material umzugehen. Diese Übungsarten werden vermutlich als tätigkeitsspezifischer wahrgenommen als andere Übungen.

Die Tab. 11.5 macht außerdem deutlich, dass nur eine Minorität der Organisationen zusätzlich zu den verhaltensorientierten Übungen im AC zumindest eine weitere Diagnosemethode einsetzt – wie etwa Wissenstests, Test von Fähigkeiten und Fertigkeiten, Persönlichkeitstests, biographische Fragebogen, Arbeitsproben oder kognitive Fähigkeitstests (z. B. Intelligenztests). Wie bereits beschrieben, begegnet man insbesondere in Westeuropa dem Einsatz psychometrischer Testverfahren mit Skeptizismus. In Nordamerika besteht ebenfalls eine Abneigung, psychometrische Testverfahren im AC einzusetzen [65, S. 103]. Dies gilt vor allem für Intelligenztests (aufgrund ihrer Unterschiede für ethnische Subgruppen und der Schwierigkeit von Validierungen für verschiedene ethnische Gruppen) und Persönlichkeitstests (aufgrund der Verletzung der Privatsphäre der Kandidaten). Ferner ist der Einsatz von psychometrischen Testverfahren im AC auch dadurch begrenzt, dass man im AC das Verhalten fokussiert. Es bleibt allerdings festzuhalten, dass Testverfahren selbstverständlich in frühen Phasen des eignungsdiagnostischen Prozesses in den USA zur Anwendung kommen (Screening). Wissenstests werden häufig als Screening im ersten Schritt bei Entscheidungen über die Beförderung von

Kandidaten des öffentlichen Sektors eingesetzt [26]. Auch Tests der verbalen Fähigkeiten und grundlegender mathematischer Fähigkeiten werden in den USA häufig als Screening für Einstiegspositionen angewandt. Integrity-Tests werden in den USA häufig als Screening für Positionen im Einzelhandel angewandt. Zusammenfassend haben wir gesehen, welche Gemeinsamkeiten und welche Unterschiede in ausgewählten Aspekten der heutigen AC-Praxis in Westeuropa und Nordamerika bestehen. Dabei blieb eine Frage ungeklärt, nämlich welche Faktoren Organisationen berücksichtigen sollten, um ihr AC lokal zu adaptieren. Dieser Frage wollen wir uns nun widmen.

Multinationale und internationale Organisationen sehen sich heute mit der Herausforderung konfrontiert, ein AC in verschiedenen Ländern anzuwenden und dabei zum einen den lokalen Gegebenheiten vor Ort Rechnung zu tragen, aber parallel das AC in gewisser Weise standardisiert zu absolvieren, um international valide Leistungsprognosen aus den AC-Resultaten ableiten zu können. Ferner kann es vorkommen, dass ein AC in einem „Ursprungsland“ entwickelt worden ist (etwa in Großbritannien), dass es nun aber in einem Land mit ganz anderem kulturellen Hintergrund eingesetzt werden soll (etwa Indonesien). Zu diesen Fragen kann die bisherige AC-Forschung bislang keine zufriedenstellenden Resultate anbieten. Allerdings beschäftigen sich Wissenschaftler und Praktiker multinationaler Unternehmen in einer Expertengruppe, der sogenannten „International Task Force on Assessment Center Guidelines“ damit, zu diesen Fragen vorläufige Antworten zu generieren. Einigkeit zwischen den Experten besteht dahin gehend, dass man das bestehende AC an lokale Gegebenheiten anzupassen hat.

- Nach den Richtlinien des Subkomitees für die interkulturelle AC-Anwendung der „International Task Force on Assessment Center Guidelines“ [62] sollten für diese Adaptierungen und den Transfer bestehender AC in andere Kulturen folgende kontextuale Randbedingungen beachtet werden:
- Ausmaß der Ähnlichkeit der Geschäftsmodelle (z. B. generelle Unternehmensstrategie, Vision, Werte und Praktiken) zwischen den Ländern bzw. kulturellen Populationen,
  - Grad zentralisierter vs. dezentralisierter Kontrolle (z. B. lokal) zwischen den Einheiten der Organisation,
  - Ausmaß der Übereinstimmung zwischen betriebswirtschaftlichen und der sozio-politischen Umwelt zwischen den Ländern (z. B. lokale Gesetzgebung, Überschneidung zwischen kulturellen Werten und Verhalten),
  - Unterschiede in nationalen Richtlinien, die durch professionelle Institutionen festgesetzt wurden (z. B. „American Psychological Association’s Ethical Guidelines“ in den USA vs. „British Psychological Association’s Professional Guidelines“ in Großbritannien),
  - Ähnlichkeit der kritischen, erfolgsrelevanten Fähigkeiten,
  - Ähnlichkeit der Leistungsstandards und des kritischen, erfolgsrelevanten Leistungsverhaltens,
  - Vergleichsstatistiken zwischen Standorten,
  - Ausmaß, in dem Personal in andere Länder/Kulturen transferiert werden muss.

Personalverantwortliche, die diese Kontextfaktoren für ihr Unternehmen spezifizieren, widmen sich damit der Anforderung nach kulturspezifischer Ausgestaltung des AC. In Abhängigkeit von der Beantwortung dieser Fragenkomplexe können dann auf der Mikroebene spezifische AC-Maßnahmen je nach Land bzw. Kultur entwickelt bzw. angepasst werden (z. B. Art der Anforderungsdimensionen, Art der Diagnosetechniken, sprachliche Adaptationen des AC, Art der Simulationen, Beobachtertraining, Zusammenstellung des Beobachterpools, Beobachtercharakteristika, Methoden der Datenintegration, Evaluation des AC).

Damit haben wir einen Überblick über die AC-Anwendung in bestimmten Regionen gewonnen, Ansatzpunkte zur Verbesserung der AC-Praxis kennen gelernt und konkrete Faktoren besprochen, die bei lokalen Adaptierungen von AC für andere Zielländer zu beachten sind. Im Folgenden wollen wir uns weiteren eignungsdiagnostischen Verfahren im internationalen Vergleich widmen, wobei nun die Graphologie im Fokus steht.

### 11.2.7 Anwendung der Graphologie und sonstige Maßnahmen im internationalen Vergleich

Graphologie schließt die Analyse der Handschrift des Bewerbers ein, sodass von der Handschrift des Kandidaten auf seine Persönlichkeit und letztlich sein zukünftiges Arbeitsverhalten geschlossen wird. Die bisherige Forschung zeigt übereinstimmend, dass graphologische Gutachten in keinem Zusammenhang mit berufsrelevanten Kriterien stehen [56, 57].

- ▶ Nach den metaanalytischen Ergebnissen von Schmidt und Hunter [50] beträgt die prädiktive Validität graphologischer Gutachten  $r = ,02$ . graphologische Gutachten sind demzufolge ineffiziente Methoden der Personalauswahl.

Im Einklang mit der mangelnden Prognosekraft graphologischer Gutachten zeigt ein Blick in die internationale Personalauswahlpraxis (s. Tab. 11.6), dass sie in fast allen Ländern nie zu Personalauswahlzwecken genutzt werden. Dies kann als erfreulicher Ist-Zustand interpretiert werden. Eine Ausnahme stellt Frankreich dar, wo graphologische Gutachten noch immer als eignungsdiagnostische Verfahren in der Personalauswahl eine Rolle spielen. Dieser Befund ist konsistent mit den Ergebnissen früherer Untersuchungen [39]. Zusammenfassend lässt sich ableiten, dass in Frankreich nach wie vor ein Lern- und Veränderungsbedarf besteht, was den Einsatz graphologischer Gutachten im Rahmen der Personalauswahl angeht. Personalverantwortliche in Frankreich sollten dabei nicht nur die mangelnde Vorhersagekraft dieser Methode der Personalauswahl zur Kenntnis nehmen, sondern dieses Wissen auch in ihrer organisationalen Praxis umsetzen.

Neben den bisher dargestellten Verfahren der Personalauswahl kommen auch medizinische Gutachten im Rahmen der Personalauswahl zur Anwendung. Die Tab. 11.6 verdeutlicht, dass medizinische Begutachtungen insbesondere vorgenommen werden in Japan, Hongkong, Irland, den Niederlanden, Großbritannien und Belgien. Für die Niederlande ist dies besonders verständlich, da Arbeitnehmer dort dafür verantwortlich

**Tab. 11.6** Graphologische Gutachten/Graphologie und medizinische Begutachtung in der Personalauswahl im internationalen Vergleich. (Quelle [45], modifizierte Darstellung)

Land	Graphologische Gutachten <i>M</i>	Medizinische Begutachtung <i>M</i>
Australien	1,07	3,42
Belgien	1,56	4,20
Kanada	1,00	2,96
Frankreich	3,26	3,61
Deutschland	1,00	3,85
Griechenland	1,21	2,57
Hongkong	1,00	4,00
Irland	1,00	4,53
Italien	1,00	2,89
Japan	1,24	5,00
Malaysia	1,02	5,00
Niederlande	1,00	4,45
Neuseeland	1,00	2,69
Portugal	1,45	3,89
Singapur	1,75	3,87
Südafrika	1,27	3,59
Spanien	1,10	4,04
Schweden	1,09	3,39
Großbritannien	1,25	4,25
USA	1,07	2,60
Alle Länder	1,56	3,57

Anmerkung:  $N = 959$  Organisationen aus 20 Ländern.  $M =$  Mittelwert  
 Antwortskala: 1 = niemals, 2 = selten (1–20 %), 3 = manchmal (21–50 %), 4 = häufig (51–80 %),  
 5 = sehr häufig oder immer (81–100 %)

sind, die Arbeitskraft anderer kranker Kollegen bis zu einem Jahr zu kompensieren. Selten werden Kandidaten medizinisch begutachtet in Neuseeland, Italien, Griechenland und Kanada. Roe und van den Berg [42] argumentieren, dass die Tatsache, dass medizinische Begutachtung in den meisten europäischen Ländern für die Personalauswahl zentral ist, vor dem Hintergrund des sozialen Schutzes des Einzelnen zu verstehen ist. Insofern spiegeln sich auch hier Unterschiede in der Gesetzgebung wider. Des Weiteren werden Kandidaten in manchen Ländern auch in Bezug auf ihren Alkohol- und Drogenkonsum hin getestet. Länder, in denen dies im Rahmen der Personalauswahl von Bedeutung ist, sind Singapur und die USA [45]. Dabei unterscheiden sich US-amerikanische Organisationen erheblich in der Häufigkeit, mit der sie Kandidaten diesen Tests unterziehen: Nach Rowe et al. [44] berichten Werte zwischen 1 % und 85 % und erklären diese Streubreite durch Merkmale der Organisation.



### 11.3 Fazit

Zusammenfassend bleibt festzuhalten, dass in der Mehrzahl der Länder die am häufigsten genutzten Verfahren der Personalauswahl keineswegs jene sind, die sich in einer Vielzahl an Einzeluntersuchungen sowie Metaanalysen als die effektivsten Formen der Auswahl erwiesen haben. Diese Diskrepanz gilt insbesondere für strukturierte Interviews, Arbeitsproben, kognitive Fähigkeitstests, Integrity-Tests und Situational-Judgment-Tests. In diesem Zusammenhang ist eine Studie von Smith und Abrahamsen [58] interessant, in der eine (negative!) Korrelation in Höhe von  $r = -,25$  zwischen Anwendung von Personalauswahlmethoden und ihrer Validität festgestellt wurde. Diese Divergenz zwischen dem wünschenswerten Soll-Zustand und dem Ist-Zustand in der Personalauswahlpraxis kann viele Gründe haben: Möglicherweise passen die validesten Verfahren nicht zu den Anforderungen der Anwender z. B. im Hinblick auf Kosten – ein Umstand, der Personalpraktiker möglicherweise dazu zwingt, weniger kostenintensive Verfahren einzusetzen. Dies bezahlt der Personalmanager mit dem Preis, auch weniger adäquate Personalauswahlverfahren einzusetzen. Zu denken ist ferner an zeitliche Restriktionen, rechtliche Rahmenbedingungen oder daran, dass manche Personalauswahlverfahren als inkompatibel im Vergleich zur Unternehmensphilosophie eingeschätzt werden. Vor diesem Hintergrund führten Terpstra und Rozell [63] eine Studie durch, die zum Ziel hatte, die Gründe für die geringe Anwendung von effektiven Verfahren der Personalauswahl zu klären. Das Ergebnis dieser Untersuchung führte zu folgendem Ranking der Gründe für den mangelnden Einsatz valider Personalauswahlverfahren: geringe Vertrautheit mit dem Verfahren, Unsicherheit im Hinblick auf die Nützlichkeit der Verfahren, rechtliche Vorbehalte, Ressourcenknappheit (in Bezug auf Zeit, Geld, Expertise im Hinblick auf die Implementierung des Verfahrens), Inkompatibilität mit der Unternehmensphilosophie, Inkompatibilität zu organisationalen Zielen, ethische/soziale Vorbehalte, Befürchtung, dass das Verfahren dem Unternehmensimage schaden könnte, mangelnde Unterstützung des Top-Managements, interne politische Gründe und Befürchtungen, den Status quo zu ändern. Dennoch wäre es anzustreben, dass Personalexperten die Erkenntnisse der wissenschaftlichen Forschung auch in ihrem Alltag umsetzen, diese genannten Ursachen überwinden und mehr Zeit- und Arbeitsstunden in das Design eines international validen Personalauswahlsystems investieren.

#### Fragen

1. Welche Gemeinsamkeiten und Unterschiede kennen Sie in Bezug auf die Anwendung der Verfahren der Personalauswahl in verschiedenen Ländern?
2. Welche Verfahren der Personalauswahl weisen die höchste Prognosegüte auf? Sind diese Verfahren auch jene, die im internationalen Kontext aktuell am häufigsten genutzt werden?
3. Welche Verbesserungen schlagen Sie für die Anwendung verschiedener Personalauswahlverfahren vor?

## Literatur

1. Anderson, B. A. (2005). Expatriate selection: Good management or good luck? *International Journal of Human Resource Management*, *16*, 567–583.
2. Arthur, W. A., Jr., 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*, 125–149.
3. Baron, H., Ryan, A. M., & Page, R. (1998). Results of an international recruitment and selection survey. Paper presented at the 24<sup>th</sup> International Congress of Applied Psychology, San Francisco CA, August.
4. Barrick, M. R., Mount, M. K., & Judge, T. (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.
5. Bissessar, A. M. (2001). Differential approaches to human resource management reform in the public services of Jamaica and Trinidad and Tobago. *Public Personnel Management*, *30*, 531–548.
6. 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.
7. Cunningham-Snell, N., & Wigfield, D. (2005). Selection of leaders in global organizations. In A. Evers, N. Anderson, & O. Smit-Voskuijl (Hrsg.), *The Blackwell handbook of personnel selection* (S. 440–457). Oxford: Blackwell.
8. Di Milia, L. (2004). Australian management selection practices: Closing the gap between research findings and practice. *Asia Pacific Journal of Human Resources*, *42*, 214–228.
9. Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III, & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, *72*, 493–511.
10. Harris, M. M., & Lievens, F. (2006). Selecting employees for global assignments: Can assessment centers measure cultural intelligence? In A. Rahim, R. Golembiewski, & K. MacKenzie (Hrsg.), *Current topics in management*. Somerset: Transaction Publishers.
11. Hough, L. M., & Oswald, F. L. (2005). They're right, well ... mostly right: Research evidence and an agenda to rescue personality testing from 1960s insights. *Human Performance*, *18*, 373–387.
12. Höft, S., & Funke, U. (2006). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 145–187). Göttingen: Hogrefe.
13. Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, *79*, 184–190.
14. Huo, Y. P., Huang, H. J., & Napier, N. K. (2002). Divergence or convergence: A cross-national comparison of personnel selection practices. *Human Resource Management*, *41*, 31–44.
15. Kanning, U., & Schuler, H. (2014). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (3. Aufl., S. 215–256). Göttingen: Hogrefe.
16. Kleinmann, M. (2003). *Assessment center*. Göttingen: Hogrefe.
17. Krause, D. E., Anderson, N., Rossberger, R. J., & Parastuty, Z. (2014). Assessment center practices in Indonesia: An exploratory study. *International Journal of Selection and Assessment*, *22*(4), 384–398.
18. Krause, D. E., & Gebert, D. (2003a). A comparison of assessment center practices in organizations in German-speaking regions and the United States. *International Journal of Selection and Assessment*, *11*, 297–312.
19. Krause, D. E., & Gebert, D. (2003b). Unterschiede in der Durchführung zwischen internen und externen Assessment Centern – Eine empirische Analyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, *47*, 87–94.

20. Krause, D. E., Gebert, D., & Thornton, G. C., III. (2007). Interkulturelle Unterschiede in der Assessment Center Anwendung. In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 359–378). Göttingen: Hogrefe.
21. Krause, D. E., Kersting, M., Heggestad, E. D., & Thornton, G. C., III. (2006). Incremental validity of assessment center ratings over cognitive ability tests. A study at the executive management level. *International Journal of Selection and Assessment*, 14, 360–371.
22. Krause, D. E., Meyer zu Kniendorf, C., & Gebert, D. (2001). Aktuelle Trends in der Assessment Center Anwendung. *Wirtschaftspsychologie*, 8, 47–55.
23. Krause, D. E., Rossberger, R. J., Dowdeswell, K., Venter, N., & Joubert, T. (2011). State of the art of assessment center practices in South Africa. *International Journal of Selection and Assessment*, 19(3), 262–275.
24. Krause, D. E., & Thornton, G. C. (2005). International research project: Cultural value orientation and assessment center practices in America, Europe, and Asia. *Journal of Personnel Psychology*, 4(4), 206–208.
25. Krause, D. E., & Thornton, G. C., III. (2007). Kulturelle Vielfalt und Assessment Center: Handlungsempfehlungen für Theorie und Praxis. *Zeitschrift für Führung und Organisation*, 4, 223–230.
26. Krause, D. E., & Thornton, G. C., III. (2009). A cross-cultural look at assessment center practices: A survey in Western Europe and Northern America. *Applied Psychology: An International Review*, 58(4), 557–585.
27. Kudisch, J. D., Avis, J. M., Thibodeaux, H., & Fallon, J. D. (2001). *A survey of assessment center practices in organizations worldwide: Maximizing innovation or business as usual? Paper presented at the 16<sup>th</sup> annual conference of the Society for Industrial and Organizational Psychology*. CA: San Diego.
28. Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, 315, 1080–1081.
29. Kühlmann, T. M. (2004). *Auslandseinsatz von Mitarbeitern*. Göttingen: Hogrefe.
30. Lavigna, R., & Hays, S. W. (2004). Recruitment and selection of public workers: An international compendium of modern trends and practices. *Public Personnel Management*, 33, 237–253.
31. Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R* (erweiterte Aufl.). Göttingen: Hogrefe.
32. Lievens, F. (2006). International situational judgment tests. In J. A. Weekley & R. E. Ployhart (Hrsg.), *Situational judgment tests* (S. 279–300). Lawrence Erlbaum: SIOP Frontier Series.
33. Lievens, F., & De Paepe, A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high-structure interviews. *Journal of Organizational Behavior*, 25, 29–46.
34. Lievens, F., Van Dam, K., & Anderson, N. (2002). Recent trends and challenges in personnel selection. *Personnel Review*, 31, 580–601.
35. Lievens, F., Harris, M. M., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural training performance: The validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology*, 88, 476–489.
36. Lievens, F., & Thornton, G. C., III. (2005). Assessment centers: Recent developments in practice and research. In A. Evers, O. Voskuil, & N. Anderson (Hrsg.), *Handbook of Personnel Selection*. London: Blackwell.
37. Marcus, B. (2006). *Inventar berufsbezogener Einstellungen und Selbsteinschätzungen (IBES)*. Göttingen: Hogrefe.
38. Marcus, B., Goffin, R., Johnston, N. G., & Rothstein, M. G. (2007). Personality and cognitive ability as predictors of typical and maximum managerial performance. *Human Performance*, 20, 275–285.

39. Newell, S., & Tansley, C. (2001). International use of selection methods. In C. L. Cooper & I. T. Robertson (Hrsg.), *International Review of Industrial and Organizational Psychology* (Bd. 21, S. 195–213). Chichester: Wiley.
40. Oakland, T. (2004). Use of educational and psychological tests internationally. *Journal of Applied Psychology*, *53*, 157–172.
41. Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005). Cognitive ability in personnel selection decisions. In A. Evers, N. Anderson, & O. Voskuijl (Hrsg.), *The Blackwell handbook of personnel selection*. New York: Blackwell.
42. Roe, R. A., & Van den Berg, P. T. (2003). Selection in Europe: Context, developments and research agenda. *European Journal of Work and Organizational Psychology*, *12*, 257–287.
43. Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, *58*, 1009–1037.
44. Rowe, P. M., Williams, M. C., & Day, A. L. (1994). Selection procedures in North America. *International Journal of Selection and Assessment*, *2*, 74–79.
45. Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology*, *52*, 359–391.
46. Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology*, *82*, 30–43.
47. Salgado, J. F. (1999). Personnel selection methods. In C. L. Cooper & L. T. Robertson (Hrsg.), *International Review of Industrial and Organisational Psychology* (Bd. 14, S. 1–54). Hoboken: Wiley.
48. Salgado, J. F., & Anderson, N. R. (2002). Cognitive and GMA testing in the European community: Issues and evidence. *Human Performance*, *15*, 75–96.
49. Salgado, J. F., Anderson, N., Moscosos, S., Bertua, C., De Fruyt, F., & Rolland, J. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology*, *88*, 1068–1082.
50. 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.
51. Schmidt, F. L., & Rothstein, H. R. (1994). Applicant decisions in the employment interview. *Journal of Applied Psychology*, *88*, 979–988.
52. Schuler, H. (2002). *Das Einstellungsinterview*. Göttingen: Hogrefe.
53. Schuler, H. (Hrsg.). (2014). *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe.
54. Schuler, H., Hell, B., Trapmann, S., Schaar, H., & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie*, *6*, 60–70.
55. Schuler, H., & Marcus, B. (2006). Biografieorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (3. Aufl., S. 257–300). Göttingen: Hogrefe.
56. Shackleton, V., & Newell, S. (1997). International assessment and selection. In N. Anderson & P. Herriot (Hrsg.), *International Handbook of Selection and Assessment*. New York: Wiley.
57. Simmer, M. L., & Goffin, R. D. (2003). A position statement by the international graphonomics society on the use of graphology in personnel selection testing. *International Journal of Testing*, *3*, 353–364.
58. Smith, M., & Abrahamsen, M. (1992). Patterns of selection in six countries. *The Psychologist*, *5*, 205–207.
59. Szychalski, A. C., Quinones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, *50*, 71–90.

60. Stahl, G. K., & Caligiuri, P. (2005). The effectiveness of expatriate coping strategies: The moderating role of cultural distance, position level, and time on the international assignment. *Journal of Applied Psychology, 90*, 603–615.
61. Taylor, P., Keelty, Y., & McDonnell, B. (2002). Evolving personnel selection practices in New Zealand organizations and recruitment firms. *New Zealand Journal of Psychology, 31*, 8–18.
62. Task Force on Assessment Center Guidelines. (2009). Guidelines and ethical considerations for assessment center operations. *International Journal of Selection and Assessment, 17*, 243–254.
63. Terpstra, D. E., & Rozell, E. J. (1997). Why some potentially effective staffing practices are seldom used. *Public Personnel Management, 26*, 483–495.
64. Thornton, G. C. III, Rupp, D. R. (2004). Simulations and assessment centers. In M. Hersen (Hrsg.), *Comprehensive handbook of psychological assessment* (Bd. 4, Industrial and Organizational Assessment). Hoboken: Wiley.
65. Thornton, G. C., III, & Rupp, D. E. (2006). *Assessment centers and human resource management*. Mahwah: Lawrence Erlbaum.
66. Van Vianen, A. E. M., De Pater, I. E., & Caligiuri, P. M. (2005). Expatriate selection: A process approach. In J. L. Farr & N. T. Tippins (Hrsg.), *Handbook of employee selection*. New York: Routledge.
67. Zee, K. I., Bakker, A. B., & Bakker, P. (2002). Why are structured interviews so rarely used in personnel selection? *Journal of Applied Psychology, 87*, 176–184.

