

9. Diagnose beruflicher Eignung und Leistung

Heinz Schuler und Uwe Funke

In diesem Kapitel wird zunächst die Bedeutung der Arbeitsanalyse als Grundlage der Bestimmung berufsbezogener Eignung und Leistung erläutert. Darauf aufbauend werden die Verfahren der Berufseignungsdiagnostik sowie die Prinzipien und Methoden der Leistungsbeurteilung dargestellt. Im letzten Teil werden aus dem umfangreichen Kontext der Eignungs- und Leistungsdiagnose zwei besonders wichtige Themen erörtert - die Gewinnung von Mitarbeitern und das Treffen personeller Entscheidungen. Vorab mögen einige Bemerkungen in die Thematik einführen.

1. Einleitende Überlegungen

Beruflicher Erfolg hängt von vielem ab. Der familiäre und soziale Hintergrund eines Menschen prägt seine Einstellungen und Erwartungen; seine Ausbildung schafft die Grundlage für die weiteren Entwicklungsmöglichkeiten; gesellschaftlicher Bedarf und Arbeitsmarktbedingungen beeinflussen die Chancen, in bestimmten Berufsfeldern tätig sein zu können; Fähigkeiten und andere Eigenschaften einer Person erleichtern oder erschweren den Erwerb von Kenntnissen und Fertigkeiten, die in berufliche Leistung umgesetzt werden können.

Was als beruflicher Erfolg angesehen wird, kann vielfältig sein: Leistung, Sinnerleben, gesellschaftlicher Status, Zufriedenheit, psychische und physische Gesundheit, das Gefühl, gefordert zu sein oder sein ruhiges Auskommen zu haben, die Überzeugung, das zu tun, was den eigenen Fähigkeiten und Interessen entspricht, der Eindruck, sich persönlich weiterzuentwickeln. Die Zusammenhänge dieser Facetten beruflichen Erfolgs mit persönlichen Merkmalen und solchen

der Lebenssituation sind vielfältig und nur in Teilen durchschaubar.

Bei der Wahl eines Berufs, einer Organisation oder einer Tätigkeit wird sich ein Mensch von mehreren Gesichtspunkten leiten lassen. Die eigenen Interessen und Wünsche, aber auch Abneigungen und Befürchtungen sind ihm dabei vielleicht besonders deutlich vor Augen. Auch die Aussichten, einen Ausbildungs- oder Studienplatz im angestrebten Fach zu bekommen, dürften zu den bewußt ins Kalkül einbezogenen Entscheidungsparametern gehören. Weniger gut erkennbar ist dem Betreffenden wahrscheinlich, in welchem Ausmaß seine Interessen von seinen Fähigkeiten abhängen oder welchen Einfluß frühere Erfahrungen und Werthaltungen in der Familie auf seine jetzigen Präferenzen haben. Die Berufsberatung, wie sie vom Arbeitsamt angeboten wird, hat den Zweck, den Ratsuchenden bei der Klärung dieser Fragen zu helfen und ihnen eine fundierte berufsbezogene Entscheidung zu ermöglichen.

Sucht ein Unternehmen oder eine andere Organisation Mitarbeiter oder Auszubildende, so stellen sich ganz ähnliche Fragen. Das Personalmarketing wird dann am erfolgreichsten sein, wenn es gelingt, das Arbeitsplatzangebot so darzustellen und den Kreis potentieller Mitarbeiter so anzusprechen, daß diejenigen gewonnen- und später gehalten - werden können, die möglichst viel zum Erfolg der Organisation beitragen können. Die Auswahlentscheidung seitens des Unternehmens wird sich primär an der zu erwartenden Leistung orientieren. Aber jedes Unternehmen wäre schlecht beraten, nicht auch die Zufriedenheit und Gesundheit der Mitarbeiter als Zielkriterien in die Auswahlentscheidung einzubeziehen. In beider Interesse ist es, Über- wie Unterforderung zu vermeiden und Möglichkeiten zur

Tätigkeit		Person
Anforderungen	—	Fähigkeiten, Fertigkeiten und Kenntnisse
Befriedigungspotential	—	Interessen und Bedürfnisse
Veränderung	—	Entwicklungspotential

Abbildung 1: Vergleiche zwischen Tätigkeit und Person bei berufsbezogenen Entscheidungen.

Entwicklung berufsbezogener Kompetenzen zu sichern. Für beide Partner der Entscheidung ergibt sich damit als Kern an gemeinsamen Zielen, Person und Tätigkeit (incl. Umfeld) so zu vergleichen, daß eine zufriedenstellende Lösung im Hinblick auf möglichst viele Zielkriterien gefunden wird. Abbildung 1 zeigt auf, daß es zumindest drei Aspekte sind, hinsichtlich derer der Vergleich zwischen Person und Tätigkeit zu erfolgen hat.

Die zum gegebenen Zeitpunkt bekannten Anforderungen der vorgesehenen Tätigkeit sind den erforderlichen Fähigkeiten, Fertigkeiten und Kenntnissen der Bewerber oder Ratsuchenden gegenüberzustellen, um die künftige Leistung zu gewährleisten. Methoden der Arbeits- und Anforderungsanalyse auf der einen Seite, Verfahren der psychologischen Eignungsdiagnostik auf der anderen sind Hilfsmittel, diesen Vergleich durchzuführen. Das Befriedigungspotential der fraglichen Tätigkeit ist mit den Interessen und Bedürfnissen der Personen zu vergleichen, um Zufriedenheit und andere Aspekte des Wohlbefindens sowie das Verbleiben in der Organisation sicherzustellen. Befriedigungspotentiale können u. a. durch Organisationsanalysen (z. B. Mitarbeiterbefragung oder Fluktuationsdaten) festgestellt werden, Interessen und Bedürfnisse z.B. mittels eines Berufsinteressentests.

Nicht alle eignungsrelevanten Merkmale einer Tätigkeit sind allerdings zum Zeitpunkt der Entscheidung bestimmbar - Arbeitsanforderungen verändern sich in einer teilweise nicht vorhersehbaren Weise. Deshalb ist es von Nutzen, zusätzlich zur Bestimmung der derzeitigen Anforderungen abzuschätzen, in welche Richtung Veränderungen zu

erwarten sind, und zusätzlich mit einem nicht bestimmbareren Anteil an Änderungen zu rechnen. Das erwünschte Entwicklungspotential einer Person sollte sowohl den absehbaren Veränderungen entsprechen als auch darüber hinaus genügende Wahrscheinlichkeit bieten, daß sie auch künftigen Entwicklungen ungewisser Art gewachsen sein wird. Die Prognose ist in diesem Bereich natürlich besonders schwierig, aber es haben sich doch einige Eigenschaften als erfolgsrelevant in unterschiedlichsten Berufskontexten herausgestellt, an die in diesem Kontext zu denken ist. Zu ihnen gehören vornehmlich Intelligenz (incl. Lernfähigkeit), allgemeine Leistungsmotivation, soziale Kompetenz und Selbstvertrauen. Ebenso wird motorisches Geschick in einigen Berufsfeldern auch künftig zu den erforderlichen Erfolgsvoraussetzungen gehören. Diese Merkmale gelten als relativ stabil und lassen sich zumindest partiell durch psychologische Instrumente, wie sie in späteren Abschnitten dargestellt werden, mit ausreichender Zuverlässigkeit erfassen - allerdings ist darauf zu achten, daß nicht alle dieser Merkmale deckungsgleich sind mit jenen, die in der neueren Persönlichkeitsforschung als generelle Faktoren unterschieden werden. Defizite hinsichtlich dieser Merkmale sind teilweise durch Training und Führung ausgleichbar (vgl. das nachfolgende Kapitel). Auch Maßnahmen zur Gestaltung der Arbeitsbedingungen (vgl. Kapitel 7) sind hierzu kein Gegensatz, sondern Ergänzungen, um eine fähigkeits- und bedürfnisentsprechende - also menschengerechte - Arbeitswelt zu gewährleisten.

2. Berufliche Anforderungen

Will man Aussagen über den Erfolg von Personen bei bestimmten beruflichen Aufgaben machen, so sind die Merkmale des Arbeitsplatzes ein zentraler Bezugspunkt. Berufliche Anforderungen sind Merkmale des Arbeitsplatzes, werden aber als «personelle Leistungsvoraussetzungen» (Hacker, 1983) dargestellt, die von seiten des Stellenin-

bers für eine erfolgreiche Bewältigung seiner Arbeitsaufgaben erforderlich sind. Zu den Anforderungen zählen: *Eigenschaftsanforderungen* (Fähigkeiten, Interessen und weitere strukturelle Merkmale der Person), *Qualifikationsanforderungen* (Kenntnisse und Fertigkeiten), *Verhaltensanforderungen* (notwendiges Verhaltensrepertoire, auch bestimmte kognitive Prozesse) und *Ergebnisanforderungen* (geforderte Aufgabenlösungen bzw. Arbeitsergebnisse). Anforderungen haben in der Regel den Status von Hypothesen über die Bedingungen beruflichen Erfolgs, die der Überprüfung in Validierungsstudien bedürfen. In vielen Fällen, aber durchaus nicht immer (vgl. 2.2), werden Anforderungen dadurch bestimmt, daß die in Frage stehenden Arbeitsplätze zunächst mit psychologischen Arbeitsanalyseverfahren untersucht werden und die Anforderungen aus den Arbeitsanalyseergebnissen abgeleitet werden. Arbeitsanalyse kann dabei allgemein definiert werden als eine hinsichtlich Anwendungszweck, theoretischer Fundierung, Inhalt und formaler Charakteristika heterogene Gruppe von Verfahren zur Sammlung und Analyse arbeitsbezogener Informationen. Ihre Anwendungsschwerpunkte im Bereich der Organisationspsychologie sind neben Eignungsdiagnostik Training/Personalentwicklung und Arbeitsgestaltung. Die Darstellung spezieller Vorgehensweisen und Verfahren für Zwecke der Arbeitsgestaltung ist in Kapitel 6 erfolgt. Im folgenden wird daher auf anwendungsspezifische Aspekte von Arbeitsanalysen als Grundlage eignungs- und leistungsdiagnostischer Maßnahmen eingegangen.

2.1 Die Funktion von Arbeitsanalysen für die Eignungsdiagnostik und Leistungsbeurteilung

Das Konzept der Eignung betrifft das Ausmaß der Übereinstimmung von Anforderungen des Arbeitsplatzes und der weiteren Arbeitsumgebung mit den Leistungsvoraussetzungen der Person. Dabei wird angenommen, daß der Grad dieser Übereinstimmung

die Erfolgswahrscheinlichkeit für eine Berufstätigkeit mitbestimmt. In der Eignungsdiagnostik ist neben der personendiagnostischen Aktivität also als zweiter Bezugspunkt zunächst Situationsdiagnostik (Arbeits- und Anforderungsanalyse) gefordert. Die Anforderungsableitung kann als Übergang von der Darstellung des Arbeitsplatzes in Situationsbegriffen zu einer Darstellung in Personbegriffen gesehen werden.

Arbeitsanalysen ist seit den Anfängen der Eignungsdiagnostik (vgl. Guion, 1976) bis heute (Algera & Greuter, 1989) hohe Bedeutung zugemessen worden. Nach Erfahrungsberichten zur Entwicklung von eignungsdiagnostischen Systemen kann der Zeitaufwand für die Arbeitsanalyse und Anforderungsableitung mehr als die Hälfte des gesamten Projektumfangs betragen. Nicht zuletzt aus diesem Grund ist in einer Reihe von Untersuchungen die Frage aufgeworfen worden, ob bzw. in welchem Differenzierungsgrad Arbeitsanalysen für eignungsdiagnostische Zwecke sinnvoll sind. Zunächst sollen Bedeutung und Möglichkeiten der Erhebung arbeitsanalytischer Daten und der Bezug zu eignungsdiagnostischen Verfahren und zu Kriterien der Leistungsbeurteilung dargestellt werden. Danach werden Alternativen diskutiert. Tabelle 1 zeigt eine Übersicht verschiedener Funktionen der Arbeitsanalyse für eignungsdiagnostische Zwecke.

Zur Taxonomie von Arbeitsanalyseverfahren für eignungsdiagnostische Zwecke sind verschiedene Vorschläge gemacht worden. Als Grundlage der weiteren Darstellung wird hier eine vereinfachte Kategorisierung der Arbeitsanalyseverfahren nach drei generellen Beschreibungsebenen für Anforderungen vorgenommen (Aufgabenebene, Verhaltensebene, Eigenschaftsebene), die mit den Beschreibungsebenen für eignungsdiagnostische Verfahren und Berufserfolgskriterien abgestimmt sind. Abbildung 2 zeigt, daß eignungsdiagnostische Verfahren schlüssiger abgeleitet und begründet werden können, wenn sie Arbeitsanalyseverfahren vergleichbarer Ebenen zur Grundlage haben.

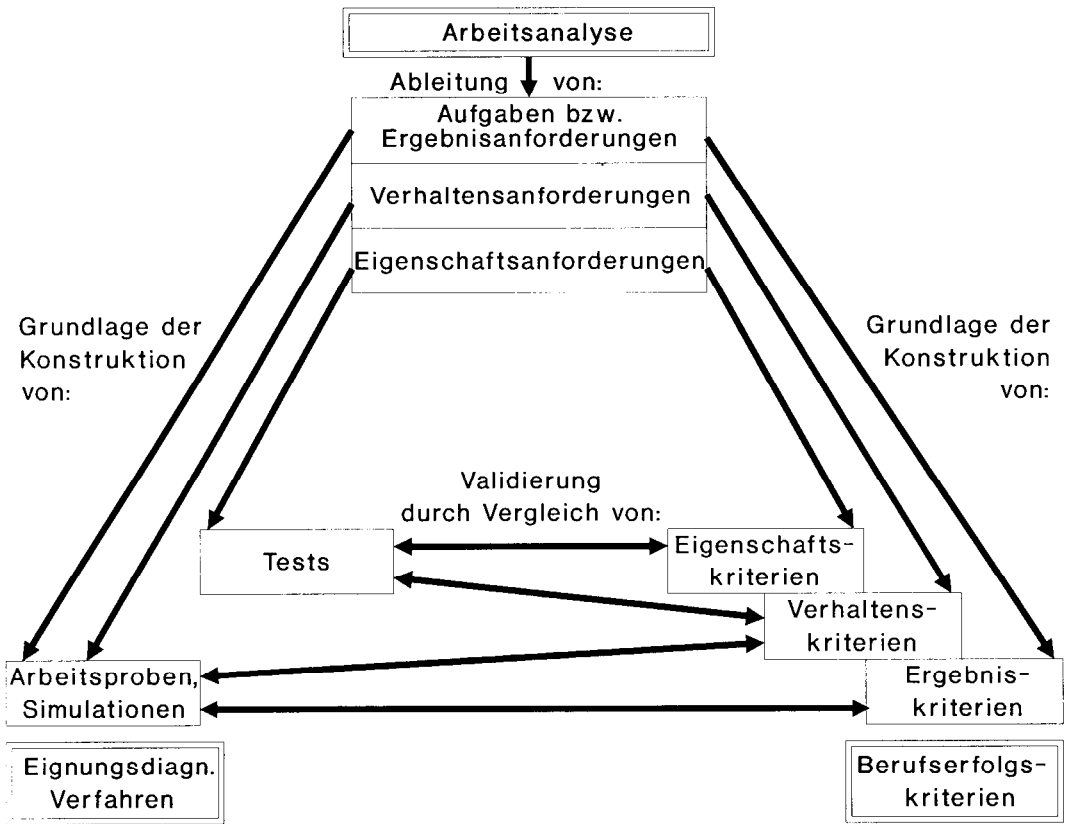


Abbildung 2: Beschreibungsebenen eignungsdiagnostisch relevanter Instrumente (vgl. Funke, in Vorbereitung).

Aufgabenebene

Schwerpunkt der aufgabenorientierten Arbeitsanalyse ist die Beschreibung des Arbeitsinhalts anhand der Aufgaben und ihrer objektiven Merkmale. Dies erfordert eine technologie- und tätigkeitsspezifische Ausrichtung dieser Arbeitsanalyseverfahren. Die Verfahren sind daher nicht voll standardisiert. Diese Spezifität ist z. B. erforderlich, wenn realitätsnahe Simulationen konstruiert werden sollen. Aufgabenorientierte Arbeitsanalysen stellen daher die Grundlage für die Konstruktion oder Auswahl eignungsdiagnostischer Arbeitsproben bzw. Simulationen dar.

Die verbreitetste Variante aufgabenorientierter Arbeitsanalyseverfahren sind sogenannte Aufgabeninventare (task inventories, TI). Hierbei wird den Arbeitsplatzinhabern oder deren Vorgesetzten eine Liste von

(oft mehreren hundert) Aufgaben vorgegeben, die jeweils im Hinblick auf einen oder mehrere Bewertungsgesichtspunkte (z.B. Bedeutung der jeweiligen Aufgabe, Häufigkeit, Schwierigkeit, Ausmaß negativer Konsequenzen bei Fehlern, Trainingsaufwand) auf Ratingskalen einzustufen sind. Die einzelnen Aufgaben werden durch einfache Sätze mit Beschreibung der jeweiligen Aktivität und des Objekts gekennzeichnet. TIs sind für Arbeitsplätze, zumindest aber für Berufe bzw. Gruppen von Arbeitsplätzen, jeweils speziell zu konstruieren, dies kann jedoch organisationsübergreifend geschehen. Eine Übersicht zur Konstruktion und Durchführung findet sich z.B. bei Drauden (1988).

Zur Konstruktion eignungsdiagnostischer Arbeitsproben werden die nach den Resultaten der TI-Durchführung wichtigsten Arbeitsaufgaben in Aufgaben-Simulationen

Tabelle 1: Funktionen psychologischer Arbeitsanalysen für eignungsdiagnostische Verfahren und Leistungsbeurteilungsverfahren (vgl. Funke, in Vorbereitung).

Konstruktion eignungsdiagnostischer Prädiktoren und Kriterien

1. Prüfung der Zielposition auf Anforderungshomogenität (reicht ein gemeinsames eignungsdiagnostisches Verfahren bzw. Leistungsbeurteilungsverfahren oder müssen verschiedene Verfahren verwendet werden?)
2. Aufnahme neuer Aufgaben, Verhaltens-, Eigenschaftsbereiche in eignungsdiagnostische Verfahren (Konstruktion/Verwendung neuer Aufgaben oder Verfahrensteile) und Leistungsbeurteilungsverfahren
3. Inhaltliche Konstruktionshinweise für neue Verfahren (Aufgabenstellung, Formulierung, Bewertung)
4. Höhere Arbeitsplatzbezogenheit der Verfahren (Ähnlichkeit der «geprüften» bzw. beurteilten Aufgaben und der Berufsaufgaben)
5. Beitrag zur Bewerberorientiertheit der Verfahren (Akzeptanz der Verfahren, Erleben der Situation, individueller Nutzen der diagnostischen Informationen)

Validierung eignungsdiagnostischer Prädiktoren und Kriterien

1. Symmetrische Zuordnung von Prädiktoren und Kriterien (inhaltliche und formale Ähnlichkeit von eignungsdiagnostischen Verfahren und Leistungsbeurteilungsverfahren bei Validierungsstudien)
2. Bestimmung/Vergroößerung der Personenstichprobe für Validierungsstudien (Aufnahme der Inhaber anforderungsähnlicher Arbeitsplätze)
3. Prüfung situationsspezifischer Validitäten (Arbeitsanalysedaten als Moderatorvariablen)
4. Definition des Geltungsbereichs von Validitätsgeneralisierungen für ein Verfahren bzw. einen Verfahrenstyp (durch Angabe der in den Einzelstudien repräsentierten Anforderungen)
5. Grundlage für synthetische Validierung (Validitätsschätzung für einen speziellen Arbeitsplatz durch die Gewichtung von Teilvaliditäten entsprechend der Bedeutung einzelner Anforderungen des Arbeitsplatzes)

Anwendung eignungsdiagnostischer Prädiktoren und Kriterien

1. Auswahl/Anwendung eines vorliegenden Verfahrens aufgrund seiner Validität für festgestellte Anforderungen (auf Basis von Einzelstudien wie auch Validitätsgeneralisierungsstudien)
2. Anwendungsausweitung auf anforderungsähnliche Arbeitsplätze
3. Anpassung von Verfahrenssystemen an Anforderungsänderungen (Neuzusammenstellung mehrerer Verfahren aus einem «Verfahrensbaukasten» entsprechend aktueller Anforderungen)

umgesetzt. Teilweise erfolgt eine weitere Bedeutsamkeitsprüfung der Aufgaben durch die Überprüfung ihres Bezugs zu übergeordneten Eigenschaftsdimensionen, die durch die Aufgaben ausreichend repräsentiert sein sollen. Eine ausführliche Darstellung dieses Vorgehens geben Schmitt und Ostroff (1986). Aus TI abgeleitete Kriterien der Leistungsbeurteilung sind vor allem ergebnisorientiert, d.h. es wird der Erfüllungsgrad der jeweiligen Aufgabe eingeschätzt.

Verhaltensebene

Schwerpunkt verhaltensorientierter Arbeitsanalyseverfahren ist der prozessuale Aspekt des Verhaltens bei der Arbeitsaus-

führung. Dabei lassen sich drei Verfahrenskategorien unterscheiden: Arbeitsplatzspezifische bzw. aufgabenspezifische Verhaltensbeschreibungen, z. B. durch die Methode kritischer Ereignisse (Critical Incident Technique, CIT, Flanagan, 1954); arbeitsplatzübergreifende Verhaltensbeschreibungen, z. B. durch den Fragebogen zur Arbeitsanalyse (FAA, Frieling & Hoyos, 1978); arbeitsplatzübergreifende Beschreibung speziell kognitiver, z.B. informationsverarbeitender oder handlungsregulierender Aktivitäten (Facaoaru & Frieling, 1985). Hauptsächlicher Vertreter arbeitsplatzspezifischer bzw. aufgabenspezifischer Verhaltensbeschreibungen ist die CIT. Ziel ist die detaillierte Beschreibung beobachtbaren leistungsrelevanten Verhaltens einschließ-

Informationsbox 1:

Konstruktion von Simulationen bzw. Arbeitsproben auf der Basis aufgaben- und verhaltensorientierter Arbeitsanalyse

137 Arbeitsplätze von Ingenieuren und Naturwissenschaftlern in der industriellen Forschung und Entwicklung (F&E) aus 14 Abteilungen von fünf Unternehmen in den drei Hauptbereichen Elektro, Kraftfahrzeug- /Maschinenbau und Chemie wurden mit einem in Anlehnung an TI und CIT entwickelten Aufgaben- und Verhaltensinventar (217 Items) durch die Arbeitsplatzinhaber beschrieben. Zu den 13 faktorenanalytisch fundierten Aufgabenbereichen wurden inhaltlich abgestimmte Arbeitsproben und Simulationen konstruiert (s. Abschnitt 5). Dargestellt werden hier sieben der 13 Aufgabenbereiche (vgl. Schuler, Funke, Moser, Donat & Barthelme, 1994).

Inhalte des Aufgaben- und Verhaltensinventars	Korrespondierende Simulationen bzw. Arbeitsproben
<i>Innovation</i> (Neuerungen erkennen, Neuerungen realisieren, Fachwissen aktualisieren) <i>Kooperation mit Vorgesetzten</i> <i>Führung</i> (Interaktion mit Mitarbeitern, Planung und Koordination)	Multimodales Interview aus situativen und biographischen Fragen mit verhaltensverankerten Ratingskalen zur Bewertung der Antworten. Getrennte Fragenbereiche zu «Innovation», «Kooperation mit Vorgesetzten» und «Führung» (vgl. Informationsbox 2).
<i>Fachliches Problemlösen</i> (Informationen beschaffen, Ziel- und Informationsunsicherheit bewältigen, Prioritäten setzen, Zusammenhänge sowie Neben- und Langzeitwirkungen berücksichtigen) <i>Experimentelle Analyse</i> (Versuchsplanung, Bedingungsvariation, Analysestrategien)	PC-gestützte komplexe Problemlöse-Simulationenaufgabe zur Analyse und Steuerung eines ökonomischen Systems (fiktives elektrotechnisches Unternehmen). Automatische Registrierung und Auswertung der Systemzustände sowie der Teilnehmer-Strategien (siehe Informationsbox 5).
<i>Kommunikation mit Kollegen</i> (Fachliche Diskussionen, Informationsaustausch, Teamarbeit) <i>Interdisziplinäre Zusammenarbeit</i> (Fachübergreifende Kontakte)	Führerlose Gruppendiskussion «Anforderungen an Projektleiter». Bewertung des Teilnehmerverhaltens während der Diskussion nach vorgegebenen Dimensionen.

lich der situativen Bedingungen, gegliedert in inhaltlich breitere Verhaltensdimensionen. Das Vorgehen dabei ist halbstandardisiert. Die Datenerhebung erfolgt bei Arbeitsplatzexperten (Arbeitsplatzinhaber, Vorgesetzte usw.) in Interview-, Workshop- oder Fragebogenform und ist zentriert auf «kritische Ereignisse» oder «Schlüsselereignisse», die besonders erfolgreiches oder besonders erfolgloses aufgabenbezogenes Verhalten darstellen. Durch die Beteiligung

mehrerer Arbeitsplatz-Experten werden in der Regel mehrere hundert kritische Ereignisse pro Arbeitsplatz beschrieben. Diese müssen anschließend bearbeitet werden, vor allem durch Zusammenfassung ähnlicher CIs und inhaltliche Gliederung in Verhaltensdimensionen. Es bleibt aber ein erheblicher Ermessungsspielraum. Eine aktuelle Beschreibung der verschiedenen Durchführungsmöglichkeiten der CIT geben Bownas und Bernardin (1988). Anwendungsmög-

lichkeiten für Auswahl- und Beurteilungsverfahren werden in nachfolgenden Abschnitten genannt. Ein Beispiel für ein Arbeitsanalyseverfahren auf TI- und CIT-Basis gibt Informationsbox 1.

Mit arbeitsplatzübergreifenden, verhaltensorientierten Analyseverfahren soll Verhalten erfaßt werden, das für die Erfüllung verschiedener Aufgaben und Teilaufgaben an unterschiedlichen Arbeitsplätzen relevant ist. Hierfür ist in den letzten 20 Jahren eine Reihe vollstandardisierter Instrumente überwiegend in Form von Beobachtungsinterviews entwickelt worden. Das verbreitetste Verfahren ist der Position Analysis Questionnaire (PAQ; McCormick & Jeanneret, 1988; deutsche Version als FAA von Frieling & Hoyos, 1978), der nach Experteneinschätzungen als breit anwendbar und ökonomisch gilt. Die verhaltensbezogenen Ergebnisse der Arbeitsanalysen werden jedoch nicht direkt für eignungsdiagnostische oder leistungsbezogene Verhaltensmaße genutzt, sondern dienen als Basis für die Ableitung von Eigenschaftsanforderungen (siehe nächster Abschnitt).

Der Einsatz arbeitsplatzübergreifender, kognitiv orientierter Verfahren als Grundlage spezieller Eignungsdiagnostica ist bisher kaum realisiert. Die vorliegenden Verfahren beziehen sich auf den Anwendungszweck der Arbeitsgestaltung (siehe Kapitel 6). Eine Ausnahme hiervon stellt das umfangreiche Tätigkeits-Analyse-Inventar dar (TAI; Frieling, Facaoaru, Benedix, Pfaus & Sonntag, 1990) das auch für differenzierte Anforderungsableitungen auf Aufgaben-, Verhaltens- und Eigenschaftsebene genutzt werden kann.

Eigenschaftsebene

Die Beschreibung des Arbeitsplatzes erfolgt hier über Eigenschaften des Arbeitsplatzinhabers, die für Aufgabenausführung und Berufserfolg als erforderlich angesehen werden (Fähigkeiten, Fertigkeiten, Persönlichkeitseigenschaften, Interessen). Die Bestimmung von Eigenschaftsanforderungen ist die Grundlage klassischer eignungsdiagno-

stischer Verfahren, meist in Form von Intelligenz- und Persönlichkeitstests. Aus der faktorenanalytischen Intelligenz- und Persönlichkeitsforschung übernommene Konstrukte sowie faktorenanalytisch bestimmte Leistungsdimensionen im psychomotorischen und physischen Bereich wurden in verschiedenen eignungsdiagnostisch orientierten «Eigenschaftslisten» zusammengefaßt (z.B. Peterson & Bownas, 1982). Die einzelnen eigenschaftsorientierten Anforderungsanalysen unterscheiden sich hauptsächlich darin, ob der Arbeitsplatz als Ganzes einer Beurteilung hinsichtlich der Eigenschaften unterzogen wird (holistisches Vorgehen), oder ob die Eigenschaften auf Grund einer vorgeschalteten aufgaben- oder verhaltensorientierten Arbeitsanalyse bezüglich der identifizierten Aufgaben oder Verhaltensweisen separat beurteilt und anschließend aggregiert werden (dekomponierendes Vorgehen). Jede Ableitung von Eigenschaftsanforderungen ist aber auf Expertenurteile über die Bedeutung der einzelnen Eigenschaften angewiesen.

Die führende Variante des holistischen Vorgehens sind die ability requirement scales (ARS) von Fleishman (Fleishman & Quaintance, 1984). Die ARS bestehen aus speziellen Ratingskalen für vorgegebene Eigenschaften. Bei den einzuschätzenden Eigenschaften, deren Zahl bei verschiedenen Veröffentlichungen zwischen 37 und 52 liegt, handelt es sich ausschließlich um kognitive, psychomotorische und physische Fähigkeiten. Ein ausführliches Anwendungsbeispiel für die Testentwicklung gibt Landy (1988). Als eigenschaftsorientierte Beurteilungskriterien können diese Merkmale direkt übernommen oder zu Gruppen zusammengefaßt werden.

Prototyp des dekomponierenden Vorgehens ist die Synthetisierung eigenschaftsbezogener Anforderungsprofile auf Basis der Ergebnisse des PAQ bzw. FAA (Frieling, 1977). Für eine Liste am Arbeitsplatz erforderlicher Verhaltensweisen (job elements), deren Ausprägung in der Arbeitsanalyse eingeschätzt wird, werden wiederum verschiedene Eigenschaften einer festgelegten Eigenschaftsliste als notwendige Vorausset-

zung angenommen. Die notwendige Ausprägung jeder dieser Eigenschaften für jedes job element des PAQ bzw. FAA wurde durch Expertenurteile allgemein festgelegt (nicht für jeden Arbeitsplatz, sondern nur einmal in der Entwicklungsphase des Arbeitsanalyseverfahrens). Dem Vorgehen liegt also die Annahme zugrunde, daß die Wichtigkeit jeder Eigenschaft *für den speziellen Arbeitsplatz insgesamt* abhängt von der allgemeinen Wichtigkeit der Eigenschaft für jedes job element und der speziellen Ausprägung jedes job elements für diesen Arbeitsplatz. Die Wichtigkeit einer Eigenschaft für den speziellen Arbeitsplatz insgesamt ergibt sich dann als Summe der allgemeinen Wichtigkeitseinschätzungen einer Eigenschaft bezüglich der job elements, die (durch Multiplikation) mit der Ausprägung der job elements für den speziellen Arbeitsplatz gewichtet wurden. Für jeden mit PAQ oder FAA analysierten Arbeitsplatz ist damit «automatisch» ein eigenschaftsorientiertes Anforderungsprofil erstellbar. Ob mit diesem Konzept generell höhere Beurteilerübereinstimmungen bei den Wichtigkeitseinschätzungen der Eigenschaften erreichbar sind als beim holistischen Vorgehen, bei dem der gesamte Arbeitsplatz «im Kopf» des Urteilers repräsentiert sein muß, ist empirisch nicht eindeutig geklärt. So konnten Cornelius und Lyness (1980) keinen Genauigkeitsgewinn der dekomponierenden Vorgehensweise nachweisen.

2.2 Alternativen zur Verwendung differenzierter Arbeits- und Anforderungsanalysen

Die hauptsächlichlichen Probleme und Kritikpunkte im Zusammenhang mit Anforderungsanalysen lassen sich zu drei Hauptpunkten zusammenfassen:

Schlecht definierbare Arbeitsaufgaben

Berufe oder Arbeitsplätze mit hoher Änderungsgeschwindigkeit bezüglich der Aufgaben bzw. Anforderungen ergeben keine sta-

bilen Arbeits- und Anforderungsdaten. Arbeitsplätze, die Teil bestimmter Karrierewege für Mitarbeiter sind und damit nur «Durchgangsstationen» darstellen, liefern ebenfalls nur Teilanforderungen. Ähnlich wirkt sich die Existenz hohen Handlungsspielraums für die Arbeitsplatzinhaber aus. Falls entsprechende Arbeits- und Anforderungsanalysen mit Konzentration auf prospektive Entwicklungen oder mit Perspektivenerweiterung auf ganze Karrierewege bzw. erweiterte Handlungsfelder nicht realisierbar sind, ist der Rückgriff auf Eigenschaften üblich, die als generell berufserfolgsrelevant angenommen werden. Dies wird häufig als «Potentialdiagnose» bezeichnet.

Inhaltliche und methodische Defizite

Mit den vorliegenden Arbeits- und Anforderungsanalyseverfahren sind z. B. interpersonale Aspekte wie Anforderungen an die «soziale Kompetenz», Bedingungen erfolgreicher Kooperation usw. bisher weniger gut analysierbar. Auch ist der Weg von der Analyse des Arbeitsplatzes bis zur Konstruktion oder Auswahl eines eignungsdiagnostischen Verfahrens mangels empirisch abgesicherter Ableitungs- und Konstruktionsschritte an Expertenurteile mit eventuellen Fehlerkomponenten gebunden. Immerhin stellt aber die im eignungsdiagnostischen Bereich eigentlich obligatorische Validierung der Auswahlverfahren eine globale empirische Kontrolle der Fehlermöglichkeiten aller Schritte dar.

Ökonomie

Sofern keine Neukonstruktion eignungsdiagnostischer Verfahren das Ziel ist, werden häufig weniger detaillierte Analysedaten zur Auswahl eignungsdiagnostischer Standardverfahren als ausreichend angesehen. Auch sind die anforderungsbezogenen Wahlmöglichkeiten unter eignungsdiagnostischen Verfahren ziemlich begrenzt. Das hat zur Erprobung weniger detaillierter Anforderungs-

rungsanalyseverfahren und globaler Expertenschätzungen geführt.

Die Alternativen zur Durchführung von Anforderungsanalysen zeichnen sich dadurch aus, daß Hypothesen über Anforderungen bzw. über valide eignungsdiagnostische Verfahren gebildet werden unter Verwendung von allgemeinen Ergebnissen eignungsdiagnostischer Forschung.

Zunächst kann man sich an *vorhandenen Validierungsstudien* orientieren, deren Methoden und Ergebnisse auf den vorliegenden Fall übertragbar sind, falls die betreffenden Arbeitsplätze ähnlich sind. Problematisch, weil unzuverlässig, sind Validierungsstudien bei zu kleinem Stichprobenumfang (ideal wären mehrere hundert Personen). Dies gilt auch für eigene *«explorative Validierungsstudien»*. Dabei werden zufällig ausgewählte oder besonders erfahrene Positionsinhaber bezüglich mutmaßlich relevanter Eignungsmerkmale untersucht (der Verfahrensauswahl liegen also Anforderungshypothesen zugrunde).

Allgemeinere Grundlagen lassen sich durch die metaanalytische Methode (vgl. Kapitel 4) der *Validitätsgeneralisierung* (Hunter & Schmidt, 1990) gewinnen. Dabei läßt sich die durchschnittliche korrigierte Validität über alle einbezogenen Studien zu einem eignungsdiagnostischen Verfahren als globales Maß der Validität dieses Verfahrens berechnen. Weiterhin kann über die relative Größe der korrigierten Streubreite aller Validitätskoeffizienten die Situationsspezifität der Validität bestimmt werden. Bis in die 80er Jahre wurde generell die Situationsspezifität der Validität unter Verweis auf die hohe (*unkorrigierte*) Streuung der Validitätskoeffizienten hervorgehoben. Konstruktion oder Auswahl sowie Validierung eignungsdiagnostischer Verfahren wären danach für *jeden* Arbeitsplatz gesondert notwendig. Für eine ganze Reihe eignungsdiagnostischer Verfahrenstypen, besonders zur Messung kognitiver Fähigkeiten, konnte dagegen mit der Methode der Validitätsgeneralisierung ein Großteil der unkorrigierten Streuung der Validitätskoeffizienten auf Artefakte, v.a. Stichprobenfehler bedingt durch kleine Validierungsstichproben, zurückgeführt wer-

den. Wenn Validitätskoeffizienten dann nach den erforderlichen Artefaktkorrekturen über sehr verschiedene Arbeitsplätze und Anwendungsbedingungen kaum noch variieren, läßt sich Validitätsgeneralisierung, d.h. Übertragbarkeit der Validität, annehmen.

Für die praktische Anwendung von Validitätsgeneralisierungsergebnissen ist zu beachten:

- die zu erwartende Validität kann trotz nachgewiesener Generalisierbarkeit gering (wenn auch > 0) sein;
- situationspezifische Konstruktion oder Auswahl von Prädiktoren kann die Validität noch erhöhen, methodische Probleme von Einzelstudien erlauben aber eventuell den Nachweis nicht;
- verlässliche Validitätsgeneralisierungsstudien setzen eine größere Zahl von Validierungsstudien voraus; diese sind für viele Auswahlverfahren, v. a. im deutschsprachigen Bereich, noch nicht vorhanden;
- Validitätsgeneralisierung ist umso eher zu erwarten, je allgemeiner die eignungsdiagnostischen Verfahren und die Erfolgskriterien ausgerichtet sind. Spezialfälle sind damit weniger bearbeitbar.

Sind keine empirischen Ergebnisse verfügbar, so muß auf *Expertenschätzungen* zurückgegriffen werden, wie sie von Hirsh, Schmidt und Hunter (1986) erprobt wurden. 28 Organisationspsychologen mit durchschnittlich vierjähriger Erfahrung in Personalauswahl gaben nach schriftlicher Information Validitätsschätzungen über sechs kognitive Tests hinsichtlich neun Berufen ab. Die Validitätseinschätzungen wurden mit den «wahren» Validitätskoeffizienten verglichen. Die durchschnittliche wahre Validität von $r = .31$ wurde im Mittel der 28 Beurteiler um 0.07 überschätzt. Eine frühere Studie der Autoren zeigte, daß Top-Experten noch genauere Validitätsschätzungen abgeben können. Eine empirische Validierungsstudie gleicher Genauigkeit hätte eine Stichprobengröße von mehr als 240 Personen erfordert.

3. Eignungsdiagnostische Auswahlverfahren

Erfolgreiche Eignungsdiagnostik erfordert im Regelfall spezielle diagnostische Verfahren - im Unterschied zur differentiellen Psychologie. Binning und Barrett (1989) unterscheiden zwischen einerseits differentialpsychologischen, meist eigenschaftsorientierten Prädiktorkonstrukten mit ihrem Bezug auf eine «natürliche» Ordnung bzw. Kovariation allgemein auftretender Verhaltensweisen und andererseits Konstrukten beruflicher Leistung, die zielorientiert oder verhaltensorientiert definiert sind. Diese Unterscheidung von Eigenschafts- und Leistungskonstrukten läßt sich auch an den Ergebnissen von Validitätsgeneralisierungsstudien veranschaulichen, die zeigen, daß Tests für Konstrukte kognitiver Fähigkeiten - diese sind stärker leistungsbezogen konstruiert - im Gegensatz zu Tests für andere Prädiktorkonstrukte (Persönlichkeitsmerkmale) meist generellere Validität besitzen. Die gleiche Erfahrung macht man mit biographischen Fragebogen. Wie in Abschnitt 2 dargestellt, bemüht man sich daher, die Prädiktorkonstrukte den zu prognostizierenden Kriterienkonstrukten anzunähern.

Dies kann erfolgen durch die Entwicklung speziellerer leistungsrelevanter Prädiktorkonstrukte, z.B. durch Spezifizierung von «Extraversion» zu «Serviceorientierung» als Basiskonstrukt für die Konstruktion eines Auswahlverfahrens für Verkaufspersonal. Prädiktorkonstrukte werden als relativ *homogene* Verhaltensmerkmale bzw. Eigenschaften konzipiert, für deren Erfassung z. B. Tests (siehe 3.4) konstruiert werden. Es handelt sich also um ein *eigenschaftsorientiertes* Prognosekonzept. Ein anderer Weg besteht darin, eine «Punkt-zu-Punkt-Übereinstimmung» zwischen dem eignungsdiagnostischen Auswahlverfahren und dem zu prognostizierenden Leistungsbereich anzustreben. Die Aufgaben des Auswahlverfahrens stellen dann eine Simulation wichtiger realer beruflicher Aufgaben unter kontrollierten Bedingungen dar, so daß man von einem *simulationsorientierten* Prognosekon-

zept sprechen kann. Der Prädiktor bezieht sich auf einen *heterogenen* Verhaltensbereich, d. h. die simulierte Aufgabe stellt im Gegensatz zum eigenschaftsorientierten Prognosekonzept eine Kombination verschiedener Anforderungen gleichzeitig an den Probanden. Das klassische Instrument simulationsorientierter Diagnostik sind Arbeitsproben (s. Abschnitt 3.6), eine Reihe weiterer eignungsdiagnostischer Verfahren kann aber als mehr oder weniger simulationsorientiert angesehen werden (s. Tabelle 2). Im Gegensatz zur Konstruktion von Tests verzichtet man darauf, die Arbeitstätigkeiten in für deren erfolgreiche Ausführung erforderliche Personmerkmale zu übersetzen. Anstelle eines «Anzeichens» (Testverhalten) oder einer «Prädisposition» (Eigenschaft) wird von einer Verhaltensstichprobe auf ähnliches künftiges Verhalten geschlossen. Dies sollte umso besser gelingen, je ähnlicher Prädiktor und Kriterium einander sind. Dementsprechend beginnt der typische Weg der Konstruktion von Arbeitsproben mit der Herleitung aus aufgaben- oder verhaltensorientierten Methoden der Arbeitsanalyse (s. Abbildung 2).

Außer im Prognosekonzept unterscheiden sich eignungsdiagnostische Verfahren z.B. auch im Medium der Präsentation der Aufgaben bzw. der Registrierung geforderter Reaktionen. Daraus ergibt sich ein Schema eignungsdiagnostischer Verfahren, wie es in Tabelle 2 dargestellt ist.

Die wichtigsten der in Tabelle 2 genannten Prädiktortypen werden im folgenden erläutert.

3.1 Auswertung der Bewerbungsunterlagen

Den ersten Schritt bei der Auswahl neuer Mitarbeiter stellt gewöhnlich die Auswertung der Bewerbungsunterlagen dar. In vielen Fällen ist dies insofern bereits der entscheidende Schritt, als nur ein Teil der Bewerber die nächste Auswahlstufe - zumeist das Vorstellungsgespräch - erreicht. Im Unterschied zu allen übrigen Auswahlverfahren werden diese beiden Informationsquellen - Bewerbungsunterlagen

Tabelle 2: Schema eignungsdiagnostischer Verfahren.

Medium der Aufgabe bzw. Reaktion	Prognosekonzept	
	eigenschaftsorientiert	simulationsorientiert
schriftlich (Papier- Bleistift)	Test	Bewerbungsunterlagen, Biograph. Fragebogen
schriftlich (PC)	computergestützter Test	computergestützte Simulation
mündlich	herkömmliches Interview	multimodales Interview
Verhalten	Videotest	Arbeitsprobe, Assessment Center, Probezeit

und Gespräch - in nahezu allen Fällen als Entscheidungsgrundlage herangezogen (Schuler, Frier & Kauffmann, 1993). Die Logik dieser (Vor-)Auswahl besteht in ihrem Kern im Versuch, aus vergangenem Verhalten auf künftiges Verhalten zu schließen, sei es auf dem Weg über Eigenschaftskonstrukte oder in direkter Analogie. Zusätzlich spielt in beiden Verfahren die Selbstpräsentation des Bewerbers eine große Rolle. Bewerbungsunterlagen werden allerdings heute oft in stark vereinheitlichter Form vorgelegt, was ihre Aussagekraft zumindest hinsichtlich formaler Aspekte einschränkt. Die Validität von Bewerbungsunterlagen wird oft in Frage gestellt. Tatsächlich errechnet sich aus hierzu durchgeführten Untersuchungen eine durchschnittliche prognostische Validität von nicht mehr als $r = 0.18$, gemessen am Kriterium Vorgesetztenbeurteilung (Reilly & Chao, 1982). Die valideste Komponente der Bewerbungsunterlagen dürften die Schul- und Examensnoten darstellen. Eine Metaanalyse aller vorliegenden Einzeluntersuchungen hat ergeben, daß Schulzeugnisse relativ gute Voraussagen weiterer Ausbildungsleistungen erlauben: Die prognostische Validität beträgt $r = 0.40$ für den Schluß von Realschulzeugnissen auf die Leistung beim Abschluß der beruflichen Ausbildung; Studienleistungen lassen sich mit $r = 0.46$ auf der Basis der Abiturnoten vorhersagen (Baron-Boldt, Schuler & Funke, 1988). Prognosen des beruflichen Erfolgs im engeren Sinn sind dagegen in weit geringerem Maße möglich, die Validitätskoeffizienten kommen auf einen Durchschnittswert von $r = 0.15$ (Samson, Graue, Weinstein & Walberg, 1984).

Eine weitere Komponente in Bewerbungsunterlagen, die einen eigenständigen Beitrag zur Validität leisten kann, sind Arbeitszeugnisse und Referenzschreiben. Zwar steht ihre Aussagekraft unter der Einschränkung, daß kaum ausdrücklich Nachteiliges über eine Person ausgesagt wird - was nach geltendem Arbeitsrecht auch nur sehr eingeschränkt zulässig wäre -, aber Abstufungsgrade in positiven Aussagen, das Weglassen wichtiger Aspekte und andere indirekte Hinweise (vgl. Weuster, 1991) erlauben zumindest Personalfachleuten vorsichtige Schlüsse. Die gelegentlich praktizierte Überprüfung der Referenzinformation liefert nach dem Eindruck vieler Personalpraktiker und Unternehmensberater wertvolle ergänzende Information. Was im Englischen mit «reference check» bezeichnet wird und nach Hunter und Hunter (1984) mit $r = 0.26$ höhere Validität verspricht als die Auswertung der Bewerbungsunterlagen selbst, besteht neben der Referenzüberprüfung zumeist im aktiven Einholen zusätzlicher Beurteilungsinformation von früheren Arbeitsstellen und Vorgesetzten.

Bedauerlicherweise wurde bisher keine systematische Information über Validitätswerte erhoben, die sich aus der Kombination der Unterlagenauswertung mit anderen Auswahlmethoden ergeben. Auch bezüglich der Urteilsbildung der Personen, die Auswahlentscheidungen treffen, ist die Verwertung von Bewerbungsunterlagen bisher nur unzulänglich erforscht. Studien mit dem Ziel, diagnostische Prozesse zu rekonstruieren und/oder die Verwertung diagnostischer Information zu verbessern, werden für diese Informationsquelle - im Unterschied zum

Interview - erst in neuester Zeit unternommen (z.B. Seibt & Kleinmann, 1991). Personalfachleute in den Personalabteilungen großer Unternehmen und in Unternehmensberatungen nennen vor allem folgende Aspekte in Bewerbungsunterlagen als beachtenswert:

- Formale Aspekte (z.B. Vollständigkeit, Korrektheit)
- Stil der Ansprache und Selbstdarstellung
- bisherige Tätigkeiten
- erreichte Position
- Ausbildung und Kenntnisse
- Schulbildung; schulische und Studienleistungen
- Arbeitszeugnisse und Referenzen
- Fort- und Weiterbildung
- Ergänzende anforderungsspezifische Aspekte (z. B. Mobilität)
- Bewerbungsmotive
- Plausibilität früherer Stellenwechsel
- Übereinstimmung von Lebenslauf und Belegen

Bewerbungsunterlagen werden nicht nur für sich genommen als Entscheidungsgrundlage verwendet, sondern auch zur Vorbereitung eventuell anschließender Einstellungsgespräche genutzt, beispielsweise indem Hinweise für die Gesprächsstrukturierung abgeleitet oder offengebliebene Fragen vorge-merkt werden.