Uwe Peter Kanning

---

## Zusammenfassung

Die Fairness von Personalauswahlverfahren kann in vielfältiger Weise eingeschränkt sein: Bewerber werden „aus dem Bauch heraus“ ausgewählt, ohne dass die Entscheidungsträger dabei bemerken, dass sie in systematischer Weise bestimmte Personengruppen diskriminieren. An sich aussagekräftige Personalauswahlmethoden werden nicht professionell umgesetzt, wodurch die Validität der getroffenen Entscheidungen sinkt. Methoden, die in starkem Maße sprachgebunden sind, setzen Zuwanderer in einen Nachteil gegenüber Muttersprachlern. Darüber hinaus können manche Bevölkerungsgruppen bei auswahlrelevanten Merkmalen niedrigere Werte aufweisen, wodurch sie mit geringerer Erfolgswahrscheinlichkeit bestimmte Auswahlverfahren positiv durchlaufen. Die Akzeptanz der Personalauswahlverfahren hängt nur bedingt von ihrer diagnostischen Qualität ab. Grundsätzlich bevorzugen Bewerber Auswahlmethoden, die einen offensichtlichen Bezug zur Arbeitsplatzrealität aufweisen. Die Diskussion verschiedener Lösungsstrategien zeigt, dass insbesondere eine Orientierung an den wissenschaftlichen Prinzipien guter Personalauswahl zu einer Reduktion entsprechender Probleme beitragen kann.

---

U.P. Kanning (✉)

Hochschule Osnabrück, Fakultät Wirtschafts- und Sozialwissenschaften  
Osnabrück, Deutschland

E-Mail: U.Kanning@hs-osnabrueck.de

## Inhaltsverzeichnis

12.1	Einleitung	272
12.2	Fairness	273
12.2.1	Probleme	275
12.2.2	Lösungsansätze	282
12.3	Akzeptanz	285
12.3.1	Modelle	285
12.3.2	Auswahlmethoden im Vergleich	290
12.3.3	Bedeutung	292
12.4	Fazit	294
	Literatur	295

### 12.1 Einleitung

Ziel der Personalauswahl ist es, vakante Arbeitsplätze möglichst optimal zu besetzen. Hierzu erfolgt ein Abgleich zwischen den Anforderungen der fraglichen Stelle und den relevanten Merkmalen der Bewerber (Fähigkeiten, Fertigkeiten, Motive etc.). Im bestmöglichen Fall passt beides zueinander wie ein Schlüssel in sein Schloss. Ausgehend vom Bewerbungsverfahren, lässt sich dann die Leistung der neu eingestellten Mitarbeiter gut prognostizieren. Ein bedeutsamer Teil der personalpsychologischen Forschung dient der (Weiter-)Entwicklung diagnostischer Methoden, mit deren Hilfe sich derartige Prognosen des beruflichen Erfolgs verbessern lassen [60]. Es geht dabei z. B. um die Frage, welche Informationen aus den Bewerbungsunterlagen aussagekräftig sind [34], wie ein guter Leistungstest zur Auswahl von Auszubildenden beschaffen sein muss [21] oder welche Prinzipien bei der Gestaltung valider Einstellungsinterviews zu berücksichtigen sind [55].

Vordergründig dient diese Forschung ausschließlich den Interessen des Arbeitgebers, der eine abgesicherte Auswahlentscheidung treffen möchte. Bei tiefer gehender Betrachtung wird jedoch schnell deutlich, dass alle beteiligten Seiten von einer qualitativ hochwertigen Personalauswahl profitieren können:

1. Die Arbeitgeber identifizieren diejenigen Kandidaten, die eine hohe Arbeitsleistung erbringen, Potenzial zur Weiterentwicklung in sich tragen und aufgrund einer guten Passung zu den Anforderungen auch wahrscheinlich dauerhaft im Unternehmen verbleiben.
2. Die neu eingestellten Arbeitnehmer finden einen Arbeitsplatz, der sie weder unter- noch überfordert. Schlechte Auswahlverfahren führen u. a. dazu, dass leistungsstarke Bewerber als solche nicht erkannt werden, weil sie z. B. aufgrund formaler Fehler in den Bewerbungsunterlagen voreilig zurückgewiesen werden [36]. Alternativ werden Schwächen ungeeigneter Kandidaten übersehen, was zur Folge hat, dass sie später im Arbeitsalltag unter besonders großem Druck stehen und ggf. sogar scheitern.



3. Die Belegschaft bekommt neue Kollegen, die sich tatkräftig an der Erfüllung der gemeinsamen Arbeitsaufgaben beteiligen. Leistungsschwache Mitarbeiter bedeuten für ihre Kollegen mitunter Mehrarbeit, da sie deren Schwächen ausgleichen müssen. Leistungsstarke Mitarbeiter helfen hingegen, die wirtschaftliche Position des Unternehmens zu stärken und tragen dadurch indirekt auch zur Sicherung der Arbeitsplätze ihrer Kollegen bei.
4. Die Kunden profitieren von besonders befähigten Mitarbeitern beispielsweise durch ein höheres Serviceniveau.

Doch erfüllen die uns heute zur Verfügung stehenden Personalauswahlmethoden tatsächlich diese Erwartungen? Führen sie in der Regel zu „weisen“ Entscheidungen, die für alle Beteiligten von Vorteil sind? Und wie erleben Bewerber die Personalauswahl? Bevorzugen sie Methoden, bei denen ihnen kritisch auf den Zahn gefühlt wird oder sind sie ganz zufrieden, wenn man sich mit ein wenig „Impression Management“ geschmei­dig durch das Auswahlverfahren lavieren kann?

Derartigen Fragen wird seit einigen Jahren in der Forschung vermehrt Aufmerksamkeit geschenkt. Dabei lassen sich zwei Forschungsfelder voneinander abgrenzen, die im Folgenden vorgestellt werden: Zum einen geht es um die primär ethisch motivierte Perspektive der Fairness [56]. Hier beschäftigen sich Studien mit der Frage, inwieweit ganze Gruppen von Menschen (z. B. Afroamerikaner) durch bestimmte Auswahlmethoden diskriminiert werden. Zum anderen geht es um die Akzeptanz verschiedener Auswahlmethoden aus der Sicht der Bewerber.

---

## 12.2 Fairness

Ohne Zweifel lässt sich sehr schnell ein Konsens darüber erzielen, dass Personalauswahlverfahren fair gestaltet sein sollen. Ungleich schwerer fällt die Einschätzung, inwieweit sie es tatsächlich sind. Dies liegt ein Stück weit in der Natur der Sache. Folgen wir der Forschung zur Verteilungsgerechtigkeit – hierbei geht es um die Frage, welche grundlegenden Prinzipien bei der Verteilung knapper Ressourcen, wie etwa monetären Leistungszuschlägen als gerecht erlebt werden –, so können drei Prinzipien unterschieden werden [10]. Je nachdem, welches dieser Prinzipien zugrunde gelegt wird, resultiert eine unterschiedliche Bewertung:

- **Need-Prinzip:** Gerechtigkeit liegt vor, wenn die Ressourcen nach den Bedürfnissen der Menschen verteilt werden. Gerecht wäre Personalauswahl demnach nur dann, wenn sie dafür sorgen würde, dass z. B. gut dotierte Stellen an besonders bedürftige Bewerber vergeben würden. Geringqualifizierte Bewerber mit hohen Schulden wären daher geradezu prädestiniert, einflussreiche Vorstandsposten einzunehmen, während der Hochschulabsolvent aus wohlhabendem Elternhaus eher an der Pforte Platz nehmen sollte.

- **Equality-Prinzip:** Gerechtigkeit ist gegeben, wenn alle dasselbe bekommen. Personalauswahl wäre in diesem Fall überflüssig. Jeder Bewerber würde eine gleichwertige Stelle mit gleicher Bezahlung und gleichwertigem Einfluss erhalten. Ob eine Organisation – zumal ein Unternehmen – überhaupt überlebensfähig wäre, wenn alle Mitglieder über den gleichen Einfluss verfügen würden, sei dahingestellt.
- **Equity-Prinzip:** Gerechtigkeit liegt vor, wenn jeder entsprechend seiner Leistung an den Ressourcen partizipiert. Dies ist das Prinzip, auf das sich die meisten Menschen in unserem Kulturkreis einigen können [10, 13]. Wie später noch zu sehen sein wird, steckt der sprichwörtliche Teufel hier jedoch im Detail. Ist es beispielsweise gerecht, wenn bei der Besetzung von Führungspositionen im öffentlichen Dienst bei gleicher Qualifikation Frauen bevorzugt eingestellt werden? Ein Blick auf die Verteilung der Geschlechter in Führungspositionen lässt diese Praxis als gerecht erscheinen. Aus der Perspektive des männlichen Individuums, das gewissermaßen die Rechnung für die tradierte Ungleichverteilung bezahlen muss, sieht dies jedoch ganz anders aus. Noch fragwürdiger wird es, wenn unter den Gleichqualifizierten fünf Männer einer Frau gegenüber stehen. Eine Regelung, die auf der abstrakten Ebene sozialer Gruppen als gerecht erlebt wird, kann auf der Ebene der Individuen durchaus mit erlebter Ungerechtigkeit einhergehen.

Görlich und Schuler [20, 21] unterscheiden fünf Facetten der Fairness von Personalauswahlentscheidungen:

1. **Statistische Fairness:** Durch den Einsatz wissenschaftlich gut abgesicherter Methoden und Instrumente lässt sich sicherstellen, dass die tatsächlichen Merkmale der Bewerber hinreichend objektiv und zuverlässig erfasst werden. Ist dann auch noch nachweislich eine Validität gegeben, sind die wichtigsten Voraussetzungen für eine faire Personalauswahl erfüllt. Doch die statistische Fairness reicht noch weiter. Sie bezieht sich auch auf die Art der Normierung, die Verwendung von Quoten u. ä., mit deren Hilfe Ungleichverteilungen im Hinblick auf bestimmte Bevölkerungsgruppen ausgeglichen werden können (s. u.).
2. **Biografische Fairness:** Ereignisse aus der eigenen Biografie, die die Bewerber selbst nicht zu verantworten haben (z. B. der Beruf der Eltern), sollten nicht direkt als Kriterium in die Auswahlentscheidung einfließen. Indirekt lässt sich dies in vielen Fällen sicherlich kaum vermeiden, da z. B. ein höheres Bildungsniveau der Eltern die Wahrscheinlichkeit erhöht, selbst mehr zu lernen. Das Ergebnis dieses Sozialisierungsprozesses – also z. B. ein abgeschlossenes Studium vs. eine Ausbildung – ist durchaus ein legitimes Kriterium der Auswahl. Zudem sollte Menschen, die zu einem früheren Zeitpunkt ein Auswahlverfahren erfolglos durchlaufen haben, auch eine Chance gegeben werden, in einem späteren Verfahren ihre ggf. verbesserte Leistungsfähigkeit unter Beweis zu stellen.
3. **Anforderungsbezogene Fairness:** Personalauswahl dient letztlich der Prognose beruflichen Erfolgs. Da der Erfolg in der Regel durch mehrere Merkmale des Mit-

arbeiters beeinflusst wird und diese Merkmale kontextspezifisch wirken, ist es zwingend notwendig, die jeweils tatsächlich relevanten Merkmale durch eine Anforderungsanalyse zu identifizieren. Nur diejenigen Merkmale, die in Bezug auf die fragliche Position bedeutsam sind, sollten auch untersucht werden. Wer potenzielle Gabelstaplerfahrer schon nach der Sichtung der Bewerbungsunterlagen aus dem Rennen nimmt, weil ihr Anschreiben Grammatikfehler aufweist, verstößt gegen das Prinzip der anforderungsbezogenen Fairness.

4. **Prozedurale Fairness:** Verfahren werden umso eher als fair erlebt, je besser die Bewerber über die Anforderungen der Stelle informiert sind, ein transparentes Auswahlverfahren durchlaufen, eine gewisse Kontrolle über die eigenen Ergebnisse besitzen und ein Feedback zu ihrem Abschneiden erhalten („soziale Validität“, s. u.).
5. **Gesellschaftliche Fairness:** Bei der Vergabe der Stellen sollte das Leistungsprinzip und nicht etwa das soziale Netzwerk eines Bewerbers die entscheidende Rolle spielen. Dabei müssen jedoch auch Handicaps Berücksichtigung finden, die es einem Menschen nicht erlauben, ein besonders hohes Leistungsniveau zu erreichen. Daher wird von Unternehmen bestimmter Größe erwartet, dass sie Menschen mit Behinderung auch dann einstellen, wenn diese im Vergleich zu ihren Mitbewerbern weniger leistungsstark sind.

Im Folgenden geht es zunächst um verschiedene Phänomene, die die Fairness vieler Personalauswahlverfahren infrage stellen. In einem zweiten Schritt werden möglichen Lösungen für diese Probleme diskutiert.

## 12.2.1 Probleme

### 12.2.1.1 Systematische Fehler der Personenbeurteilung

Viele Unternehmen arbeiten in der Praxis der Personalauswahl mit einem sehr geringen Grad an Strukturierung, lassen den Entscheidungsträgern also viel Freiheit, sich selbst einen subjektiven Eindruck von den Merkmalen der Bewerber zu verschaffen. So fand z. B. Kanning [36], dass fast 47 % von 244 Unternehmen keine verbindlichen Kriterien zur Sichtung von Bewerbungsunterlagen festlegen. Eine Befragung von mehr als 500 Bewerbern ergab, dass in nur etwa 30 % der Einstellungsinterviews Interviewleitfäden zum Einsatz kamen [37]. In nur 10 % der Fälle gab es offenkundig klare Kriterien zur Bewertung der Antworten. Die große Freiheit, die den Verantwortlichen hier eingeräumt wird, öffnet systematischen Fehlern der Personenbeurteilungen, wie etwa dem Halo-Effekt oder dem Einfluss sozialer Stereotype, Tür und Tor [28, 35]. Entsprechende Effekte sind vielfach belegt worden (vgl. Tab. 12.1). Sie reichen von der Überschätzung gut aussehender Menschen über den Kurzschluss, dass Menschen mit maskulinem Körperbau bessere Führungskräfte seien, bis hin zur Abwertung von Menschen mit ausländischem Namen. Das viel beschworene „Bauchgefühl“ ist offenkundig ein schlechter Ratgeber, wenn es darum geht, Bewerber hinsichtlich ihrer Eignung für eine Stelle

**Tab. 12.1** Systematische Fehler der Personenbeurteilung

Merkmalsbereich	Urteilsverzerrung
Physische Attraktivität	<ul style="list-style-type: none"> <li>• Gut aussehende Bewerber werden positiver in solchen Merkmalen eingeschätzt, die nichts mit ihrem Aussehen zu tun haben und eher zum Einstellungsinterview eingeladen als weniger gut aussehende Menschen [45, 59, 67]</li> <li>• Besonders gut aussehende Frauen werden negativer eingeschätzt, wenn sie sich auf Führungspositionen in männlich dominierten Branchen bewerben [7, 23]</li> </ul>
Körperbau	<ul style="list-style-type: none"> <li>• Menschen mit maskulinem Körperbau (groß und breitschultrig) wird im Einstellungsinterview eine höhere Führungskompetenz bescheinigt [63]</li> <li>• Übergewichtige Menschen werden hinsichtlich ihrer Leistungsfähigkeit unterschätzt [17, 47]</li> </ul>
Ethnischer Hintergrund	<ul style="list-style-type: none"> <li>• Menschen mit ausländischem Hintergrund bzw. ausländischem Namen werden negativer bewertet als „einheimische“ Bewerber [12, 27, 51]</li> <li>• Menschen, die mit Akzent sprechen, werden negativer bewertet als Menschen, bei denen dies nicht der Fall ist [16]</li> </ul>

einzuschätzen. In dem Maße, in dem sich die Verantwortlichen unreflektierten Heuristiken der Personenbeurteilung hingeben, ist mit einer systematischen Diskriminierung bestimmter Personengruppen zu rechnen. Der Einsatz kaum strukturierter Verfahren steht somit einer Realisierung von Fairness im Sinne von „equity“ deutlich im Wege, da die Stärken und Schwächen der Bewerber nur teilweise oder verzerrt erkannt werden.

### 12.2.1.2 Einsatz diagnostisch fragwürdiger Personalauswahlmethoden

Eine zentrale Voraussetzung für eine leistungsgerechte Besetzung vakanter Stellen ist der Einsatz objektiver, reliabler und valider Methoden [29, 56]. Nur so kann sichergestellt werden, dass die arbeitsplatzrelevanten Merkmale der Bewerber auch zutreffend diagnostiziert werden, und zum einen ein Abgleich mit den Anforderungen der Stelle, zum anderen ein Vergleich zwischen den Kandidaten erfolgen kann.

Die wenigen Studien, die sich mit der Praxis der Personalauswahl beschäftigen, zeigen deutliche Defizite auf. Nur vergleichsweise selten liegen den Verfahren aussagekräftige Anforderungsanalysen zugrunde [36]. Die Sichtung der Bewerbungsunterlagen erfolgt in der Mehrzahl der Fälle, ohne dass zuvor verbindliche Kriterien festgelegt wurden [36], wobei in starkem Maße formale Kriterien herangezogen werden, deren Validität zumindest fragwürdig ist (z. B. übersichtlich gestalteter Lebenslauf [36]). Sofern Testverfahren zum Einsatz kommen, orientiert man sich bei der Auswahl stärker an ihrer Verbreitung statt an ihrer diagnostischen Qualität [42]. Obwohl Intelligenztests zu den validesten Personalauswahlverfahren zählen, spielen sie bei der Auswahl von Führungskräften in Deutschland kaum eine Rolle [62]. Im Mittelstand werden selbst bei der

Besetzung von Führungspositionen überwiegend unstrukturierte und sehr gering strukturierte Einstellungsinterviews eingesetzt [65]. Assessment-Center erfüllen oft grundlegende Qualitätskriterien nicht [40], wobei selbst in großen Unternehmen bisweilen sogar Verfahren angewandt werden, deren prognostische Validität bei Null liegt [5].

Alles in allem werden die Potenziale der unterschiedlichen Personalauswahlmethoden in der Praxis lediglich ansatzweise genutzt. Die Folge ist, dass viel zu oft die tatsächliche Eignung einzelner Bewerber nicht erkannt wird und weniger geeignete Kandidaten zum Zuge kommen. Das Phänomen des „Derailment“ – also des Scheiterns von Managern, die erst wenige Monate ihre Position innehaben – ist dabei ein besonders markanter Indikator schlechter und damit auch unfairer Personalauswahl [33].

### 12.2.1.3 Sprach- und Kulturabhängigkeit

Es gibt kein Personalauswahlverfahren, in dem die sprachlichen Fähigkeiten der Bewerber nicht zumindest indirekt eine Rolle spielen. Es beginnt bei der Abfassung der Bewerbungsunterlagen über das Bearbeiten von Fragebögen und Testverfahren bis hin zum Einstellungsinterview oder interaktiven Übungen, wie sie vor allem in Assessment-Centern eingesetzt werden. Nicht selten kommt es auf sprachliche Feinheiten an. Personen aus bildungsfernen Schichten sind hier ebenso im Nachteil wie Bewerber, die Deutsch als Fremdsprache gelernt haben [14]. Wie stark sich dieser Nachteil auswirkt und wie „unfair“ dies ist, hängt letztlich von zwei Fragen ab:

1. Wie wichtig ist der sprachliche Umgang für den späteren Berufserfolg auf der vakanten Stelle? Bei einem Arzt, der seine Diagnosen u. a. auf verbalen Informationen aufbaut, ist die Sprache weitaus wichtiger als bei einem Maschinenführer, der die meiste Zeit für sich allein tätig ist.
2. Wie stark werden die unterschiedlichen sprachlichen Ausgangsbedingungen im Rahmen des Auswahlverfahrens berücksichtigt? So könnte man z. B. auf die Interpretation von Rechtschreib- und Grammatikfehlern bei der Sichtung von Bewerbungsunterlagen verzichten und das Interview wahlweise auch in englischer Sprache abhalten, wenn bestimmte Bewerber sich im Englischen sicherer fühlen als im Deutschen.

Bei realistischer Einschätzung wird sich der Nachteil, der aus einer schlechteren Sprachkenntnis erwächst, kaum vollständig eliminieren lassen, da es praktisch nicht möglich – und aus Gründen der Validität oft auch nicht sinnvoll – ist, jeden Bewerber in seiner Muttersprache (oder gar seiner Milieusprache) zu untersuchen.

Ein ganz ähnliches Problem stellt die Kulturabhängigkeit der Personalauswahlverfahren dar [14]. Die umfangreiche Ratgeberliteratur für Bewerber definiert bestimmte Kulturstandards im Bewerbungsprozess (sich als perfekter Kandidat darstellen, Angaben im Lebenslauf zum eigenen Vorteil manipulieren, Interviewern nach dem Munde reden etc.) [35], die Bewerbern aus anderen Kulturkreisen möglicherweise unbekannt sind oder deren Umsetzung ihnen mehr Schwierigkeiten bereiten könnte. Ein Stück weit ließe sich

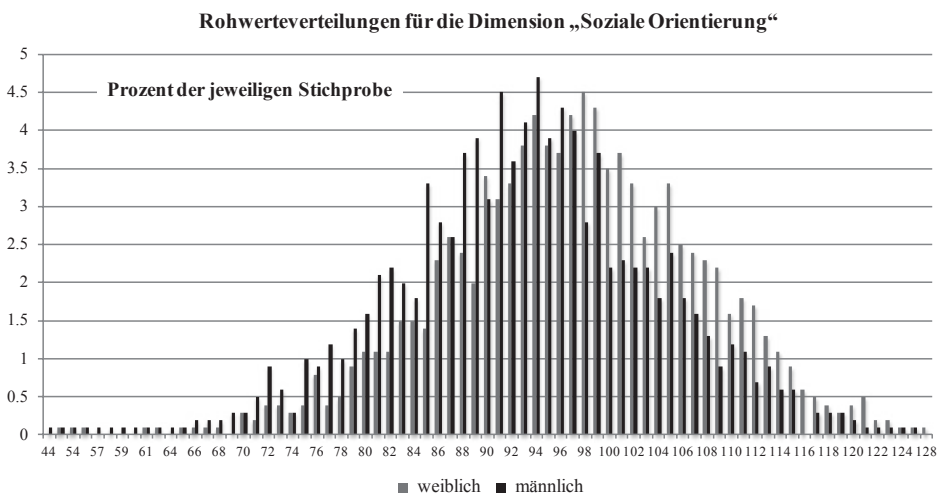
der Effekt dadurch reduzieren, dass die Personalverantwortlichen die Selbstdarstellung der Bewerber von den tatsächlichen Fähigkeiten trennen oder kulturfaire Testverfahren einsetzen (s. u.). Auch hier stellt sich jedoch immer die Frage nach der Validität. In manchen Berufen mag die offensive Selbstdarstellung ja gerade auch ein Prädiktor beruflichen Erfolgs sein [31, 56]. Eine negative Bewertung der offensiven Selbstdarsteller würde in diesem Fall einer Diskriminierung gleichkommen, die noch dazu den Arbeitgeber und die indirekt betroffenen Kollegen schädigt.

#### 12.2.1.4 Adverse Impact

Üblicherweise wird der sogenannte „adverse impact“ über Unterschiede der Mittelwerte verschiedener Personengruppen definiert. Unterscheiden sich z. B. Frauen und Männer signifikant im Mittelwert der Ergebnisse eines Fragebogens zur Messung berufsrelevanter Eigenschaften, so wird dies als Ausdruck von „adverse impact“ interpretiert. Schneiden Frauen schlechter ab als Männer, so werden sie wahrscheinlich das Auswahlverfahren negativer als ihre männlichen Konkurrenten durchlaufen.

Diese Sichtweise ist stark vereinfacht, da sie die Verteilung der einzelnen Werte ignoriert. Das verdeutlicht die Abb. 12.1: Dargestellt werden die Verteilungen der Rohdaten einer Skala zur Messung der sozialen Orientierung von mehr als 3500 Menschen, differenziert nach Frauen und Männern [30]. Beide Gruppen unterscheiden sich im Mittelwert signifikant voneinander. Die Häufigkeitsverteilungen zeigen jedoch, dass sich die Ergebnisse beider Gruppen weitgehend überschneiden. Selbst in den Extrembereichen liegen beide Gruppen sehr eng beieinander. Nehmen wir an, für eine bestimmte Stelle wären Menschen mit hoher Ausprägung des Merkmals „soziale Orientierung“ von Vorteil, so würden Männer schlechtere Chancen haben als Frauen.

Wie stark die „Benachteiligung“ der Männer in einem realen Auswahlverfahren ausgeprägt ist, hängt jedoch von (mindestens) drei Faktoren ab:



**Abb. 12.1** Verteilung der Rohwerte einer Skala zur Messung der Sozialen Orientierung. (Daten [30])

1. An welchem Punkt wird die Mindestanforderung (Cut-off) festgelegt? – Im Beispielfall führt ein Cut-off beim Punktwert 96 dazu, dass Frauen massiv im Vorteil wären. Läge der Cut-off bei einem Punktwert von 123, so wäre die Erfolgswahrscheinlichkeit für beide Gruppen fast gleich.
2. Inwieweit werden ab einem bestimmten Cut-off höhere Ausprägungen zunehmend positiver bewertet? Es könnte sein, dass eine zu stark ausgeprägte Soziale Orientierung ( $> 96$ ) ebenso negativ bewertet wird wie eine zu geringe ( $< 85$ ). In diesem Fall hätten die Männer eine größere Erfolgswahrscheinlichkeit. Ebenso gut könnte man den Cut-off auf einem sehr niedrigen Niveau (z. B. 80 Punkte) festlegen, ohne sich dafür zu interessieren, ob ein Kandidat 90 oder 120 Punkte erreicht hätte. In diesem Fall sind die Chancen für Frauen und Männer wieder in etwa gleich.
3. Welche Individuen aus der Population bewerben sich auf die ausgeschriebene Stelle? – Bei einem Cut-off von 115 Punkten würden deutlich mehr Männer eingestellt werden, falls die meisten Frauen, die diesen hohen Wert erzielen, kein Interesse an der ausgeschriebenen Stelle hätten.

Ob und inwieweit in einem konkreten Auswahlverfahren eine Benachteiligung bestimmter Personengruppen vorliegt, lässt sich mithin erst dann sicher sagen, wenn neben den Verteilungen der realen Bewerberpopulation auch die Entscheidungsregeln des Auswahlverfahrens bekannt sind.

Die Tab. 12.2 gibt einen Überblick über zahlreiche Befunde zum „adverse impact“. In den meisten Fällen zeigen sich zwar signifikante, aber doch nur geringe bis moderate Mittelwertunterschiede. Dabei sind die Effekte bezogen auf Geschlecht und Alter in der Regel geringer als die Effekte bezogen auf verschiedene ethnische Gruppen.

Im Vergleich zur primär biologisch determinierten Körperkraft fallen die Unterschiede zwischen Frauen und Männern in anderen Bereich winzig aus. Eine systematische Benachteiligung von Frauen durch bestimmte personaldiagnostische Methoden lässt sich hier nicht nachweisen. Hinsichtlich der Ergebnisse im Intelligenztest (g-Faktor) schneiden sie im Mittelwert ebenso gut oder schlecht ab wie Männer. Bei Integritätstests, Arbeitsproben, Situational-Judgment-Tests sowie im Assessment-Center sind ihre Ergebnisse im Mittelwert sogar geringfügig besser als die ihrer männlichen Konkurrenten. Allein im Persönlichkeitsfragebogen ergeben sich auf manchen Dimensionen leichte Vorteile für die Männer. Demnach sind Männer emotional stabiler und verträglicher.

Unterschiede zwischen den Altersgruppen werden seltener untersucht. Der größte Effekt ergibt sich beim logischen Denken. Hier schneiden jüngere Bewerber besser ab als ältere ( $d = ,40$ ). In den übrigen Feldern sind die Unterschiede marginal. Jüngere Menschen sind im Mittelwert ein wenig emotional stabiler, verträglicher und gewissenhafter.

Deutlich anders sehen die Befunde im Hinblick auf den Vergleich zwischen verschiedenen ethnischen Gruppen in den USA aus. Afroamerikaner erzielen sehr viel schlechtere Ergebnisse im College und schneiden schlechter in Testverfahren zur Messung des logischen Denkens ab. Ähnlich verhält es sich in Situational-Judgment-Tests, in Einstel-



**Tab. 12.2** Gruppenunterschiede in verschiedenen Merkmalen und diagnostischen Methoden

Merkmal/Methode	Größe des Unterschiedes (d)
<b>Muskelkraft</b>	
Frauen → Männer [25]	1,7
<b>college grade point average</b>	
Black → White [50]	,78
<b>Intelligenztest (g-Faktor)</b>	
Black → White [46]	,38/46
Black → White [25]	1,0
Hispanic → White [25]	,50
Asian → White [25]	,20
Männer → Frauen [25]	,00
ältere → jüngere Menschen [25]	,40
<b>Persönlichkeitsfragebogen</b>	
Black → White [15]	
Emotionale Stabilität	,09
Extraversion	,16
Offenheit	,10
Verträglichkeit	,03
Gewissenhaftigkeit	−,07
Männer → Frauen [25]	
Emotionale Stabilität	−,24
Extraversion	−,09
Offenheit	−,07
Verträglichkeit	−,39
Gewissenhaftigkeit	,08
ältere → jüngere Menschen [25]	
Emotionale Stabilität	,12
Extraversion	,01
Offenheit	−,02
Verträglichkeit	,21
Gewissenhaftigkeit	,12
<b>Integritätstest [48]</b>	
Männer → Frauen	,16
ältere → jüngere Menschen	,09
Black → White	−,04
Hispanic → White	,14
Asian → White	,04
American Indians → White	,08
<b>Arbeitsprobe (role-play work sample)</b>	
Männer → Frauen [25]	,13

(Fortsetzung)

**Tab. 12.2** (Fortsetzung)

Merkmal/Methode	Größe des Unterschiedes (d)
<b>Situational-Judgment-Test</b>	
Black → White [68]	,38
Hispanic → White [68]	,24
Asian → White [68]	,29
Männer → Frauen [68]	,11
<b>Einstellungsinterview</b>	
Black → White [26]	,25
Black → White; geringe Strukturierung [26]	,32
<b>Assessment-Center</b>	
Black → White [11]	,52
Hispanic → White [11]	,28
Männer → Frauen [11]	,19

Anmerkung: Ein positives d bedeutet, dass die zweite Gruppe einen höheren Mittelwert aufweist

lungsinterviews sowie in Assessment-Centern, wenngleich hier die Unterschiede zwischen beiden Gruppen deutlich geringer ausfallen. Nur marginalste Differenzen lassen sich hingegen bei Integritäts- und Persönlichkeitstests finden. Diese Ergebnisse erklären, warum der „adverse impact“ in den USA, vor allem bezogen auf ethnische Gruppen, intensiv diskutiert wird.

In Deutschland haben die bislang untersuchten ethnischen Gruppen rein quantitativ eine sehr viel geringere Bedeutung als in den USA. Für Deutschland wären Studien mit anderen Bevölkerungsgruppen (z. B. türkischstämmigen Bewerbern) wünschenswert.

Vor dem Hintergrund des Equity-Prinzips wäre eine Bevorzugung von Menschen, deren Gruppe einen höheren Mittelwert aufweist, kein prinzipielles Problem. Letztlich würde man ja die Person einstellen wollen, die für den fraglichen Arbeitsplatz die beste Wahl wäre. Ob dies tatsächlich so ist, hängt letztlich von der Validität ab.

Kaum jemand käme auf die Idee, die Unterschiede in der Körperkraft als Diskriminierung von Frauen zu interpretieren und, damit einhergehend, die geringe Häufigkeit von weiblichen Bauarbeitern, Stahlkochern oder Möbelpackern anzuprangern. Dies hat u. a. damit zu tun, dass die Körperkraft in diesen Fällen ganz offensichtlich eine zwingende Voraussetzung für die Berufsausübung darstellt.

Beim logischen Denken verhält es sich ein klein wenig anders. Zwar wissen wir, dass die Intelligenz der beste Einzelprädiktor beruflicher Leistung ist [54], doch gilt dies nicht für jeden Beruf in gleicher Weise [52]. Mehr noch, eine Metaanalyse von Martocchio und Whitener [46] zeigt, dass die Unterschiede im logischen Denken zwischen US-Amerikanern verschiedener Hautfarbe größer sind als die Unterschiede in der beruflichen Leistung. In einer anderen Metaanalyse konnten Berry et al. [4] zeigen, dass sich die prognostische Validität der Intelligenz für verschiedene Bevölkerungsgruppen in den

USA unterscheidet (zwischen 0,24 für Afroamerikaner und 0,33 für Menschen mit asiatischem Hintergrund). Der unreflektierte Einsatz von Intelligenztests kann also dazu führen, dass Menschen einer bestimmten Gruppe (z. B. Afroamerikaner, ältere Menschen) im Auswahlverfahren zurückgewiesen werden, obwohl sie am Arbeitsplatz gute Leistungen erbracht hätten. Im Sinne des Equity-Prinzips wäre dies nicht fair. Ob dies auch für ältere Bewerber gilt, die ebenfalls schlechter in entsprechenden Verfahren abscheiden, bleibt einstweilen unklar. Gleiches gilt für schlechtere Ergebnisse von Afroamerikanern im College, im Einstellungsinterview sowie im Assessment-Center.

### 12.2.2 Lösungsansätze

Um potenzielle Ungerechtigkeit im Prozess der Personalauswahl zu vermeiden, lassen sich verschiedene Strategien einsetzen, die im Folgenden beschrieben werden.

#### **Anforderungsbezug herstellen**

Der Dreh- und Angelpunkt eines jeden Auswahlverfahrens ist sein Anforderungsbezug. Nur wenn im Zuge einer Anforderungsanalyse [57] ermittelt wurde, welche Merkmale der Bewerber für den beruflichen Erfolg verantwortlich sind, können die passenden Personen ausgewählt werden. Dies ist im Interesse aller Betroffenen. Die Durchführung einer Anforderungsanalyse hat u. a. zu Folge, dass nicht einseitig ein einzelnes Kriterium (z. B. die Intelligenz) in das Zentrum des Verfahrens gestellt wird und auch deutlich wird, welche Mindest- und ggf. auch welche Maximalausprägung eines bestimmten Merkmals zielführend ist. Je nachdem wie die Cut-off-Werte festgelegt werden, stellt sich das Problem der systematischen Diskriminierung bestimmter Gruppe sehr unterschiedlich dar (s. o.). In der Praxis fehlen jedoch häufig aussagekräftige Anforderungsanalysen [36, 65].

#### **Objektivität herstellen**

Die Objektivität ist das grundlegendste Kriterium aussagekräftiger Personalauswahlverfahren. Sie besagt, dass das Ergebnis einer Untersuchung so weit wie möglich unabhängig von der Person des Diagnostikers ist [29, 56]. Die Objektivität setzt klare und verbindliche Kriterien zur Bewertung der Kandidaten voraus. Zumindest in Bezug auf die Sichtung von Bewerbungsunterlagen [37] und die Interviewpraxis in kleinen und mittelständischen Unternehmen [56], zeichnet sich hier ein großer Entwicklungsbedarf ab.

#### **Einsatz reliabler Methoden**

Personalauswahlentscheidungen können nur dann sinnvoll zum Ziel führen, wenn die Merkmale der Bewerber zuverlässig erfasst werden. Abgesicherte Erkenntnisse zum Status quo der Personalauswahl in Deutschland existieren in diesem Bereich nicht. Die zahlreichen anekdotischen Hinweise, dass oftmals einzelne Informationen über einen Kandidaten zu dessen Charakterisierung genutzt werden (Händedruck, optischer Eindruck der Bewerbungsmappe, Einzelfragen im Interview) [35], lässt jedoch nichts Gutes erahnen.

### **Einsatz valider Methoden**

Gute und damit auch faire Personalauswahlentscheidungen basieren letztlich auf Methoden, mit deren Hilfe sich die Merkmale der Kandidaten gültig erfassen lassen und die darüber hinaus die berufliche Leistung prognostizieren können. Auch hier ist noch viel nachzuholen [36, 42, 56], wenngleich langfristig die Qualität der eingesetzten Verfahren in großen Unternehmen zunimmt [62]. Eine Besonderheit stellen dabei Testverfahren dar. Im Gegensatz zu Einstellungsinterviews oder Assessment-Centern, die oftmals nur für ein einzelnes Auswahlverfahren konzipiert und aufgrund einmaliger Durchführung kaum vor Ort zu validieren sind, werden Testverfahren über viele Jahre hinweg unverändert eingesetzt. Hier ist eine Überprüfung der Validität eine selbstverständliche Aufgabe der Testentwickler. Die Tatsache, dass in der Praxis häufig Tests zum Einsatz kommen, deren Validität (der deutschsprachigen Variante) bestenfalls fragwürdig ist [24], verheißt nichts Gutes. Selbstverständlich müssen auch die Übersetzungen eines ursprünglich validen Verfahrens durch eigene Validitätsüberprüfungen abgesichert werden [8].

Darüber hinaus wäre zu überprüfen, ob die Validität eines Auswahlverfahrens für unterschiedliche Gruppen von Menschen in gleicher Weise gegeben ist [56]. Wenn ein Verfahren für die Gruppe A weniger valide ist als für die Gruppe B, so sollte dies bei der Aggregation der einzelnen Befunde zu einer Gesamtentscheidung berücksichtigt werden.

### **Einsatz selbst kontrollierbarer biografischer Auswahlkriterien**

Kaum ein Auswahlverfahren kommt ohne die Berücksichtigung biografischer Kriterien aus. Sie beziehen sich z. B. auf die Leistung in Schule, Ausbildung und auf vorherigen Arbeitsplätzen [58]. Bei diesen Kriterien haben die Kandidaten die Möglichkeit, durch eigene Anstrengung und individuelle Fähigkeiten bessere Ergebnisse zu erzielen („biographische Fairness“) [21, 56]. Dies gilt nicht für alle biografischen Kriterien, wie z. B. die Herkunftsfamilie. Letztlich ist für die Besetzung einer neuen Stelle weniger entscheidend, unter welchen Bedingungen ein Mensch aufgewachsen ist, als vielmehr die Frage, was er daraus selbst gemacht hat. Eine Orientierung an der Leistung bzw. den Ergebnissen der Sozialisierung wird dem einzelnen Bewerber sehr viel gerechter als eine Auswahl nach Fakten, von denen nicht bekannt ist, wie sie sich letztlich ausgewirkt haben.

### **Einsatz von Methoden mit geringer Sprach- und Kulturabhängigkeit**

Sofern die Sprachkenntnisse eines Menschen für seinen beruflichen Erfolg keine Rolle spielen, ist bei nicht-muttersprachlichen Bewerbern darauf zu achten, dass die Bewertung ihrer Leistung im Auswahlverfahren möglichst wenig mit deren Sprachfähigkeit konfundiert ist. Beim Einsatz von Testverfahren ist sprach- bzw. kulturfairen Instrumenten der Vorzug zu geben [44]. Bei Einstellungsinterviews oder in interaktiven Übungen, wie Rollenspiele, sollte der verbale Ausdruck nicht bewertet werden. Zudem sind mitunter ausführlichere Erläuterungen und ggf. mehr Zeit einzuplanen.

### **Einsatz statistischer Prozeduren**

Die schlechteren Chancen, die sich für bestimmte Gruppen von Menschen in manchen Auswahlverfahren ergeben können (vgl. Tab. 12.2), lassen sich durch verschiedene Umrechnungsprozeduren reduzieren. Görlich und Schuler [21, 56] sprechen in diesem Zusammenhang von statistischer Fairness.

Bei Testverfahren ist es üblich, das Ergebnis der einzelnen Bewerber mit denen einer großen Normstichprobe zu vergleichen. Dies ist allerdings nur dann sinnvoll, wenn die Normen nicht veraltet sind und sich auf eine vergleichbare Gruppe von Menschen beziehen. Liegen zwischen zwei Gruppen signifikante Mittelwertunterschiede vor, so können diese zum Teil durch eine gruppenspezifische Normierung ausgeglichen werden [56]. Hierdurch steigt die Wahrscheinlichkeit für Mitglieder der schwächeren Gruppe, im Auswahlverfahren positiv bewertet zu werden. Ein Bewerber aus der schwächeren Gruppe, der bezogen auf eben diese Gruppe einen Standardwert von 110 Punkten erzielt, würde demnach einem Bewerber aus der stärkeren Gruppe mit gleichem Standardwert (normiert an der stärkeren Gruppe) gleichgesetzt, obwohl er absolut gesehen eine schwächere Leistung erbracht hat. Aus Sicht der Mitglieder der stärkeren Gruppe, die absolut gesehen fähiger sind, durch die gruppenspezifische Normierung aber einen schlechteren Standardwert erhalten, ist dieses Vorgehen aber sicherlich nicht fair.

Eine ähnliche Methode stellt das „banding“ dar [56]. Hierbei werden Menschen, die nah beieinander liegende Punktwerte aufweisen, als gleich stark bewertet. Hierdurch könnten z. B. Unterschiede im IQ zwischen 100 bis 110 eliminiert werden. Gleiches könnte für den Bereich zwischen 111 und 121 gelten. Innerhalb jeder Leistungsgruppe wird eine künstliche Homogenität erzeugt, wohingegen an den Grenzen zwischen den Gruppen ein starkes Gefälle entsteht. So ist es z. B. sehr entscheidend, ob ein Bewerber einen IQ von 99 oder 100 hat, während der eigentlich viel bedeutsamere Unterschied zwischen 100 und 109 keine Rolle spielt. Im Ergebnis werden Leistungsunterschiede zwischen Gruppen (z. B. hell- und dunkelhäutige US-Amerikaner) reduziert, wodurch die Chancen auf einen sozialen Aufstieg für die schwächere Gruppe steigen. Personen mit einem IQ von 99 dürften dieses Vorgehen allerdings kaum als fair erleben.

Ähnlich verhält es sich mit der Festlegung von Quoten bezogen auf bestimmte Personengruppen [14]. Beträgt z. B. der Anteil von zwei Gruppen in der Bevölkerung (oder unter den Bewerbern) je 50 %, so könnte man 50 % der Arbeitsplätze pro Gruppe reservieren. In jeder Gruppe werden die passenden Personen ausgewählt, ohne dabei zu berücksichtigen, dass in der anderen Gruppe vielleicht noch ein besserer Kandidat existiert, der nur deshalb nicht eingestellt wird, weil die freien Plätze seiner Gruppe bereits besetzt waren.

Bei all diesen Strategien bleibt das Problem letztlich immer dasselbe: Die größere Chancengleichheit für ein Kollektiv wird auf Kosten einzelner Individuen hergestellt.

### **Einsatz anonymer Bewerbung**

Die Tatsache, dass Bewerber mit bestimmten Merkmalen (Geschlecht, Körperbau, Hautfarbe, ausländischer Name; vgl. Tab. 12.1) häufig bereits bei der Sichtung der

Bewerbungsunterlagen diskriminiert werden, führt zu der Überlegung, dass Bewerbungsunterlagen besser anonym abgefasst sein sollten. Bei anonymen Bewerbungen haben die Entscheidungsträger keine Informationen über Aussehen, Geschlecht, ethnischen Hintergrund etc. und sind daher gezwungen, sich mit den leistungsbezogenen Merkmalen auseinanderzusetzen. Erzielt wird dies entweder durch Online-Bewerbungsformulare, in denen von vornherein nur die diagnostisch relevanten Daten erfasst werden oder durch das vorübergehende Blindschalten (verdecken) der störenden Informationen im Prozess der Unterlagensichtung [43]. Die Ergebnisse eines deutschen Pilotprojektes zeigen, dass durch anonyme Bewerbungen die systematische Diskriminierung zumindest im Prozess der Vorauswahl reduziert werden kann [43].

### **Qualifizierung des Auswahlpersonals**

Alle zuvor genannten Punkte laufen letztlich darauf hinaus, dass die Entscheidungsträger in den Unternehmen in hohem Maße diagnostisch qualifiziert sein müssen. Da insbesondere in kleinen und mittelständischen Unternehmen die Personalabteilungen in aller Regel nicht über Personalpsychologen verfügen, ist hier eine entsprechende Weiterbildung anzuraten.

---

## **12.3 Akzeptanz**

Während sich die Fairness auf die Frage bezieht, inwieweit bestimmte Gruppen von Bewerbern durch Auswahlverfahren in systematischer Weise benachteiligt werden, geht es bei der Akzeptanz um das individuelle Erleben der Personalauswahl aus Sicht der Bewerber. In der englischsprachigen Literatur ist folgerichtig auch von „wahrgenommener Fairness“ die Rede [22]. Dabei spielt die Zugehörigkeit zu einer bestimmten Gruppe zunächst keine Rolle. Die Forschung zeigt, dass die Akzeptanz nicht nur für das subjektive Erleben der Bewerber von Bedeutung ist, sondern sich auch auf deren Verhalten im Auswahlprozess, auf Einstellungen gegenüber dem Arbeitgeber und die Entscheidung, ein Stellenangebot anzunehmen, auswirkt. Insofern ist es auch für die alltägliche Auswahlpraxis wichtig, sich mit diesem Thema zu beschäftigen.

### **12.3.1 Modelle**

Mehrere psychologische Modelle beschreiben die Variablen, die der Akzeptanz eines Auswahlverfahrens zugrunde liegen. Die Abb. 12.2 gibt einen Überblick.