Zur (Teil-)Formalisierung von Bewerbungsunterlagen werden von vielen Organisationen, vor allem für einfachere Tätigkeiten, Personalfragebogen verwendet. Sie erlauben eine teilstandardisierte Auswertung, gegebenenfalls sogar eine Gewichtung von Einzelitems entsprechend ihrer empirisch bestimmten Validität. Die Weiterverfolgung dieses Gedankens führt zu biographischen Fragebogen (s. Abschnitt 3.3).

3.2 Einstellungsgespräche

Vorstellungsgespräche oder Einstellungsgespräche werden in den verschiedensten Formen durchgeführt, von der völlig freien Gesprächsform über teilstrukturierte bis zu vollstrukturierten Varianten bzw. standardi-

sierten Abläufen und Fragestellungen. Die gestellten Fragen beziehen sich v. a. auf Berufserfahrung und Berufsausbildung, auf Aspekte des Lebenslaufs und deren subjektive Verarbeitung, gelegentlich auch auf persönliche Bereiche wie den des familiären Hintergrunds. Die Antworten des Bewerbers wie auch weitere Eindrücke aus dem Gesprächsverlauf, beispielsweise nonverbales Verhalten betreffend, werden gewöhnlich zu einem «klinischen» Urteil - also in intuitiver Kombination und Gewichtung - zusammengefaßt.

Die primäre Funktion des Interviews ist die gleiche, wie sie mit anderen Methoden der Personalauswahl angestrebt wird - Mitarbeiter auszuwählen, die in der betreffenden Organisation erfolgreich tätig sein werden. Daneben werden mit dem Interview allerdings auch Zielsetzungen verbunden, die über die Möglichkeiten der meisten anderen Auswahlverfahren hinausgehen, darunter die Information der Bewerber über Unternehmen, Arbeitstätigkeit, Arbeitsplatz und Arbeitsanforderungen; Kennenlernen der Erwartungen der Bewerber; Gewinnen von Information über den Arbeitsmarkt; persönliches Kennenlernen (Aufbau von Kontakt, Sympathie, Identifikation, Verpflichtung); (vorteilhafte) Darstellung des Unternehmens; Vereinbaren der Vertragsbedingungen.

Angesichts dieser vielfältigen Zielsetzungen ist die ubiquitäre Verwendung trotz methodischer Zweifel nicht verwunderlich. Gleichzeitig hat sie die Forschung und speziell in den letzten Jahren Versuche der methodischen Verbesserung des Einstellungsgesprächs gefördert. Gegenstand der Interviewforschung sind zur Zeit Fragen der Informationsverarbeitung und Urteilsbildung, der Interaktionsprozesse, der Strategien der handelnden Personen, der Fairneß, des Interviewkontexts und der optimalen Fragegestaltung (Eder, Kacmar & Ferris, 1989), vor allem aber die Validität des Interviews, in deren Dienst auch die genannten Fragestellungen direkt oder indirekt stehen.

Die Validierungsforschung zum Einstellungsgespräch ist insofern besonders interessant, als sie schon zu Anfang des zwanzig-

sten Jahrhunderts zu dem für viele Verwender kontraintuitiven Schluß kam, daß Leistungsprognosen auf seiner Basis nur sehr eingeschränkt möglich sind (Scott, 1916). Mehrere Sammelreferate in den letzten Jahrzehnten faßten eine große Zahl von Einzeluntersuchungen zusammen und tatierten die prognostische Validität auf etwa $r = 0.20$ (z. B. Arvey & Campion, 1982; Webster, 1982). Die Metaanalyse von Hunter und Hunter (1984) kommt aus der Zusammenfassung von 15 Studien auf einen Durchschnittswert von $r = 0.14$. Tendenziell noch geringer, nämlich kaum über der Zufallswahrscheinlichkeit, liegt die Vorhersageleistung der aus Interviews gewonnenen Beurteilungen für Studienleistungen (Trost, 1986).

Diese geringen Validitätswerte gelten für konventionell geführte Einstellungsgespräche. Für strukturierte Interviews fanden demgegenüber Wiesner und Cronshaw (1988) einen metaanalytisch korrigierten Durchschnittswert von $r = 0.40$; McDaniel, Whetzel, Schmidt, Hunter, Maurer und Russell (1986) berichteten für anforderungsbezogene Interviews eine durchschnittliche Validität von $r = 0.30$.

An Gründen für die relativ geringe Validität des konventionellen Gesprächs wurden in der Literatur (Arvey & Campion, 1982; Neuberger, 1974; Schmitt, 1976) vor allem die unzulängliche Informationsverarbeitung und die mangelnde testtheoretische Fundierung des Verfahrens genannt. Wesentliche aufklärende Arbeiten wurden schon in den sechziger Jahren im Rahmen der McGill-Studien geleistet (Webster, 1982). Die neuere Forschung ist bei Eder und Ferris (1989) dokumentiert.

Die wichtigsten Gründe für die geringe Validität konventionell geführter Einstellungsgespräche dürften in der Informationsverarbeitung durch die Interviewer liegen. Beispielsweise hat sich gezeigt, daß die meisten Informationen, die Interviewer bei ihren Entscheidungen berücksichtigen, irrelevant sind und daß frühe Eindrücke im Gespräch für die weitere Informationsverarbeitung dominierendes Gewicht haben und das Verhalten der Interaktionspartner steuern. Dar-

über hinaus sind emotionale Urteilskomponenten (beispielsweise beeinflusst durch Attraktivität, subjektive Ähnlichkeit und andere sympathiebedingende Parameter) für einen Großteil der Urteilsvarianz verantwortlich.

Zudem spricht vieles für die unterschiedliche Fähigkeit verschiedener Interviewer, valide Urteile abzugeben, was aber noch nicht zulänglich erforscht ist. Dreher, Ash und Hancock (1988) weisen darauf hin, daß durch unterschiedliche Urteilstendenzen der Beurteiler die Validität von Interviews unterschätzt wird, die ja gewöhnlich auf aggregiertem Niveau berechnet wird, wodurch selbst verlässliche Beurteilerunterschiede als Fehlervarianz in die Berechnung eingehen. Versuche, das Interview durch geeignete Fragengestaltung zu verbessern, wurden u.a. von Janz (1982) und von Latham, Saari, Pursell und Campion (1980) unternommen. Das von Janz konzipierte *Behavior Description Interview* beruht auf dem Prinzip biographiebezogener Fragestellung. Anders als beim *Situativen Interview* von Latham et al. werden allerdings keine präzisen Richtlinien zur Bewertung der Antworten vorgegeben. Letzteres besteht in der kurzen Schilderung von Arbeitssituationen, häufig Konfliktsituationen, wie sie aus Sammlungen von *Critical Incidents* durch geringe Umformulierung gewonnen werden können (ein Beispiel gibt Informationsbox 2). Das Situative Interview basiert auf der Zielsetzungstheorie, die annimmt, daß Verhaltensintentionen ein guter Prädiktor realen Verhaltens sind (Locke & Latham, 1990). Für beide Typen strukturierter Interviews wurden gute Validitätswerte, allerdings an relativ kleinen Stichproben, berichtet, für das Situative Interview auch gute Reliabilitätswerte. Neben der Strukturierung haben diese beiden neueren Interviewkonzepte die Orientierung am Simulationsprinzip gemeinsam. Strukturierte Interviews, die mehrere Fragenprinzipien kombinieren, wurden von Pursell, Campion und Gaylord (1980) und von Schuler (1989a) entwickelt. Das Interview von Pursell et al. enthält neben situativen Fragen auch noch Kenntnisfragen und kleine Arbeitsproben. Das *Multimodale*

Informationsbox 2:*Aufbau des Multimodalen Einstellungsinterviews*

1. *Gesprächsbeginn.* Kurze informelle Unterhaltung; Bemühen um angenehme und offene Atmosphäre; Skizzierung des Verfahrensablaufs; keine Beurteilung.
2. *Selbstvorstellung des Bewerbers.* Bewerber sprechen einige Minuten über ihren persönlichen und beruflichen Hintergrund. Beurteilung von sechs Dimensionen auf einer dreistufigen Skala.
3. *Freies Gespräch.* Interviewer stellt offene Fragen in Anknüpfung an Selbstvorstellung und Bewerbungsunterlagen. Summarische Eindrucksbeurteilung.
4. *Biographiebezogene Fragen.* Biographische (oder «Erfahrungs-»)Fragen werden aus Anforderungsanalysen abgeleitet oder anforderungsbezogen aus biographischen Fragebogen übernommen. Die Antworten werden anhand einer dreistufigen verhaltensverankerten Skala beurteilt.
5. *Realistische Tätigkeitsinformation.* Ausgewogene Information seitens des Interviewers über Arbeitsplatz und Unternehmen. Überleitung zu situativen Fragen.
6. *Situative Fragen.* Auf critical incident-Basis konstruierte situative Fragen werden gestellt, die Antworten werden auf fünfstufigen verhaltensverankerten Skalen (Beispiel s. unten) beurteilt.
7. *Gesprächsabschluß.* Fragen des Bewerbers; Zusammenfassung; weitere Vereinbarungen.

Beispiel für eine situative Frage

Die Leistung eines Ihrer Mitarbeiter hat nachgelassen. Anlässlich Ihrer jährlichen Gehaltsgespräche müssen Sie ihm erklären, daß seine Gehaltserhöhung geringer ausfällt als die Zulage, die die meisten seiner Kollegen bekommen. Wie gehen Sie vor?

Beispielantwort 0 Punkte:

Ich sage dem Mitarbeiter, daß ich ihm gerne mehr gegeben hätte, daß aber die Geschäftsleitung keinen weiteren Rahmen offenläßt.

Beispielantwort 2 Punkte:

Ich erkläre dem Mitarbeiter, daß er seine Ziele nicht erreicht hat, und stelle ihm bei Verbesserung eine Gehaltsüberprüfung in Aussicht.

Beispielantwort 4 Punkte:

Ich sage dem Mitarbeiter, daß ich mir Gedanken über seine nachlassende Leistung mache, derentwegen die Zulage geringer ausfällt. Ich versuche, gemeinsam mit ihm die Gründe herauszufinden. Dann besprechen wir Maßnahmen, die Leistung wieder zu verbessern, und vereinbaren neue Ziele.

Einstellungsinterview (Schuler, 1989a, 1992) besteht aus einer feststehenden Abfolge von insgesamt sieben Komponenten und wurde mit dem Ziel konstruiert, möglichst viele der in der Interviewforschung festgestellten Defizite zu überwinden. Aufgrund der Konstruktvalidierung (Schuler & Funke, 1989) konnte belegt werden, daß strukturierte Interviews nicht zwangsläufig bestimmte Konstrukte erfassen (Hunter & Hirsh, 1987) sondern anforderungs- bzw. konstruktorientiert gestaltet werden können. Die Interviewergebnisse zeigen mittlere bis hohe Übereinstimmung mit Assessment Center-Ergebnissen sowie den Daten aus Persönlichkeitstests und biographischen Fragebogen. Informationsbox 2 informiert über den Aufbau dieses Interviewverfahrens. Die Komponenten des Multimodalen Einstel-

lungsinterviews stellen eine Abfolge von standardisierten und frei geführten Gesprächsteilen dar. Im Unterschied zu vollständig standardisierten Verfahren (z. B. Latham & Finnegan, 1993) wird diese Gesprächsform von Bewerbern positiv bewertet.

3.3 Biographische Fragebogen

Biographische Fragebogen in der Eignungsdiagnostik (überblick bei Schuler & Stehle, 1990) beziehen sich auf berufserfolgsrelevante Ausschnitte der Lebensgeschichte. Die eignungsdiagnostische Verwendung dieser Ausschnitte wird meist damit begründet, daß sich zukünftiges Arbeitsverhalten und zukünftiger Berufserfolg aus vergangenem Verhalten und Erfahrung, bzw. deren sub-

jektiver Verarbeitung und Wiedergabe, prognostizieren lassen. Unter dieser Grundannahme von Verhaltenskonsistenz könnten biographische Fragebogen zum Teil als eine Form «retrospektiver Arbeitsproben» interpretiert werden.

Inhaltlich handelt es sich um standardisierte Selbstbeschreibungen. Die einzelnen Fragen beziehen sich im wesentlichen auf das, was auch Bewerbungsunterlagen und Einstellungsinterview an prognostisch relevanter Information enthalten. Von konventionellen Personalfragebogen unterscheiden sie sich gewöhnlich durch ihren Umfang, gelegentlich auch durch die angesprochenen Erfahrungsbereiche, vor allem aber durch ihre empirische Validierung. Biographische Fragebogen werden gelegentlich als eine Form von Persönlichkeitstests angesehen, unterscheiden sich aber von diesen durch konkretere, meist erfahrungsbezogene und prinzipiell überprüfbare Frageformulierungen, die den Befragten persönlich als relevanter erscheinen. Die Frageformulierung wird als ein Grund dafür angesehen, daß biographische Daten in Auswahl-situationen vergleichsweise wenig verfälscht werden.

Die Konstruktion bzw. Validierung biographischer Fragebogen erfolgt meist auf Einzelfragenbasis in Form einer Itemanalyse an einem Berufserfolgskriterium, wie sie für konstrukt-heterogene Fragebogen im Gegensatz zu homogenen, eindimensionalen Skalen erforderlich ist. Bei kategorial skalierten Antwortmöglichkeiten wird jede Antwortmöglichkeit separat validiert. Das Ergebnis im biographischen Fragebogen wird als Summe der entsprechend der Itemanalyse gewichteten Antwortausprägungen (s. Informationsbox 3) bestimmt. Einen Überblick dieser Verfahren der Itemvalidierung gibt Hollmann (1991). Das resultierende Verfahren ist stark an die jeweilige Stichprobe angepaßt, ebenso an das verwendete Außenkriterium. Dies führt zu verminderter Generalisierbarkeit, d.h. der Erfordernis organisations-spezifischer Validierung und auch relativ häufiger Überprüfung während längerer Verwendungsdauer (Funke & Schuler, 1990).

Als eignungsdiagnostisches Verfahren ha-

ben biographische Fragebogen - insbesondere in den Vereinigten Staaten - eine lange Tradition, während sie im deutschen Sprachraum bis vor kurzem auf andere Bereiche beschränkt waren. Owens (1976) datiert den ersten als Auswahlverfahren verwendeten Fragebogen auf das Jahr 1894. Vor allem die Auswahl von Außendienstmitarbeitern wurde zur Domäne dieses Verfahrens, daneben wurde es aber auch bei gewerblichen und kaufmännischen Arbeitnehmern, bei Führungskräften und im militärischen Bereich eingesetzt. Auch zur Auswahl von Wissenschaftlern und Ingenieuren, z.B. bei der NASA, wurden mehrere Studien veröffentlicht (zusammenfassend Funke & Schuler, 1986).

Biographische Fragebogen erreichen vergleichsweise hohe Validitätswerte. Die Metaanalyse von Hunter und Hunter (1984) resultiert in einem Wert von $r = 0.37$, gemessen an der Leistungsbeurteilung durch Vorgesetzte, und $r = 0.26$, wenn die Prognose an der späteren Beförderung gemessen wird. Schmitt, Gooding, Noe und Kirsch (1984) kommen für diese Kriterien zu vergleichbaren Ergebnissen und für das Gehalt auf $r = 0.52$, während sich für Ausbildungsleistungen nur eine mäßige ($r = 0.23$), für die Beschäftigungsdauer praktisch keine Validität findet. In einer Metaanalyse von Funke, Krauß, Schuler und Stapf (1987) über 50 Studien zur Vorhersage wissenschaftlicher Leistungen erwiesen sich biographische Fragebogen mit einer durchschnittlichen korrigierten Validität von 0.47 als validestes Verfahren. Speziell für Jugendliche finden sich dagegen niedrige Prognosewerte (U. Funke, 1986). Dies dürfte an der noch geringen Verhaltensstabilität in jüngeren Jahren, aber auch an der geringen Datenbasis, also der kürzeren Biographie, liegen.

Eine konsequentere Simulationsorientierung kommt in einer Variante biographischer Fragebogen, dem «accomplishment record» zum Ausdruck (Schmidt, Caplan, Bemis, Decuir, Dunn & Antone, 1979). Hierbei werden frühere Leistungen mit Bezug zu arbeitsanalytisch begründeten Arbeitsanforderungen erhoben. Zu jeder Leistungskategorie wird vom Bewerber die

Informationsbox 3:*Biographische Frage*

Beispielfrage für den Versicherungsaußendienst (Barthel & Stehle, 1990):

			Antwort
Wie wichtig	<input type="radio"/>	sehr großer Einfluß	(a)
war Unabhängigkeit	<input type="radio"/>	großer Einfluß	(b)
als Grundlage für	<input type="radio"/>	gewisser Einfluß	(c)
Ihre Berufswahl?	<input type="radio"/>	geringer Einfluß	(d)
	<input type="radio"/>	gar kein Einfluß	(e)

Validität der Frage (biseriale Rangkorrelation zwischen Wichtigkeit der Unabhängigkeit und Berufserfolg) $r = 0.21$

Gewichtung der Antwortalternativen (England, 1971, S. 27):

Ant- wort	erfolgreiche Mitarbeiter		nicht erfolgreiche Mitarbeiter		Häufigk.- differenz	Netto- gewicht	Grob- gewicht
	abs.	rel.	abs.	rel.			
(a)	110	43,7%	38	21,5%	22,2%	+6	+1
(b)	76	30,1%	71	41,5%	-11,4%	-4	-1
(c)	28	11,0%	35	20,5%	-9,5%	-3	0
(d)	19	7,5%	8	4,7%	2,8%	+1	0
(e)	19	7,7%	19	11,1%	-3,4%	-1	0

In der Fragebogenauswertung werden nur die Antwortalternativen a) und b) berücksichtigt, da nur sie die Gruppen ausreichend differenzieren.

Nennung von mindestens zwei Leistungsbeispielen in CI-ähnlicher Form verlangt sowie von Zeugen für diese Darstellung. Die Schilderungen werden von Experten bewertet. Hunter und Hunter (1984) berichten eine metaanalytisch korrigierte mittlere Validität aus fünf Studien von $r = 0.49$, die allerdings überschätzt ist, da die Bewerber schon Berufserfahrung hatten. Hough (1984) kommt in einer Einzelstudie bei Rechtsanwälten auf $r = 0.25$.

Ein anderer Ansatz bezieht sich auf die Forderung, in biographischen Fragebogen psychologische Konstrukte umzusetzen (Funke & Schuler, 1990). Eine generelle und umfassende Theorie biographischer Daten existiert allerdings nicht. Ansätze bieten Lern- und Sozialisationstheorien (Jäger, 1988) und an Persönlichkeitstheorien angelehnte Konstruktvalidierungen (Funke & Schuler, 1990; Hollmann, 1991). Das «developmental-inte-

grative model» von Owens und Schoenfeldt (vgl. Owens, 1976) mit dem Nachweis, daß nach biographischen Konstellationen gebildete Gruppen sich auch in einer Vielzahl anderer psychologischer Variablen unterscheiden, stellt hierzu ebenfalls einen wichtigen Beitrag dar.

3.4 Psychologische Tests

Psychologische Testverfahren sind die klassischen und am häufigsten verwendeten Instrumente der psychologischen Eignungsdiagnostik im engeren Sinn (Brambring, 1983). Psychologische Eignungsdiagnostik kann jedoch keinesfalls mehr mit der Durchführung von Tests gleichgesetzt werden, vielmehr sind alternative Verfahren im Begriff, Tests zu ergänzen und in manchen Bereichen sogar abzulösen.

Inhaltlich bezieht sich ein Test oder Subtest in aller Regel auf einen *definierten* Verhaltensbereich bzw. die entsprechende Eigenschaft. Methodisch ist er durch testtheoretische Fundierung der Konstruktion mit jeweils mehreren ähnlichen Items pro Merkmalsbereich gekennzeichnet (umgekehrt existiert für viele alternative Verfahren bisher keine befriedigende testtheoretische Basis), sowie durch strenge psychometrische Überprüfung. Hinzu kommt die Standardisierung des Frageninhalts, der Durchführung und Auswertung. Im Vergleich zu einigen anderen Verfahrenstypen ist damit die Grundlage hoher Objektivität, also geringen Einflusses subjektiver Beobachtungs- und Urteilsfehler, gegeben. Im Unterschied beispielsweise zum traditionellen Einstellungsinterview sind beim Test Durchführung und Urteilsbildung getrennt, und der Verfahrensablauf ist voll strukturiert. Tests haben v.a. im Hinblick auf die psychometrische Analyse Standards für alle anderen Verfahren gesetzt. Der Begriff «Test» wird deshalb oft auch als Überbegriff für alle eignungsdiagnostischen Verfahren verwendet, die psychometrisch überprüft wurden.

Testkompendien mit der Beschreibung und z. T. auch kritischen Bewertung veröffentlichter deutschsprachiger Tests liegen z.B. von Brickenkamp (1975, 1983) vor. Eignungstests im speziellen sind bei Schneider, Heim und Wacker (1975) und Horn (1986) zusammengestellt. In der Berufseignungsdiagnostik finden Verwendung vor allem Tests der allgemeinen Intelligenz und ihrer Komponenten, Leistungstests zur Prüfung allgemeiner Fähigkeiten wie Aufmerksamkeit und Konzentration, Tests sensorischer und motorischer Funktionen sowie Tests, die spezielle Leistungen wie technisches Verständnis erfassen. Dazu kommen Persönlichkeitstests, zu denen auch Interessen- und Motivationstests gerechnet werden, in der Mehrzahl in der Form von Fragebogenverfahren, selten als projektive Methoden. Was bisher nicht vorliegt, sind generell brauchbare Zuordnungen von Anforderungen und Testverfahren. Die Entscheidung im Einzelfall dürfte häufig zu nicht-optimalen Verfahrensanwendungen führen, vor al-

lem da für spezielle Anforderungen kaum zuverlässige Verfahren existieren. Entsprechende Defizite werden in der Erhebung von Schorr (1991) von ca. 50% der befragten Psychologen genannt.

Brambring (1983) hat Umfragen über die Anwendungshäufigkeit psychologischer Testverfahren in Unternehmen, Behörden und verwandten Einrichtungen durchgeführt. Seinen Aufstellungen zufolge gehören zu den meistverwendeten Intelligenz- und Leistungstests: Intelligenz-Struktur-Test, Leistungs-Prüf-System, Aufmerksamkeits-Belastungs-Test d2, Mechanisch-technischer Verständnis-Test, Differentieller Wissens-Test; hinsichtlich Persönlichkeitstests das Freiburger Persönlichkeitsinventar und der Berufs-Interessen-Test. In den letzten Jahren werden bevorzugt Tests zur Messung spezieller Persönlichkeits- und Fähigkeitsmerkmale entwickelt, die entsprechend geringere Verwendungshäufigkeit aufweisen (vgl. Schuler & Funke, 1991).