Die Gemeinsamkeit der Modelle besteht darin, dass sie jeweils fokussieren, wie der Arbeitgeber im Zuge des Auswahlprozesses mit den Bewerbern umgehen sollte. Positiv wirkt sich nach allen Modellen die Transparenz aus. Arbeitgeber sollten demnach deutlich machen, wie sie im Verlaufe des Auswahlverfahrens zu einer Entscheidung gelangen. Darüber hinaus wird in jeweils mindestens zwei Modellen explizit darauf hin-

<p><b>Soziale Validität [61] [56]</b></p> <ol style="list-style-type: none"> <li>1. Information: Bewerber werden informiert über die Arbeitsaufgaben und Anforderungen der vakanten Stelle, zudem über die Organisation und individuelle Entwicklungsmöglichkeiten. Hierdurch wird die Möglichkeit geschaffen, sich selbst bewusst für oder gegen die Stelle zu entscheiden (Selbstselektion).</li> <li>2. Partizipation/Kontrolle: Beteiligung der Arbeitnehmervertretung an der Gestaltung des Verfahrens und Möglichkeit der Bewerber, durch eigenes Verhalten das Ergebnis des Auswahlverfahrens beeinflussen zu können.</li> <li>3. Transparenz: Aufklärung der Bewerber über den Ablauf des Verfahrens, der Entscheidungsfindung, der an der Auswahl beteiligten Personen u. ä.</li> <li>4. Urteilkommunikation/Feedback: Erklärung, warum die Entscheidung für oder gegen einen Kandidaten gefallen ist und Aufzeigen etwaiger Entwicklungsmöglichkeiten.</li> </ol>
<p><b>Modell von Gilliland [18] [19]</b></p> <p>Regeln distributiver Gerechtigkeit:</p> <ul style="list-style-type: none"> <li>• Equity</li> <li>• Equality</li> <li>• Need</li> </ul> <p>Regeln prozeduraler Gerechtigkeit:</p> <ul style="list-style-type: none"> <li>• Anforderungsbezug des Auswahlverfahrens</li> <li>• Möglichkeit, Leistung zu zeigen</li> <li>• Möglichkeit zur Korrektur</li> <li>• Konsistenz der Untersuchungsbedingungen</li> <li>• Feedback</li> <li>• Transparenz im Hinblick auf Auswahlprozedur</li> <li>• Ehrlichkeit gegenüber den Bewerbern</li> <li>• Respektvoller Umgang mit den Bewerbern</li> <li>• Zweiweg-Kommunikation</li> <li>• Angemessenheit der Fragen, die im Laufe des Auswahlverfahrens gestellt werden</li> </ul>
<p><b>Modell von Hausknecht et al. [23]</b></p> <ol style="list-style-type: none"> <li>1. Merkmale des Bewerbers: Berufserfahrung, Auswahlerfahrung, Persönlichkeit, Demographie</li> <li>2. Merkmale des Auswahlverfahrens: prozedurale Gerechtigkeit, respektvoller Umgang mit dem Bewerber, Information, Länge des Verfahrens, Wahrung der Privatsphäre, Transparenz, Schwierigkeitsniveau, Ergebnis des Verfahrens</li> <li>3. Merkmale der Stelle: Anforderungen der Stelle, Job-Stereotype, Attraktivität des Arbeitsplatzes, Einhaltung von Bestimmungen</li> <li>4. Merkmale der Organisation: Geschichte, Ressourcen, Selektionsrate</li> </ol>

**Abb. 12.2** Drei Modelle im Vergleich: Bedingungen der Akzeptanz von Personalauswahlverfahren

gewiesen, dass Auswahlverfahren einen Bezug zu den tatsächlichen Anforderungen der Arbeitsstelle aufweisen müssen, die Bewerber respektvoll behandelt werden sollten und das Ergebnis des Verfahrens dem Bewerber möglichst in einem Feedback erklärt wird. Auch wenn sich die verwendeten Konstrukte mitunter auf sehr unterschiedlichem Abstraktionsniveau bewegen, zeigt sich über die einzelnen Modelle hinweg ein weitgehender



Konsens. Das Modell von Hausknecht et al. [22] stellt insofern eine Besonderheit dar, als dass es über die wahrgenommenen Merkmale des Auswahlverfahrens hinaus auch Merkmale der Bewerber, der zu besetzenden Stelle sowie der Organisation berücksichtigt.

Schauen wir uns das prominente „Modell der wahrgenommenen Fairness von Gilliland“ [18, 19] einmal näher an. Zunächst wird zwischen zwei Grundformen der wahrgenommenen Fairness unterschieden: distributive und prozedurale Gerechtigkeit.

Die „distributive Gerechtigkeit“ (Verteilungsgerechtigkeit) wurde weiter oben bereits erläutert. In der Regel erleben Menschen, die in der westlichen Welt sozialisiert wurden ein Verteilungssystem als gerecht bei dem die Leistung des Einzelnen Berücksichtigung findet (Equity-Prinzip). Wer mehr Leistung erbringt soll demnach auch ein höheres Gehalt beziehen. Für die Personalauswahl bedeutet dies, dass die Stellen nach der Leistung bzw. der Eignung der Bewerber und nicht etwa nach Sympathie oder aufgrund von Netzwerken besetzt werden sollten.

Wurde dem Auswahlverfahren das Leistungsprinzip zugrunde gelegt, so stellt sich nun die Frage wie die konkrete Ausgestaltung des Personalauswahlverfahrens aussieht. Die einzelnen Punkte, die nach dem Modell von Gilliland [18, 19] hierbei zu berücksichtigen wären sind unter dem Begriff der „prozeduralen Gerechtigkeit“ (oder Verfahrensgerechtigkeit) zusammengefasst (vgl. Abb. 12.2). Ein konkretes Auswahlverfahren wird demzufolge u. a. dann von den Bewerbern als gerecht erlebt, wenn sich die Auswahlkriterien auf die tatsächlichen Anforderungen der Stelle beziehen (Anforderungsbezug). Wer bei der Auswahl eines Gabelstaplerfahrers dessen Grammatikfehler im Anschreiben zur Vorauswahl der Kandidaten heranzieht, verstößt gegen dieses Prinzip, da der berufliche Erfolg des Gabelstaplerfahrers gewiss nicht von seinen Grammatikkenntnissen abhängt.

Gilt das Leistungsprinzip, so müssen die Bewerber im Auswahlverfahren zudem die Möglichkeit haben, ihre Eignung unter Beweis zu stellen (Möglichkeit, Leistung zu zeigen). Wer im Einstellungsinterview primär nach seinen Hobbys und seinen Kenntnissen über den Arbeitgeber befragt wird, kann seine fachliche Eignung als Ingenieur nicht unter Beweis stellen und wird das Verfahren kaum akzeptieren können. Ebenso problematisch wäre ein Assessment-Center, bei dem die Übungen so leicht sind, dass alle Bewerber immer die volle Punktzahl erzielen, obwohl sie sich untereinander in ihrer Leistungsfähigkeit unterscheiden.

Hat ein Bewerber nach der Beantwortung einer Frage im Interview das Bedürfnis, sich noch einmal zu korrigieren, sollte dies auch möglich sein. Diese „Möglichkeit zur Korrektur“ unterscheidet sich grundlegend von einer Hilfestellung durch den Interviewer. Letzteres setzt einzelne Bewerber in einen Vorteil gegenüber den Konkurrenten. Erkennt der Bewerber von allein die Notwendigkeit zur Korrektur, so ist dies Ausdruck seiner Leistungsfähigkeit.

Selbstverständlich müssen die Untersuchungsbedingungen für alle Bewerber gleich gestaltet sein (Konsistenz der Untersuchungsbedingungen). Dies spricht u. a. für den Einsatz von Leitfäden im Einstellungsinterview. Dabei geht es nicht nur darum, dass den Bewerbern die gleichen Fragen gestellt werden, sondern auch, dass verbindliche Kriterien zur Bewertung der Antworten existieren. Dies ist der Praxis meist nicht der Fall

[35], obwohl eine solche Strukturierung des diagnostischen Vorgehens sowohl unter dem Gesichtspunkt der Bewerberakzeptanz als auch im Hinblick auf die Validität der Auswahlverfahrens geboten ist [29, 55].

Nachdem das Auswahlverfahren abgeschlossen und die Auswahlentscheidung gefällt wurde, erscheint es angebracht, den Bewerbern ein Feedback über ihr Ergebnis zu geben. Dieses Feedback sollte nicht nur das Ergebnis mitteilen, sondern auch erklären, warum ein Bewerber ggf. nicht genommen wurde. Gerade im Assessment-Center scheint dies angeraten zu sein. Die Bewerber investieren sehr viel Zeit und Mühe. Da ist es nur fair, ihnen auch ein Feedback zurückzugeben. Ein solches Feedback kann dem Bewerber helfen, die Entscheidung zu akzeptieren, auch wenn er sich wahrscheinlich ein anderes Ergebnis gewünscht hätte. Darüber hinaus mag es Anstoß für einen Entwicklungsprozess geben. Leider ist ein differenziertes Feedback die große Ausnahme in deutschen Unternehmen [35].

Bereits vor Beginn des Verfahrens sollte Transparenz bzgl. der Auswahlprozedur hergestellt werden:

- Welche Methoden kommen zum Einsatz?
- Welche Personen sind aufseiten des Unternehmens beteiligt?
- Existieren klare Auswahlkriterien?
- Wann genau wird eine Entscheidung getroffen und wie kommuniziert?

Dies sind Beispiele für Fragen, die in diesem Zusammenhang zu beantworten wären.

Die Ehrlichkeit gegenüber den Bewerbern sollte dabei eine Selbstverständlichkeit sein. Hierzu gehört z. B. auch, dass die Bewerber in Pausen des Assessment-Center oder im Warteraum vor einem Einstellungsinterview nicht verdeckt beobachtet werden. Wer ein professionelles Verfahren einsetzt, braucht Ehrlichkeit nicht zu scheuen.

Damit geht ein respektvoller Umgang mit den Bewerbern einher. Bewerber dürfen durchaus in belastende Leistungssituationen gebracht werden, sofern dies den Anforderungen der Stelle entspricht. Stress im Auswahlverfahren ist jedoch kein Selbstzweck.

Gerade im Einstellungsinterview ist zudem auf eine „Zwei-Wege-Kommunikation“ zu achten. Natürlich geht es darum, genügend Informationen über einen Bewerber zu sammeln, damit dessen Eignung gut eingeschätzt werden kann. Es geht jedoch nicht nur um die Frage, welcher Bewerber aus der Perspektive des Arbeitgebers geeignet erscheint (Fremdselektion), sondern auch um die Frage, ob der Bewerber glaubt, dass er zu dem Arbeitsplatz bzw. zum Arbeitgeber passt (Selbstselektion). Der Bewerber muss daher die Möglichkeit haben, seinerseits Fragen stellen zu können.

Auch wenn jedes Auswahlverfahren eine wichtige Investitionsentscheidung für einen Arbeitgeber darstellt, so gilt es doch eine gewisse Angemessenheit nicht aus dem Blick zu verlieren. Je einflussreicher die Stelle im Unternehmen ist, desto mehr muss der Bewerber auch von sich erzählen. Der sprichwörtlich gläserne Kandidat ist für eine fundierte Auswahlentscheidung aber nicht notwendig.

Im Gesamteindruck der verschiedenen Modelle, die in Abb. 12.2 skizziert wurden, ergibt sich ein Geflecht zahlreicher Variablen, die für die Akzeptanz von Personalauswahlverfahren relevant sind. In der Metaanalyse von Hausknecht et al. [22] zeigt sich, dass den Merkmalen der Bewerber im Vergleich zu Merkmalen des Auswahlverfahrens eine eher untergeordnete Bedeutung zukommt. Ihr Einfluss ist in der Regel nicht einmal signifikant. Bei den Merkmalen des Auswahlverfahrens bewegen sich die Effektstärken in einem Bereich von  $p = ,17$  bis  $,63$ . Die zwei wichtigsten Variablen sind der wahrgenommene Bezug des Auswahlverfahrens zu den Anforderungen des Arbeitsplatzes ( $p = ,61$ ) sowie die wahrgenommene prognostische Validität ( $p = ,63$ ). Dies spricht für den Einsatz simulationsorientierter Verfahren der Personalauswahl [38] sowie für situative Fragen im Einstellungsinterview [58].

Dabei darf nicht außer Acht gelassen werden, dass die Akzeptanz in nennenswertem Maße auch davon abhängt, wie die befragten Personen in einem Auswahlverfahren abschneiden. Je positiver das eigene Ergebnis ausfällt, desto besser erscheint den Betroffenen das Verfahren:

- $r = ,29$ ; Felduntersuchung über verschiedene AC-Übungen hinweg [32],
- $r = ,79$ ; Laboruntersuchung bezogen auf Leistungs- und Integritätstests [49],
- $r = ,27$  bis  $,42$ ; Felduntersuchung; überwiegend Einstellungsinterviews [53],
- $r = ,58$  bis  $,79$ ; Laboruntersuchung, Auswahltest [3].

Zur Aufrechterhaltung des Selbstwerts werden solche Verfahren, die selbstwertbedrohliche Ergebnisse produzieren mit Skepsis betrachtet. Im Einklang hierzu fanden Schinkel et al. [53], dass abgelehnte Bewerber sich wohler fühlen, wenn sie das Auswahlverfahren als unfair bewerten können, während erfolgreiche Bewerber sich wohler fühlen, wenn sie das Verfahren als fair erlebt haben. In beiden Fällen haben die Bewerber dasselbe Auswahlverfahren durchlaufen, wählen aber jeweils entgegengesetzte Strategien zur Absicherung eines positiven Selbstbildes.

Wer schlecht abschneidet, neigt dazu die Verantwortung hierfür in der Umwelt (dem Auswahlverfahren, dem Auswahlpersonal etc.) und nicht bei sich selbst zu sehen. Dies ist weniger selbstwertgefährdend als wenn die betroffene Person sich selbst für das Ergebnis verantwortlich sieht und annimmt, sie wäre z. B. nicht hinreichend intelligent oder leistungsmotiviert.

Wer positiv abschneidet interpretiert die Situation genau entgegengesetzt. In diesem Fall ist es positiv für den Selbstwert, wenn die Person glaubt, sie habe das gute Ergebnis aufgrund der eigenen Anstrengung und Fähigkeit zu Recht erzielt. Selbstwertgefährdend wäre hingegen, das positive Abschneiden als Zufall oder Fehler der Diagnostik anzusehen. In diesem Sinne wählt der Betroffene also jeweils die Ursachenzuschreibung (= Attribution), die für ihn die vorteilhafteste ist.

### 12.3.2 Auswahlmethoden im Vergleich

Seit fast 30 Jahren gibt es Studien, in denen die Akzeptanz verschiedener Auswahlmethoden miteinander verglichen werden. Die Tab. 12.3 gibt einen Überblick über die Befunde. Über die Jahrzehnte hinweg ergibt sich dabei eine erstaunliche Konsistenz. Zu den Methoden, die durchweg die höchste Akzeptanz finden, gehört das Einstellungsinterview sowie die Arbeitsprobe. Beide Methoden sind etablierte Bestandteile vieler Auswahlverfahren, wobei insbesondere die Arbeitsprobe einen offensichtlichen Bezug zu den Arbeitsplatzanforderungen aufweist. Im Mittelfeld bewegen sich Leistungstests und Persönlichkeitsfragebogen, während Glaubwürdigkeitstests, persönliche Beziehungen sowie die Grafologie am unteren Ende der Skala liegen. Bewerber wollen offenkundig, dass man sich individuell mit ihrer Person auseinandersetzt und dass ein Bezug zu den Arbeitsplatzanforderungen besteht. Dies zeigt auch die stärkere Präferenz für AC-Übungen (Selbstvorstellung, Präsentation, Verhaltensübungen), die den Berufsalltag simulieren (Rollenspiel, Gruppendiskussion). Testverfahren sind im Vergleich hierzu abstrakt und für Bewerber schwer zu durchschauen. Das wiederholte Abfragen ähnlicher Sachverhalte im Persönlichkeitsfragebogen mag zudem den Eindruck der Kontrolle verstärken. Das schlechte Abschneiden von Glaubwürdigkeitstests passt zu dieser Interpretation. Am schlechtesten schneiden die Grafologie – eine nachweislich aussageleere Methode – sowie der Einsatz persönlicher Beziehungen ab. Letzteres kann als Hinweis darauf gedeutet werden, dass auch Bewerber mehrheitlich der Meinung sind, dass tatsächlich die besten Kandidaten zum Zuge kommen sollten (Equity-Prinzip).

Interessant sind die Unterschiede zwischen USA und Frankreich [64]. Während in den USA Interviews, Lebenslauf, biografische Fragebogen und Glaubwürdigkeitstests signifikant positiver gesehen werden als in Frankreich, gilt für Persönlichkeitstests und grafologische Gutachten die Umkehrung. Der bei weitem größte Unterschied ergibt sich in Bezug auf die Grafologie, eine Methode, die in Frankreich auch heute noch weit verbreitet ist. Allerdings täuscht der hohe Mittelwert ein wenig. In der Studie von Steiner und Gilliland [64] wurde eine bipolare Skala eingesetzt. Der Mittelwert von 3,23 liegt im Bereich der Ablehnung. Die Grafologie wird somit in beiden Ländern von den Bewerbern abgelehnt, in den USA jedoch weitaus stärker als in Frankreich.

Mehrere Studien beschäftigen sich tiefer gehend mit einzelnen Auswahlmethoden. So konnten z. B. Klingner und Schuler [41] zeigen, dass sogenannte Hybridverfahren von Bewerbern positiver bewertet werden als klassische Intelligenztests. Hybridverfahren erfassen die kognitive Leistungsfähigkeit, verwenden dabei jedoch Aufgabentypen und Inhalte, die einer Arbeitsprobe nahe kommen. Kanning et al. [39] konnten zeigen, dass Bewerber situative Testverfahren (Situational-Judgment-Tests) positiver bewerten, wenn diese mit Videofilmen arbeiten, die Situationen aus dem Berufsalltag zeigen. Zudem schnitten interaktive Items – also Items, bei denen sich eine Situation in Abhängigkeit von der Antwort des Bewerbers verändert – positiver ab als nicht-interaktive Item-Formate. Beide Studien unterstreichen die Erkenntnis, dass Verfahren vor allem dann eine

**Tab. 12.3** Akzeptanz verschiedener Auswahlmethoden

Schüler [56] (Rang- ordnung)	Steiner & Gilliland [64] <sup>a</sup>			Kanning [32] (bez. auf Assessment- Center)	
		USA	Frankreich		
Interview	Interview	5,39	>	4,56	Interview (5,94)
Arbeitsprobe	Lebenslauf	5,37	>	4,54	Selbstvorstellung (5,80)
Praktikumsleistung	Arbeitsprobe	5,26		5,26	Präsentation (5,44)
Zeugnisnoten	Biogr. Frag.	4,59	>	3,91	Gruppendiskussion (5,30)
Eignungstest	Leistungstest	4,50		4,21	Rollenspiel (5,23)
Lebenslauf	Referenz	4,38		4,12	Planungsaufgabe (4,67)
Handschrift	Pers.-Test	3,50	<	3,96	Stegreifrede (4,61)
Losverfahren	Glaubw.-T.	3,41	>	2,54	Pers.-Test (4,15)
	Beziehungen	3,29		2,92	Leistungstest (4,19)
	Grafologie	1,95	<	3,23	Konstruktionsauf- gabe (4,03)
Hausknecht et al. [22] <sup>b</sup> (Metaanalyse)		Anderson et al. [2] <sup>a</sup> (Metaanalyse)			
Interview (3,84)		Arbeitsprobe (5,38)			
Arbeitsprobe (3,61)		Interview (5,22)			
Lebenslauf (3,57)		Lebenslauf (4,97)			
Referenz (3,29)		Leistungstest (4,59)			
Leistungstest (3,11)		Referenzen (4,36)			
Pers.-Test (2,83)		Biografischer Fragebogen (4,28)			
Biografischer Fragebogen (2,81)		Pers.-Test (4,08)			
Pers. Kontakt (2,51)		Glaubwürdigkeitstest (3,69)			
Glaubwürdigkeitstest (2,47)		Beziehungen (2,59)			
Grafologie (1,69)		Grafologie (2,33)			

Anmerkung: <sup>a</sup>siebenstufige Skala (1–7)

<sup>b</sup>fünfstufige Skala (1–5)

< bzw. > = signifikanter Unterschied zwischen der amerikanischen und der französischen Stichprobe

hohe Akzeptanz bei den Bewerbern finden, wenn sie einen offenkundigen Bezug zur Arbeitsrealität aufweisen.

Ein weiterer Punkt der sich positiv auf die Akzeptanz auswirkt, ist die Vertrautheit der Bewerber mit einer diagnostischen Methode. Je häufiger Bewerber bestimmte Formen des Einstellungsinterviews, des Assessment-Center oder der Testdiagnostik erlebt haben, desto größer fällt die Akzeptanz für diese Methoden aus [37]. An dieser Stelle

**Tab. 12.4** Zusammenhang zwischen der Akzeptanz von Auswahlmethoden und ihrer Verbreitung sowie ihrer prognostischen Validität. (Nach Anderson [2])

Facetten der Akzeptanz von Auswahlmethoden	Verbreitung der Auswahlmethoden in der Praxis	Validität der Auswahlmethoden
Bevorzugung durch die Bewerber	,65 <sup>a</sup>	,66 <sup>a</sup>
Wahrgenommene Wissenschaftlichkeit	,49	,70 <sup>a</sup>
Berücksichtigung von Rechten der Bewerber	,68 <sup>a</sup>	,57
Möglichkeit eigene Leistungsfähigkeit unter Beweis zu stellen	,66 <sup>a</sup>	,64 <sup>a</sup>
Interpersonelle Wärme	,55	,44
Subjektiv eingeschätzte Validität (Augenscheinvalidität)	,69 <sup>a</sup>	,63 <sup>a</sup>
Eingeschätzte Verbreitung der Methoden	,78 <sup>a</sup>	,34
Respekt vor der Privatsphäre der Bewerber	,42	,43

Anmerkung: <sup>a</sup>p < ,05

wirkt wahrscheinlich der Mere-Exposure-Effekt [6, 69], der besagt, dass mit zunehmender Vertrautheit eines Stimulus dessen Bewertung positiver ausfällt. Manche Methoden, die trotz guter prognostischer Validität bislang keine hohe Akzeptanz finden – wie z. B. Intelligenztests – könnten in Zukunft sehr viel positiver abschneiden, sofern Arbeitgeber sie häufiger einsetzen würden.

Letzteres findet sich auch in den Daten von Anderson et al. [2] wieder. Auf einem abstrakteren Analyseniveau gehen sie der Frage nach, inwieweit die Akzeptanz von Auswahlverfahren im Allgemeinen mit ihrer Verbreitung sowie mit ihrer Aussagekraft (prognostische Validität) korrelieren. Tab. 12.4 gibt die Ergebnisse wieder. Sofern signifikante Zusammenhänge vorliegen, zeigt sich dabei, dass die Methoden insgesamt umso positiver bewertet werden, je häufiger sie zum Einsatz kommen und je höher ihre prognostische Validität ausfällt.

### 12.3.3 Bedeutung

Die Bedeutung der Akzeptanz eines Auswahlverfahrens reicht weit über das subjektive Wohlbefinden der Bewerber hinaus. Selbst wenn einem Arbeitgeber das subjektive Erleben seiner Bewerber nicht interessieren sollte, gibt es mehrere gute Gründe, sich über die

Akzeptanz des eigenen Auswahlverfahrens Gedanken zu machen. In einem Literaturüberblick beschreiben Truxillo et al. [66] die folgenden Effekte:

- Die Akzeptanz überstrahlt das Auswahlverfahren. Bewerber, die das Auswahlverfahren als fair erleben, bewerten auch den Arbeitgeber insgesamt positiver. Dieser Effekt ist jedoch nicht von langer Dauer.
- Organisationen mit akzeptablen Auswahlverfahren werden von den Bewerbern anderen Menschen als potenzielle Arbeitgeber empfohlen.
- Die Akzeptanz des Verfahrens geht mit einer höheren Selbstwirksamkeit der Bewerber sowohl in Bezug auf ihr Abschneiden im Auswahlprozess als auch in Bezug auf die spätere Arbeitsleistung im Berufsalltag einher.
- Bewerber, die ein Verfahren als unfair erleben, sind eher bereit, ihre Bewerbung zurückzuziehen.
- Nach einer Absage steigt die Wahrscheinlichkeit, sich beim selben Arbeitgeber erneut zu bewerben, wenn die betroffenen Personen das erste Auswahlverfahren als fair erlebt haben.
- Ein offensichtlicher Bezug zu den Anforderungen des Arbeitsplatzes führt bei Leistungstests zu besseren Ergebnissen. Für Persönlichkeitsfragebogen gilt dies nicht.
- Wahrgenommene Fairness des Auswahlverfahrens geht, zumindest vorübergehend, mit einem höheren Commitment der eingestellten Bewerber einher.

Hausknecht et al. [22] fanden in ihrer Metaanalyse substanzielle Zusammenhänge zwischen verschiedenen Maßen der Akzeptanz und folgenden Variablen:

- Leistung im Auswahlverfahren ( $p = ,09$  bis  $,31$ ),
- wahrgenommene Attraktivität des Arbeitgebers ( $p = ,37$  bis  $,51$ ),
- Bereitschaft, den Arbeitgeber anderen zu empfehlen ( $p = ,40$  bis  $,52$ ),
- Bereitschaft, ein Stellenangebot anzunehmen ( $p = ,30$  bis  $,34$ ),
- Selbstwirksamkeit ( $p = ,10$  bis  $,33$ ),
- Selbstwert ( $p = ,29$ ).

In diese Metaanalyse fließen sowohl Laboruntersuchungen mit simulierten Auswahlverfahren als auch Felduntersuchungen ein. Die Effekte der Felduntersuchungen sind in den meisten Fällen geringer als die der Laboruntersuchungen, was darauf zurückgeführt werden kann, dass in realen Auswahlverfahren viele weitere Variablen Einfluss nehmen. Beispielsweise bewerben sich viele Menschen nicht nur in Unternehmen, bei denen sie unbedingt arbeiten möchten. Erscheint der Arbeitgeber nicht attraktiv, so besteht auch wenig Veranlassung, ihn weiterzempfehlen oder sich in besonderem Maße anzustrengen.

Schinkel et al. [53] fanden in ihrer Feldstudie Zusammenhänge von  $r = ,05$  bis  $,37$  für verschiedene Maße der Akzeptanz und der erlebten Attraktivität des Arbeitgebers. Bei Ababneh, Hackett und Schat [1] ergab sich in einer Laboruntersuchung ein Zusammenhang



von ,30, wobei die Bereitschaft, den Arbeitgeber zu empfehlen, zu ,62 von der wahrgenommenen Fairness des Verfahrens abhing.

All dies verdeutlicht, dass es wichtig ist, sich als Arbeitgeber mit diesem Thema zu beschäftigen. Mitunter mag dabei die gute Validität eines Auswahlverfahrens (z. B. Intelligenztest) im Widerspruch zu den Bemühungen um Bewerberakzeptanz stehen. Die Lösung für dieses Dilemma kann selbstverständlich nicht sein, sich einseitig um die Akzeptanz zu sorgen und dabei die prognostische Validität des Auswahlverfahrens aus dem Blick zu verlieren. Die Studie von König et al. [42] zeigt, dass viele Arbeitgeber bei der Gestaltung ihrer Auswahlverfahren mehr Wert auf die (vermeintliche) Bewerberakzeptanz als auf die prognostische Validität legen. Hier wird das Kind mit dem Bade ausgeschüttet. Niemandem ist letztlich mit einem Verfahren gedient, bei dem die Bewerber sich rundum wohl fühlen, das aber seine primären Auftrag nicht erfüllt. Im Zentrum einer für alle Seiten aussagekräftigen und damit auch gerechten Personalauswahl steht die prognostische Validität. Es geht darum, valide Auswahlverfahren derart zu gestalten, dass sie so weit wie möglich auch Akzeptanz finden. Ist die Akzeptanz gering, so muss man sich um Aufklärung bemühen. Der Intelligenztest liegt wahrscheinlich nur deshalb im Mittelfeld der Akzeptanz, weil die Aufgaben per Augenschein nichts mit den beruflichen Aufgaben zu tun haben. Hier ist zum einen anzuraten, den Bewerbern die Funktion und Aussagekraft der Verfahren zu erklären, zum anderen könnte so weit wie möglich auf Hybridverfahren ausgewichen werden [56].

Grundsätzlich sollten sich die Verantwortlichen nicht davor fürchten, leistungsstarke Bewerber einem fordernden Verfahren auszusetzen. Die Metaanalyse von Chapman et al. [9] zeigt zwar, dass die Bereitschaft, ein Stellenangebot anzunehmen in starkem Maße vom Auswahlprozess und dem Verhalten des Auswahlpersonals abhängt, viel bedeutsamer ist aber die Möglichkeit, Leistung im Auswahlverfahren zeigen zu können. Leistungsstarke Kandidaten wollen nicht eingestellt werden, weil sie einem Personalverantwortlichen gefallen, sondern weil sie ihre Leistungsfähigkeit erfolgreich unter Beweis stellen konnten.

---

## 12.4 Fazit

Fairness in der Personalauswahl ist ein vielschichtiges Phänomen. Dabei ist zu unterscheiden zwischen mangelnder Fairness, die sich aus unprofessioneller Personalauswahlpraxis ergibt, und mangelnder Fairness, die aus der unterschiedlichen Verteilung auswahlrelevanter Eigenschaften in bestimmten Personengruppen entstehen kann.

Im ersten Fall haben Bewerber keine Chance, ihre Eignung für eine bestimmte Stelle unter Beweis zu stellen, da sie z. B. aufgrund nicht-valider Kriterien, frühzeitig ausscheiden. Andere Bewerber werden trotz guter Eignung nicht eingestellt, da ihre Eignung in einem unstrukturierten Interview oder einem schlechten Persönlichkeitsfragebogen nicht erkannt wurde. Die Lösung für diese Art von Problemen ist sehr einfach. Die Verantwortlichen in den Unternehmen müssten Forschungsergebnisse zur Kenntnis nehmen und praktisch umsetzen. Der Alltag der Personalauswahl in Deutschland ist von diesem Ideal noch sehr weit entfernt [34].

Im zweiten Fall ist die Sachlage komplizierter. Wenn z. B. Menschen mit Migrationshintergrund im Mittelwert schlechtere Ergebnisse in einem auswahlrelevanten Merkmal aufweisen oder grundsätzlich bei einer bestimmten Methode schlechter abschneiden, so ist die Wahrscheinlichkeit, ein entsprechendes Verfahren positiv zu durchlaufen, für Vertreter dieser Gruppe reduziert. Amerikanische Studien lassen erwarten, dass derartige Probleme in Deutschland nicht im Hinblick auf Geschlechtsgruppen, sondern eher bei ethnischen Gruppen auftreten (vgl. Tab. 12.2). Dabei stehen individuelle und kollektive Interessen einander nicht selten gegenüber. Verwendet man z. B. gruppenspezifische Normen und erhöht damit die Wahrscheinlichkeit für Vertreter der schwächeren Gruppe A, ein Auswahlverfahren erfolgreich zu durchlaufen, so führt dies dazu, dass einzelne Vertreter der im Mittelwert stärkeren Gruppe B absolut gesehen mehr leisten müssen, um sich gegen einen an sich schwächeren Kandidaten durchsetzen zu können. Die Förderung des einen Kollektivs geht mit einer Diskriminierung einzelner Individuen aus dem anderen Kollektiv einher. Ein Ausweg aus diesem Dilemma wäre die Ableitung von gruppenspezifischen Entscheidungsregeln, die auf umfangreichen Studien zur differenziellen prognostischen Validität beruhen müssten. Ließe sich nachweisen, dass beispielsweise für eine Gruppe A die erfassten Daten in anderer Weise zu einer Auswahlentscheidung aggregiert werden müssten als bei Gruppe B, um dadurch für Vertreter beider Gruppen zu einer gleichermaßen guten Prognose der beruflichen Leistung gelangen zu können, wäre dies ein großer Schritt in Richtung einer für alle Beteiligten fairen Lösung. Hiervon sind die Forschung und noch sehr viel mehr die alltägliche Auswahlpraxis jedoch sehr weit entfernt. Ein zweiter Lösungsansatz ist eher im gesellschaftspolitischen Bereich anzusiedeln. Es stellt sich die Frage, warum bestimmte Bevölkerungsgruppen ggf. im Mittelwert geringere Merkmalsausprägungen aufweisen und wie Vertreter dieser Gruppen bereits in der Schule gefördert werden können, damit in der nächsten Generation derartige Unterschiede verschwinden.

---

### Fragen

1. Inwiefern führen Bauchentscheidungen in gering strukturierten Auswahlverfahren zu systematischen Fehlern bei der Beurteilung von Bewerbern?
2. Inwieweit führt der Einsatz bestimmter diagnostischer Methoden zu signifikant unterschiedlichen Ergebnissen bei Angehörigen verschiedener Bevölkerungsgruppen?
3. Welche Auswahlmethoden finden eine besonders hohe Akzeptanz bei Bewerbern?
4. Wie lässt sich der vorhandene Widerspruch zwischen der hohen prognostischen Validität eines Intelligenztests und seiner geringe Akzeptanz in der Praxis auflösen?

---

### Literatur

1. Ababneh, K. I., Hackett, R. D., & Schat, A. C. H. (2014). The role of attributions and fairness in understanding job applicant reactions to selection procedures and decisions. *Journal of Business Psychology, 29*, 111–129.

2. Anderson, N., 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.
3. Bernerth, J. B., Feild, H. S., Goles, W. F., & Cole, M. S. (2006). Perceived fairness in employee selection: The role of applicants personality. *Journal of Business and Psychology*, 20, 545–563.
4. Berry, C. M., Clark, M. A., & McClure, T. K. (2011). Racial/ethnic differences in the criterion-related validity of cognitive ability tests: A qualitative and quantitative review. *Journal of Applied Psychology*, 96, 881–906.
5. Boltz, J., Kanning, U. P., & Hüttemann, T. (2009). Qualitätsstandards für Assessment Center – Treffende Prognosen durch Beachtung von Standards. *Personalführung*, 10, 32–37.
6. Bornstein, R. F. (1989). Exposure and Affect: Overview and Meta-Analysis of Research, 1968-1987. *Psychological Bulletin*, 106, 65–289.
7. Braun, S., Peus, C., & Frey, D. (2012). Is beauty beastly? Gender-specific effects of leader attractiveness and leadership style in followers' trust and loyalty. *Zeitschrift für Psychologie*, 220, 98–108.
8. Byrne, Z. S., Mueller-Hanson, R. A., Cardador, J. M., Thornton, G. C., III, Schuler, H., Frintrup, A., & Fox, S. (2004). Measuring achievement motivation: Tests of equivalency for English, German, and Israeli versions of the achievement motivation inventory. *Personality and Individual Differences*, 37, 203–217.
9. Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., & Jones, D. A. (2005). Applicant attraction to organizations and job choice: A Meta-Analytic review of the correlates of recruiting outcome. *Journal of Applied Psychology*, 90, 928–944.
10. Conlon, D. E., Porter, C. O. L. H., & Parks, J. M. (2004). The fairness of decision rules. *Journal of Management*, 30, 329–349.
11. Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center Ratings. *Journal of Applied Psychology*, 93, 685–691.
12. Deros, E., Nguyen, H. H., & Ryan, A. M. (2009). Hiring discrimination against Arab minorities: Interactions between prejudice and job characteristics. *Human Performance*, 22, 297–320.
13. Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, 31, 137–149.
14. Flubacher, B., Schuler, H., & Frintrup, A. (2011). *Menschen mit Migrationshintergrund in der Bundesverwaltung: Modelentwurf zur Gestaltung personaldiagnostischer Prozesse*. Stuttgart: Unveröffentlichtes Gutachten. S & F Personalpsychologie.
15. Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analysis comparing five US racial groups. *Personnel Psychology*, 61, 579–616.
16. Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta analysis of the effect of speakers' accents on interpersonal evaluation. *European Journal of Social Psychology*, 42, 120–133.
17. Giel, K. E., Zipfel, S., Alizadeh, M., Schäfferle, N., Zahn, C., Wessel, D., Hesse, F. W., Thiel, S., & Thiel, A. (2012). Stigmatization of obese individuals by human resource professionals: an experimental study. *BMC public health*, 12(1), 1–9.
18. Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, 18, 694–734.
19. Gilliland, S. W. (1995). Fairness from the applicant's perspective: Reactions to employee selection procedures. *International Journal of Selection and Assessment*, 3, 11–19.
20. Görlich, Y., & Schuler, H. (2010). *Arbeitsprobe zur berufsbezogenen Intelligenz (AZUBI-BK)*. Göttingen: Hogrefe.

21. Görlich, Y., & Schuler, H. (2014). Personalentscheidungen, Nutzen und Fairness. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 1137–1199). Göttingen: Hogrefe.
22. 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.
23. Heilmann, M. E. (1983). Sex bias in working settings: The lack of fit model. *Research in Organizational Behavior, 5*, 269–298.
24. Höft, S., & Muck, P. M. (2009). Golden Profiler of Personality (GPOP) Deutsche Adaptation des Golden Personality Type Profiler von John P. *Golden. Report Psychologie, 7*(8), 322–323.
25. Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lesson learned. *International Journal of Selection and Assessment, 9*, 152–194.
26. Huffcutt, A. I., & Roth, P. L. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology, 83*, 179–189.
27. Kaas, L., & Manger, E. (2010). *Ethnik discrimination in Germany's labour market: A field experiment* (Unveröffentlichter Forschungsbericht Nr. 4741). Bonn: Forschungsinstitut zur Zukunft der Arbeit.
28. Kanning, U. P. (1999). *Die Psychologie der Personenbeurteilung*. Göttingen: Hogrefe.
29. Kanning, U. P. (2004). *Standards der Personaldiagnostik*. Göttingen: Hogrefe.
30. Kanning, U. P. (2009). *Inventar sozialer Kompetenzen (ISK/ISK-K)*. Göttingen: Hogrefe.
31. Kanning, U. P. (2011a). *Inventar zur Messung der Glaubwürdigkeit in der Personalauswahl (IGIP)*. Göttingen: Hogrefe.
32. Kanning, U. P. (2011b). *Akzeptanz von Assessment-Center-Übungen bei AC-Teilnehmern*. *Wirtschaftspsychologie, 13*, 89–101.
33. Kanning, U. P. (2014). Managementversagen – Eine diagnostische Perspektive. *Wirtschaftspsychologie, 3*, 13–20.
34. Kanning, U. P. (2015a). Welche Aussagekraft besitzen biographische Daten bei der Sichtung von Bewerbungsunterlagen? – Ein Überblick über aktuelle Studien. *Wirtschaftspsychologie, 17*, 42–50.
35. Kanning, U. P. (2015b). *Personalauswahl zwischen Anspruch und Wirklichkeit – Eine wirtschaftspsychologische Analyse*. Heidelberg: Springer.
36. Kanning, U. P. (2016a). Über die Sichtung von Bewerbungsunterlagen in der Praxis der Personalauswahl. *Zeitschrift für Arbeits- und Organisationspsychologie, 60*, 18–32.
37. Kanning, U. P. (2016b). Wie Bewerberinnen und Bewerber die Praxis der Personalauswahl erleben und bewerten. *Report Psychologie, 4*, 56–66.
38. Kanning, U. P., & Schuler, H. (2014). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 215–256). Göttingen: Hogrefe.
39. Kanning, U. P., Grewe, K., Hollenberg, S., & Hadouche, M. (2006). From the subjects' point of view: Reactions to different types of situational judgment items. *European Journal of Psychological Assessment, 22*, 168–176.
40. Kanning, U. P., Pöttker, J., & Gelléri, P. (2007). Assessment-Center-Praxis in deutschen Großunternehmen – Ein Vergleich zwischen wissenschaftlichem Anspruch und Realität. *Zeitschrift für Arbeits- und Organisationspsychologie, 51*, 155–167.
41. Klingner, Y., & Schuler, H. (2004). Improving participants' evaluations while maintaining validity by a work sample-intelligence test hybrid. *International Journal of Selection and Assessment, 12*, 120–134.
42. König, C. J., Klehe, U. C., Berchtold, M., & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment, 18*, 17–27.

43. Krause, A., Rinne, U., Zimmermann, K. F., Böschen, I., & Alt, R. (2012). *Pilotprojekt „Anonymisierte Bewerbungsverfahren“ – Abschlussbericht* (IZA Research Report No. 44). Berlin: Institut zur Zukunft der Arbeit (IZA).
44. Kubinger, K. D. (2010). Testfairness für globalisierte türkische Kinder. Die Intelligenz-Testbatterie AID 2-Türkisch. *Report Psychologie*, *35*, 72–81.
45. Marlowe, C. M., Schneider, S. L., & Nelson, C. E. (1996). Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology*, *81*, 11–21.
46. Martocchio, J. J., & Whitener, E. M. (1992). Fairness in personnel selection: A meta-analysis and policy implication. *Human Relations*, *45*, 489–506.
47. O'Brien, K. S., Latner, J. D., Ebner, D., & Hunter, J. A. (2012). Obesity discrimination: The role of physical appearance, personal ideology, and anti-fat prejudice. *International Journal of Obesity*, *1*, 1–6.
48. Ones, D. S., & Viswesvaran, C. (1998). Gender, age, and race differences on overt integrity tests: results across four large-scale job applicant data sets. *Journal of Applied Psychology*, *83*, 35–42.
49. Rolland, F., & Steiner, D. D. (2007). Test-taker reactions to the selection process: Effects of outcome favorability, explanations, and voice on fairness perceptions. *Journal of Applied Social Psychology*, *37*, 2800–2826.
50. Roth, P. L., & Bobko, P. (2000). College grade point average as a personnel selection device: Ethnic group differences and potential adverse impact. *Journal of Applied Psychology*, *85*, 399–406.
51. Sachverständigenrat deutscher Stiftungen. (2014). *Diskriminierung am Ausbildungsmarkt: Ausmaß. Ursachen und Handlungsperspektiven*: Ohne Verlag.
52. Salgado, J. F., Moscoso, S., Fruyt, F. de, Anderson, N., Bertua, C., & Rolland, J. P. (2003). A Meta-Analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology*, *88*, 1068–1081.
53. Schinkel, S., Vianen, A. v., & Dierendonck, D. v. (2013). Selection fairness and outcomes: A field study of interactive effects on applicants reactions. *International Journal of Selection and Assessment*, *21*, 22–31.
54. Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practice and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.
55. Schuler, H. (2002). *Das Einstellungsinterview*. Göttingen: Hogrefe.
56. Schuler, H. (2014a). *Psychologische Personalauswahl: Eignungsdiagnostik für Personalausscheidungen und Berufsberatung*. Göttingen: Hogrefe.
57. Schuler, H. (2014b). Arbeits- und Anforderungsanalyse. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 61–97). Göttingen: Hogrefe.
58. Schuler, H. (2014c). Biographieorientierte Verfahren der Personalauswahl. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 257–299). Göttingen: Hogrefe.
59. Schuler, H., & Berger, W. (1979). Physische Attraktivität als Determinante von Beurteilung und Einstellungsempfehlung. *Psychologie und Praxis*, *23*, 59–70.
60. Schuler, H., & Kanning, U. P. (Hrsg.). (2014). *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe.
61. Schuler, H., & Stehle, W. (1983). Neuere Entwicklungen des Assessment-Center-Ansatzes – beurteilt unter dem Aspekt der sozialen Validität. *Psychologie und Praxis. Zeitschrift für Arbeits- und Organisationspsychologie*, *27*, 33–44.

62. Schuler, H., Hell, B., Trapmann, S., Schaar, H., & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. *Zeitschrift für Personalpsychologie, 6*, 60–70.
63. Sczesny, S., & Stahlberg, D. (2002). *Geschlechtsstereotype Wahrnehmung von Führungskräften. Wirtschaftspsychologie, 9*, 35–40.
64. Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology, 81*, 134–141.
65. Stephan, U., & Westhoff, K. (2002). Personalauswahlgespräche im Führungskräftebereich des deutschen Mittelstandes: Bestandsaufnahmen und Einsparungspotential durch strukturierte Gespräche. *Wirtschaftspsychologie, 3*, 3–17.
66. Truxillo, D. M., Steiner, D. D., & Gilliland, S. W. (2004). The importance of organizational justice in personnel selection: Defining when selection fairness really matters. *International Journal of Selection and Assessment, 12*, 39–53.
67. Watkins, L. M., & Johnston, L. (2000). Screening job applicants: The impact of physical attractiveness and applicants quality. *International Journal of Selection and Assessment, 8*, 76–84.
68. Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A Meta-Analysis. *Human Performance, 21*, 291–309.
69. Zajonc, R. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1–27.

---

## Zusammenfassung

In Zeiten turbulenter, durch die globale Rezession ernsthaft angeschlagener Märkte ist es für Organisationen bedeutsam, zur Erlangung bzw. Aufrechterhaltung ihrer Wettbewerbsfähigkeit kontinuierlich zu innovieren. Ein Weg, organisationale Innovation zu steigern, besteht in der Rekrutierung und Auswahl innovativer Talente. Gemeint sind organisationsexterne Personen, die in ihren neuen Arbeitsrollen und Arbeitskontexten zukünftig kreativ und innovativ sein werden. Dieses Kapitel beschreibt und integriert vorliegende Forschungsergebnisse zu unterschiedlichen Variablen, die individuelle, innovative Leistung am Arbeitsplatz vorherzusagen vermögen. Auf diesen Überblick aufbauend geben wir Handlungsempfehlungen für die zukünftige Forschung und Praxis. Insbesondere werden wir der Frage Beachtung schenken, welche spezifischen Personalauswahlverfahren zur Diagnose des innovativen Potenzials von Personen besonders geeignet sind.

---

K. Potočnik (✉)

The University of Edinburgh Business School, Edinburgh, UK

E-Mail: Kristina.Potocnik@ed.ac.uk

N. Anderson

Brunel University London, Uxbridge, UK

E-Mail: Neil.Anderson@brunel.ac.uk



## Inhaltsverzeichnis

13.1	Einleitung.....	300
13.2	Individuelle Unterschiede und innovative Leistung.....	302
13.2.1	Persönlichkeitsmerkmale.....	302
13.2.2	Kognitive Fähigkeiten und Denkstile.....	306
13.2.3	Arbeitsbezogenes Wissen, Kompetenzen und Fähigkeiten.....	308
13.3	Empfehlungen für die zukünftige Forschung.....	309
13.4	Empfehlungen für die Praxis.....	311
13.5	Fazit.....	313
	Literatur.....	314

---

### 13.1 Einleitung

Dass Innovation am Arbeitsplatz einen wichtigen Faktor für organisationale Leistung, Erfolg und nachhaltiges Wachstum darstellt, ist unbestritten [7]. Daher überrascht es nicht, dass Organisationen unterschiedlicher Industriezweige sich mit der Frage beschäftigen, wie sie ihre Innovationskapazität steigern können. Ein Weg für Organisationen, organisationale Innovationen zu fördern, besteht in einer effektiven Rekrutierung und einer effektiven Personalauswahl. Der Grund hierfür liegt darin, dass mit zunehmender Anzahl an innovativ veranlagten Beschäftigten auch die Wahrscheinlichkeit steigen dürfte, organisationale Innovationen zu generieren [7, 41, 52].

Die innovationsförderliche Personalauswahl ist in der Literatur zu Personalauswahl und -rekrutierung als wichtiges Teilgebiet anerkannt. Einige Studien zeigen sogar ein wachsendes Interesse an diesem Thema [15]. So wurden diverse psychometrische Tests für die Personalauswahl entwickelt, um die innovationsbezogene Leistung individueller Bewerber vorherzusagen, z. B. der Innovation Potenzial Indicator (IPI) [47] und das Team Selection Inventory (TSI) [2]. Diese Instrumente haben sich als valide und reliable Verfahren zur Prädiktion der Passung von Stelleninhaber und Arbeitsstelle („person-job fit“) sowie der Passung von Stelleninhaber und Team („person-team fit“) erwiesen, wenn das innovative Potenzial der Bewerber ein Hauptkriterium bei der Besetzung der jeweiligen Stelle war [15]. Des Weiteren hat die Innovations- und Kreativitätsforschung die Bedeutung verschiedenster individueller Faktoren für die Innovationskraft der Mitarbeiter/-innen untersucht [30, 48] – Faktoren, die auch ein Licht auf die Prädiktoren für innovationsbezogene Leistungen bei der Personalauswahl werfen können. Das Dilemma hier liegt auf der Hand: Zu Kreativität und Innovation gehören per Definition Mitarbeiter, die neue und unvorhergesehene Verhaltensweisen zeigen, die für die Organisation einen Wert darstellen, aber naturgemäß vorher nicht bekannt sind. Daher ist eine an sehr spezifischen Verhaltensweisen orientierte Personalauswahl äußerst problematisch. Möglich ist hingegen, Personal auf der Grundlage einer großen Anzahl von

Persönlichkeitsmerkmalen und Fähigkeiten auszuwählen, für die allgemein eine Korrelation mit innovativen Verhaltensweisen und Kreativitätsanstrengungen festgestellt wurde. Das Ziel dieses Kapitels ist, das Thema organisationale Innovation mit der Forschungsliteratur zu Personalauswahl und -rekrutierung zu verknüpfen, um daraus Implikationen für die Auswahl von Bewerbern unter dem Gesichtspunkt innovationsbezogener Leistungen abzuleiten.