Daneben werden von verschiedenen Wirtschaftsunternehmen und Forschungsinstituten, vor allem aber der Bundesanstalt für Arbeit (Eckardt, 1991; Husted, 1991) und der Bundeswehr (Steege & Aschenbrenner, 1991), eigene Testverfahren entwickelt und oft bereits computergestützt angewandt. Ein Beispiel hierfür ist die Testbatterie EUB bzw. deren Nachfolgeverfahren Berufswahltest (BWT), die vom Psychologischen Dienst der Bundesanstalt für Arbeit zur Unterstützung der Berufsberatung verwendet wird (Engelbrecht, 1991). Das Verfahren besteht aus Leistungstests zur Erfassung von Einzelfähigkeiten, aus Selbstbeurteilungs- und Interessenskalen. Die Leistungsergebnisse jedes Ratsuchenden werden mit Normwerten aus 15 verschiedenen Berufsbereichen verglichen. Außerdem werden Berufsvorschläge ausgedruckt, die auf Profilvergleichen des Interessenprofils des Ratsuchenden mit den Interessen von Berufsvertretern beruhen. Beispiele für Testverfahren aus dem spezifischen Tätigkeitsfeld «Forschung und Entwicklung» zeigt Informationsbox 4.

Zur Validität von Intelligenztests im weiteren Sinne bzw. «kognitiven Fähigkeitstests»

Informationsbox 4:*Psychologische Tests für die industrielle Forschung und Entwicklung*

Aus 13 Aufgabenbereichen (siehe Informationsbox 1) wurden 28 Eigenschaftsanforderungen mit dem dekomponierenden Verfahren abgeleitet (siehe Abschnitt 2). Entsprechend wurden 15 Untertests zu Persönlichkeitsmerkmalen und 13 kognitive Untertests entwickelt bzw. adaptiert, von denen sich drei kognitive und drei Persönlichkeitstests als valide erwiesen. Die Gesamtvalidität des Auswahlsystems aus den drei kognitiven und den drei Persönlichkeitstests kombiniert mit den neun validen Arbeitsproben beträgt $R=0.48$. Dieser Wert steigt bei Ausschluß von Mitarbeitern, die ihrem Vorgesetzten weniger als zwei Jahre bekannt sind (Gefahr ungenauer Leistungsbeurteilung), auf $R=0.63$.

Beispiel für eine kognitive Testaufgabe (Einfallsflüssigkeit):

Ihre Aufgabe besteht darin, möglichst viele Hypothesen aufzustellen bzw. viele mögliche Ursachen anzugeben, die folgende Behauptung erklären könnten:

Es sei statistisch festgestellt worden, daß Menschen, die mehr Geld verdienen, sich häufig weniger glücklich fühlen. Warum?

Beispiel für ein Persönlichkeitsitem (intrinsische Leistungsmotivation):

Wenn ich arbeite, dann . . .

- schweifen meine Gedanken manchmal ab
- bin ich nicht leicht abzulenken
- arbeite ich in einem Stück durch

(allgemeine und spezielle Intelligenztests, Leistungstests usw.) liegt eine größere Zahl von Untersuchungen vor, v. a. im angloamerikanischen Raum. Als Hauptergebnis der darauf aufbauenden Studien zur Validitätsgeneralisierung bzw. Metaanalyse ergibt sich, daß es praktisch keinen Beruf gibt, für den allgemeine und spezielle Intelligenztests nicht zur Leistungsprognose beitragen könnten. Dabei lassen sich Ausbildungslei-

stungen besser vorhersagen als andere Kriterien der Berufsleistung, was nicht überraschen dürfte, da kognitive Fähigkeiten im größerem Maße Lernfähigkeit betreffen. Im Durchschnitt über alle Berufe kommen Hunter und Hunter (1984) auf eine Validität allgemeiner kognitiver Fähigkeitstests für Berufserfolg von $r = 0.45$. Bestätigt wird das Ergebnis durch den Wert von $r = 0.41$ für dieselben Tests auf Basis anderer Untersuchungen, den Hunter und Hirsh (1987) nach Anwendung der üblichen Korrekturen für Einflüsse von Streuungseinschränkung und Kriterienunreliabilität aus den Daten von Schmitt, Gooding, Noe und Kirsch (1984) errechneten. Hunter und Hunter (1984) berichten auch eine metaanalytische Reanalyse der klassischen Validitätssammlung von Ghiselli (1973), die sie um Artefakte korrigierten, und kommen je nach Berufsgruppe (Manager, Verkäufer, Handwerker, Fahrer usw.) auf Validitätskoeffizienten von $r = 0.27$ bis 0.61. Die Validität steigt dabei, je höher die Komplexität der Leistungsanforderungen ist. Den gleichen Einfluß der Aufgabenkomplexität auf die Validität der kognitiven Subtests zeigen sie am Beispiel der vom U.S. Employment Service verwendeten Testbatterie GATB (der wohl bestuntersuchten und meistverwendeten eignungsdiagnostischen Testbatterie) mit Validitäten von $r = 0.23$ bis 0.56.

In *speziellen Berufsklassen* mit spezifischen Anforderungen kann die Validität allgemeiner kognitiver Fähigkeiten vergleichsweise niedrig sein. Eine Metaanalyse eignungsdiagnostischer Studien zur Personalauswahl von Wissenschaftlern und Ingenieuren (Funke et al., 1987) ergab eine mittlere korrigierte Validität von nur $r = 0.16$. Nicht ausreichend berücksichtigt werden konnten dabei allerdings vermutete Streuungseinschränkungen aufgrund der im Ausbildungs- und Berufsweg von Wissenschaftlern und Ingenieuren impliziten Vorauswahl nach kognitiven Fähigkeiten.

Auch bezogen auf das Kriterium Ausbildungserfolg ergaben Validitätsgeneralisierungsstudien die Übertragbarkeit der Validität allgemeiner kognitiver Fähigkeitstests auf andere Berufsgruppen, denn signifikan-

te Validitätsdifferenzen treten dabei nicht auf. Im Durchschnitt über alle Berufe geben Hunter und Hunter (1984) die Validität für Ausbildungserfolg mit $r = 0.54$ an.

Niedriger als für Intelligenz- und Leistungstests ist die Validität von *Persönlichkeitstests*. Hunter und Hirsh (1987) beziffern die mittlere Validität auf $r = 0.27$ bezogen auf das Vorgesetztenurteil. Für die geringere Validität werden meist zwei Hauptgründe verantwortlich gemacht: Zu globale Konstrukte und Verfälschbarkeit der Antworten.

Zunächst wird darauf verwiesen, daß viele Persönlichkeitstests *nicht für eignungsdiagnostische Zwecke konstruiert* seien, sondern z. B. als klinisch-psychologische Diagnoseverfahren oder als persönlichkeitspsychologische Forschungsinstrumente mit dem Ziel der Messung möglichst breiter, umfassender Konstrukte (Schneider & Schmitt, 1986). Einen Hinweis hierauf ergibt die Metaanalyse von Barrick und Mount (1991), wo sich zeigt, daß Konstrukte wie Extraversion keine allgemeine Validität über alle Berufsgruppen aufweisen, sondern nur für Manager ($r = 0.18$) und Vertrieb ($r = 0.15$). Auch mit dem Gesamturteil im Assessment Center sind die allgemeinen Merkmale Extraversion und Neurotizismus (psychische Stabilität) nur gering korreliert, wohingegen die spezifischeren Merkmale Dominanz und Selbstvertrauen als Subkonstrukte der beiden generellen Merkmale hohe Korrelationen aufweisen (Scholz & Schuler, 1993).

Verbesserungen erzielt man, wenn man diese Tests durch Ausschluß nicht eignungsrelevanter Items mit Hilfe der Itemvalidierung (siehe biographische Fragebogen) situationsspezifisch anpaßt (Funke & Schuler, 1990). Oder Persönlichkeitstests werden von vornherein stärker zielorientiert konstruiert, wie dies z. B. von McHenry, Hough, Toquam, Hanson und Ashworth (1990) im Rahmen eines umfangreichen Testentwicklungsprogramms im Auftrag der US Army durchgeführt wurde. Die Validierung der regressionsstatistisch kombinierten Gesamtwerte an einer Stichprobe von 4000 Personen ergab zwar für kognitive Fähigkeiten erwartungsgemäß die höchsten Validitäten

bezogen auf aufgabenspezifische und allgemeine Berufsleistung ($R = 0.65$, $R = 0.69$). Die Kriterienbereiche Führung/Engagement, Disziplin und rollenentsprechendes militärisches Verhalten, die eher zeigen, wie gut eine Person in eine Arbeitssituation paßt, waren aber durch die Persönlichkeitsvariablen besser vorhersagbar ($R = 0.38$, 0.35 , 0.38).

Als zweites Hauptproblem wird die Verfälschbarkeit von Persönlichkeitstests in Auswahl-situationen durch positive Selbstdarstellung («impression management») genannt. Sie ist im Prinzip der Selbstbeschreibung begründet, das diesen Verfahren zugrundeliegt. Das muß nicht heißen, daß verfälschbare Tests in Auswahl-situationen auch tatsächlich verfälscht werden. Während Dunnette, McCartney, Carlson und Kirchner (1962) erhebliche Validitätseinbußen für Verkaufspersonal konstatierten, kommen Hough, Eaton, Dunnette, Kamp und McCloy (1990) anhand einer sehr großen Stichprobe im Militär (2.500 Personen mit hohen Selbstdarstellungswerten im Vergleich zu ca. 6.000 mit geringen Selbstdarstellungswerten) zu keinem Einfluß der positiven Selbstdarstellung auf die Validität.

Für projektive Tests, deren Hauptverwendung die Erfassung von Leistungs- und Machtmotiv zur Prognose von Management-erfolg ist, geben Reilly und Chao (1982) die mittlere Validität aus fünf Studien mit $r = 0.18$ an. Die Validitäten von Interessentests bezogen auf Erfolgskriterien liegen noch etwas darunter ($r = .10$). Dabei ist jedoch zu bedenken, daß Interessentests häufiger zu Beratungs- als zu Selektionszwecken Verwendung finden, und daß das hier angemessene Kriterium weniger die Berufsleistung als die Zufriedenheit mit der seinerzeit getroffenen Berufswahl ist.

3.5 Computergestützte Eignungsdiagnostik

Alle Phasen der Datenerhebung und -verarbeitung in der Eignungsdiagnostik können computergestützt erfolgen. Von computergestützter Eignungsdiagnostik (CED) wird hier aber nur gesprochen, wenn dabei die

Anwendung von (Personal-)Computern als *diagnostisches Instrument* mit direktem Probandenkontakt erfolgt. Reine Auswertungs- und Interpretationshilfen für herkömmliche Verfahren, automatische Gutachtengeneratoren u. ä. bleiben daher außer Betracht.

Die wachsende Bedeutung von CED schlägt sich in der Zahl entsprechender Verfahren nieder (vgl. Schuler & Funke, 1991). Als generelle Vorteile von CED lassen sich nennen:

- volle Standardisierung der Durchführung und Auswertung sowie Kontrolle der Zeitvorgabe für Items und Verfahrensteile;
- Möglichkeit zur Registrierung von Zusatzdaten wie Latenzzeiten, Fehlerreaktionen und Korrekturen, allerdings z.T. mit Problemen bei der Validierung und Interpretation;
- Rationalisierung bei Durchführung und Auswertung, schnellere Ergebnisverfügbarkeit, Einsparung separater Dateneingaben für spätere Analysen;
- erhöhte Schutzmöglichkeiten gegen unzulässige Verbreitung des Verfahrens (Kopierschutz);
- Akzeptanz computergestützter Verfahren bisher meist höher als die vergleichbarer Papier-Bleistift-Verfahren. Computererfahrung hat in der Mehrzahl der Studien positiven Einfluß;
- reduziertes «impression management» der Bewerber mit allerdings teilweise widersprüchlichen Ergebnissen und noch unklarer Verursachung.

Schwerpunkte der CED sind *computergestützte Tests und computergestützte Simulationen*. In jeder der beiden Kategorien ist weiter zu unterscheiden zwischen herkömmlichen Verfahren, die von Papier-Bleistift-Basis auf EDV übertragen wurden, und neukonzipierten Verfahren, die die speziellen Möglichkeiten von Computern vor allem durch das diagnostische Konzept, die Art und Präsentation der Aufgabe(n) oder deren Auswertung für die Gewinnung neuer Prädiktoren nutzen.

Die Mehrzahl der vorliegenden Verfahren sind auf Computer übertragene herkömmli-

che Tests mit Darstellung der Items auf dem Monitor (Beispiele bei Booth, 1991) und Eingabe der Antwort über Tastatur, Lichtgriffel oder Maus. Hierbei stellt sich die Frage der Äquivalenz beider Formen, also etwa der Gültigkeit der Normwerte. Neben den unvermeidlichen, technisch bedingten Durchführungsunterschieden z. B. hinsichtlich der Itembehandlungsmöglichkeiten (die Computerversion erlaubt meist keine Überblicksgewinnung über alle Items, keine individuelle Bearbeitungsreihenfolge, kein Zurückstellen usw.) z.B. bei Zeitdruck (aufgrund von Ergonomieunterschieden der Formen), wie auch aufgrund unterschiedlicher PC-Vorerfahrung dürfte die Äquivalenz umso geringer werden, je mehr die computergestützte Form durch komfortable Instruktionsmethoden, Menütechniken, Farbdarstellung usw. gegenüber der Urform aufgewertet wird (Jäger, 1990). Dennoch gehen Überblicksarbeiten (Davis & Cowles, 1989; Klieme & Stumpf, 1990) von einer weitgehenden Äquivalenz beider Formen aus.

Eine weitergehende Nutzung der von Computern bereitgestellten Möglichkeiten wird durch speziell konstruierte adaptive Tests angestrebt. Hier geht es hauptsächlich um eine Optimierung der Itemauswahl. Das Grundprinzip (Einführung z. B. bei Kisser, 1988) besteht darin, der Person nur die für sie angemessenen Items zu präsentieren, sie also im Falle eines Fähigkeitstests weder mit zu einfachen Aufgaben zu langweilen, noch mit zu schwierigen zu überfordern. Als theoretische Basis muß dem adaptiven Test in der Regel eine probabilistische Testtheorie zugrundegelegt sein, um die Vergleichbarkeit der Ergebnisse zu ermöglichen, obwohl Testpersonen unterschiedliche Itemzusammenstellungen vorgegeben wurden. Die Realisierung in der Praxis wird aufgrund des Aufwands - u. a. ist ein großer und homogener Itempool erforderlich - auf Großorganisationen beschränkt bleiben, wobei Wildgrube (1990) für die bisherigen Erprobungen und Realisierungsmöglichkeiten im Rahmen der Bundeswehr, die hier eine Vorreiterrolle spielt, zu einer eher kritischen Zwischenbilanz kommt.

Computergestützte Versionen herkömmlicher Simulationsaufgaben bzw. Arbeitsproben sind ebenfalls entwickelt worden. Als Beispiele können Versionen der als Assessment Center-Aufgabe verbreiteten Postkorbübung (z.B. Roest & Horn, 1990) sowie der Dispositionsaufgabe (z. B. Routen-Planung des Handlungsreisenden) gelten. Neben den allgemeinen Vorteilen der CED kann in diesen PC-Versionen der Ablauf der Aufgabe durch laufende Rechnerreaktionen auf Teilnehmer-Entscheidungen, durch zusätzlich eingeführte Ereignisse usw. dynamisch gestaltet werden. Validitätsdaten wurden zu diesen Verfahren bisher nicht veröffentlicht. Betriebswirtschaftliche Simulationen (business games) sind ein weiteres Beispiel. Sie weisen nach bisherigen Untersuchungen kaum Validität für das Kriterium Karriereerfolg auf (Kluwe, 1990), wobei zu beachten ist, daß sie Fachwissen voraussetzen.

Aus der kognitionspsychologischen Grundlagenforschung stammende computergestützte Simulationen komplexer Systeme werden inzwischen eignungsdiagnostisch erprobt. Zur Begründung wird meist global auf die Ähnlichkeit der simulierten Problemmerkmale (Komplexität, Dynamik, Vernetztheit, Intransparenz der beteiligten Einflußgrößen) mit Anforderungen z. B. im Management-Bereich verwiesen. Der Nachweis hierfür in Form von Anforderungsanalysen liegt bisher nur in Ausnahmefällen vor (siehe Informationsbox 1). Diagnostiziert werden sollen Leistungen oder Strategien der «operativen Intelligenz» (Dörner, 1986) von Problemlösekompetenz oder heuristischem Wissen. Eine Charakterisierung verschiedener Simulationen, insbesondere des verbreiteten Systems «Schneiderwerkstatt/Tailorshop» findet sich in J. Funke (1986). Einige Aspekte verdienen für den eignungsdiagnostischen Einsatz verstärkte Beachtung:

- Wie realitätsnah sollen Simulationen sein, wenn es um eher allgemeine, übertragbare Problemlösekompetenz geht, nicht aber um Fach- und Spezialwissen? Ist die Simulation kognitiver Anforderungen von komplexen Problemlöse- und Ent-

scheidungssituationen weitgehend fach-/vorwissensunabhängig möglich?

- Bisher übliche grobe Maße für die Steuerleistung («Systemendzustand») sollten ergänzt werden durch Strategiemasse für das *Verhalten* der Problemlöser in der Problemsituation (Putz-Osterloh, 1991). Hierzu sind möglichst ökonomische Strategieregistrierungen nötig.
- Die Konstruktvalidität der Systeme ist weiter klärungsbedürftig, z. B. die widersprüchlichen Ergebnisse zum Zusammenhang der Steuerleistung mit den Strategien, mit Testintelligenz und mit Vorwissen.
- Es fehlen aussagekräftige Studien zur kriterienbezogenen Validität. Vorliegende Experten-Novizen-Vergleiche der Steuerleistung und der Strategien (z. B. Manager vs. Studenten) zeigen zwar in der Regel eine Überlegenheit der Problemlöse-Experten, die Validierung sollte aber an Berufserfolgskriterien erfolgen. Hohe Validitätsdaten für Strategien in einer zweitägigen Simulation bezogen auf subjektive Kriterien (Selbsteinschätzungen der Teilnehmer) wurden von Streufert, Pogash und Piasecki (1988) berichtet. Erste Validitätsdaten einer knapp einstündigen Simulation gemessen an Vorgesetztenurteilen zur beruflichen Problemlöseleistung berichtete Funke (1991); vgl. Informationsbox 5.

3.6 Arbeitsproben

Unter Arbeitsproben werden standardisierte Aufgaben verstanden, die inhaltlich valide und erkennbar äquivalente Stichproben des erfolgsrelevanten beruflichen Verhaltens darstellen. Ihre Entwicklung geht auf die deutsche Eignungsdiagnostik der zwanziger Jahre zurück (Giese, 1924). Die Abgrenzung von Arbeitsproben gegenüber Tests ist schwierig und wird uneinheitlich gehandhabt: Gelegentlich wird von Arbeitsprobe nur dann gesprochen, wenn es sich um motorische Aufgaben handelt, häufig werden Arbeitsproben als Tests bezeichnet, sobald sie in standardisierter und normierter

meter erhoben (z. B. Informationsbedarf oder Zahl der benötigten Versuche; s. hierzu auch Computersimulationen im vorangegangenen Abschnitt) oder ein Leistungsmaß nach Abschluß der Trainingsphase. Durch diese Verfahrensweise sind Trainierbarkeitsverfahren auch geeignet für Personen ohne Vorerfahrung mit der betreffenden Tätigkeit (Robertson & Downs, 1989).

Für psychometrisch fundierte Arbeitsproben werden relativ hohe Validitätskoeffizienten berichtet. Die Metaanalyse von Schmitt et al. (1984) resultiert in einem durchschnittlichen Validitätskoeffizienten von $r = 0.38$, Hunter und Hunter (1984) kommen gar auf $r = 0.54$. Die errechneten Validitätswerte kovariieren u. a. mit der Art der Arbeitsproben und der Art der Kriterien. So werden für «motorische Arbeitsproben» höhere Werte berichtet als für «verbale Arbeitsproben», wenn am Kriterium Vorgesetzten-Beurteilung gemessen wird; umgekehrt ermöglichen verbale Arbeitsproben bessere Prognosen, wenn als Kriterium die Ausbildungsleistung gewählt wird (Robertson & Kandola, 1982).

Neben der relativ hohen Validität wird die Akzeptanz von Arbeitsproben seitens der Betroffenen hervorgehoben (Robertson & Kandola, 1982) - vermutlich aufgrund ihrer Augenscheinigkeit sowie der in ihnen enthaltenen Information über Arbeitsanforderungen. Dies scheint der Selbstselektion entgegenzukommen: In einer Studie von Downs, Farr und Colbeck (1978) wurde allen Bewerbern im Anschluß an eine Arbeitsprobe die Anstellung zur Probe angeboten; von den in der Arbeitsprobe Leistungsstärksten nahmen 91% das Angebot an, von den Leistungsschwächsten nur 23 %. Auch konnte in mehreren Untersuchungen gezeigt werden, daß die Ergebnisse für Angehörige von Minoritätengruppen relativ günstiger sind als beim Einsatz von Fähigkeitstests (Cascio, 1987).

Den Vorzügen von Arbeitsproben steht ein relativ hoher Aufwand bei der Konstruktion gegenüber, wenn sie anforderungsbezogen erstellt werden. Oft ist die Zahl der Items geringer als bei Tests und ihre Unabhängigkeit nur beschränkt gegeben (z. B. bei praktisch-

technischen Montageaufgaben). Geht man von der Richtigkeit der theoretischen Argumentation ihrer Proponenten aus, so müßte die Generalisierbarkeit der Validität von Arbeitsproben wesentlich geringer sein als die von Fähigkeitstests. Auch der Vorzug der Augenscheinigkeit beschränkt sich auf die Verwendung als Simulation der in Frage stehenden Tätigkeit. Darüber hinaus ist mit einer rascheren Verminderung der Validitätskoeffizienten im zeitlichen Verlauf zu rechnen. Die Metaanalyse von Robertson und Downs (1989) bestätigt diese Vermutung für Arbeitsproben, die als Trainierbarkeitstests ausgelegt sind.

Neuerdings werden Arbeitsproben vielfach zur Simulation von Führungsaufgaben eingesetzt, zumeist in der multiplen Form des Assessment Centers (vgl. den nächsten Abschnitt). Für drei der verbreitetsten Einzelverfahren - führerlose Gruppendiskussion, Postkorbübung und Wirtschaftsspiele (Business Games) - stellt Cascio (1987) Bewährungsdaten zusammen, die durchschnittlich bei etwa $r = 0.25$ bis 0.30 liegen. Die Konstruktion eines kompletten Auswahlverfahrens auf Arbeitsprobenbasis demonstrieren Schmitt und Ostroff (1986).

3.7 Assessment Center

Die Diagnosemethode «Assessment Center» nimmt insofern eine Sonderstellung ein, als es sich hierbei um eine multiple Verfahrenstechnik handelt, zu der mehrere eignungsdiagnostische Instrumente oder leistungsrelevante Aufgaben zusammengestellt werden (monographische Darstellungen bei Lattmann, 1989; Schuler & Stehle, 1992; Thornton & Byham, 1982). Ihr Einsatzbereich ist die Einschätzung aktueller Kompetenzen oder die Prognose künftiger beruflicher Entwicklung und Bewährung, sie wird deshalb sowohl zur Auswahl künftiger Mitarbeiter wie auch als organisationales Beurteilungs- und Förderungsinstrument eingesetzt. Charakteristisch für Assessment Centers ist, daß mehrere Personen (etwa 6-12) gleichzeitig als Beurteilte daran teilnehmen und daß auch die Einschätzung

gen von mehreren unabhängigen Beurteilern (im Verhältnis etwa 1 : 2 zur Zahl der Beurteilten) vorgenommen werden, darunter vor allem Linienvorgesetzte (typischerweise zwei Hierarchieebenen über der Zielebene der zu Beurteilenden) sowie Psychologen und Mitarbeiter der Personalabteilung.