Bevor wir uns mit den verschiedenen Auswahlkriterien zur Vorhersage von Innovation beschäftigen, sollten wir definieren, was wir mit diesem Kriterium meinen, d. h. wie wir innovative Leistung definieren. Aus der Perspektive des Personalmanagements wurde Innovation am Arbeitsplatz am häufigsten definiert als „the intentional introduction and application (within a role, group or organization) of ideas, processes, products or procedures, new to the relevant unit of adoption, designed to significantly benefit the individual, the group, the organization or wider society“ [63, S. 9]. Gemäß dieser Definition betrachten wir die Kreativität der Mitarbeiter (verstanden als neue und nützliche Ideen) als ersten Schritt im Innovationsprozess und somit als Teil innovationsbezogener Leistung [48]. Diese Konzeptualisierung des Kriteriums innovationsbezogene Leistung stellt die Theorie und Praxis der Personalauswahl vor einige wichtige Herausforderungen. Erstens handelt es sich bei innovativen Leistungen um etwas Neues und Unerwartetes, das jederzeit geschehen kann und daher nur sehr schwierig vorherzusagen ist. Zweitens wird nicht jede Person in jeder Situation und jedem Kontext die optimale Innovationsleistung erbringen [3]. Drittens wurde vorgeschlagen, dass Organisationen eine Risikoperspektive einnehmen und ein innovationsförderliches Betriebsklima entwickeln sollten, um individuelle Talente anzuregen, Innovationen voranzutreiben [4, 62]. Das Ausmaß innovativer Verhaltensweisen hängt also zu großen Teilen von der jeweiligen Organisationskultur ab. Vor dem Hintergrund dieser Argumente besteht die Herausforderung, mit der sich viele Organisationen konfrontiert sehen, darin, herauszufinden, wie man genau die Personen auswählt, die Innovationskraft am Arbeitsplatz zukünftig entfalten werden. Aufgrund all dieser ziemlich problematischen Aspekte haben die Verfasser vor kurzem die folgende, integrative Definition von Kreativität und Innovation vorgelegt.

- Creativity and innovation at work are the process, outcomes, and products of attempts to develop and introduce new and improved ways of doing things. The creativity stage of this process refers to idea generation, and innovation to the subsequent stage of implementing ideas toward better procedures, practices, or products. Creativity and innovation can occur at the level of the individual, work team, organization, or at more than one of these levels combined but will invariably result in identifiable benefits at one or more of these levels of analysis [7].

Das Ziel dieses Kapitel ist, diese Aspekte anhand einer Durchsicht der vorliegenden Forschungsergebnisse zu individuellen Unterschieden (z. B. Persönlichkeitsmerkmale und kognitive Fähigkeiten) bei innovativen Leistungen am Arbeitsplatz näher zu beleuchten.

## 13.2 Individuelle Unterschiede und innovative Leistung

Organisationen, die bestrebt sind, Kandidaten einzustellen, die in ihren jeweiligen Arbeitsrollen und Teams hohe innovationsbezogene Leistungen zeigen werden, sollten ihre Personalauswahl- und Rekrutierungspraktiken sorgfältig planen, um jene Bewerber herauszufiltern, die mit geringerer Wahrscheinlichkeit innovative Verhaltensweisen zeigen werden. Die bisherige Forschung kommt zu dem Schluss, dass eine Reihe unterschiedlicher Persönlichkeitsmerkmale, kognitiver Fähigkeiten und Denkstile sowie arbeitsplatzrelevantes Wissen und Kompetenzen innovativen Leistungen zuträglich sind. Im Folgenden werden diese Faktoren im Einzelnen näher betrachtet.

### 13.2.1 Persönlichkeitsmerkmale

In neueren Metaanalysen wurden mehrere Persönlichkeitsdimensionen identifiziert, die individuelle, innovative Verhaltensweisen vorherzusagen vermögen [6, 30]. Diese lassen sich in den folgende Kategorien gruppieren: 1) Big-Five-Persönlichkeitsmerkmale (oder Fünf-Faktoren-Modell, FFM), 2) kreative und proaktive Persönlichkeit sowie 3) Selbstevaluationen.

#### 13.2.1.1 Big-Five-Persönlichkeitsmerkmale

Bei der Analyse der Zusammenhänge zwischen Big-Five-Persönlichkeitsmerkmalen und Kreativität bzw. Innovation hat sich die Forschung bislang hauptsächlich auf die Dimensionen Offenheit für neue Erfahrungen und Gewissenhaftigkeit konzentriert [30, 53]. Im Allgemeinen erwies sich dabei der Zusammenhang zwischen Offenheit für neue Erfahrungen und innovativer Leistung konsistent als positiv [6, 9, 13, 26, 28, 30, 38, 53]. Mitarbeiter mit hohen Offenheitswerten lassen sich als Personen mit großer Imaginationskraft, Neugierde, vielseitigem Interesse und künstlerischer Sensibilität charakterisieren und sind in höherem Maße durch divergentes Denken geprägt [40, 43] als Personen mit niedrigen Werten in der Dimension Offenheit für neue Erfahrungen. Es dürfte daher kaum überraschen, dass genau diese Personen auch am Arbeitsplatz mit größerer Wahrscheinlichkeit innovative Verhaltensweisen zeigen. All diese Eigenschaften tragen zu einer höheren Flexibilität und einer aktiven Suche nach neuen Erfahrungen bei, was zu einer höheren innovativen Leistung führen kann.

Demgegenüber sind die Befunde zu den Auswirkungen von Gewissenhaftigkeit auf die innovative Leistung uneinheitlich. Die Dimension Gewissenhaftigkeit bezieht sich auf das Ausmaß, in dem Personen verlässlich, fleißig, zielorientiert, selbstdiszipliniert und ausdauernd arbeiten [53, 61]. Obwohl Gewissenhaftigkeit als einer der verlässlichsten Prädiktoren für die allgemeine Arbeitsleistung gelten kann [14, 33, 54], ist der Zusammenhang dieses Persönlichkeitsmerkmals mit innovationsbezogenen Leistungen alles andere als empirisch in homogener Weise belegt: Die vorhandenen Studien zeigen die volle Bandbreite an möglichen Effekten zwischen Gewissenhaftigkeit und

Innovationsleistung. Manche Untersuchungen konstatieren einen negativen Zusammenhang zwischen Gewissenhaftigkeit und Innovationsleistung [24] und erklären diesen damit, dass sehr gewissenhafte Mitarbeiter in erster Linie bestrebt sind, die ihnen übertragenen Aufgaben effizient und auf die konventionelle, erprobte Weise zu erledigen, statt nach neuen Lösungswegen zu suchen oder sogar neue Elemente in ihre Arbeit einzuführen [28, 53].

Einige Studien haben die Randbedingungen aufgedeckt, unter denen sich der negative Effekt von Gewissenhaftigkeit auf die innovationsbezogene Leistung am stärksten zeigt. George und Zhou [28] zum Beispiel fanden heraus, dass sehr gewissenhafte Mitarbeiter geringe kreative Leistungen zeigen, wenn sie von ihren Kollegen am Arbeitsplatz nicht unterstützt werden und unter der engen Aufsicht ihrer Vorgesetzten stehen. Demgegenüber wird in anderen Studien die Auffassung vertreten, dass neue und nützliche Ideen – sprich: Innovationen – sich ohne selbstdisziplinierte und fleißige Mitarbeiter nur schwerlich erfolgreich umsetzen lassen [61]. Einer kürzlich erschienenen Studie zufolge hatte Gewissenhaftigkeit nur dann einen positiven Effekt auf die innovative Leistung, wenn die Organisationskultur insgesamt sowohl ergebnis- als auch innovationsorientiert war [61].

In den meisten Studien konnte jedoch kein signifikanter Zusammenhang zwischen Gewissenhaftigkeit und Innovation festgestellt werden [13, 26, 42, 49, 53]. Das Ausbleiben eines signifikanten Zusammenhangs zwischen Gewissenhaftigkeit und innovativer Leistung wurde kürzlich auch in einer Metaanalyse mit 97 unabhängigen Stichproben bestätigt [6]. Anders als im Fall der allgemeinen Arbeitsleistung legen diese Ergebnisse den Schluss nahe, dass Gewissenhaftigkeit als Persönlichkeitsmerkmal von geringer Bedeutung ist, wenn eine Stelle vor allem unter dem Aspekt des Innovationspotenzials des Kandidaten besetzt werden soll. Wahrscheinlich wird hier fälschlicherweise davon ausgegangen, dass sehr gewissenhafte Personen häufig stärker regelgeleitet handeln und folglich weniger geneigt sind, etablierte Abläufe in vielen betrieblichen Kontexten infrage zu stellen. Dem wäre jedoch entgegenzuhalten, dass ein gewisses Maß an Gewissenhaftigkeit unabdingbar ist, damit neue Ideen auch tatsächlich umgesetzt und erfolgreich im Betriebsablauf implementiert werden – d. h. dass ein hohes Maß an Gewissenhaftigkeit in der späteren Phase der Ideenimplementierung wertvoll ist, während es in der frühen Phase der Ideengenerierung möglicherweise ein Hindernis darstellt [7].

Auch für die Dimension Extraversion finden sich in der bisherigen Forschung keine eindeutigen Ergebnisse, auch wenn die vorhandenen Studien mehrheitlich auf einen positiven Effekt von Extraversion auf die Innovationsleistung hindeuten [26, 53]. Extravertierte Menschen werden als gesprächig, energetisch, optimistisch, durchsetzungsstark, gesellig, neugierig, ehrgeizig und begeisterungsfähig charakterisiert [49, 53]. Diese Eigenschaften stehen in engem Zusammenhang mit der Tendenz der Extravertierten, nach neuen Wegen für die Erledigung beruflicher Aufgaben zu suchen und proaktiv mit neuen Ideen zu experimentieren, die ihre innovative Leistung erhöhen könnten.

Verträglichkeit – definiert als „tendency to be flexible, forgiving, trusting, good-natured, generous, and cooperative“ [40] – scheint keinen signifikanten Effekt auf die innovationsbezogene Leistung zu haben [6, 13, 49]. Verschiedentlich ist jedoch ein negativer

Zusammenhang zwischen Verträglichkeit und Innovation postuliert worden mit der Begründung, dass verträgliche Individuen bestrebt sind, Harmonie und positive Beziehungen innerhalb ihrer Teams oder Organisationen zu sichern, und dass innovationsförderliches Verhalten diese Harmonie gefährden könnte, weil es zu Streit und Konflikten führen kann. Mit dieser Argumentation konstatieren Furnham et al. [26] einen negativen Zusammenhang zwischen Verträglichkeit und divergentem Denken.

Schließlich ist auch die Dimension der emotionalen Stabilität im Hinblick auf ihren Effekt auf die innovative Leistung untersucht worden. Emotional stabile Menschen werden als entspannt und selbstsicher charakterisiert und zeigen mit höherer Wahrscheinlichkeit innovative Verhaltensweisen als emotional instabile Personen, die als unsicher, ängstlich und vorsichtig beschrieben werden [53]. Raja und Johns [49] stützen diese These, allerdings nur bei Positionen mit einem großen Aufgabenbereich. Andere Studien zur Bedeutung dieses Persönlichkeitsmerkmals für die innovative Leistung konnten keine signifikanten Zusammenhänge feststellen [6, 26, 53].

Zusammengefasst zeigen diese Befunde sehr deutlich, dass die FFM-Persönlichkeitsmessungen für die innovationsförderliche Personalauswahl äußerst hilfreich sind. Es scheint, dass Bewerber, die Werte in den folgenden Richtungen erzielen, am ehesten innovative Leistungen am Arbeitsplatz erbringen:

- Hohe Werte in der Dimension Offenheit für neue Erfahrungen,
- Mittlere Werte in der Dimension Gewissenhaftigkeit,
- Mittlere Werte in der Dimension Extraversion,
- Geringe Werte in der Dimension Verträglichkeit,
- Hohe Werte in der Dimension emotionale Stabilität.

### 13.2.1.2 Kreative und proaktive Persönlichkeit

Neben den Big-Five-Dimensionen hat sich die Forschung auch mit der Bedeutung anderer Persönlichkeitsmerkmale für die innovationsbezogene Leistung befasst, zum Beispiel im Hinblick auf kreative und proaktive Persönlichkeiten [6, 30]. Dabei wurde das Merkmal kreative Persönlichkeit am häufigsten mithilfe der von Gough (zitiert in Hammond) [30] entwickelten kreativen Persönlichkeitsskala („creative personality scale“) operationalisiert. Dieses Instrument enthält eine Liste von Adjektiven, die häufig auf außergewöhnlich kreative Menschen zutreffen (z. B. erfinderisch, originell, selbstsicher, unkonventionell und clever) und kreative von weniger kreativen Individuen unterscheiden. Der positive Zusammenhang zwischen einer kreativen Persönlichkeit und innovativen Leistungen ist in der Forschung gut belegt [30, 45, 51, 66, 65]. Obwohl dieser Ansatz selten im Kontext der Personalauswahl zum Einsatz kommt, scheint er doch ein vielversprechendes Hilfsmittel bei der Auswahl kreativerer Mitarbeiter zu sein.

Mit dem Konzept der proaktiven Persönlichkeit ist die Verhaltensdisposition gemeint, bestehende Möglichkeiten zur Initiierung von Veränderungsprozessen zu erkennen [46]. Die Forschung hat gezeigt, dass Menschen mit hohen proaktiven Persönlichkeitswerten eine positive Einstellung zu Veränderungen haben und daher mit höherer

Wahrscheinlichkeit ein innovatives Verhalten zeigen. Diese Annahme wurde durch empirische Studien bestätigt [6, 29]. Insgesamt deuten diese Befunde zwar darauf hin, dass Bewerber mit hohen kreativen und proaktiven Persönlichkeitswerten mit höherer Wahrscheinlichkeit innovative Verhaltensweisen zeigen werden, aber Organisationen gut beraten sind, beide Messinstrumente genau auf ihre Validität, Reliabilität und das Fehlen negativer Auswirkungen im Falle heikler Personalauswahlentscheidungen zu prüfen.

### 13.2.1.3 Selbstevaluationen und Selbstkonzepte

Eine weitere Kategorie von Persönlichkeitsmerkmalen, die oft mit innovativen Leistungen am Arbeitsplatz in Verbindung gebracht wird, ist die der Selbstevaluationen. Hierbei geht es um Konzepte wie Selbstwirksamkeitserwartung, kreative Selbstwirksamkeitserwartung, kreative Rollenidentität und zentrale Selbstbewertungen.

Eine große Anzahl wissenschaftlicher Studien hat sich mit der Bedeutung der Selbstwirksamkeitserwartung („self-efficacy“) und hier insbesondere von kreativen Selbstwirksamkeitserwartungen beschäftigt [8, 25]. Immer wieder wurde ein positiver Zusammenhang zwischen kreativen Selbstwirksamkeitserwartungen – definiert als die individuelle Überzeugung einer Person in die eigene Fähigkeit, neue und nützliche Ideen hervorbringen zu können [56] – und innovationsbezogenen Leistungen festgestellt [16, 21, 56, 57, 58]. Ein ähnliches Konzept ist das der kreativen Rollenidentität, definiert als das Ausmaß, in dem Mitarbeiter sich mit ihrer kreativen Rolle am Arbeitsplatz identifizieren und Kreativität als einen zentralen Teil ihrer Identität [23]. In der Forschung finden sich Belege für einen positiven Zusammenhang zwischen diesem Persönlichkeitsmerkmal und verschiedenen Dimensionen innovationsbezogener Leistungen [23, 58, 60].

Schließlich wollen wir noch kurz auf die Bedeutung von zentralen Selbstbewertungen („core self-evaluations“, CSE) für die innovationsbezogene Leistung eingehen. Unter CSE werden die Selbsteinschätzungen verstanden, die Menschen hinsichtlich ihrer Selbstachtung und ihrer Fähigkeiten abgeben [34]. Dieses Konzept wird als Persönlichkeitsmerkmal höherer Ordnung verstanden, das mehrere Mechanismen umfasst, z. B. Selbstwert („self-esteem“), allgemeine Selbstwirksamkeitserwartung („self-efficacy“), Kontrollüberzeugung („locus of control“) und Neurotizismus („neuroticism“) bzw. emotionale Stabilität. Personen mit hohen CSE-Werten sind solche, die ein positives Selbstwertgefühl haben und sich als sehr selbstwirksam, innengeleitet und emotional stabil wahrnehmen. Es wird angenommen, dass diese Eigenschaften die Wahrscheinlichkeit innovativer Verhaltensweisen erhöhen, da Mitarbeiter, die ein hohes Zutrauen in ihr eigenes Leistungsvermögen haben, hohe Selbstachtung besitzen und für ihr Handeln Verantwortung übernehmen, eher geneigt sind, bestehende oder etablierte Betriebsabläufe infrage zu stellen und neue Ideen zu entwickeln, auch wenn es dadurch zu Konflikten mit anderen Personen kommt. Anderson et al. [6] konnten diese Hypothese in ihrer jüngsten Metaanalyse bestätigen, die einen starken positiven Zusammenhang zwischen zentralen Selbstevaluationen einer Person und innovativer Leistung bestätigte.

Angesichts dieser Befunde schlagen wir vor, dass die Erweiterung von Persönlichkeitstests um CSE sowie kreative und proaktive Persönlichkeitsskalen hilfreich wäre,

um HR-Praktiker bei der Auswahl von Mitarbeitern zu unterstützen, die mit höher Wahrscheinlichkeit innovative Leistungen zeigen werden.

Zusammengefasst zeigt unser Überblick, dass einige Persönlichkeitsmerkmale, wie z. B. kreative und proaktive Persönlichkeit und positive Selbstevaluationen, wichtige Prädiktoren für innovative Leistungen am Arbeitsplatz sind. Allerdings mussten wir ebenfalls feststellen, dass manche Eigenschaften, die allgemein als die stärksten Prädiktoren für eine insgesamt hohe Arbeitsleistung gelten (z. B. Gewissenhaftigkeit und Verträglichkeit) im Hinblick auf die innovationsbezogene Leistung keine so große Relevanz haben.

Im nächsten Schritt kommen wir von der Analyse der Bedeutung von Persönlichkeitsmerkmalen zur Analyse von kognitiven Fähigkeiten und Denkstilen im Hinblick auf innovationsbezogene Leistungen.

## 13.2.2 Kognitive Fähigkeiten und Denkstile

Forschungsarbeiten zur Rolle von kognitiven Fähigkeiten und Denkstilen bei innovativen Leistungen haben sich mit der Bedeutung von Intelligenz, kreativen Fähigkeiten und unterschiedlichen Denkstilen befasst. Im Folgenden wollen wir jeden dieser Punkte etwas detaillierter betrachten.

### 13.2.2.1 Allgemeine Intelligenz

Die allgemeine Intelligenz („general mental ability“, GMA) gehört zu den verlässlichsten positiven Prädiktoren der Arbeitsleistung. Daher sollten wir davon ausgehen, dass sie auch mit innovationsbezogenen Leistungen in einem positiven Zusammenhang steht, obwohl Arbeitsleistung und innovationsbezogene Leistung natürlich nicht deckungsgleich sind [6]. Tatsächlich betrachten manche Autoren Kreativität als Teilgebiet der Intelligenz in Guilfords Intelligenz-Strukturmodell [11]. Einige Wissenschaftler vertreten jedoch auch die Auffassung, dass Intelligenz zwar eine notwendige, aber keine hinreichende Bedingung für Kreativität darstellt. Diese Argumentationslinie wird durch Studien bestätigt, die in diversen Stichproben nur eine mäßig positive Korrelation zwischen Intelligenz und divergentem Denken gezeigt haben [27]. Obwohl manche Studien einen positiven Effekt mentaler Fähigkeiten (z. B. numerische Fähigkeiten, verbales Schlussfolgern, fluide Intelligenz) auf einige Indikatoren innovationsbezogener Leistungen wie Kreativität und divergentes Denken bestätigt haben [13, 27, 26], konnten Anderson et al. [6] in ihrer Metaanalyse keinen signifikanten Effekt der allgemeinen Intelligenz auf die innovationsbezogene Leistung feststellen. Diese Befunde könnten darauf hindeuten, dass es nicht die allgemeine Intelligenz ist, die mit innovativem Potenzial in Verbindung steht, sondern spezifischere, kreativitäts- und innovationsrelevante Denkfähigkeiten.

### 13.2.2.2 Kreative Fähigkeiten

Kreative Fähigkeiten – definiert als Fähigkeiten oder Kompetenzen, die insbesondere für kreative und innovative Leistungen relevant sind, z. B. intuitives Denken und Imaginationskraft [17] – wurden als wichtiger individueller Faktor für die Steigerung innovativer



Leistungen gesehen. Wir könnten die kreativen Fähigkeiten in der Kategorie der kreativitätsrelevanten Fähigkeiten nach dem Drei-Komponenten-Modell der individuellen Kreativität verorten [1]. Diese Fähigkeiten wurden definiert als „a cognitive style favourable to taking new perspectives on problems, an application of techniques (or ‘heuristics’) for the exploration of new cognitive pathways, and a working style conducive to persistent, energetic pursuit of one’s work“ [1, S. 43].

Die bisherige Forschung bestätigt konsistent einen positiven Zusammenhang zwischen kreativen Fähigkeiten und verschiedenen Indikatoren innovationsbezogener Leistungen [18, 20, 32]. Choi et al. [20] berichten auch von einer isolierenden Wirkung kreativer Fähigkeiten in dem Sinne, dass Mitarbeiter mit hohen kreativen Fähigkeiten unabhängig von hemmenden Faktoren, wie inkompetenten Kollegen oder fehlender Unterstützung, ein stabiles Niveau innovativer Leistung zeigen.

### 13.2.2.3 Denkstile

Bei der Untersuchung der Bedeutung unterschiedlicher Denkstile für die innovationsbezogene Leistung kam in der bisherigen Forschung hauptsächlich das „Kirton Adaptation-Innovation Inventory“ zum Einsatz (KAI) [35]. Dieses Instrument umfasst 32 Items zur Messung des hinreichenden Vorhandenseins von Originalität, Effizienz (umgekehrt gepolt) und Rollen-/Gruppenkonformität (umgekehrt gepolt). Die Befragten müssen bei jedem Item angeben, wie „leicht“ oder „schwer“ sie es finden, sich als eine bestimmte Art von Person zu präsentieren. Diejenigen, die niedrige Werte auf allen drei Subskalen erzielen, werden als Adaptoren beschrieben (d. h. sie bevorzugen einen adaptiven kognitiven Stil), wohingegen die Personen mit hohen Werten als Innovatoren definiert werden (d. h. sie bevorzugen einen innovativen kognitiven Stil).