Vorläufer des Assessment Centers gab es zur Zeit der Weimarer Republik in der Offiziersauswahl der deutschen Streitkräfte, später auch der britischen Armee und des amerikanischen Nachrichtendienstes. Bestandteile der militärischen Beurteilungsprogramme waren neben individuellen und gruppenbezogenen Arbeitsproben auch Interviews, Intelligenz- und Persönlichkeitstests, biographische Daten sowie die Einstellung zur eigenen Biographie. Eine noch größere Verfahrensvielfalt findet sich in der großen klassischen Assessment Center-Untersuchung, der Management Progress Study, die ab 1956 in der American Telephone and Telegraph Company (AT&T) durchgeführt wurde (Bray, Campbell & Grant, 1974). Damals wurden 422 bereits beim Unternehmen beschäftigte Nachwuchs-Führungskräfte mit einer großen Anzahl psychologischer Tests untersucht und mit Aufgaben konfrontiert, die auch heute noch zum Assessment-Standardrepertoire zählen, darunter Postkorb, Wirtschaftsspiel und führerlose Gruppendiskussion.

Die Ergebnisse wurden mit dem späteren Karriereerfolg der Kandidaten verglichen und zeigten hohe Vorhersageleistungen des Gesamtverfahrens, wobei die prognostische Validität vor allem auf die Arbeitsproben und die kognitiven Leistungstests zurückging, während Persönlichkeitstests und Interviews nur einen geringen Beitrag leisteten.

Die heutige vielfältige Verwendung des Assessment Centers beruft sich nicht immer zu Recht auf das berühmte Vorbild-wenn Verfahrensvielfalt und Sorgfalt der Durchführung nicht dem Standard des Modells entsprechen, ist auch die dort gefundene hohe prognostische Validität nicht zu erwarten. Thornton, Gaugler, Rosenthal und Bentson (1992) errechneten mittels metaanalytischer

Zusammenfassung von 107 Validitätswerten aus 50 Studien einen Durchschnittswert von $r = 0.37$ bei einer Streubreite von $r = -0.25$ bis $r = +0.78$. Die Metaanalyse von Schmitt et al. (1984) ergab eine mittlere Validität von $r = 0.40$.

An Vorteilen dieser Methode verspricht man sich jedoch nicht nur und nicht in allen Fällen eine bessere Vorhersagemöglichkeit beruflicher Bewährung. Vorteile des Verfahrens werden auch in «latenten» Funktionen gesehen, wie im Gewinnen eines Überblicks über den Nachwuchs, über Leistungsstand und Defizite im Unternehmen (und zwar nicht nur im Hinblick auf Personen, sondern auch auf Organisationseinheiten, Programme, Führungsstile etc.), in der Gelegenheit, ein gemeinsames Verständnis von Anforderungen und Leistungen zu entwickeln und dafür operationale verhaltensbezogene Formulierungen zu finden, in der Betonung der Bedeutung von Personalplanung und Personalentwicklung, in der Möglichkeit, Aspekte der «Unternehmenskultur» zu diskutieren und zu inszenieren, die Teilnehmer mit den Anforderungen - auch sozialpsychologischer Art - einer Führungstätigkeit vertraut zu machen, ihre Selbsteinschätzung zu verbessern und ihnen die Gelegenheit zum sozialen Vergleich zu bieten. Schließlich scheint die Aufgabe des Beobachtens im Assessment Center nicht nur, wie schon länger vermutet, ein gutes Beurteilertraining darzustellen, sondern sogar der Erfüllung weiterer Aufgaben einer Führungskraft dienlich zu sein (Lorenzo, 1984) und überdies deren Selbstverständnis entgegenzukommen. Die wichtigste «manifeste» Zielsetzung des Assessment Centers ist die Auswahl oder Förderung von Führungskräften. Aber auch zur Auswahl und Entwicklung von Personen für eine Vielzahl anderer Tätigkeiten wurde die Methode bereits eingesetzt, darunter Verkäufer und Trainees, Auszubildende, Beamte und in Einzelfällen auch Journalisten und Studienplatzbewerber. In den USA werden häufig auch Polizisten, Schulleiter und Feuerwehrleute mit dieser Methode ausgewählt.

Eine Vielzahl von Einzelaufgaben wurde im Assessment Center-Kontext entwickelt oder

eingesetzt. Eine Zusammenstellung der wichtigsten im Assessment Center gebräuchlichen Einzelverfahren findet sich in Tabelle 3.

Die zwei am häufigsten eingesetzten Aufgaben gehören zu den beiden erstgenannten Kategorien; es sind die Postkorbübung und die führerlose Gruppendiskussion. Diese beiden Aufgabenarten weisen bereits relativ gute Einzelvaliditäten auf. Da Assessment Center vorwiegend zur sogenannten Potentialbeurteilung herangezogen werden, also zur Einschätzung weiterer Einsatzmöglichkeiten und künftiger Leistungen, ist für die Verfahrenselemente charakteristisch, daß sie Verhaltensstichproben erfassen, die während der bisherigen Tätigkeit nicht in gleicher Weise beobachtbar waren.

Die Leistungen der Teilnehmer im Assessment Center werden von den Beurteilern üblicherweise - soweit es sich nicht um Testverfahren oder andere objektiv auswertbare Instrumente handelt - anhand vorgegebener Skalen eingeschätzt, gewöhnlich in Form von Einstufungen oder Checklisten, gelegentlich auch der Rangreihenbildung. Wie generell bei der Leistungsbeurteilung, sind die Registrierungen der Beurteiler Beschreibungen oder Interpretationen des Verhaltens, Nennungen von Verhaltensergebnissen oder Zuschreibungen von Fähigkeiten und anderen Eigenschaften. Die Urteilkonkordanz der Beobachter liegt üblicherweise zwischen $r = 0.50$ und 0.90 . Die Einzelurteile werden entweder rechnerisch aggregiert oder in einer Beurteilerkonferenz

diskutiert, um Gesamteinschätzungen festzulegen, die Grundlage des häufig im Anschluß geführten Feedback-Gesprächs sind. Vergleiche zeigen, daß von der Diskussion der Beurteilungen keine Verbesserung ihrer Validität zu erwarten ist (Sackett & Wilson, 1982). Andere Funktionen des gemeinsamen Urteilsprozesses (s. die oben genannten «latenten Funktionen») sind allerdings durch die statistische Datenkombination nicht ersetzbar.

Die Vielfalt der als Beschreibungsdimensionen verwendeten Begriffe ist groß, wobei die prognostische Validität hiervon weitgehend unbeeinflusst zu sein scheint. Untersuchungen zur Konstruktvalidität von Assessment Center-Beurteilungen haben gezeigt, daß die Beobachter in ihrer Einstufung der verschiedenen Dimensionen innerhalb einer Aufgabe nicht wirklich differenzieren - oder, anders gesagt, daß die angebotenen Beschreibungsbegriffe menschlichen Urteilsprozessen nicht wirklich angemessen sind. Hier findet sich wieder, was die Forschung zur Leistungsbeurteilung gezeigt hat, daß nämlich Beurteilungen gewöhnlich in sehr globalen Kategorien vorgenommen werden, innerhalb derer sich Beschreibung von Bewertung nicht trennen läßt. Dementsprechend ergab eine Metaanalyse (Scholz & Schuler, 1993), deren Ziel die Aufklärung des Gesamturteils im Assessment Center war, daß Intelligenz, soziale Kompetenz, Leistungsmotivation, Dominanz und Selbstvertrauen die deutlichsten persönlichkeitspsychologischen Korrelate des Abschnei-

Tabelle 3: Im Assessment Center gebräuchliche Einzelverfahren.

-
- Individuell auszuführende Arbeitsproben und Aufgabensimulationen (v. a. Organisations-, Planungs-, Entscheidungs-, Controlling- und Analyseaufgaben)
 - Gruppendiskussionen mit und ohne Rollenvorgabe
 - Sonstige Gruppenaufgaben mit Wettbewerbs- und/oder Kooperationscharakteristik
 - Vorträge und Präsentationen
 - Rollenspiele (meist dyadisch, z. B. Einstellungsinterview, Verkaufsgespräch)
 - Einzel- und Gruppeninterviews
 - Selbstvorstellung
 - Wirtschaftsspiele
 - Fähigkeits- und Leistungstests
 - Persönlichkeits- und Interessentests
 - Biographische Fragebogen
 - Lockerungsübungen
-

dens im Assessment Center darstellen. Diese Merkmalsliste kommt den in der Einleitung zu diesem Kapitel genannten Determinanten des Berufserfolgs sehr nahe.

Thornton et al. (1992) konnten in ihrer Metaanalyse aller bisher durchgeführten Untersuchungen einige Aspekte finden, deren Beachtung eine Validitätsverbesserung erwarten läßt. Hierzu gehören die Verwendung einer möglichst großen Zahl von Aufgaben (die Unterschiedliches erfassen) und die Beteiligung von Psychologen als Beurteiler. Auch die Einbeziehung gegenseitiger Beurteilungen von Teilnehmern hat sich als validitätsfördernd erwiesen, ist allerdings gegen mögliche Nachteile abzuwägen; zweckmäßigerweise wird man sich hierbei auf Positivnennungen beschränken.

Verbesserungen des Verfahrens sind häufig auch beim sogenannten Feedback-Gespräch möglich. Nach der Durchführung und Auswertung aller Aufgaben wird üblicherweise mit jedem Teilnehmer ein Gespräch über sein Abschneiden und eventuell auch bereits über künftige Entwicklungsmöglichkeiten oder Aufgaben geführt. Dieses Gespräch sollte, entsprechend den generellen Prinzipien des Leistungsfeedbacks (Farr, 1993), gleichzeitig so informativ und rücksichtsvoll geführt werden, daß die Teilnehmer die Möglichkeit haben, ihr Selbstbild mit den Eindrücken kompetenter anderer zu vergleichen, und daß in offener partnerschaftlicher Haltung mit ihnen mögliche Konsequenzen für die künftige Arbeitsgestaltung erörtert werden können. Dabei sollte nicht vergessen werden, was bei jeder Diskussion unter den Beurteilern sehr deutlich wird, daß es sich bei den gesammelten Eindrücken um Hinweise mittlerer, aber keinesfalls perfekter Validität handelt.

3.8 Weitere Verfahren

über die im einzelnen besprochenen Methoden hinaus wird eine ganze Reihe weiterer Quellen genutzt, um tatsächlich oder auch nur vermeintlich erfolgsrelevante Informationen über Personen zu sammeln. Dazu gehört beispielsweise die *Probezeit*,

die, wo durchführbar, einen der validesten Prädiktoren späteren Berufserfolgs darstellt und auch den Bewerbern Einblick in Tätigkeit und Organisation gibt. Ihr diagnostischer Charakter als zeitlich ausgedehnte Arbeitsprobe kombiniert mit Lerntestaspekten kommt bei entsprechender Planung der übertragenen Aufgaben besonders zum Tragen. Die Validität von Leistungsdaten in der Probezeit wird von Hunter und Hunter (1984) mit $r = 0.44$ angegeben.

Für Karriereentscheidungen wird häufig auf *Leistungsbeurteilungen* bisherigen Berufserfolgs zurückgegriffen, deren Validität etwa in gleicher Höhe liegt. Sie werden im nachfolgenden Abschnitt ausführlicher besprochen. Für die Brauchbarkeit von *Selbstbeurteilungen* konnten eine Reihe von Einflußfaktoren nachgewiesen werden, darunter Situationseinflüsse wie die Art des Arbeitsplatzes, Ziele, Vertraulichkeit und Überprüfbarkeit der Selbstbeurteilung sowie Merkmale der Person (Ausprägung von Selbstachtung, Intelligenz und internaler Kontrollüberzeugung). Für Auswahl-situationen wird die Validität von Selbstbeurteilungen von Reilly und Chao (1982) mit $r = .015$ angegeben.

Zur Brauchbarkeit *ausdrucksdiagnostischer Methoden* gibt es kaum methodisch akzeptable Untersuchungen. Trotz fehlender Validitätsnachweise haben Empfehlungen zur Deutung von Mimik, Gestik, Körperhaltung und physiognomischen Merkmalen Eingang in «Bewerberberater» wie auch «Auswahl-ratgeber» für Einstellungsgespräche gefunden. Von wissenschaftlicher Seite fand dieses Gebiet seit den Tagen der älteren Ausdruckspsychologie keine Beachtung (vgl. allerdings Brandstätter, Kunkel & v. Rosenstiel, 1968). Erst in neuester Zeit mehrten sich wieder die Hinweise darauf, daß hier möglicherweise eine ergiebige Erkenntnisquelle vernachlässigt wurde. So fanden Albricht, Kenny und Malloy (1988) mittelhohe Übereinstimmungen zwischen Selbst- und Fremdbeurteilungen, wobei letztere allein aufgrund kurzer Beobachtung des Verhaltens und äußerer Merkmale zustande gekommen waren. Eine ähnliche Untersuchung mit bestätigendem Ergebnis wurde

von Borkenau (1991) durchgeführt. Auch wenn hierfür keine Belege vorliegen, ist nicht auszuschließen, daß ein erheblicher Anteil (laien-)diagnostischer Prozesse (z.B. im Assessment Center) in intuitiver Ausdrucksdeutung besteht.

Nicht besser ist es um die wissenschaftliche Legitimation der *Graphologie* bestellt. Klimoski und Raffaelli (1983) referieren die geringe Zahl von Untersuchungen, die zu dieser Frage vorliegen. Mit einer einzigen Ausnahme geben diese Arbeiten, deren methodische Qualität allerdings größtenteils zu wünschen übrig läßt, keine Hinweise auf Prognoseleistungen von überzufälliger Trefferrhäufigkeit. Auch hinsichtlich Konstruktvalidität zeigen sich keine systematischen Zusammenhänge zwischen Handschriftmerkmalen und Persönlichkeitsmerkmalen des Schreibers (Heinze, 1990). Als eine Ursache für die Validitätsüberschätzung der Graphologie konnten Ben-Shakhar, Bar-Hillel, Bilu, Ben-Abba und Flug (1986) den Einfluß «harter Fakten» in handgeschriebenen Lebensläufen demonstrieren. Nichtgraphologen (klinische Psychologen) kamen zu ebenso validen Eigenschaftsratings über die Schreiber wie Graphologen (Korrelation jeweils $r = 0.20$ mit Vorgesetztenurteilen), aber beide wurden übertroffen durch ein statistisches, lineares Modell, in das die harten (biographischen) Fakten der Lebensläufe eingegeben wurden. Weitere Ursachen möglicher Validitätsüberschätzung sind zu sehen in der unabweisbaren Individualität und Stabilität der Handschriftenmerkmale und deren fälschlicher Gleichsetzung mit Validität und schließlich in der Grundrate, da bei weitgehender Vorselektion (alle Bewerber geeignet) selbst Zufallsauswahl erfolgreich ist.

4. Leistungsbeurteilung

Einschätzungen des beruflichen Erfolgs werden in den meisten Arbeitsorganisationen vorgenommen, auch wenn nicht immer förmliche Beurteilungsmethoden dazu verwendet werden. Im Kontext der Auswahl

von Mitarbeitern besteht die primäre Funktion von Leistungsdaten darin, Kriterium für die Angemessenheit der getroffenen Personalentscheidungen zu sein. Speziell in diesem Fall ist das Prinzip einleuchtend, sowohl die Auswahlverfahren als auch die Beurteilungskriterien aus der Analyse der Anforderungen abzuleiten - man kann die Arbeits- und Anforderungsanalyse geradezu als eine Theorie der Leistung in der betreffenden Tätigkeit ansehen. Schon Thorndike (1949) allerdings hat festgestellt, daß das schwierigste Problem der eignungsdiagnostischen Forschung darin besteht, befriedigende Maße zu finden, die als Kriterien der Validierung von Auswahlverfahren dienen.

4.1 Funktionen von Leistungsbeurteilungen

Leistungsdaten können aber auch vielfältige andere Funktionen erfüllen. Eine Zusammenstellung der wichtigsten Verwendungszwecke gibt Tabelle 4.

Entsprechend ihrer spezifischen Funktion haben Leistungsdaten unterschiedlichen Ansprüchen zu genügen: Werden beispielsweise Beurteilungen als Kriterien für die Validierung von Auswahlverfahren verwendet, so müssen sie ihrerseits valide sein und quantitative Differenzierungen zwischen den Beurteilten ermöglichen. Dienen sie der Personalentwicklung, steht die Erkennbarkeit individueller Stärken und Schwächen im Vordergrund. Für die Bemessung von Gehältern und Zulagen ist besonders die Lösung des Attributionsproblems wichtig, also die Beantwortung der Frage, inwieweit ein beobachtetes Ergebnis von der Person selbst zu verantworten bzw. auf andere Einflußgrößen zurückzuführen ist. Steht der Orientierungscharakter der Beurteilung im Vordergrund, so ist auf den Anforderungsbezug zu achten, und wird Verhaltenssteuerung bezweckt, kommt der Gestaltung der rückkoppelnden Leistungsinformation (Feedback) besondere Bedeutung zu.

Nicht alle dieser Zielsetzungen sind untereinander kompatibel und mittels der gleichen Methoden zu erreichen; einige haben sich sogar als antagonistisch erwiesen; z.B.

Tabelle 4: Funktionen der Leistungsbeurteilung (modifiziert aus Schuler, 1991b, S. 14).

1. Leistungsverbesserung durch Verhaltenssteuerung (Feedback für die Beurteilten)
2. Planung, Auswahl und Gestaltung von Maßnahmen der Personalentwicklung
 - individuell: Maßnahmen zur Erhöhung der individuellen Bewährungswahrscheinlichkeit und Einsatzbreite (u.a. Aufgabenerweiterung, Verhaltenstraining, Fort- und Weiterbildung)
 - kollektiv: Personalentwicklungsplanung
3. Personelle Entscheidungen auf individuellem und kollektivem Niveau
 - individuell: Plazierung, Beförderung, Versetzung, Übernahme, Kündigung
 - kollektiv: Personalplanung
4. Gestaltung von Arbeitsbedingungen (Arbeitsplatz und Arbeitsumgebung), Ausgangspunkt von Organisationsdiagnose und Organisationsentwicklung
5. Gehalts- und Lohnbestimmung
6. Verbesserung der Führungskompetenz der Vorgesetzten
7. Evaluation von Selektionskonzepten, personellen Entscheidungen, Maßnahmen der Personalentwicklung, Programmen der Organisationsentwicklung, Anreiz- und Verstärkungssystemen
8. Artikulation von Anforderungen an Arbeitstätigkeit und soziales Verhalten (Leistungsbeurteilung als Instrument der betrieblichen Sozialisation, der Motivierung und Verhaltenssteuerung)
9. Hervorhebung der Bedeutung leistungsorientierter Personalentwicklung in der Organisation

ist eine gehaltsbezogene Beurteilung im allgemeinen nicht gleichzeitig als Instrument der Personalentwicklung verwendbar (weil es dann im Interesse des Beurteilten liegt, seinen Entwicklungsbedarf - zumindest den «on the job» - herunterzuspielen), und eine Beurteilung, bei der individuelle Aspekte der Beratung und Unterstützung im Vordergrund stehen, wird in vielen Fällen keine Information liefern, die Vergleichbarkeit über Organisationseinheiten hinweg gewährleistet.

Die aufgelisteten Funktionen zeigen, daß die Attribution mangelnder Leistungen keineswegs zwangsläufig der betreffenden Person gelten muß, sondern daß beispielsweise auch die Konsequenz gezogen werden kann, Arbeitsbedingungen zu verbessern. Ebenso wenig wie der Begriff der Leistung an Verhaltensergebnisse einzelner Personen gebunden ist, ist die Leistungsbeurteilung auf die Feststellung dieser Ergebnisse durch eine hierarchisch übergeordnete Instanz beschränkt. Leistung kann vielmehr auf jedem Niveau der Kollektivität erfaßt werden - als Beitrag einer Organisationseinheit zu den Zielen der Gesamtorganisation -, und als Methoden der Leistungsbeurteilung kommen praktisch alle Verfahren der Evaluation empirisch erfassbarer Sachverhalte in Betracht.

4.2 Arten von Kriterien

Zur Klassifikation von Kriterien wurde eine Vielfalt von Kategorien vorgeschlagen (vgl. Schuler, 1989b). Wichtig im Sinne der Funktionalität, der Begründbarkeit und der Akzeptabilität (vgl. Abschnitt 6) ist die Unterscheidung der Beschreibungsebenen Person, Verhalten, Ergebnisse (Campbell, Dunnette, Lawler & Weick, 1970).

Ergebnisorientierte Beurteilungsmethoden werden oft besonders befürwortet und etwa von jedem zweiten (amerikanischen) Unternehmen tatsächlich eingesetzt (Lazer & Wikstrom, 1977). Ergebniskriterien können in Verkaufs- oder Umsatzzahlen bestehen, in Produktionseinheiten, Fluktuationsdaten und Fehlzeiten, Ausschußzahlen, Reklamationen, Marktanteilen oder auch in Beurteilungen des Organisations- bzw. Arbeitsklimas durch die unterstellten Mitarbeiter. Häufig ist der Bezug zur Arbeitsanalyse gering, die Formulierung der Kriterien orientiert sich vielmehr an den angestrebten Zielen der Organisation oder der Arbeitseinheit. Ergebnisorientierte Beurteilungen kommen deshalb vor allem dort zum Einsatz, wo eine Vielfalt unterschiedlichen Verhaltens zum Erfolg führen kann, wo Verhaltensbeobachtungen nicht möglich sind oder wo verhaltensbezogene Intervention unan-

gemessen scheint. So besteht z.B. die Führungsmethode Management by Objectives (MBO), die besonders häufig bei Führungskräften Verwendung findet (Odiorne, 1965), in der Orientierung und Messung der Leistung an individuell vereinbarten Zielen.

Zu den Vorteilen ergebnisbezogener Beurteilungen zählen die deutliche Repräsentation der eigentlichen Leistungsziele, der Verhaltensspielraum (es mag mehrere Wege zum Leistungsziel geben, die von Mitarbeitern individuell genutzt werden können) und die Anwendbarkeit sowohl auf individuelle wie auf Gruppen- und Unternehmensziele. Probleme bestehen u.a. in der Schwierigkeit der individuellen Zurechnung der Ergebnisse, in der Überbewertung des Zähl- und Meßbaren gegenüber weniger konkretisierbaren Zielsetzungen sowie in unzulänglicher Steuerungsmöglichkeit bei negativen Zielabweichungen: Wie mehrfach belegt wurde (Locke & Latham, 1990), sind verhaltensbezogene Zielsetzungen ergebnisbezogenen im Informationsgehalt und in der motivationalen Wirkung überlegen. Reliabilitätsprobleme ergeben sich, wenn man seltene Ereignisse (z. B. Arbeitsunfälle) als Ergebniskriterien verwendet.

Verhaltensbezogene Beurteilungen werden ebenfalls allgemein als wichtig erachtet (Eichel & Bender, 1984) und sind weit verbreitet. Viele Verfahren der Arbeitsanalyse erlauben die direkte Ableitung von Verhaltenskriterien aus der Beschreibung der Tätigkeiten. Sie können als Kernbereich der Leistungsbeurteilung angesehen werden, setzen allerdings die Kenntnis von Verhaltens-Erfolgs-Zusammenhängen voraus.

Als Vorteil verhaltensbezogener Beurteilungen wird vor allem ihr Informationsgehalt für den Beurteilten und damit ihr Bezug zur Verhaltenssteuerung und Leistungsverbesserung gesehen, zumal Verhaltenseinheiten unterschiedlichen Umfangs und Konkretisierungsgrades betrachtet werden können. Schwierigkeiten können in der tatsächlichen Beobachtbarkeit des Verhaltens liegen, in seinem Bezug zu den eigentlichen Zielen und in der Einengung des Verhaltensspielraums. Nicht immer wird wirklich Verhalten beurteilt, wenn Beurteiler verhal-

tensbezogene Urteile abzugeben glauben; Beurteiler neigen vielmehr zu globalen Eigenschaftseinschätzungen (Ilgen & Schneider, 1991).

Eigenschaftsbezogene Beurteilungen werden vornehmlich dort vorgenommen, wo Verhalten nicht beobachtbar und Ziele nicht erfassbar sind oder wo relativ anforderungsunabhängiges «Potential» festgestellt werden soll. Beispielsweise beziehen sich Einschätzungen aus Anlaß von Karriereentscheidungen üblicherweise auf Fähigkeiten und andere als relativ stabil angenommene Personmerkmale (also Verhaltensdispositionen in bezug auf breite Situationsklassen).

Als Vorteil eigenschaftsbezogener Urteile gilt besonders ihr hoher Allgemeinheitsgrad und damit ihre tatsächliche oder vermeintliche Relevanz für eine Vielfalt künftiger Anforderungssituationen und ihre Verwendbarkeit für eine große Zahl unterschiedlicher Arbeitsplätze. Nachteilig ist die hohe Urteilssubjektivität, die Exzessivität (Urteile sind nicht auf leistungsrelevante Aspekte beschränkt) und die geringe Tauglichkeit eigenschaftsbezogener Beurteilungen für Feedback und Verhaltenssteuerung.

4.3 Beurteilungsverfahren

Zur Einschätzung von Leistung auf den Beschreibungsebenen Person - Verhalten - Ergebnisse wird eine Vielfalt formeller und informeller Techniken eingesetzt. Die meisten der formellen Beurteilungsverfahren sind einer der folgenden Methoden zuordenbar.

Freie Eindruckschilderungen

Besonders verbreitet zur Beurteilung von Führungskräften-teilweise in Kombination mit Zielsetzungsverfahren - ist die Methode der freien Eindruckschilderung. Ohne oder mit geringen formalen Vorgaben werden dabei Eindrücke über erfolgsrelevantes Verhalten oder über entsprechende Eigenschaften und Kompetenzen in freien Worten festgehalten. Der Einsatz freier Eindruckschilderungen ist vor allem dort vertretbar, wo es nicht entscheidend auf die Vergleich-

barkeit von Beurteilungsergebnissen ankommt, etwa als Grundlage von Fördergesprächen. Um ein Mindestmaß an Vergleichbarkeit zwischen verschiedenen Beurteilern sicherzustellen, können Urteilsaspekte oder Begriffslisten als Auswahlhilfe angeboten werden. Mitunter, z.B. in Bereichen des öffentlichen Dienstes, ist eine freie Eindruckschilderung als Ergänzung gebundener Skalierungen vorgesehen.

Vorteile freier Eindruckschilderungen liegen im geringen Konstruktionsaufwand, in der Möglichkeit individueller Charakterisierung einer Person bzw. ihrer Leistungsergebnisse und in der Anpassung der Beurteilung an die kognitive Kategorisierung des Beurteilers (statt an ein vorgefertigtes Schema). Nachteile sind ihre geringe Reliabilität und die Möglichkeit für den Beurteiler, auf das Ansprechen heikler - und gleichwohl wichtiger - Aspekte zu verzichten.

Einstufungsverfahren

Unter den gebundenen Urteilstechniken sind für alle Gruppen beurteilter Mitarbeiter Einstufungsverfahren die verbreitetste Verfahrensklasse (Cascio, 1987). Methodisches Prinzip ist die Zuordnung («Einstufung») von Verhaltensbeobachtungen, Ergebnis- oder Merkmaleinschätzungen zu einer mehrstufigen Skala. Die Skalenstufen können durch Zahlenwerte und verbal verankert sein, die Leistungsbeschreibung erfolgt mit etwa fünf bis 20 Skalen. Die einfachste Form ist die *graphische Einstufungsskala*, die in der Mehrzahl der Fälle mit fünf bis neun Skalenstufen verwendet wird, worin ein Reliabilitätsoptimum zu liegen scheint. Die Beurteilungsdimensionen werden bei diesem Skalentyp zumeist so allgemein formuliert, daß sie für verschiedene Arbeitsplätze Verwendung finden können. Anspruchsvollere Varianten sind *Verhaltensbeobachtungsskalen* (Latham & Wexley, 1977) und *Verhaltensverankerte Einstufungsskalen* (Smith & Kendall, 1963; ein Beispiel wird in Abbildung 3 vorgestellt). Beide basieren auf Arbeitsanalysen - vorzugsweise nach der Methode der kritischen Ereignisse (vgl. Abschnitt 2.1) - und können damit für

Organisationsgrad der Vorlesung: Eine gute Vorlesung ist klar strukturiert; die Beziehung zwischen den Themen wird deutlich; der Dozent hält sich an den Ablauf der Vorlesung.

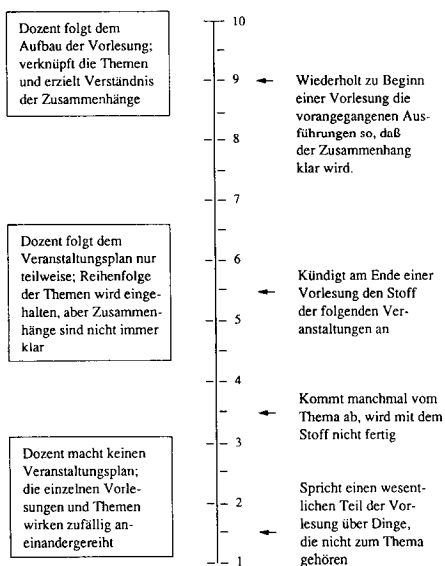


Abbildung 3: Verhaltensverankerte Einstufungsskala (verändert nach Bernardin, 1977).

sich in Anspruch nehmen, nur erfolgsrelevantes Verhalten in die Beurteilung einzubeziehen. Dies kommt der Akzeptanz und Rechtmäßigkeit der Beurteilungen zugute, ihrer Validität zumindest in einer inhaltlichen Form. Trotz relativ großen Konstruktionsaufwands, speziell im Falle der verhaltensverankerten Einstufungsskalen, ist allerdings zweifelhaft, ob damit wirklich eigenschaftsbezogene Beurteilungen, wie propagiert, vermieden werden (Murphy, Martin & Garcia, 1982): Die Forschung zur Informationsverarbeitung bei Beurteilungsprozessen weist darauf hin, daß sich Beurteiler nur in Ausnahmefällen an einzelne Beobachtungen erinnern und stattdessen die Einstufungen entsprechend ihrem in globaleren Urteilskategorien formulierten Bild von dem zu Beurteilenden vornehmen (Feldman, 1981).

Kennzeichnungs- und Auswahlverfahren

In geringerem Umfang werden Kennzeichnungs- und Auswahlverfahren eingesetzt. Die wichtigsten Varianten, die *Gemischte*

Aussagenliste mit freier Wahl und die *Gruppierte Aussagenliste mit Wahlzwang*, werden bei Brandstätter (1970) eingehend beschrieben. Ihrer Konstruktion liegt die Absicht zugrunde, dezidiert auf Verhaltensniveau zu beschreiben und Urteilstendenzen (v. a. Mittelwerts- und Streuungstendenzen) zu vermeiden. Neuere Untersuchungen mit diesen Skalentypen werden von Bernardin und Beatty (1984) berichtet. Mit der Zielsetzung, die Objektivität von Beurteilungen zu erhöhen, entwickelten Flanagan und Burns (1955) ein Beurteilungsverfahren, das auf der Technik der kritischen Ereignisse (Flanagan, 1954) basiert. Methodisches Prinzip ist die Erfassung solcher Verhaltensweisen, die zu Erfolg oder Mißerfolg im Arbeitsvollzug führen. Damit sind zwei Grundprinzipien betont, die auch in der heutigen Diskussion als bedeutsam hervorgehoben werden - der Anforderungsbezug und die Verhaltensorientierung. Aufgabe der Beurteiler ist die laufende Registrierung «erfolgskritischer» Verhaltensweisen.

Rangordnungsverfahren

Ebenfalls mehr in informellem Gebrauch denn als elaboriertes Beurteilungssystem dürfte die Bildung von Rangordnungen Verwendung finden. Explizite Rangbildung wird vor allem im Vertrieb, aber auch im militärischen Bereich und zur Charakterisierung von Ausbildungsleistungen vorgenommen.

Rangreihen können sowohl als globale Bewertung als auch differenziert nach Leistungsbereichen gebildet werden. Die Reliabilität der Beurteilung wird erhöht, wenn die Rangreihe über *Paarvergleiche* hergestellt wird. Verbreitet ist die *Quotenvorgabe* bei Verwendung von Einstufungsskalen (z. B. 10% - 20% - 40% - 20% - 10% der Werte auf den Ausprägungen einer fünfstufigen Skala). Hierbei handelt es sich um eine Rangordnung, deren mittlere Rangplätze mehrfach zu vergeben sind. Hierdurch soll unterschiedlichen Urteilsmaßstäben und vor allem der Mittelwertstendenz entgegen gewirkt werden. Speziell bei geringer Personenzahl wirken sich Verteilungszwang und

der Nullsummencharakter dieser Methode häufig in geringer Akzeptanz aus. Ein Verfahren, das Rangreihenbildung und Einstufung sequentiell verbindet, wurde von Miner (1988) vorgestellt.

Mit dem Ziel, einen direkten Vergleich zwischen Arbeitsanforderungen und Personalmerkmalen zu ermöglichen und gleichzeitig mangelnder Differenzierung zwischen Urteilsaspekten entgegenzuwirken, wurde von Brandstätter und Schuler (1974) das *Verhaltensrangprofil* entwickelt. Dieses Beurteilungsverfahren hat die Form einer ipsativen Rangreihe, bei der nicht Personen, sondern Verhaltensweisen oder Fähigkeiten für jede Person in eine Rangordnung nach dem Grad ihres relativen Zutreffens gebracht werden. Um zusätzlich verankernde Leistungsvergleiche zwischen Personen vornehmen zu können, wurde diesem Verfahren als Ergänzung die *Sequentielle Prozentrangskalierung* an die Seite gestellt. Deren asymmetrischer Aufbau ermöglicht verbesserte Differenzierung im oberen Leistungsbereich.

4.4 Qualität von Leistungsbeurteilungen

Im Hinblick auf Objektivität und Reliabilität wurden verschiedene Modelle vorgeschlagen, den Prozeß der Urteilsbildung darzustellen (für einen Überblick s. Landy & Farr, 1983; Schuler, 1989b). Hierzu gehören das Prozeßmodell von Wherry (Wherry & Bartlett, 1982), das Verhaltens-Eindrucks-Aussage-Modell von Brandstätter (1969) das Beobachtungs-Bewertungs-Gewichtungs-Modell von Borman (1978) und der Versuch von Ilgen und Feldman (1983) eine Darstellung des Beurteilungsprozesses in ein umfangreicheres «systemisches» Modell leistungsbedingender Faktoren zu integrieren.

über längere Zeit bestand die theoriebezogene Erforschung von Beurteilungen hauptsächlich in der Erörterung von Urteilstendenzen. Brandstätter (1970) faßte die wichtigsten Urteilstendenzen zu Mittelwerts-, Streuungs- und Korrelationstendenzen zusammen. Insbesondere die Ursachen reali-

tätsgerechter oder fehlerhafter Interkorrelation verschiedener Urteilsdimensionen («Halo» oder «Überstrahlung») wurde in vielen Einzelstudien untersucht (zusammenfassend Cooper, 1981). Generell zeigt sich für Urteilstendenzen ein eher geringer Zusammenhang mit Maßen der Genauigkeit (accuracy), Reliabilität und Validität (Cooper, 1981; Lord, 1985). Urteile können frei sein von erkennbaren Urteilstendenzen und trotzdem falsch, weshalb zunehmend dafür plädiert wird, differenzierte Messungen der «Richtigkeit» von Urteilen vorzunehmen (z. B. Ilgen & Feldman, 1983).

Zur Vermeidung von Urteilstendenzen, speziell von Korrelationstendenzen, kann empfohlen werden: die Verwendung eindeutiger, verhaltensbezogener und operational abgrenzbarer (also nicht nur sprachlich unterscheidbarer) Urteilkriterien, häufige und repräsentative Beobachtung, Kenntnis der Anforderungen, laufende Registrierung und Einsatz mehrerer Beurteiler. Die Wirkung der häufig zu diesem Zweck durchgeführten Beurteilertrainings ist umstritten (vgl. Lord, 1985 vs. Latham, 1986).

Speziell wenn Leistungskriterien zur Überprüfung von Personalauswahlverfahren verwendet werden, wird man die Frage nach der *Validität* dieser Daten stellen. In vielen Fällen muß man damit rechnen, daß die Prädiktordaten von höherer Qualität sind als die Kriteriendaten, die ihnen zum Maßstab dienen sollen. Dabei sind inhaltliche und konstruktbezogene Aspekte der Validität von Leistungsdaten sehr häufig in Frage zu stellen: Beispielsweise sind Vorgesetzten-, Kollegen- und Selbstbeurteilungen nur etwa in Höhe von $r = 0.20$ bis 0.30 untereinander korreliert (Schuler & Schmitt, 1987). Die prognostische Validität von Vorgesetztenurteilen liegt mit $r = 0.44$ etwa in gleicher Höhe wie die Validität von Kollegenurteilen ($r = 0.49$) (Hunter & Hunter, 1984). Schmitt et al. (1984) fassen die Validität von Vorgesetzten- und Kollegenurteilen metaanalytisch zum Wert von $r = 0.43$ zusammen, wobei die Prognose von Statusänderungen am besten (0.51) die der Gehaltshöhe am schlechtesten (0.21) gelingt. Schmidt und Hunter (1981) geben als durchschnittliche

Korrelation der Beurteilung von Trainings- und Arbeitsleistung $r = 0.77$ an.

Auch bei intensiveren Bemühungen um eine anforderungsbezogene Leistungsbeurteilung ist mit speziellen situativen Einflußfaktoren zu rechnen. So wurde der Einfluß «maximaler» gegenüber «typischer» Leistungsergebnisse von Sackett, Zedeck und Fogli (1988) untersucht. Dazu wählten sie die Tätigkeit der Kassiererin im Supermarkt, deren Leistung als Geschwindigkeit und Fehlerzahl einfach zu messen ist. Verglichen wurden die Leistungsergebnisse mittels standardisierter Warenkörbe, die einmal nach Ankündigung (maximale Leistung), das andere Mal ohne Wissen der Beobachteten (typische Leistung) an die Kasse gebracht wurden. Die Übereinstimmung der beiden Leistungsmessungen lag für erfahrene Kassiererinnen bei $r = 0.32$, für Neuestellte sogar nur bei $r = 0.14$. Nachdem es sich in beiden Fällen um reliable Messungen handelte, scheint mit beiden Leistungsmaßen unterschiedliches erfaßt zu werden. Wenn wir beispielsweise annehmen, daß im Assessment Center eher maximales als typisches Verhalten erfaßt wird, so müßte sich

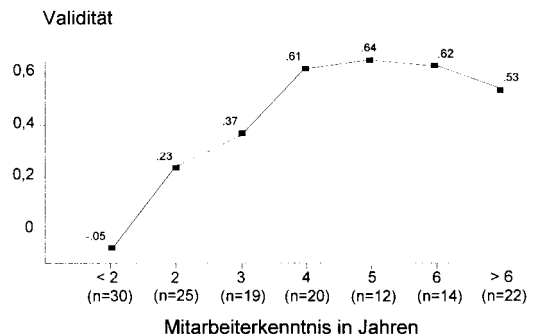


Abbildung 4: Einfluß des Zeitraums der Vorgesetzten-Mitarbeiter-Zusammenarbeit auf die Validität von Auswahlverfahren (nach Schuler et al., 1994).

höhere Validität errechnen lassen, wenn auch auf der Kriterienseite maximales Verhalten (z.B. der Erfolg in Sondereinsätzen) erfaßt wird, als wenn typisches Verhalten Gegenstand der Bewertung ist. Andererseits könnte eine praktische Konsequenz sein, sich in eignungsdiagnostischem Kontext verstärkt um die Evokation «typischen» Verhaltens zu bemühen.

Wie wichtig es speziell bei Validierungsstudien ist, über angemessene und verlässliche Kriterien zu verfügen, zeigte die in Informationsbox 1 bereits erwähnte Untersuchung, in deren Rahmen ein komplexes Auswahlverfahren für Wissenschaftler und Ingenieure entwickelt wurde. Die Überprüfung ergab wachsende Validitätskoeffizienten zum einen bei höherer Bedeutung der Anforderungen am entsprechenden Arbeitsplatz, zum anderen für Vorgesetztenurteile bei längerer Zeitspanne der Zusammenarbeit zwischen Vorgesetzten und Mitarbeitern (s. Abbildung 4).

Hätte man sich in dieser Studie auf die Erhebung von Kriterien nach kurzfristiger Zusammenarbeit beschränkt - wie nicht selten in Validierungsstudien -, wäre die Validität des Auswahlverfahrens grob unterschätzt worden. Die nur partielle Überschneidung verschiedener Kriterientypen, insbesondere die unterschiedlichen Beziehungen zwischen Prädiktoren und Kriterien, sprechen oft dafür, Leistungsbeurteilungen multimodal zu gestalten, d.h. sie aus verschiedenen Kriterientypen zu kombinieren (Schuler & Schmitt, 1987).

5. Personelle Entscheidung

Die Funktion der Diagnose von Eignung und Leistung besteht im allgemeinen darin, nicht nur personelle Leistungsvoraussetzungen festzustellen, sondern darauf aufbauend eine Personalentscheidung zu ermöglichen, zumindest zu unterstützen. Basis der personellen Entscheidung können in allen Fällen eignungsdiagnostische Daten (meist bei externen Bewerbern) oder Leistungsdaten

(meist bei organisationsinternen Bewerbern) sein. Eine Personalentscheidung kann angesehen werden als Zuweisung von Personen zu bestimmten Bedingungen, beispielsweise verschiedenen Arbeitsplätzen, aber eventuell auch der Bedingung «Ablehnung des Bewerbers». Zu beachten ist, daß an dieser Stelle die Qualität des gesamten Auswahl- und Entscheidungsprozesses trotz valider Auswahlverfahren oder Leistungsindikatoren durch Entscheidungsfehler zunichte gemacht werden kann. Allerdings sind Personalentscheidungen nicht nur unter dem Gesichtspunkt der Eignung, d.h. unter Bezug auf den eignungsdiagnostisch prognostizierten Berufserfolg zu treffen, sondern es sind in der Regel auch eignungsfremde institutionelle Gesichtspunkte wie Personalbedarf und Arbeitsmarktlage mit zu berücksichtigen. Dies kann Zuordnungskompromisse und die Ablehnung prinzipiell Geeigneter zur Folge haben.

5.1 Zuordnung und Veränderung

Entsprechend der *inhaltlichen* Ausgestaltung der Bedingungen (d. h. der Art der Personalmaßnahmen), denen Personen zugewiesen werden, wird zwischen Zuordnungs- und Veränderungsstrategien unterschieden. Zuordnungsstrategien umfassen die möglichst optimale Zuordnung von Personen und Arbeitsplätzen, entweder als Personalauswahl, bei der für einen bestimmten Arbeitsplatz die geeignetste Person zu finden ist, oder z.B. als Berufsberatung mit der Aufgabe, für eine bestimmte Person die geeignetsten Berufswege aufzuzeigen. Veränderungsstrategien sind auf eine Modifikation der betreffenden Personen (z.B. Training, Personalentwicklung) oder der sie umgebenden Situation (Arbeitsgestaltung, Organisationsentwicklung) ausgerichtet. Eignungsdiagnostische Verfahren sind v. a. für Zuordnungsstrategien wichtig. Bei Veränderungsstrategien sind eignungsdiagnostische Verfahren speziell zur Feststellung der (Grenzen der) Trainierbarkeit von Personen einsetzbar. Dabei ist die Diagnose der aufgabenspezifischen Lernfähigkeit («Potenti-

aldiagnose») von der Feststellung des (durch Training zu verbessernden) gegenwärtigen Qualifikationsniveaus zu unterscheiden. Letztere soll möglichst verhaltensbezogen erfolgen, damit darauf z. B. die Auswahl spezifischer Trainingseinheiten aufgebaut werden kann.

Die Alternative «Training oder Auswahl» stellt sich damit kaum, sondern meist werden beide in sequentieller Weise kombiniert. Auch sind fundierte Entscheidungen zwischen Auswahl- und Trainingsmaßnahmen bisher erschwert durch die schwache Wissensbasis im Hinblick auf anforderungs- und merkmalspezifische Trainierbarkeit bzw. Stabilität. Immerhin kann man aber die isolierte Wirkung der beiden Alternativen Personalauswahl und Training in bezug auf erreichte Leistungsverbesserungen z.B. im Managementbereich vergleichen (Funke, 1989): In einer Metaanalyse über 70 Studien zu Führungstrainings fanden Burke und Day (1986) mittlere Trainingsverbesserungen, die in Korrelationskoeffizienten umgerechnet zwischen $r = 0.24$ (für Vorgesetzten-/Kollegenurteil) und $r = 0.32$ (für objektive Resultate) liegen. Die mittlere korrigierte Validität des Assessment Centers über ebenfalls 70 Studien beträgt dagegen nach Thornton et al. (1992) $r = 0.37$.

5.2 Entscheidungen mit Hilfe statistischer Modelle

Eine Vielzahl personeller Entscheidungen wird in der Praxis von Personalfachleuten unter Verwendung ihres persönlichen Fachwissens, ihrer Erfahrung und Intuition auf der Basis diagnostischer Einzeldaten getroffen (individuelle Entscheidung). Die diesen diagnostischen Urteilen zugrundeliegenden impliziten Datenintegrations- und Entscheidungsregeln können mit Methoden der diagnostischen Urteilsforschung (Jäger, 1982) analysiert und mit speziellen computergestützten Modellierungsverfahren auch explizit gemacht werden (Wottawa & Hossiep, 1987). Dabei zeigt sich in der Regel eine, geringe Nutzung der erhobenen diagnosti-

schen Information für die Personalentscheidung.

Unter dem Gesichtspunkt der Genauigkeit ist diese «Klinische Urteilsbildung» in Form von Expertenurteilen einer «statistischen Urteilsbildung» mit der Anwendung abgesicherter Berechnungsvorschriften vielfach als - zumindest im Durchschnitt - unterlegen nachgewiesen worden (Übersicht bei Jäger, 1982). Vergleichbare Ergebnisse fanden sich bei Assessment Center-Prognosen. Die geringe Leistungsfähigkeit «klinischer» Urteilsbildung drückt sich z.B. aus in Einstellungsempfehlungen von Experten, die mit der späteren Vorgesetztenbeurteilung im Durchschnitt nur zu $r = 0.17$ korrelierten (Reilly & Chao, 1982).

Für personelle Entscheidungen auf empirisch-statistischer Basis (institutionelle Entscheidungen) müssen Daten für Prädiktoren und für Kriterien vorliegen. Dann kann zunächst der prognostizierte Berufserfolg von Personen durch ein entsprechendes mathematisch-statistisches Verfahren berechnet werden. Um zu einer Zuweisung der Personen zu bestimmten Bedingungen zu kommen, z.B. Annahme oder Ablehnung als Mitarbeiter, wird zusätzlich eine *Zuordnungsregel* benötigt, die bestimmt, mit welchen prognostizierten Berufserfolgswerten welcher Bedingung zugewiesen wird. Die Zuordnungsregel umfaßt meist einen *kritischen Wert*, d.h. einen Grenzwert (Cutoff), der unter bestimmten Annahmen mathematisch-statistisch bestimmbar ist. Wichtig ist, daß der kritische Wert so gewählt bzw. berechnet werden kann, daß er die je nach Entscheidungssituation unterschiedlich gravierenden Entscheidungsfehler (z. B. fälschlich Abgelehnte gegenüber fälschlich Angenommenen) unterschiedlich stark berücksichtigt oder z.B. auch die Gesamtzahl der Falschzuordnungen über alle Bedingungen hinweg im Durchschnitt minimiert. Personelle Entscheidungen sind somit meist zweistufige Prozesse aus Datenverdichtung zu einer Kriterien- bzw. Berufserfolgsprognose und aus einer regelgeleiteten Entscheidung. Drei grundlegende Typen von Personalentscheidungen lassen sich dabei unterscheiden:

- *Selektion* (Personalauslese) in Form der Zuweisung zu einer der beiden Bedingungen «Annahme» und «Ablehnung» auf der Basis univariater oder entsprechend kombinierter Daten;
- *Plazierung* als Zuweisung auf ordenbaren Bedingungen (z.B. in Rangreihe nach ansteigenden Anforderungen geordnete Arbeitsplätze oder Ausbildungsgänge) ohne Möglichkeit der Zurückweisung von Personen,
- *Klassifikation* als Zuweisung zu unterschiedlichen, nicht ordenbaren Bedingungen (z.B. unterschiedliche Ausbildungsgänge oder Karrierewege); dies erfordert multivariate Daten.