Adaptoren neigen allgemein dazu, sich an vorgegebene Abläufe und Arbeitsweisen zu halten, während Innovatoren eher zum riskanteren Weg tendieren, bewährte Methoden zur Erledigung von Aufgaben zu ignorieren, um neue und nützliche Lösungen zu entwickeln [51]. Daher ist zu erwarten, dass Innovatoren mit höherer Wahrscheinlichkeit innovative Verhaltensweisen zeigen. Diese Annahme wird durch die Forschung weitgehend bestätigt [37, 55].

Ein etwas anderer Ansatz findet sich bei Clegg et al. [21], die zwischen einem intuitiven Denkstil und einem systematischen Denkstil unterscheiden. Der intuitive Denkstil weist konzeptuelle Ähnlichkeiten zum innovativen Stil des KAI auf, während der systematische Denkstil dem adaptiven Stil des KAI entspricht. Diese Autoren konnten einen positiven Zusammenhang zwischen dem intuitiven Denkstil und der Generierung von Ideen feststellen.

Schließlich gibt es einige erste Studien zur Bedeutung des Kognitionsbedürfnisses („need for cognition“) – definiert als „individual dispositional tendency to engage in and enjoy in thinking“ (Cacioppo und Petty zitiert in Wu, Parker und de Jong) [64] – für die innovationsbezogene Leistung. Personen mit einem hohen Kognitionsbedürfnis zeigen eher innovative Verhaltensweisen, weil sie mehr Energie für die kognitive Ausarbeitung aufbringen; sie entwickeln bessere Lernfähigkeiten und sind selbstsicherer beim Äußern und Vorantreiben ihrer Ideen [64]. Die bisherige Forschung hat den positiven Effekt des

Kognitionsbedürfnisses auf die innovationsbezogene Leistung bestätigt, und zwar insbesondere für die Mitarbeiter, die an ihrem Arbeitsplatz wenig Autonomie und geringen Zeitdruck verspürten [64]. Auch Hülsheger et al. [32] konstatieren über drei verschiedene Stichproben hinweg einen positiven Zusammenhang zwischen Kognitionsbedürfnis und innovativer Leistung. Darüber hinaus stellten sie fest, dass dieser positive Zusammenhang bei Personen mit hohen CSE-Werten noch stärker ausgeprägt war.

Im Kontext der Personalauswahl und -rekrutierung sind kognitive Fähigkeitstests zur Bewertung der allgemeinen Intelligenz notwendig und unerlässlich, da die GMA-Werte einen starken Prädiktor für die Arbeitsleistung darstellen [50]. Für die Suche nach besonders innovativ veranlagten Bewerbern empfehlen wir jedoch auch den Einsatz von Inventaren, die stärker auf kreativitäts- und innovationsspezifische Fähigkeiten abzielen, um so die Wahrscheinlichkeit zu erhöhen, innovative Mitarbeiter für die Organisation zu gewinnen.

Als nächstes soll der Zusammenhang zwischen arbeitsbezogenem Wissen/Kompetenzen und innovationsbezogener Leistungen am Arbeitsplatz diskutiert werden.

### 13.2.3 Arbeitsbezogenes Wissen, Kompetenzen und Fähigkeiten

Amabile [1] zufolge besteht ein entscheidender Faktor für die Steigerung der individuellen Kompetenz in einem Wissen, das Faktenkenntnis, technische Kompetenzen und technische Fähigkeiten umfasst. Es wäre zu erwarten, dass diese Komponenten auch für innovationsbezogene Leistungen erforderlich sind, weil Mitarbeiter bei der Erledigung ihrer arbeitsbezogenen Aufgaben ein solides Wissen und solide Kompetenzen benötigen, um in der Lage zu sein, neue Ideen hervorzubringen, deren Nützlichkeit zu prüfen und sie schließlich umzusetzen [19, 61]. Die bisherige Forschung bestätigt diese Annahme und stellt einen positiven Zusammenhang zwischen verschiedenen Dimensionen des arbeitsbezogenen Wissens bzw. arbeitsbezogener Fähigkeiten und den Indikatoren für innovative Leistung fest [19, 31, 36, 44]. Kürzlich kamen auch Anderson et al. [6] in ihrer Metaanalyse zu dem Ergebnis, dass arbeitsbezogenes Wissen und entsprechende Fähigkeiten einen starken positiven Effekt auf die innovationsbezogenen Leistungen der Mitarbeiter haben.

Die meisten Studien verwenden bei ihren Analysen das Ausbildungsniveau und die Dauer der Organisationszugehörigkeit als Proxyvariablen für arbeitsbezogenes Wissen und Fähigkeiten, denn Innovationsforscher würden argumentieren, dass sowohl das Ausbildungsniveau als auch die Dauer der Organisationszugehörigkeit arbeitsbezogenes Wissen darstellen, dass die Mitarbeiter durch „formal training or experience on the job“ erwerben [30, S. 92]. Dieser Argumentation folgend wäre jeweils ein positiver Zusammenhang zwischen Ausbildungsniveau, Dauer der Organisationszugehörigkeit und innovationsbezogener Leistung zu erwarten. Unsere Durchsicht der Literatur hat ergeben, dass diese beiden Variablen in den meisten Fällen als Kontrollvariablen behandelt wurden [30], d. h., dass die Autoren dieser Studien kein besonderes Interesse an der

Analyse der Auswirkungen dieser Variablen auf die innovationsbezogene Leistung hatten. Unsere Literaturanalyse zeigte, dass die Zusammenhänge zwischen Innovation und Ausbildungsniveau bzw. Dauer der Organisationszugehörigkeit heterogen ausfallen: Während einige Studien tatsächlich einen positiven Zusammenhang zwischen diesen beiden Variablen und innovationsbezogener Leistung belegen [30, 58], konnten andere Studien keinerlei Effekte von Ausbildungsniveau und Dauer der Organisationszugehörigkeit auf die innovationsbezogene Leistung feststellen [27, 39, 61]. Schließlich finden sich auch Fälle eines negativen Zusammenhangs zwischen Innovation und Ausbildungsniveau bzw. Dauer der Organisationszugehörigkeit. Baer et al. [10] zum Beispiel stellten einen positiven Zusammenhang zwischen Ausbildungsniveau und Kreativität fest; der Zusammenhang zwischen Dauer der Organisationszugehörigkeit und Kreativität hingegen war negativ. Dieser negative Zusammenhang impliziert, dass Mitarbeiter, die bereits seit längerer Zeit in ihren jeweiligen Positionen arbeiten, mit geringerer Wahrscheinlichkeit neue und nützliche Ideen hervorbringen und im Vergleich zu denen, die erst kürzlich zur Organisation gestoßen sind, eher dazu neigen, sich an die bewährten Arbeitsabläufe zu halten.

Diese Befunde legen den Schluss nahe, dass arbeitsgebietsbezogenes Wissen und Arbeitserfahrung bei der Suche nach innovativ veranlagtem Personal berücksichtigt werden sollten. In der Praxis sollten verschiedene Auswahlverfahren zur Bewertung dieser Variablen zum Einsatz kommen, z. B. Arbeitsproben, strukturierte Interviews, Lebensläufe und biografische Daten [5].

Zusammengefasst deuten die bisherigen Ergebnisse darauf hin, dass arbeitsbezogenes Wissen und arbeitsbezogene Fähigkeiten wichtige Faktoren bei der Steigerung von innovationsbezogenen Leistungen sein können. Die zukünftige Forschung sollte jedoch noch überzeugendere Belege für diese Zusammenhänge bereitstellen.

Im nächsten Abschnitt geben wir einige spezifischere Empfehlungen sowohl für die Praxis als auch für die zukünftige Forschung auf dem Gebiet der innovationsförderlichen Personalauswahl.

---

### 13.3 Empfehlungen für die zukünftige Forschung

Unser Forschungsüberblick hat zahlreiche interessante Stoßrichtungen für die zukünftige Forschung zur innovationsförderlichen Personalauswahl aufgezeigt. Wie wir gesehen haben, sind die wissenschaftlichen Befunde zu den meisten der untersuchten individuellen Prädiktoren innovationsbezogener Leistungen uneinheitlich und nicht eindeutig. Weitere Forschung auf diesem Gebiet ist daher vonnöten, um zu validen Implikationen für Praktiker zu kommen, die sich beruflich mit der Rekrutierung und Auswahl von innovativen Talenten beschäftigen.

Eines der vielleicht verblüffendsten Ergebnisse, die sich in der Literatur zeigen, ist das Fehlen signifikanter Korrelationen zwischen Gewissenhaftigkeit und innovativen Leistungen. Gewissenhaftigkeit hat zwar einen starken positiven Effekt auf die Arbeitsleistung, aber innovationsförderliches Verhalten scheint diese Eigenschaft überhaupt

nicht vorherzusagen. In zukünftigen Studien könnte dem weiter nachgegangen werden, indem Gewissenhaftigkeit auf der Facettenebene untersucht wird. Anknüpfend an Dudley et al. [22], die zu dem Ergebnis kamen, dass die verschiedenen Subdimensionen von Gewissenhaftigkeit auf unterschiedliche Weise spezifische Leistungsdimensionen wie Aufgabenerledigung, Engagement, Förderung der interpersonalen Kommunikation sowie kontraproduktives Verhalten am Arbeitsplatz prädiktieren (innovationsbezogene Leistungen wurden bei der Analyse nicht berücksichtigt), sind wir der Auffassung, dass eine solche differenzierte Analyse mehr Licht auf den Zusammenhang zwischen Gewissenhaftigkeit und Innovation werfen könnte.

Eine Analyse auf der Facettenebene ließe sich auch für andere Dimensionen des Fünf-Faktoren-Modells durchführen, z. B. Extraversion und Verträglichkeit, denn auch für diese beiden Persönlichkeitsmerkmale sind die Auswirkungen auf die innovationsbezogene Leistung (noch) nicht eindeutig geklärt. Interessanterweise hat die bisherige Forschung die Analyse einzelner Facetten des Persönlichkeitsinventars und ihrer Effekte auf die innovationsbezogene Leistung vollkommen vernachlässigt.

Ein weiteres Betätigungsfeld für die zukünftige Forschung besteht in der Untersuchung der Bedeutung von Intelligenz (GMA) für die innovationsbezogene Leistung. Wie unser Literaturüberblick gezeigt hat, ist der Beziehung von GMA und innovationsbezogener Leistung bislang vergleichsweise wenig wissenschaftliche Beachtung geschenkt worden und die wenigen Befunde, die es gibt, haben nur geringe Aussagekraft. Bei den meisten Studien wurden ferner Studierendenstichproben verwendet, um die Bedeutung von numerischen Fähigkeiten und verbalem Schlussfolgern für das divergente Denken zu untersuchen (und nicht speziell für die innovationsbezogene Leistung). Diesen Studien mangelt es an ökologischer Validität. Daher lautet unsere Empfehlung, in zukünftigen Forschungsarbeiten die Intelligenz und ihr prädiktives Potenzial für innovationsförderliches Verhalten am Arbeitsplatz in den Blick zu nehmen und Mitarbeiterstichproben zu gewinnen.

Allgemein empfehlen wir, mehr praxisnahe Forschung in realen Personalauswahl- und Rekrutierungskontexten durchzuführen, um Ergebnisse zu erzielen, die Praktikern bei der Suche nach innovativen Talenten unmittelbare Implikationen an die Hand geben. Z. B. fehlt es auf diesem Gebiet bislang an wissenschaftlichen Untersuchungen zur Brauchbarkeit der verschiedenen Methoden und Techniken zur Auswahl innovativ veranlagter Bewerber. Die empirische Forschung sollte die Validität verschiedener Verfahren zur innovationsförderlichen Personalauswahl überprüfen als ein Kriterium, anhand dessen in der Praxis entschieden werden kann, welche Verfahren für die Auswahl innovativen Potenzials zum Einsatz kommen sollten. Anknüpfend an neuere Entwicklungen in der Messung innovationsbezogener Leistungen [48] empfehlen wir darüber hinaus der Forschung auf diesem Gebiet, objektive oder unabhängige Bewertungen für innovative Leistungen zu verwenden, um die mit Selbsteinschätzungen einhergehenden Verzerrungen zu vermeiden.

Schließlich ist es wichtig zu betonen, dass die innovative Dimension der Arbeitsleistung bislang kaum Eingang in die theoretischen Modelle der Arbeitsleistung gefunden

hat (mit Ausnahme von [12, 59]). Um zu bestimmen, welche Eigenschaften der Persönlichkeit innovative Leistungen prädiktieren und um zu Schlussfolgerungen bezüglich der potenziell unterschiedlichen Bedeutung von Prädiktoren über verschiedene Leistungsdimensionen hinweg zu kommen, sollte die zukünftige Forschung damit beginnen, die bestehenden Leistungsmodelle dahin gehend zu modifizieren, dass auch die innovationsbezogenen Aspekte der Arbeitsleistung einbezogen werden.

Dies waren lediglich einige Anregungen für die zukünftige Forschung auf diesem Gebiet. Wir sind der Auffassung, dass das Verfolgen dieser Forschungslinien wichtige und dringend benötigte Ergebnisse liefern könnte, um Praktikern auf diesem Gebiet fundierte Hinweise zur Verfügung zu stellen, wie sich eine innovationsförderliche Personalauswahl am effektivsten gestalten lässt.

---

### 13.4 Empfehlungen für die Praxis

Die bisherige Forschung liefert zwar begrenzte, aber dennoch wichtige Empfehlungen für die Praxis. Als gesichertes Ergebnis der bisherigen Studien darf beispielsweise gelten, dass Offenheit für neue Erfahrungen ein eindeutiger Prädiktor für innovationsbezogene Leistungen ist. Auch Extraversion hängt mit innovativem Verhalten am Arbeitsplatz zusammen. Unsere Empfehlung für die Praxis lautet daher, Bewerber nach diesen beiden Big-Five-Persönlichkeitsmerkmalen zu selektieren, wenn es hauptsächlich darum geht, Kandidaten mit hohem Innovationspotenzial auszuwählen. Neben der Berücksichtigung der Big-Five-Persönlichkeitsmerkmale empfehlen wir HR-Praktikern, bei der Suche nach innovativem Potenzial besonders auf Bewerber mit hohen CSE-Werten und einer kreativen und proaktiven Persönlichkeit zu achten. Was die Fähigkeiten betrifft, gilt es, besonders auf kreative Fähigkeiten und intuitive Denkstile zu achten.

Wie bereits mehrfach in diesem Kapitel erwähnt, können die Personalabteilungen auf einige sehr beliebte und einige weniger häufig verwendete Auswahlverfahren zurückgreifen, wenn es darum geht, Bewerber vor allem im Hinblick auf ihr Innovationspotenzial für eine bestimmte Position oder Vakanz auszuwählen. Tab. 13.1 fasst auf Grundlage der bisherigen Forschung [41] unsere Einschätzungen bezüglich des zu erwartenden Wertes der verschiedenen prädiktiven Verfahren der Personalauswahl zusammen, wenn das Innovationspotenzial von Bewerbern für die rekrutierende Organisation eine hohe Priorität darstellt. Natürlich ist der Kontext für jede spezifische Auswahlsituation unterschiedlich, und natürlich sind die Anforderungen an Kreativität und Innovationskraft bei jeder zu besetzenden Position verschieden, aber dennoch gilt, dass diese Verfahren unter den beschriebenen Umständen geeignet sind, zusätzliche Anhaltspunkte zu liefern.

Nachdem wir diese potenziellen Einsatzmöglichkeiten aufgezeigt haben, scheint es uns angebracht, auf einige Aspekte hinzuweisen, die Organisationen bedenken sollten, bevor sie nur noch innovativ veranlagte Bewerber auswählen, ohne auch andere Arbeitsinhalte zu berücksichtigen.

**Tab. 13.1** Mögliche Auswahlverfahren und ihr Wert für eine innovationsförderliche Personalauswahl

Auswahlverfahren	Mögliche Eignung	Kommentare
Kognitive Fähigkeitstests	Hoch	GMA-Tests gehören zum Standard bei allen Auswahlkontexten, aber Kreativität und innovationsspezifische Fähigkeiten (z. B. kreative Fähigkeiten) sollten ebenfalls bewertet werden
Persönlichkeitstests	Hoch	FFM-Tests können sehr nützlich sein, aber CSE-Werte sowie Fragebogen zur kreativen und proaktiven Persönlichkeit sollten ebenfalls zum Einsatz kommen
Strukturierte Interviews	Mittel	Informationen zur Innovationsneigung und zu bereits gezeigten Innovationsanstrengungen und -erfolgen können geprüft werden. Auch die Selbsteinschätzung der Kandidaten hinsichtlich ihrer Kreativität kann gemessen werden (z. B. „Im Vergleich zu meinen Kollegen würde ich mich bezüglich meines innovativen Potenzials in den Top ___ % einordnen“)
Situativer Urteilstest	Mittel	Das Innovationspotenzial der Kandidaten kann in spezifischen Situationen bewertet werden, denen die Kandidaten höchstwahrscheinlich im Rahmen ihrer Tätigkeit begegnen werden
Assessment-Center	Hoch	Es kann eine Kombination verschiedener Verfahren verwendet werden; gleichzeitig können verschiedenen Leistungsdimensionen überprüft werden. Kreatives und innovatives Potenzial kann mithilfe von schriftlichen Analyse-Übungen, Interviews, simulierten Präsentationen, Teamübungen, Rollenspielen (bei denen die Kandidaten neue Ideen für die Problemlösung entwickeln sollen usw.) bewertet werden
Arbeitsproben	Mittel	Arbeitsgebietsbezogenes Wissen kann überprüft werden
Lebenslauf, biografische Daten	Mittel (aber nur spezielle Elemente)	Informationen zu den bisherigen Leistungen, zum Ausbildungsniveau und (mit Einschränkungen) zum Fachwissen können überprüft werden

Erstens sind innovationsbezogene Leistungen nur ein Teil der Arbeitsleistung insgesamt [48]. Keine Arbeit besteht ausschließlich daraus, innovativ zu sein – zu jeder Arbeitstätigkeit gehören immer auch Routinearbeiten, die es erforderlich machen, Kandidaten gleichzeitig auch nach anderen Kompetenzen, Fähigkeiten und persönlichkeitsbezogenen Faktoren auszuwählen.

Zweitens sollten Organisationen und Personalverantwortliche nicht in dem Versuch, das Innovationspotenzial zu maximieren, gewissermaßen nach dem Idealbild des

perfekten Profils Ausschau halten. Zählt man alle oben diskutierten Faktoren zusammen, wäre es mit großer Sicherheit ein Fehler für jede Organisation zu versuchen, nur noch Mitarbeiter einzustellen, deren Werte bei allen genannten Messinstrumenten im Extrembereich liegen. Solche Personen bringen eventuell sogar andere Probleme und Leistungsdefizite mit sich.

Drittens ist es bei vielen kreativen Aufgaben notwendig, dass der/die Stelleninhaber/-in Teil eines innovativen Teams ist. Werden Kandidaten nur nach individuellen Faktoren ausgewählt, vernachlässigt man unweigerlich diese eminent wichtigen Faktoren auf der Teamebene. Stattdessen kommt es in vielen Kontexten darauf an, Bewerber (auch) danach auszusuchen, ob sie ins Team passen, teamorientiert arbeiten und bei kollektiven Kreativitätsaufgaben teamorientierte Beiträge leisten. Den eigensinnigsten und halsstarrigsten kreativen Kopf einzustellen kann demgegenüber dazu führen, dass die Anstrengungen des Teams, neue und bessere Produkte, Prozesse und Arbeitsabläufe zu entwickeln, unterminiert werden. Hier ist also ein gewisses Maß an Ausgleich gefragt: Personen mit hohem Innovationspotenzial einzustellen ist schön und gut, aber Organisationen müssen auch daran denken, dass solche Personen ihre Fähigkeiten und kreativen Verhaltensweisen oft innerhalb eines Teamkontexts und zum erweiterten Nutzen der Innovation auf Teamebene einsetzen müssen [2].

---

## 13.5 Fazit

In diesem Kapitel wurden neuere Forschungsarbeiten zur innovationsförderlichen Personalauswahl vorgestellt und Wege aufgezeigt, wie HR-Praktiker in Organisationen dieses Ziel sinnvoll und auf der Grundlage evidenzbasierter Praxis verfolgen können. Es wurden mehrere Faktoren (Persönlichkeitsmerkmale, kognitive Fähigkeiten, Denkstile, arbeitsbezogenes Wissen und Erfahrungen) herausgearbeitet, die bei der Personalauswahl hilfreich sein können, wenn es darum geht, die Wahrscheinlichkeit innovativer Arbeitsleistungen zu erhöhen. Ausgleichend wurde auch darauf hingewiesen, dass Organisationen nicht versuchen sollten, diese Faktoren zu „maximieren“. Stattdessen sollte jedes sinnvoll gestaltete Auswahlverfahren auch andere Aspekte der allgemeinen Arbeitsleistung und der zu erwartenden Beiträge zur Innovationssteigerung auf Teamebene berücksichtigen. Wir haben das Potenzial der gängigen Personalauswahlverfahren im Hinblick auf eine innovationsförderliche Auswahl beschrieben und unsere Ansichten dazu in Tab. 13.1 zusammengefasst. Unsere Empfehlungen für die Praxis folgen aus diesen wissenschaftlichen Erkenntnissen, und wir haben auf die Notwendigkeit weiterer Forschung auf mehreren Gebieten hingewiesen. Insbesondere möchten wir an diesem Gebiet interessierte Wissenschaftler dazu ermutigen, präzise Untersuchungen in realen Auswahlkontexten durchzuführen, um dadurch zu Erkenntnissen bezüglich der Validität von Auswahlverfahren in Bezug auf innovationsbezogene Leistung als Kriterium zu generieren.



### Fragen

1. Was sind die wichtigsten Herausforderungen, denen sich Praktiker heute stellen müssen, wenn sie Personalauswahl mit Blick auf Innovationen betreiben wollen?
2. Beschreiben Sie Prädiktoren, die sich in empirischen Studien konsistent als signifikant für innovative Leistungen bestätigt haben.
3. Welche Rolle spielte Gewissenhaftigkeit in unterschiedlichen Phasen des Innovationsprozesses?
4. Welche Personalauswahlverfahren sind im Hinblick auf die Auswahl von Personen mit hoher innovativer Leistung aus welchen Gründen am besten geeignet?

### Literatur

1. Amabile, T. M. (1997). Motivating creativity in organizations: On doing what you love and loving what you do. *California Management Review*, 40, 39–58.
2. Anderson, N., & Burch, G. S. J. (2003). *The team selection inventory*. Windsor: ASE/NFER-Nelson.
3. Anderson, N., & Gasteiger, R. M. (2008). Innovation and creativity in organisations: Individual and work team research findings and implications for government policy. In B. Nooteboom & E. Stam (Hrsg.), *Micro-foundations for innovation policy* (S. 249–271). Amsterdam: Amsterdam University Press/WRR.
4. Anderson, N., De Dreu, C. K. W., & Nijstad, B. A. (2004). The routinization of innovation research: A constructively critical review of the state-of-the-science. *Journal of Organizational Behavior*, 25, 147–173.
5. Anderson, N., Salgado, J., Schinkel, S., & Cunningham-Snell, N. (2008). Staffing the organization: An introduction to personnel selection and assessment. In N. Chmiel (Hrsg.), *An introduction to work and organizational psychology: A European perspective* (S. 257–280). Oxford: Blackwell.
6. Anderson, N., Hülsheger, U., & Salgado, J. (2010). Selecting for innovation. Unpublished manuscript.
7. Anderson, N., Potočnik, K., & Zhou, J. (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary and guiding framework. *Journal of Management*, 2014, 1–37.
8. Axtell, C. M., Holman, D. J., Unsworth, K. L., Wall, T. D., Waterson, P. E., & Harrington, E. (2000). Shopfloor innovation: Facilitating the suggestion and implementation of ideas. *Journal of Occupational and Organizational Psychology*, 73, 265–285.
9. Baer, M., & Oldham, G. R. (2006). The curvilinear relation between experienced creative time pressure and creativity: Moderating effects of openness to experience and support for creativity. *Journal of Applied Psychology*, 91, 963–970.
10. Baer, M., Oldham, G. R., & Cummings, A. (2003). Rewarding creativity: When does it really matter? *The Leadership Quarterly*, 14, 569–586.
11. Barron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology*, 32, 439–476.
12. Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, 9, 1185–1203.
13. Batey, M., Furnham, A., & Safiullina, X. (2010). Intelligence, general knowledge and personality as predictors of creativity. *Learning and Individual Differences*, 20, 532–535.

14. Barrick, R. M., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.
15. Burch, G. S. J., Pavelis, C., & Port, R. L. (2008). Selecting for creativity and innovation: The relationship between the innovation potential indicator and the team selection inventory. *International Journal of Selection and Assessment, 16*, 177–181.
16. Carmeli, A., & Schaubroeck, J. (2007). The influence of leaders' and other referents' normative expectations on individual involvement in creative work. *The Leadership Quarterly, 18*, 35–48.
17. Choi, J. N. (2004a). Individual and contextual predictors of creative performance: The mediating role of psychological processes. *Creativity Research Journal, 16*, 187–199.
18. Choi, J. N. (2004b). Person-environment fit and creative behavior: Differential impacts of supplies-values and demands-abilities versions of fit. *Human Relations, 57*, 531–552.
19. Choi, J. N. (2004c). Individual and contextual dynamics of innovation-use behavior in organizations. *Human Performance, 17*, 397–414.
20. Choi, J. N., Anderson, T. A., & Veillette, A. (2009). Contextual inhibitors of employee creativity in organizations: The insulating role of creative ability. *Group and Organization Management, 34*, 330–357.
21. Clegg, C., Unsworth, K., Epitropaki, O., & Parker, G. (2002). Implicating trust in the innovation process. *Journal of Occupational and Organizational Psychology, 75*, 409–422.
22. 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.
23. Farmer, S. M., Tierney, P., & Kung-Mcintyre, K. (2003). Employee creativity in Taiwan: An application of role identity theory. *Academy of Management Journal, 46*, 618–630.
24. Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review, 2*, 290–309.
25. Frese, M., Teng, E., & Wijnen, C. J. D. (1999). Helping to improve suggestion systems: Predictors of making suggestions in companies. *Journal of Organizational Behavior, 20*, 1139–1155.
26. Furnham, A., Batey, M., Anand, D., & Manfield, J. (2008). Personality, hypomania, intelligence and creativity. *Personality and Individual Differences, 44*, 1060–1069.
27. Furnham, A., & Niderstrom, M. (2010). Ability, demographic and personality predictors of creativity. *Personality and Individual Differences, 48*, 957–961.
28. George, J. M., & Zhou, J. (2001). When openness to experience and conscientiousness are related to creative behavior: An interactional approach. *Journal of Applied Psychology, 86*, 513–524.
29. Gong, Y., Cheung, S., Wang, M., & Huang, J. (2012). Unfolding the proactive process for creativity: Integration of the employee proactivity, information exchange, and psychological safety perspectives. *Journal of Management*. doi:10.1177/0149206310380250 (in press).
30. Hammond, M. M., Neff, N. L., Farr, J. L., Schwall, A. R., & Zhao, X. (2011). Predictors of individual-level innovation at work: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts, 5*, 90–105.
31. Howell, J. M., & Boies, K. (2004). Champions of technological innovation: The influence of contextual knowledge, role orientation, idea generation, and idea promotion on champion emergence. *The Leadership Quarterly, 15*, 123–143.
32. Hülsheger, U. R., Lang, J. W. B., Anderson, N., & Kersting, M. (2011). Beyond the Big Five: The role of need for cognition and core self-evaluations for innovative work performance. Presentation at the 15th European Congress of Work and Organizational Psychology (EAWOP), Maastricht.
33. Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869–879.
34. Judge, T. A., Locke, A. E., & Durham, C. C. (1997). The dispositional causes of job satisfaction: A core evaluations approach. *Research in Organizational Behavior, 19*, 151–188.

35. Kirton, M. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology, 61*, 622–629.
36. Krause, D. E. (2004). Influence-based leadership as a determinant of the inclination to innovate and of innovation-related behaviors: An empirical investigation. *The Leadership Quarterly, 15*, 79–102.
37. Lowe, E. A., & Taylor, W. G. K. (1986). Creativity in life sciences research. *R und D Management, 16*, 45–61.
38. Madjar, N. (2008). Emotional and informational support from different sources and employee creativity. *Journal of Occupational and Organizational Psychology, 81*, 83–100.
39. Madjar, N., Oldham, G. R., & Pratt, M. G. (2002). There's no place like home? The contributions of work and nonwork creativity support to employees' creative performance. *Academy of Management Journal, 45*, 757–767.
40. McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology, 52*, 1258–1265.
41. McEntire, L. E., & Greene-Shorridge, T. M. (2011). Recruiting and selecting leaders for innovation: How to find the right leader. *Advances in Developing Human Resources, 13*, 266–278.
42. Miron, E., Erez, M., & Naveh, E. (2004). Do personal characteristics and cultural values that promote innovation, quality, and efficiency compete or complement each other? *Journal of Organizational Behavior, 25*, 175–199.
43. Mount, M. K., & Barrick, R. M. (1995). The Big Five personality dimensions: Implications for research and practice in human resource management. *Research in Personnel and Human Resources Management, 13*, 153–200.
44. Obstfeld, D. (2005). Social networks, the tertiusiungens orientation, and involvement in innovation. *Administrative Science Quarterly, 50*, 100–130.
45. Oldham, G. R., & Cummings, A. (1996). Employee creativity: Personal and contextual factors at work. *Academy of Management Journal, 39*, 607–634.
46. Parker, S. K., Williams, H. M., & Turner, N. (2006). Modeling the antecedents of proactive behavior at work. *Journal of Applied Psychology, 91*, 636–652.
47. Patterson, F. (2000). *The Innovation Potential Indicator: Test manual and user's guide*. Oxford: Oxford Psychologists Press.
48. Potočnik, K., & Anderson, N. (2012). Assessing Innovation: A 360-degree appraisal study. *International Journal of Selection and Assessment, 20*, 497–509.
49. Raja, U., & Johns, G. (2010). The joint effects of personality and job scope on in-role performance, citizenship behaviors and creativity. *Human Relations, 63*, 981–1005.
50. 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 finding. *Psychological Bulletin, 124*, 262–274.
51. Shalley, C. E., Zhou, J., & Oldham, G. R. (2004). The effects of personal and contextual characteristics on creativity: Where should we go from here? *Journal of Management, 30*, 933–958.
52. Shipton, H., Fay, D., West, M., Patterson, M., & Birdi, K. (2005). Managing people to promote innovation. *Creativity and Innovation Management, 14*, 118–128.
53. Sung, S. Y., & Choi, J. N. (2009). Do Big Five personality factors affect individual creativity? The moderating role of extrinsic motivation. *Social Behavior and Personality, 37*, 941–956.
54. Tett, P. R., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703–742.
55. Tierney, P., Farmer, S. M., & Graen, G. B. (1999). An examination of leadership and employee creativity: The relevance of traits and relationships. *Personnel Psychology, 52*, 591–620.
56. Tierney, P., & Farmer, S. M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management Journal, 45*, 1137–1148.

57. Tierney, P., & Farmer, S. M. (2004). The Pygmalion process and employee creativity. *Journal of Management*, 30, 413–432.
58. Tierney, P., & Farmer, S. M. (2011). Creative self-efficacy development and creative performance over time. *Journal of Applied Psychology*, 96, 277–293.
59. Welbourne, T. M., Johnson, D. E., & Erez, A. (1998). The role-based performance scale: Validity analysis of a theory-based measure. *Academy of Management Journal*, 41, 540–555.
60. Wang, A., & Cheng, B. (2010). When does benevolent leadership lead to creativity? The moderating role of creative role identity and job autonomy. *Journal of Organizational Behavior*, 31, 106–121.
61. Wang, H., Begley, T., Hui, C., & Lee, C. (2012). Are the effects of conscientiousness on contextual and innovative performance context specific? Organizational culture as a moderator. *International Journal of Human Resource Management*, 23, 174–189.
62. West, M. A., & Anderson, N. (1996). Innovation in top management teams. *Journal of Applied Psychology*, 81, 680–693.
63. West, M. A., & Farr, J. L. (1990). Innovation at work. In M. A. West & J. L. Farr (Hrsg.), *Innovation and creativity at work: Psychological and organizational strategies* (S. 3–13). Chichester: Wiley.
64. Wu, C., Parker, S. K., & Jong, J. P. J. de. (2014). Need for cognition as an antecedent of individual innovation behavior. *Journal of Management*, 40, 1511–1534. doi:10.1177/0149206311429862.
65. Zhou, J. (2003). When the presence of creative coworkers is related to creativity: Role of supervisor close monitoring, developmental feedback, and creative personality. *Journal of Applied Psychology*, 88, 413–422.
66. Zhou, J., & Oldham, G. R. (2001). Enhancing creative performance: Effects of expected developmental assessment strategies and creative personality. *Journal of Creative Behavior*, 35, 151–167.

---

## Verzeichnis der Mitwirkenden

**Prof. Neil Anderson, Ph.D.**, ist Professor für Human Resource Management und Direktor des Forschungszentrums Human Resource Management und Organizational Behavior (WORC) der Brunel University (England). Nachdem er seine Dissertation an der Aston University im Jahr 1989 abschloss, hatte er Lehrstühle inne an der University of London (Goldsmiths College) und der University of Amsterdam in den Niederlanden. Prof. Anderson publizierte in verschiedenen Gebieten im Bereich Human Resource Management und Organisationspsychologie und ist heute einer der fünf am häufigsten zitierten Personal- und Organisationspsychologen in Europa. Er ist Fellow der British Psychological Society, der American Psychological Association und der Division 14 der American Psychological Association (Society for Industrial-Organizational Psychology, SIOP). Prof. Anderson publizierte mehr als 70 wissenschaftliche Artikel in Zeitschriften mit Peer-Review, 20 Bücher und 30 Buchkapitel. Seine wissenschaftlichen Arbeiten erschienen in top-tier Journals wie z. B. *Journal of Applied Psychology*, *Journal of Organizational Behavior*, *Personnel Psychology*, *Human Relations*, *Human Performance*, *Journal of Occupational and Organizational Psychology*. Ferner gab Prof. Anderson mehrere Handbücher im Bereich Human Resource Management und Organisationspsychologie heraus wie etwa das *Blackwell Handbook of Selection (2005)*, *Handbook of Industrial, Work, and Organizational Psychology*. Prof. Anderson ist ferner in diversen internationalen editorial boards tätig (z. B. *Journal of Applied Psychology* [beratender Herausgeber], *Journal of Management*, *Journal of Business and Psychology*, *Personnel Psychology*, *Journal of Occupational and Organizational Psychology*, *Human Relations*, *European Journal of Work and Organizational Psychology*). Prof. Anderson ist derzeit Gastprofessor am Institut für Arbeitspsychologie der University of Sheffield sowie der City University in London, England. Darüber hinaus war er Gastprofessor an der University of Minnesota (USA) und der School of Management, Zhejiang University (China). Ein starkes Interesse an praktischen Anwendungen von Forschungsergebnissen im Bereich des Human Resource Management beibehaltend, berät er diverse Organisationen in England, Europa und den USA bei Fragestellungen des Personalpsychologie. Zu diesen Organisationen zählen z. B. auch British Aerospace, the British Army, Shell, Saville Consulting

und NASA in den USA (Jupiter Mission). Prof. Andersons spezielle Expertise betrifft die Gebiete Personalauswahl und Leistungsbeurteilung, Innovation und Kreativität am Arbeitsplatz.

E-Mail: Neil.Anderson@brunel.ac.uk

**Arleta Franczukowska**, Mag., ist wissenschaftliche Mitarbeiterin an der Alpen-Adria Universität Klagenfurt (Österreich). Sie beschäftigt sich im Rahmen ihres Promotionsvorhabens mit Korruption in Organisationen und lehrt im Bachelor-Studiengang Projektmanagement und Teamarbeit.

E-Mail: Arleta.Franczukowska@aau.at

**Luise Franke-Bartholdt**, Dipl.-Psych., ist Wissenschaftliche Mitarbeiterin am Lehrstuhl Persönlichkeitspsychologie und Diagnostik an der Technischen Universität Chemnitz. Sie forscht zu Einflussfaktoren auf die Qualität von Anforderungsanalysen sowie zu Bedingungen und Konsequenzen (un-)moralischen Verhaltens von Mitarbeitern in Organisationen und arbeitet für einen direkten Transfer ihrer wissenschaftlichen Erkenntnisse intensiv mit Partnern aus der Industrie zusammen.

E-Mail: luise.franke-bartholdt@psychologie.tu-chemnitz.de

**Prof. Dr. Rüdiger Hossiep** ist seit 1990 an der Fakultät für Psychologie der Ruhr-Universität Bochum tätig und leitet dort seit 1994 das Projektteam Testentwicklung. Zuvor war er u. a. Betriebspsychologe bei der Deutsche Bank AG, Frankfurt, und Personalberater bei der Unternehmensberatung Dieter Schröder & Partner, Düsseldorf. Sein Forschungsinteresse gilt der Entwicklung von Instrumenten für den personalpsychologischen Kontext. Rüdiger Hossiep ist Management-Diagnostiker, Führungskräftetrainer, Coach sowie Autor einschlägiger Fachbücher und psychologischer Testverfahren.

E-Mail: ruediger.hossiep@rub.de

**Prof. Dr. Uwe Peter Kanning**, Dipl.-Psych., ist Professor für Wirtschaftspsychologie an der Hochschule Osnabrück. Der Schwerpunkt in seiner Forschung liegt im Bereich der Personaldiagnostik – insbesondere Entwicklung von Testverfahren zur Messung sozialer Kompetenzen und Arbeitsmotiven, Assessment-Center, Urteilsfehler, sozial erwünschtes Antwortverhalten – sowie der Auseinandersetzung mit fragwürdigen Methoden der praktischen Personalarbeit. Darüber hinaus berät er seit fast 20 Jahren Unternehmen und Behörden bei personalpsychologischen Fragestellungen. Er ist Autor und Herausgeber von mehr als zwei Dutzend Fachbüchern und psychologischen Testverfahren. Er wurde im Jahr 2013 und 2015 von der Praxiszeitschrift Personalmagazin unter die 40 einflussreichsten Köpfe des Personalwesens gewählt, und 2015 von der Zeitschrift Unikum mit dem 3. Platz im Wettbewerb „Professor des Jahres“ ausgezeichnet.

E-Mail: U.Kanning@hs-osnabrueck.de

**Prof. Dr. Martin Kersting** beschäftigt sich wissenschaftlich und praktisch mit Personalpsychologie. Der Diplom-Psychologe war über 10 Jahre als Personalberater tätig und hat unter anderem das Kienbaum Institut für Management Diagnostik geleitet. Am Bildungs- und Wissenschaftszentrum der Bundesfinanzverwaltung leitete Kersting fast drei Jahre

den Bereich „Führung und Steuerung“ und war u. a. für die Führungskräfte trainings verantwortlich. Im Jahr 2015 wurde Kersting vom Personalmagazin zu den „40 führenden Köpfen“ des Personalwesens gewählt. Wissenschaftliche Stationen waren die FU Berlin und die RWTH Aachen. Seit 2011 ist Martin Kersting Professor für Psychologische Diagnostik an der Justus-Liebig-Universität Gießen.

E-Mail: [Martin.Kersting@psychol.uni-giessen.de](mailto:Martin.Kersting@psychol.uni-giessen.de)

**Prof. Dr. Diana E. Krause** beschäftigt sich mit den Themen Führung, Macht und Einfluss, Assessment-Center und Förderung organisationaler Innovation. Sie ist Vizepräsidentin der Tropical Education Foundation und war Professorin und Leiterin der Abteilung für Personal, Führung und Organisation an der Alpen-Adria Universität Klagenfurt (Österreich). Zuvor war sie Assistant Professor an der University of Western Ontario in Kanada und Post doc Fellow an der Colorado State University in den USA. Ferner hatte sie diverse Gast- und Vertretungsprofessuren inne an der Universität Paderborn, der Universität Fribourg in der Schweiz, der DHV in Speyer, der UMIT in Tirol und der Humboldt Universität zu Berlin. Im Jahr 2005 habilitierte sie sich an der Ludwig-Maximilians-Universität in München (venia legendi für Psychologie). Ihre Promotion in Betriebswirtschaftslehre schloss sie 2003 an der Technischen Universität Berlin mit summa cum laude ab. Sie publizierte sechs Bücher, mehr als 40 Forschungsartikel mit Peer-Review, mehr als 50 Buch- und Enzyklopädieartikel und gab mehr als 100 Präsentationen auf internationalen Kongressen sowie Konferenzen und an Forschungsinstitutionen.

E-Mail: [diana.eva.krause@gmail.com](mailto:diana.eva.krause@gmail.com)

**Prof. Dr. Klaus Melchers** ist Professor an der Universität Ulm und leitet dort die Abteilung für Arbeits- und Organisationspsychologie. Seine Forschungsschwerpunkte liegen vor allem im Bereich Personalauswahl und umfassen unter anderem Assessment-Center, Einstellungsinterviews, Bewerberreaktionen und unterschiedliche Aspekte von Selbstdarstellung in Auswahl-situationen. Weitere Arbeitsfelder betreffen die Themen Persönlichkeitserfassung im Arbeitskontext und Beurteilertraining.

E-Mail: [klaus.melchers@uni-ulm.de](mailto:klaus.melchers@uni-ulm.de)

**Prof. Dr. Friedemann W. Nerdinger**, Dipl.-Psych., ist emeritierter Professor für Wirtschafts- und Organisationspsychologie der Universität Rostock. Seine Forschungsinteressen gelten den Feldern Psychologie der Dienstleistung, Arbeitsmotivation und Arbeitszufriedenheit, Extra-Rollenverhalten in Unternehmen, Auswirkungen des demografischen Wandels auf die Personalarbeit. Er ist geschäftsführender Herausgeber der Zeitschrift für Arbeits- und Organisationspsychologie und im Herausgeberbeirat u. a. der Managementforschung.

E-Mail: [friedemann.nerdinger@uni-rostock.de](mailto:friedemann.nerdinger@uni-rostock.de)

**Michael Ott**, Dipl.-Psychologe absolvierte sein Studium an der Universität Erlangen-Nürnberg und ist wissenschaftlicher Mitarbeiter und Doktorand in der Abteilung für Psychologische Diagnostik der Justus-Liebig-Universität Gießen. Er ist in der Abteilung für die Entwicklung des Gießener Self-Assessments „Ready for Justus“ mitverantwortlich.



Zu seinen Forschungsschwerpunkten gehören die Themen Anforderungsanalyse, Self-Assessment, Leistungstests und Testängstlichkeit.

E-Mail: Michael.Ott@psychol.uni-giessen.de

**Dr. Carolin Palmer**, Diplom-Ökonomin, arbeitet seit 2011 als Wissenschaftliche Mitarbeiterin in der Abteilung für Psychologische Diagnostik an der Justus-Liebig-Universität Gießen. Neben eignungsdiagnostischen Fragestellungen und der Beschäftigung mit den Möglichkeiten psychometrischen Testens liegen ihre aktuellen Tätigkeitsschwerpunkte auf der berufsbezogenen Kreativitätsdiagnostik, auf der Erforschung von Dominanz als genderübergreifende Führungsanforderung sowie auf der Verbindung psychologischer Diagnostik mit Entrepreneurship-Themen.

E-Mail: carolin.palmer@psychol.uni-giessen.de

**Dr. Kristina Potočnik** ist Assistant Professor für Human Resource Management an der University of Edinburgh Business School (England). Nach Abschluss ihres Psychologiestudiums an der University of Ljubljana (Slowenien) im Jahr 2004, promovierte sie im Jahr 2009 in Arbeits- und Organisationspsychologie an der University of Valencia (Spanien). Ihre Forschungsinteressen sind Kreativität und Innovation am Arbeitsplatz, ältere Arbeitnehmer/-innen sowie Personalauswahl und Personalbeurteilung. Sie publizierte verschiedene Forschungsartikel zu diesen Themen in internationalen Zeitschriften mit Peer-Review wie etwa Journal of Occupational and Organizational Psychology, International Journal of Selection and Assessment, Journal of Applied Social Psychology, European Journal of Work and Organizational Psychology. Dr. Potočnik leitet ferner das von der British Academy geförderte Forschungsprojekt „Innovationsförderliche Personalauswahl: Individuelle Merkmale und Innovation im Arbeitsleben“. Das Kapitel in diesem Buch wurde partiell von diesem Drittmittelprojekt unterstützt. Sie ist Mitglied der British Psychological Society (BPS), Chartered Institute of Personnel and Development (CIPD), International Association of Applied Psychology (IAAP), European Association of Work and Organizational Psychology (EAWOP), Society for industrial and Organizational Psychology (SIOP).

E-Mail: Kristina.Potocnik@ed.ac.uk

**Prof. Dr. Anja Strobel** ist Professorin für Persönlichkeitspsychologie und Diagnostik an der Technischen Universität Chemnitz. Sie setzt sich unter anderem als Mitglied im Arbeitskreis Assessment-Center e. V. seit vielen Jahren intensiv für eine hohe Qualität der Eignungsbeurteilung und eine enge Verzahnung von Forschung und Praxis ein, wobei die Aussagekraft von Eignungsinterviews und Auswahlprozessen und die Kompetenz von Interviewenden wesentliche Schwerpunkte ihrer Arbeit darstellen. In der Forschung beschäftigt sich Prof. Dr. Strobel weiterhin mit dem Persönlichkeitsmerkmal kognitiver Motivation („need for cognition“), dessen Bedeutung in Schule und Beruf, seinen Determinanten und der Entwicklung über die Lebensspanne sowie mit Bedingungen und Konsequenzen (un-)moralischen Verhaltens von Mitarbeitern in Organisationen.

E-Mail: anja.strobel@psychologie.tu-chemnitz.de

**Anna-Sophie Ulfert**, M.Sc., ist wissenschaftliche Mitarbeiterin und Doktorandin in der Abteilung für Psychologische Diagnostik der Justus-Liebig-Universität Gießen. In der Abteilung ist sie für die Entwicklung des Gießener Self-Assessments „Ready for Justus“ mitverantwortlich. Zuvor absolvierte sie ihren Bachelor und Master in Psychologie an der Universität Maastricht mit Studien- und Forschungsaufenthalten an der University of Toronto und der Richard Ivey School of Business. Zu ihren Forschungsschwerpunkten zählen Themen wie Anforderungsanalyse, Self-Assessment, Priming, Selbstwirksamkeit sowie Zielsetzungsverhalten.

E-Mail: Anna-Sophie.Ulfert@psychol.uni-giessen.de

**Dr. Sabine Weiß** ist seit 2005 wissenschaftliche Mitarbeiterin im Projektteam Testentwicklung an der Ruhr-Universität Bochum. Zuvor absolvierte sie an der Bochumer Fakultät für Psychologie ihr Diplom mit Schwerpunkt Arbeits- und Organisationspsychologie. Seit 2005 ist sie auch als Beraterin in Personalauswahl- und Personalentwicklungsprozessen tätig. 2014 erfolgte die Promotion zum Einsatz persönlichkeitsorientierter Testverfahren im beruflichen Kontext vor dem Hintergrund des Allgemeinen Gleichbehandlungsgesetzes. Ihr Forschungsinteresse gilt der Entwicklung berufsbezogener Testverfahren.

E-Mail: s.weiss@rub.de