Die grundlegenden Prognosemodelle für Selektion und Plazierung sind lineare Modelle (Brandstätter, 1982). Der prognostizierte Kriterienwert ergibt sich dabei als Summe der mit speziell errechneten Gewichten versehenen einzelnen Prädiktoren, wodurch sich die Prädiktoren untereinander kompensieren können. Standardmäßig verwendet wird die multiple Regression, bei der an einer Personenstichprobe die optimalen Gewichte bestimmt werden. Auf diese Weise lassen sich Einzelverfahren zu Auswahlbatterien kombinieren. Lineare Modelle, insbesondere die multiple Regression, werden für personelle Entscheidungen überwiegend als sehr effiziente und robuste Methoden angesehen (Cascio, 1987). Selbst starke Vereinfachungen durch die Verwendung von Einheitsgewichten anstatt errechneter Optimalgewichte sind in vielen Fällen erfolgreich. Methodische Verfeinerungen durch die Verwendung nichtlinearer Modelle haben sich dagegen bisher kaum bewährt. Personalentscheidungen als Selektion beruhen meist auf dem Prinzip der *Bestenauslese*. Bewerber werden nach prognostiziertem Berufserfolg oder nach dem Prädiktorwert in eine Rangreihe gebracht, und die Einstellungen erfolgen von oben nach unten. Der Cutoff liegt damit implizit an der Stelle des letzten eingestellten Bewerbers. In anderen Fällen ist ein solches Vorgehen nicht möglich, z.B. wenn Bewerber einzeln und auf größere Zeiträume verteilt auftreten. Als

Entscheidungsgrundlage ist dann die *Bestimmung des Cutoffs* nötig (Übersicht bei Cascio, Alexander & Barrett, 1988).

Die empirisch gestützte Bestimmung eines Cutoffs ist nur bei Vorliegen von Validitätsdaten, also Prädiktor- und Kriterienwerten möglich. Dann läßt sich eine Mindestberufsleistung festlegen (z.B. aufgrund von Leistungsbeurteilungsdaten), aus der mittels Regression der Cutoff für den Prädiktor abgeleitet werden kann. Ebenso kann aufgrund des bivariaten Streudiagramms von Prädiktor und Kriterium die Berufserfolgswahrscheinlichkeit für mögliche Cutoffs bestimmt werden. Ohne Kriteriendaten müssen unsicherere Ersatzlösungen in Form von Expertenurteilen über erforderliche Mindestleistungen gefunden werden.

Liegen mehrere Prädiktoren vor, die nicht (linear) kombiniert werden sollen, weil eine Basiskompetenz in jedem Prädiktor erforderlich ist und die gegenseitige Kompensation unterschiedlicher Prädiktorenwerte nicht zugelassen werden kann, dann muß für jeden Prädiktor ein eigener Cutoff bestimmt werden (multiple Cutoffs). Beispielsweise könnte gefordert werden, daß Assessment Center-Teilnehmer in allen Übungsteilen mindestens mittlere Ergebnisse erzielen und schlechte Ergebnisse der einen Übung nicht durch besonders gute Ergebnisse anderer Übungen ausgleichbar sind.

Für Zwecke der Klassifikation von Personen können verschiedene multivariate Verfahren zum Einsatz kommen (Übersicht bei Janke, 1982). Umfangreiche Systeme personeller Entscheidungen, v.a. der Klassifikation, werden aber selten eingeführt. Sie setzen längerfristige Personalplanungszeiträume, umfangreichere Stellen- und Bewerberzahlen sowie die Erhebung und Verarbeitung größerer Datenmengen voraus. Nicht selten wird die Gefahr einer gewissen Inflexibilität im Hinblick auf einschränkende und sich ändernde Randbedingungen, z. B. Zahl und Qualifikation der Bewerber, Mobilität, Karrierevorstellungen usw. gesehen.

5.3 Nutzen

Cronbach und Gleser (1965) schlugen vor, Personalentscheidungen auch nach ihren Konsequenzen zu bewerten, d.h. die Gesamtheit aller im Zusammenhang mit der Eignungsdiagnostik und Personalentscheidung auftretenden positiven (Nutzen) und negativen (Kosten) Konsequenzen zu erfassen und den Nettotonutzen (Gewinn) in Geldeinheiten anzugeben. Den Übergang von einer auf Präzision orientierten Evaluation personeller Entscheidungen zur Evaluation unter Nutzungsgesichtspunkten markiert die klassische Arbeit von Taylor und Russel (1939). Sie demonstrierten anhand der von ihnen entwickelten Tabellen, daß die Erfolgsquote (auch als «selektiver Eignungsquotient» bezeichnet), die den Prozentsatz der Erfolgreichen unter den insgesamt ausgewählten Bewerbern angibt, von folgenden drei Größen abhängt (siehe Abbildung 5):

- Validitätskoeffizient,
- Selektionsquote (Anteil Ausgewählter unter den Bewerbern),
- Grundquote (Anteil Geeigneter unter den Bewerbern).

Jeder Bewerber ist mit seiner spezifischen Kombination von Prädiktorwert und Berufserfolg innerhalb der ellipsenförmigen bivariaten Verteilung repräsentiert. Die Erfolgsquote ist das Verhältnis der eingestellten Geeigneten zu allen Eingestellten, d.h. $EG/(EG+EU)$. Dieses Verhältnis verändert sich durch höhere Validitäten (schlankere Ellipse). Die Erfolgsquote steigt mit höherer Validität, weil dabei sämtliche Zuordnungsfehler reduziert werden. Sie steigt, je kleiner die Selektionsquote im Vergleich zur Grundquote ist (d.h. je strenger die Auswahl ist), wobei die Selektionsquote durch die Wahl unterschiedlich strenger Cutoffs beeinflussbar ist. Durch strenge Cutoffs wird ein Zuordnungsfehler (Auswahl Ungeeigneter) verringert auf Kosten des anderen Zuordnungsfehlers (Ablehnung Geeigneter), der traditionell bei Personalauswahlentscheidungen keine Berücksichtigung findet. Die Grundquote ist Ausdruck der vorgegebenen Fähigkeitsverteilung in der Bewerber-

berstichprobe und durch Selektionsmaßnahmen nicht veränderbar. Personalauswahl bei hoher Grundrate (horizontale Linie liegt tief) ist vergleichsweise weniger nützlich, da nur wenige Ungeeignete abgelehnt werden, d.h. die Erfolgsquote ist bereits ohne Personalauswahl hoch.

Bereits seit 40 Jahren existieren die ersten Kalkulationsmodelle für den finanziellen Nutzen von Personalentscheidungen. Einen Überblick und Anwendungsbeispiele geben Funke und Barthel (1990) die bei diesen Nutzenmodellen drei Funktionen hervorheben:

- Evaluation geplanter oder vorhandener Auswahlprogramme unter Berücksichtigung und Konkretisierung ökonomischer Einflußfaktoren,
- Kommunikationsverbesserung durch Verdeutlichung der Ergebnisse organisationspsychologischer Maßnahmen gegenüber Vertretern ökonomischer Fachrichtungen in *finanziellen* Begriffen,
- Legitimationsfunktion durch Darstellung organisationspsychologischer Maßnahmen als *Investition* anstatt als reine Ausgaben und damit eine innerbetrieblich erhöhte Konkurrenzfähigkeit um Budgets.

Brogden (1949) und Cronbach und Gleser (1965) propagierten das folgende klassische Nutzenmodell, das die Grundlage aller weiteren Entwicklungen darstellt:

$$\Delta U = N_A \cdot T \cdot r_{xy} \cdot (\Phi/p) \cdot SD_y - C \cdot N_B$$

Bestimmt wird dabei der finanzielle Gewinn, der durch die Leistungsverbesserung der «behandelten» Gruppe entsteht. Die Leistungsverbesserung wird erkennbar am mittleren, z-standardisierten Kriterienwert dieser Gruppe, \bar{z}_y . Der finanzielle Wert dieser Leistungsverbesserung wird ausgedrückt durch die Standardabweichung der Leistung (gemessen in Geldeinheiten!), SD_y , die mit \bar{z}_y zu multiplizieren ist. Durch ein Eignungsdiagnosticum kann \bar{z}_y geschätzt werden als $r_{xy} \cdot \bar{z}_x$, und die durchschnittliche Testleistung \bar{z}_x ist wiederum von der Selektionsquote p abhängig: $\bar{z}_x = \Phi/p$ (Φ ist die Ordinatehöhe der Standardnormalverteilung

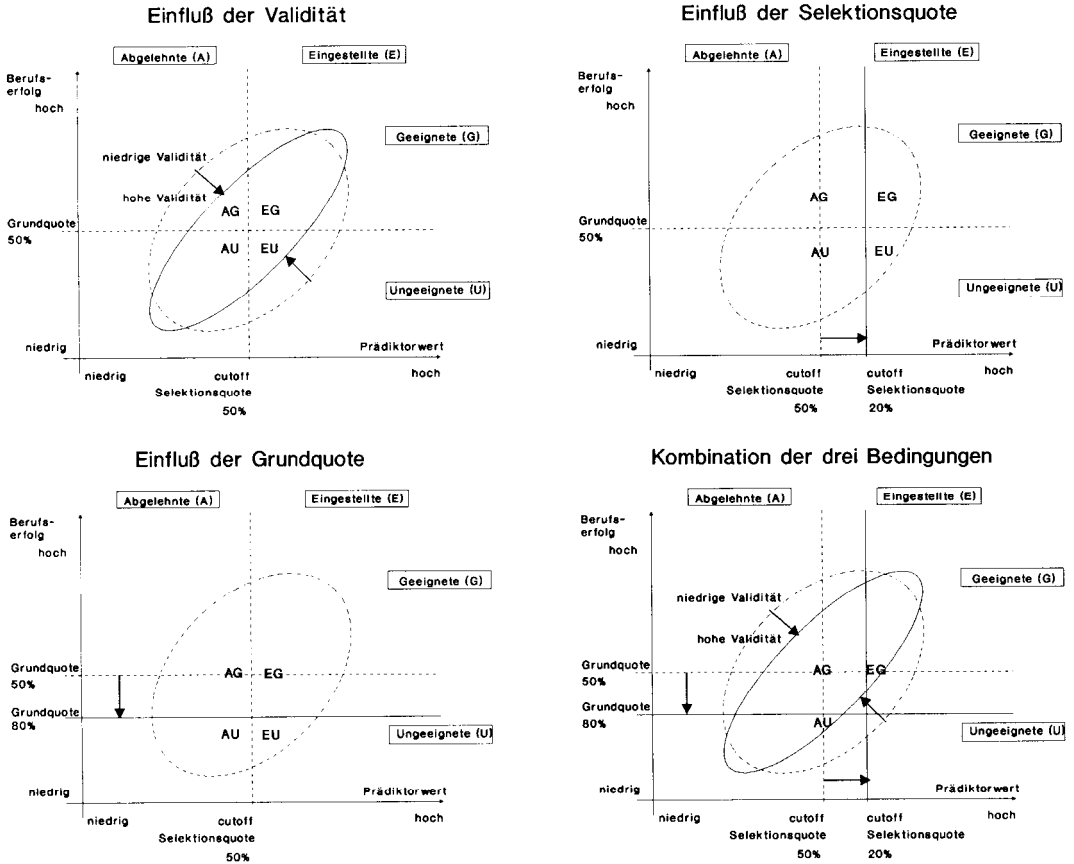


Abbildung 5: Abhängigkeit der Erfolgsquote bei Personalentscheidungen von verschiedenen situativen Bedingungen.

am cutoff für die Selektionsrate p). Der Nutzen steigt proportional zur Zahl der Angenommenen NA und der Verweildauer T dieser leistungsstärkeren Personen. Nach Abzug der Kosten C für die Bewerberzahl Na ergibt sich der Gewinn (Nettonutzen ΔU). Als einflußreichste Variablen der Gleichung sind nach Sensitivitätsanalysen r_{xy} , p und SD_y anzusehen, worunter gerade SD_y nur schwer zu bestimmen ist. Dies hat dazu geführt, daß Nutzenmodelle bis vor 10 Jahren kaum zur Anwendung kamen und erst seit der Einführung verbesserter Schätzmethoden für SD_y größere Verbreitung finden. Als erste praktikable Alternative, die gegenwärtig die verbreitetste Methode der SD_y -Bestimmung ist, wurde von Schmidt, Hunter, McKenzie und Muldrow (1979) ein Prozent-

rangschätzverfahren vorgeschlagen. Mehrere Experten, in der Regel Vorgesetzte, sollen den Geldwert der Jahresleistung von Mitarbeitern schätzen, die sich am Mittelwert einer Leistungsverteilung (Prozentrang [PR] 50), eine Standardabweichung darunter (PR 15) oder darüber (PR 85) befinden. Die Differenzen PR 85 - PR 50 und PR 50 - PR 15 sind dann Schätzwerte für SD_y . Während das bisherige Nutzenmodell im Grunde genommen bereits einer statischen Gewinnvergleichsrechnung entsprach, wurde es später durch verschiedene betriebswirtschaftliche Parameter v. a. der dynamischen Investitionsrechnung ergänzt (Boudreau, 1989):

- Diskontierung zukünftiger erwarteter Einnahmen und Kosten;

Informationsbox 6:*Schätzung des Nutzens am Beispiel eines biographischen Fragebogens*

Geprüft wird die Wirkung eines *zusätzlich* zum traditionellen Einstellungsinterview für die Bewerberauswahl eingesetzten biographischen Fragebogens in einer norddeutschen Versicherungsgesellschaft (Barthel & Schuler, 1989). Das einfache Nutzenmodell nach Brogden (1949) und Cronbach und Gleser (1965) (siehe Text) wird erweitert um:

- Diskontierung der erst in späteren Jahren erwarteten Einnahmen und Ausgaben (Faktor $1/(1+i)^k$)
- Berücksichtigung mehrerer Anwendungsperioden (Aufsummierung über die Anwendungsperioden $k = 1$ bis F)
- Kohortenbetrachtung der Mitarbeiterzu- und -abgänge als Differenzierung der Zahl Angenommener N_A und der Verweildauer T (Aufsummierung über die Wirkungsdauer $t = 1$ bis k)
- Fixkosten für Entwicklung, Schulung und Lizenzgebühren des biographischen Fragebogens

Das Modell läßt sich damit folgendermaßen darstellen:

$$\Delta U = \sum_{k=1}^F \left[\sum_{t=1}^k (N_{At} - N_{St}) \right] [1/(1+i)^k] r_{xy} (\Phi/p) SD_y - \sum_{k=1}^F [C_k \cdot 1/(1+i)^{k-1}]$$

Bestimmung der Parameter:

ΔU : Gewinn bzw. Nettonutzen durch zusätzlichen Einsatz des biographischen Fragebogens

F : Zahl der Jahre, in denen das Verfahren eingesetzt wird, $F = 3$

k : Wirkungsdauer eines Verfahrenseinsatzes, entspricht der durchschnittlichen Verweildauer Angenommener, $k = 4$

N_{At} : Anzahl der Zugänge pro Jahr

N_{St} : Anzahl der Abgänge pro Jahr

Die Zahl der im Unternehmen verbliebenen Mitarbeiter, auf die das Verfahren angewandt wurde, steigt über die drei Anwendungsjahre an und fällt über die restlichen drei Jahre ab:

Jahr	$N_{At} - N_{St}$
1	150
2	235
3	299
4	202
5	117
6	53

i : Kalkulationszinsfuß, hier 7%

r_{xy} : Validitätsvorsprung des biographischen Fragebogens gegenüber dem traditionellen Interview, $r = 0.18$

Φ/p : bei Selektionsrate 50%: $0.3989/0.50$

SD_y : Standardabweichung der Leistung in DM, bestimmt mittels Vorgesetzten-schätzungen der Beitragseinnahmen durch Mitarbeiter (Prozentrangschätzverfahren), $SD_y = 4.800$ DM

C_k : Gesamtkosten des biographischen Fragebogens pro Jahr:

1. Jahr: Fixkosten 75.000,- DM (Entwicklung, Lizenzen)
variable Kosten 10.000,- DM (237 Bewerber a DM 42,-)
2. Jahr: Fixkosten 10.000,- DM
variable Kosten 10.000,- DM
3. Jahr: Fixkosten 10.000,- DM
variable Kosten 10.000,- DM

Ergebnis:

Nettonutzen bzw. Gewinn durch zusätzliche Anwendung des biographischen Fragebogens bei dreijährigem Einsatz (ca. 700 Anwendungen): 473.265,- DM (Gesamtnutzen 594.424,- DM, Gesamtkosten 121.159,- DM)

- Berücksichtigung fixer Kosten (Entwicklungskosten für neue eignungsdiagnostische Verfahren) neben variablen Kosten der kontinuierlichen Anwendung;
- Berücksichtigung mehrerer Anwendungsperioden (z. B. Jahre) und des kontinuierlichen Mitarbeiterzu- und -abgangs statt statischer Zahlen (Kohortenbetrachtung);
- variable Kosten höherer Leistung, z.B. müssen leistungsstärkere Mitarbeiter oft besser bezahlt werden;
- Steuern auf den Gewinn;
- Nutzenminderung bei Absage der Stellenangebote durch besonders leistungsstarke Bewerber.

Bereits Schmidt et al. (1979) wiesen darauf hin, daß der Nutzen psychologischer Eignungsdiagnostik weit unterschätzt wurde. Boudreau (1989) berichtet von 39 bisher durchgeführten empirischen Kosten-Nutzen-Analysen. In jedem Fall ergab sich ein finanzieller Gewinn, selbst für weniger valide Verfahren wie das traditionelle Interview. Der wesentlich aussagekräftigere Vergleich alternativer Auswahlverfahren durch Gegenüberstellung des Gewinns der Alternativen ging jeweils zugunsten des valideren Verfahrens aus, trotz dessen oft wesentlich höheren Kosten. Der finanzielle Gewinn liegt in der Regel in vier- bis fünfstelliger Größenordnung pro besetztem Arbeitsplatz. Ein empirisches Beispiel hierzu gibt Informationsbox 6. Wird der Nutzen valider Personalauswahl national hochgerechnet, so kann er Milliardenhöhe erreichen (Hunter & Schmidt, 1982).

6. Kontext und Wirkung

Im Zusammenhang mit Assessment Center und Leistungsbeurteilungen wurden bereits Funktionen genannt, die über den engeren Kontext der diagnostischen Methode hinausweisen: Aufschluß über Leistungsstand und Defizite der Organisation - nicht nur der Personen - zu ermöglichen, Anforderungen und Ziele zu operationalisieren und

der Diskussion zugänglich zu machen, die Bedeutung von Personalplanung und Personalentwicklung zu unterstreichen, Beiträge zur Sozialisation und Verhaltenssteuerung, zur Arbeitsgestaltung, zu Führung und Zusammenarbeit sowie zu anderen Aspekten der Organisation und ihrer «Kultur» zu leisten. Dies gilt im Prinzip für jede wirksame Methode und erinnert daran, daß auch das Messen oder Diagnostizieren bereits eine Intervention, einen Eingriff in die Organisation darstellt. Auch bestätigt es die im Einleitungskapitel vertretene Auffassung, daß sich Menschen nicht nur «in Organisationen verhalten», sondern auch durch ihr Verhalten oder Handeln Organisationen in ihren jeweiligen Eigenheiten konstituieren (vgl. hierzu Porter, Lawler & Hackman, 1975, und Neuberger, 1991).

Nicht alle Aspekte der Wechselbeziehung zwischen Methoden, Verhalten und Organisationen können hier angesprochen werden. Besonders wichtig und naheliegend aber ist die Frage, wie die Eignungs- und Leistungsdiagnose mit der Gewinnung von Mitarbeitern zusammenhängt, mit der Auswahl von Organisationen durch Individuen und mit deren Reaktion auf diagnostische Verfahren. Diesen Fragen sind die nächsten Abschnitte gewidmet.

6.1 Gewinnung von Mitarbeitern

Ein wichtiger Teil personalpsychologischer Arbeit besteht darin, Mitarbeiter zu finden und zu gewinnen, die bereit sind, der Organisation beizutreten und Arbeit zu den angebotenen Bedingungen zu leisten. Diese Aufgabe wird gewöhnlich mit *Personalwerbung* bezeichnet. Von *Personalmarketing* wird zumeist dann gesprochen, wenn sich die Austauschprozesse nicht nur auf den organisationsexternen, sondern auch auf den internen Arbeitsmarkt beziehen und eine Vielfalt hierzu dienlicher Maßnahmen umfassen (Staffelbach, 1986). Faßt man Personalmarketing gar so weit wie Wunderer (1991, S.119), nämlich als «die Kunst, attraktive und effiziente Arbeitsbedingungen zu analysieren, zu gestalten und zu kommu-

nizieren», könnte man einen erheblichen Teil der in diesem Lehrbuch besprochenen Problemfelder auch als Marketingprobleme auffassen.

Aus marktwirtschaftlicher Sicht ist der grundsätzliche Bedarf an Personalmarketing vor allem durch das Verhältnis von Angebot und Nachfrage geregelt, d.h. die Bemühungen, Mitarbeiter zu gewinnen und zu halten, sind dort größer, wo die Nachfrage nach Arbeitsplätzen geringer ist. Des Weiteren ist Personalmarketing eher dort erforderlich, wo ein großer Personalersatzbedarf vorliegt, wo eine ausgeprägte Konkurrenzsituation der Organisation gegeben ist und wo die individuellen Leistungsunterschiede der Mitarbeiter bzw. Bewerber groß sind (Moser, 1992). Betriebswirtschaftliche und organisationspsychologische Sichtweisen decken sich darin, daß Personalmarketing eine dezidiert mitarbeiterorientierte Personalpolitik impliziert (Bartscher & Fritsch, 1992) und für die Organisation bedeutet, attraktiv genug sein zu müssen, um die qualifiziertesten Bewerber zu interessieren und Mitarbeiter zu motivieren und zu behalten (Schneider & Schmitt, 1986). Hierfür ist nicht nur die Gestaltung des Arbeitsplatzes im engeren Sinn von Bedeutung, sondern auch alle anderen - im Teil «Arbeit» dieses Lehrbuchs bereits angesprochenen - Aspekte der Organisation, die zu Interessen und Bedürfnissen von Bewerbern und Mitarbeitern in Beziehung stehen; dies sind vor allem die allgemeinen Aspekte Arbeitsinhalt, Kommunikation und Zusammenarbeit, Bezahlung und Entwicklungsmöglichkeiten.

Im Zusammenhang mit diagnostischen Fragen naheliegender sind die individuellen Besonderheiten an Interessen und Arbeitsmotiven; hier kann z.B. das Bedürfnis nach Abwechslung, nach geistiger oder körperlicher Beanspruchung, nach Geltung und Anerkennung, nach Wettbewerb und Karriere von besonderer Bedeutung sein. Mittels Interessententest oder Interview können motivationale Besonderheiten sowie solche der Lebensbedingungen - Alter, Familienstand, Gesundheit, soziale Verpflichtungen außerhalb der Organisation etc. - erfaßt werden, allerdings erst zum Zeitpunkt der ersten

Kontakte zwischen Bewerber und Organisation. Um externe Bewerber interessengemäß anzusprechen, orientiert man sich deshalb häufig an gruppenspezifischen Charakteristika und am vermuteten Selbstbild der Zielgruppen. In einer Untersuchung an Hochschulabsolventen (Schwaab & Schuler, 1991) zeigte sich dementsprechend ein deutlicher Zusammenhang der Attraktivität eines potentiellen Arbeitgebers mit der Übereinstimmung zwischen Selbstbild der potentiellen Bewerber und ihrem Bild vom betreffenden Unternehmen.

Je besser es gelingt, Zielgruppen fähigkeits- und motivationsgerecht anzusprechen, desto eher kann dadurch bereits ein Teil der Selektion wahrgenommen werden. Die Kenntnis der Zusammenhänge zwischen Fähigkeiten und Bedürfnissen bei Mitarbeitern der Organisation mit beruflichem Erfolg ermöglicht also eine verbesserte Personalgewinnung bereits im Vorfeld der Auswahl und verminderten Aufwand beim Auswahlprozeß als solchem.

Die Methoden, potentielle Stellenbewerber anzusprechen, sind vielfältig. Eine Zusammenstellung der verbreitetsten Instrumente, die von rekrutierenden Unternehmen auf dem externen Personalmarkt der Hochschulabsolventen eingesetzt werden, gibt Tabelle 5.

Der Wirkungsgrad der Methoden zur Ansprache potentieller Bewerber ist sehr unterschiedlich. Die Wirkung schriftlicher Information wird oft überschätzt; demgegenüber sind persönliche Kontakte, wie sie durch Praktika und Hochschulkontakte zustandekommen, sowie Empfehlungen bereits beschäftigter Mitarbeiter von nachhaltigerem Einfluß (Powell, Kirman, Farley & Geisinger, 1989). Die persönliche Ansprache potentieller Mitarbeiter wird auf höheren Positionsebenen für noch wichtiger gehalten: Während Führungskräfte der mittleren Ebene zu 35% über Zeitungsanzeigen des Unternehmens gesucht werden, wird dieser Weg bei der obersten Führungsebene nur noch in 9,6% der Fälle gewählt. Statt dessen wird dort die persönliche Ansprache durch das Unternehmen oder durch beauftragte Personen zum dominierenden Ver-

Table 5: Verbreitetste Instrumente der Ansprache von Hochschulabsolventen (aus Schwaab & Schuler, 1991).

- Stellenanzeigen
- Broschüren, die das Unternehmen und die Einstiegsmöglichkeiten darstellen
- Entsenden von Lehrbeauftragten an Hochschulen
- Annoncen und Firmenpräsentationen in von der Zielgruppe häufig gelesenen Zeitungen und Zeitschriften
- Vorträge und Unternehmenspräsentationen an Hochschulen
- Präsenz bei an Hochschulen organisierten Firmenkontaktgesprächen
- Anbieten von Praktikantenplätzen (Praxissemester) an Studenten
- Anbieten von Ferienjobs und studienbegleitenden Teilzeitbeschäftigungen
- Informationsveranstaltungen wie Workshops in Unternehmen
- Kontakte zu Lehrstuhlinhabern und deren Assistenten
- Veranstaltung von Seminaren für Studenten
- Kooperation mit studentischen Vereinen
- Unterstützung von Diplomanden und Doktoranden
- Einrichtung von Förderkreisen für ausgewählte Studenten.

fahren (Gaugler, 1980, zit. n. Pulling, 1990, S. 60). Persönliche Ansprache ist auch in der Betreuungsphase nach Vertragsabschluß wichtig: So konnten Ivancevich und Donnelly (1971) durch telefonische Kontakte die Rücktrittsquote von Führungskräften von 12,5% auf 2,6% senken.

Bei allem Bemühen, externe Bewerber zu gewinnen, sollte nicht übersehen werden, daß die überwiegende Zahl personeller Entscheidungen interne Mitarbeiter betrifft. In diesem Fall kann sich der Fehler, Geeignete abzulehnen, in vermindertem «Commitment» auswirken, d.h. in geringerer Bindung oder Verpflichtung gegenüber dem Unternehmen, in der Zurücknahme des Engagements und des Einsatzes über den «Dienst nach Vorschrift» hinaus. Commitment gehört zu den wichtigsten Korrelaten der Fluktuation (Schneider & Schmitt, 1986) und ist nach Scheins Konzept des «psychologischen Kontrakts» (Schein, 1965) an die Wahrnehmung eines reziproken Commitment seitens der Organisation gebunden. Bereits mit der Entscheidung, eine verfügbare Stelle intern oder extern zu besetzen, steht die Organisation vor der Gefahr, sich gegenüber den bereits Beschäftigten als illoyal zu erweisen. Mangelnde Qualifikation unter ihnen weist auf die Mitverantwortung des Unternehmens zurück, für ihre angemessene Auswahl und Entwicklung frühzeitig zu sorgen. Breitere Auswahlmöglichkeit, innovative Impulse und auch die Anerkennung speziell von Führungskräften mögen für Orientierung am externen Arbeitsmarkt

sprechen, während die Erfahrung und Betriebskenntnis, die geringeren Beschaffungskosten und die Bedeutung von Aufstiegsmöglichkeiten Argumente für die interne Lösung sind (Kompa, 1984). Überdies steht diese Entscheidung natürlich auch im Spannungsfeld der unternehmenspolitischen Kräfte.

6.2 Individuelle Organisationswahl

Die Entscheidung einer Person, in einer bestimmten Organisation zu arbeiten, ist zu einem großen Teil durch Einflüsse ihres sozialen Umfelds, durch die Erreichbarkeit des Arbeitsplatzes und andere Bedingungen ihrer individuellen Lebenswelt bestimmt (Seifert, 1982). Weitere personinterne und -externe Ursachen beschränken die individuelle Wahlmöglichkeit beträchtlich. Dennoch bleibt in vielen Fällen ein Entscheidungsspielraum bei der Wahl der Arbeitsorganisation - im Falle hochqualifizierter Bewerber, an denen diesen Organisationen speziell gelegen ist, sogar ein besonders großer Spielraum. Was sind nun die Gründe der Bevorzugung einer Organisation gegenüber einer anderen?

Die meisten der vorliegenden Theorien der Organisationswahl thematisieren den Vergleich zwischen individuellen Bedürfnissen und der (vermuteten) Befriedigungskapazität der Organisation (Wanous, 1980) sowie den zwischen wichtigen Facetten des Selbst-

bilds und entsprechenden der Organisation zugeschriebenen Merkmalen.

Wanous (1989) bezeichnet diese Variante von Organisationswahltheorien als "prospektiv rational", denn sie unterstellen einen Prozeß des Abwägens vor der Entscheidung, in den eine größere Anzahl von Gesichtspunkten einbezogen wird, die nach persönlicher Bedeutung gewichtet werden, und innerhalb dessen mehrere alternativ verfügbare Stellen gleichzeitig in Betracht gezogen werden. Ein Prototyp «prospektiv rationaler» Theorien ist das Erwartungs-Wert-Modell, das die «Valenz» der Arbeitsplatzmerkmale mit der «Instrumentalität» des Arbeitsplatzes zur Bedürfnisbefriedigung und der Wahrscheinlichkeit, ein Stellenangebot zu bekommen, multiplikativ verbindet (kritische Darstellung bei Schuler & Moser, 1992).

In Kontrast hierzu verweist Wanous auf Theorien der Organisationswahl, die er als «retrospektiv rational» bezeichnet. Sie nehmen an, daß tatsächlich nur wenig sachliche Information in die Entscheidung eingeht, daß die Präferenzbildung auf dem Wege aufeinanderfolgender Paarvergleiche stattfindet und daß nach erfolgter Entscheidung - wie von der Dissonanztheorie beschrieben - ihre Richtigkeit durch Argumente aufgewertet, also rationalisiert wird. Wanous zählt eine weitere Auffassung zu den «retrospektiv rationalen» Theorien, die allerdings von diesen gut abgrenzbar scheint - die Annahme, daß Stellungs-suchende sich einen globalen Eindruck von einer Organisation bilden und diesen mit ihrem Selbstbild vergleichen. Diese Sichtweise kann durch Aussagen wie «das ist der richtige Betrieb für mich» oder «ich könnte mir nicht vorstellen, dort zu arbeiten» illustriert werden.

Die empirische Forschung bietet bislang nur wenig Hilfe zur Entscheidung zwischen diesen theoretischen Ansätzen; zwar liegt eine größere Zahl von Ergebnissen vor, die eher für die «prospektiv rationale» Vorgehensweise spricht als für die «retrospektiven» Varianten, aber das liegt vor allem an der geringen Zahl empirischer Prüfungen der letzteren - und an der Schwierigkeit, sie zu prü-

fen. Inhaltlich sind für die Entscheidung eines Bewerbers natürlich alle bekannten oder erwarteten motivationsrelevanten Aspekte bedeutsam, wie sie weiter oben erwähnt und im Teil «Arbeit» erörtert wurden. In begrenztem Umfang sind Anreize oder Motivatoren untereinander kompensierbar (z. B. Einkommenshöhe und Sicherheit des Arbeitsplatzes, vgl. Schwoerer & Rosen, 1989).

Bedeutung für die individuelle Entscheidung, einer Organisation beizutreten, wird auch der Art der Vorinformation zugeschrieben, die Bewerber über die Organisation und ihre künftige Tätigkeit erhalten. Der Befürchtung, durch Offenlegen auch negativer Aspekte qualifizierte Bewerber abzuschrecken, wurde die Erwartung entgegengestellt, daß insbesondere diese Gruppe Offenheit als Stärke honorieren werde und deshalb die Organisation anziehender finde (im Sinne der zweiseitigen Kommunikation in der Einstellungsforschung). Insbesondere aber erhoffte man sich durch freimütige Information eine verminderte Gefahr des «Berufseintrittsschocks», größere Rollenklarheit, höhere Arbeitszufriedenheit und stärkere Bindung an das Unternehmen, zumal bei neueingetretenen Organisationsmitgliedern die höchsten Fluktuationsraten auftreten (Wanous, 1980). Die zur Wirkung der Vorinformation durchgeführten Untersuchungen (vgl. die Metaanalyse von Premack & Wanous, 1985) kommen zum Schluß, daß der Einfluß auf die Annahme eines Stellenangebots geringfügig ist, daß aber im späteren Berufsalltag weniger enttäuschte Erwartungen verzeichnet werden und die Fluktuation etwas reduziert wird, insbesondere bei komplexeren Tätigkeiten. Möglichkeiten zu realitätsgerechter Vorinformation bieten sich u.a. durch das Einstellungsgespräch sowie durch das Kennenlernen des Arbeitsplatzes, der Kollegen und Vorgesetzten.

6.3 Reaktionen auf Auswahl und Beurteilung

Daß eignungsbezogene Personalauswahl nicht nur der Organisation, sondern auch

dem Kandidaten zugutekommt, wird seit Münsterberg (1912) immer wieder kolportiert. Tatsächlich ist es nur zu plausibel, daß jemand in einer Position, die seinen Fähigkeiten entspricht und die ihn weder über- noch unterfordert, nicht nur erfolgreicher ist - das ist belegt -, sondern auch zufriedener und gesünder. Hierfür findet sich mancherlei Evidenz aus der Streßforschung und aus Alltagsbeobachtungen, aber es fehlen direkte Belege im Kontext der Eignungsdiagnostik. Zu wenige Validierungsstudien bedienen sich anderer Erfolgskriterien als solcher der Berufsleistung, insbesondere werden kaum subjektive Erfolgsmaße erhoben. Eine andere Argumentation bezieht sich darauf, daß Organisationen mit eignungsbezogenen Auswahlprogrammen wirtschaftlich erfolgreicher sind und deshalb ihren Mitgliedern bessere und sicherere Arbeitsplätze bieten können. Aber auch dies gilt unmittelbar nur für die ausgewählten Bewerber, für die abgelehnten hingegen allenfalls in einem allgemeinen volkswirtschaftlichen Sinn; ob sie an anderer Stelle eine besser passende Tätigkeit finden, ist ungewiß. Empirische Forschung fehlt hier, vor allem fehlen Modelle der Nutzensschätzung, zumal solche, die auch subjektive, nicht-monetäre Nutzenelemente einbeziehen.

Was aber vorliegt, sind verschiedene Forschungsergebnisse zur Reaktion auf bestimmte Auswahlverfahren sowie auf Leistungsbeurteilungen und auf Leistungsfeedback. Zusammenstellungen dieser Ergebnisse finden sich bei Schuler und Moser (1992) und bei Schuler, Farr und Smith (1993); an dieser Stelle müssen einige beispielhaft ausgewählte Ergebnisse genügen.

Daß Assessment Center bei kompetenter Durchführung von den Teilnehmern positiv eingeschätzt werden, ist seit längerem bekannt. Holling und Leippold (1991) verglichen die Einschätzungen vor und nach der Verfahrensdurchführung und stellten fest, daß sich die Einstellungen nach der Durchführung noch weiter verbessert hatten. Dies galt umso mehr, je positiver die Beurteilungen der Kandidaten ausfielen.

Studierende verschiedener Fachgebiete wurden von Fruhner, Schuler, Funke und

Moser (1991) ersucht, verschiedene Auswahlinstrumente vergleichend zu bewerten. Als generelle Präferenz ergab sich folgende Rangreihe, die sich als weitgehend stabil über alle Gruppen hinweg erwies:

1. Vorstellungsgespräch
2. Arbeitsprobe
3. Praktikum
4. Schulnoten
5. Psychologische Tests
6. Lebenslauf
7. Handschrift
8. Losverfahren

Als Ursache der unterschiedlichen Bewertung stellten sich in dieser und verwandten Untersuchungen u. a. die Faktoren Transparenz, Handlungskontrolle, Streß und Fairneß heraus. Fähigkeits- und Leistungstests wurden positiver beurteilt als Persönlichkeitstests.

Auch innerhalb einzelner Verfahrenstypen lassen sich Unterschiede finden, die im wesentlichen mit dem Verhaltensstil der Auswählenden zusammenhängen. So wurden in einer bei Schuler (1990) berichteten Studie nondirektiv vorgehende Interviewer als angenehmer und als leichter beeinflussbar beurteilt als direktiv vorgehende Interviewer. Dies korrespondiert dem von Latham und Finnegan (1993) berichteten Ergebnis, daß vollstandardisierte Interviews zwar von Interviewern und Arbeitsrichtern, nicht aber von Bewerbern positiv beurteilt werden. Schmitt und Coyle (1976) hatten gefunden, daß die Entscheidung, ein Arbeitsplatzangebot anzunehmen, wesentlich vom Eindruck abhing, den die Bewerber vom Interviewer gewonnen hatten. Der ideale Repräsentant einer Organisation in den Augen der Bewerber ist nach Rynes (1993) eine warmherzige, offene Persönlichkeit, die detaillierte Information darüber gibt, inwieweit die Fertigkeiten und Interessen des Bewerbers den Anforderungen der Tätigkeit entsprechen, die glaubwürdig wirkt, hohen Status hat und im übrigen dem Bewerber in Ausbildung oder Position ähnlich ist.

Bezüglich der Wirkung der Beurteilung auf beurteilte Personen scheint es zweckmäßig, zwei Hauptfaktoren zu unterscheiden: das

Wissen um die eigene Leistung und die Kritik durch eine andere Person. Isoliert betrachtet, führen Zielsetzung und Maßnahmen des Verhaltens-Ziel-Vergleichs fast durchgängig zu positiven Leistungskonsequenzen (Landy & Farr, 1983; Locke & Latham, 1990). Die Wirkung ist um so förderlicher, je kleiner die Einheiten sind und je verhaltensbezogener die Information ist.

Weit differenzierter ist die Reaktion auf die Bewertung durch eine andere Person zu sehen. Schon die Studien von Meyer, Kay und French (u. a. 1965) hatten ergeben, daß Lob uneindeutige und Kritik negative Effekte auf Leistung und Befinden hatten und daß Kritik häufig defensive Reaktionen provoziert. Die Schlußfolgerungen hieraus werden auch von den meisten neueren Autoren geteilt: Zielsetzung sollte gemeinsam vorgenommen werden; «Coaching» erfolgt am besten unmittelbar und direkt verhaltensbezogen, nicht summarisch nach langen Zeitabständen; die Form der Rückmeldung ist von der Zielsetzung abhängig zu machen; vor allem sollten Gespräche, die der Leistungsverbesserung dienen, nicht gleichzeitig Bezahlung und Beförderung zum Inhalt haben.

Im einzelnen ist noch zu wenig bekannt, wovon die Reaktion auf Beurteilung abhängt. Spekulativ können genannt werden: «Vorzeichen» der Beurteilung, Konsequenzen der Beurteilung, überdauernde Bewertungsangst/Selbstsicherheit, Übereinstimmung mit Selbstbeurteilung, Vertrauen in Beurteiler und Organisation, Macht und Kompetenz des Beurteilers, Beurteilungsverfahren, Validitätseinschätzung, Transparenz, Partizipation, Kompatibilität mit eigenen Erwartungen und Zielsetzungen, Vertraulichkeit, Art des Feedback, organisationaler und sozialer Kontext. Eine Zusammenfassung der existierenden Literatur zur Reaktion auf systematische Beurteilung gibt Dickinson (1993) eine Zusammenstellung der Forschung zum Beurteilungsgepräch findet sich bei Farr (1993) Ratschläge für das Führen von Mitarbeitergesprächen allgemein gibt Neuberger (1991).

7. Literatur

- Albright, L., Kenny, D.A. & Malloy, T.E. (1988). Consensus in personality judgements at zero acquaintance. *Journal of Personality and Social Psychology*, 55, 387-395.
- Algera, J.A. & Greuter, M.A.M. (1989). Job analysis for personnel selection. In M. Smith & I. Robertson (Eds.), *Advances in selection and assessment* (pp. 7-30). Chichester: Wiley.
- Arvey, R.D. & Campion, J.E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology*, 35, 281-322.
- Baron-Boldt, J., Schuler, H. & Funke, U. (1988). Prädiktive Validität von Zeugnisnoten - Eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 2, 79-90.
- Barrick, M.R. & Mount, M.K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barthel, E. & Schuler, H. (1989). Nutzenkalkulation eignungsdiagnostischer Verfahren am Beispiel eines biographischen Fragebogens. *Zeitschrift für Arbeits- und Organisationspsychologie*, 33, 73-83.
- Barthel, E. & Stehle, W. (1990). Biographisches Profil erfolgreicher Mitarbeiter im Versicherungsaußendienst. In H. Schuler & W. Stehle (Hrsg.), *Biographische Fragebogen als Methode der Personalauswahl* (2. Aufl., S.80-90). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Bartscher, T.R. & Fritsch, S. (1992). Personalmarketing. In E. Gaugler & W. Weber (Hrsg.), *Handwörterbuch des Personalwesens* (2. Aufl., Spalte 1747-1758). Stuttgart: Poeschel.
- Ben-Shakhar, G., Bar-Hillel, M., Bilu, Y., Ben-Abba, E. & Flug, A. (1986). Can graphology predict occupational success? Two empirical studies and some methodological ruminations. *Journal of Applied Psychology*, 71, 645-653.
- Bernardin, H.J. (1977). Behavioral expectation scales versus summated scales: A fairer comparison. *Journal of Applied Psychology*, 62, 422-427.
- Bernardin, H.J. & Beatty, R.W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston: Kent.
- Binning, J.F. & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, 74, 478-494.
- Booth, J.F. (1991). Die Anwenderschnittstelle - Schlüssel zum anwenderfreundlichen und validen computergestützten Testen. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 70-77). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Borkenau, P. (1991). Sind erste Eindrücke valide, und wenn ja, warum? In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 252-255). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Borman, W.C. (1978). Exploring upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology*, 63, 135-144.
- Boudreau, J.W. (1989). Selection utility analysis. In M. Smith & I. Robertson (Eds.), *Advances in selection and assessment* (pp. 227-257). New York: Wiley.
- Bownas, D.A. & Bernardin, H.J. (1988). Critical incident technique. In S. Gael (Eds.), *The job analysis handbook for business, industry, and government* (Vol. 2, pp. 1120-1137). New York: Wiley.

- Brambring, M. (1983). Spezielle Eignungsdiagnostik. In K.J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik. Enzyklopädie der Psychologie, B/11/2* (S. 414-481). Göttingen: Hoefefe.
- Brandstädter, H. (1969). *Soziale Urteilsbildung in Organisationen*. Unveröff. Habilitationsschrift. Universität München.
- Brandstädter, H. (1970). Die Beurteilung von Mitarbeitern. In A. Mayer & B. Herwig (Hrsg.), *Betriebspsychologie* (S. 558-734). Göttingen: Hogrefe.
- Brandstädter, H. (1982). Psychologische Grundlagen personeller Entscheidungen. In H. Schuler & W. Stehle (Hrsg.), *Psychologie in Wirtschaft und Verwaltung* (S. 19-47). Stuttgart: Poeschel.
- Brandstädter, H., Kunkel, W. & Rosenstiel, L. v. (1968). Zur Diagnostik des Gefühls als des Zentrums der Integration. In: J. Tenzler (Hrsg.), *Wirklichkeit der Mitte* (S.504-550). Freiburg: Alber.
- Brandstädter, H. & Schuler, H. (1974). *Overcoming halo and leniency: A new method of merit rating*. Paper presented at the 18 International Congress of Applied Psychology, Montreal.
- Bray, D.W.: Campbell, R.J. & Grant, D.L. (1974). *Formative years in business: A long-term AT& T study of managerial lives*. New York: Wiley.
- Brickenkamp, R. (1975). *Handbuch psychologischer und pädagogischer Tests*. Göttingen: Hogrefe.
- Brickenkamp, R. (1983). *Erster Ergänzungsband zum Handbuch psychologischer und pädagogischer Tests*. Göttingen: Hogrefe.
- Brogden, H.E. (1949). When testing pays off. *Personnel Psychology*, 2, 171-183.
- Burke, M.J. & Day, R.R. (1986). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology*, 71, 232-245.
- Campbell, J.P., Dunnette, M.D., Lawler, E.E. & Weick, K.E. (Eds.). (1970). *Managerial behavior, performance, and effectiveness*. New York: McGraw-Hill.
- Cascio, W.F. (1987). *Applied psychology in personnel management* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Cascio, W.F., Alexander, R.A. & Barrett, G.V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. *Personnel Psychology*, 41, 1-24.
- Cooper, W.H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90, 218-244.
- Cornelius, E.T. & Lyness, K.S. (1980). A comparison of holistic and decomposed judgment strategies in job analysis by job incumbents. *Journal of Applied Psychology*, 65, 155-163.
- Cronbach, L.J. & Gleser, G.C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana: University of Illinois Press.
- Davis, C. & Cowles, M. (1989). Automated psychological testing: Method of administration, need for approval, and measures of anxiety. *Educational and Psychological Measurement*, 49, 311-320.
- Dickinson, T.L. (1993). Attitudes about performance appraisal. In H. Schuler, J.L. Farr & M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 141-161). Hillsdale, NJ: Erlbaum.
- Dörner, D. (1986). Diagnostik der operativen Intelligenz. *Diagnostica*, 32, 290-308.
- Downs, S., Farr, R.M. & Colbeck, L. (1978). Self appraisal: A convergence of selection and guidance. *Journal of Occupational Psychology*, 51, 271-278.
- Drauden, G.M. (1988). Task inventory analysis in industry and the public sector. In S. Gael (Ed.), *The job analysis handbook for business, industry, and government* (Vol. 2, pp. 1051-1071). New York: Wiley.
- Dreher, G.F., Ash, R.A. & Hancock, P. (1988). The role of traditional research design in underestimating the validity of the employment interview. *Personnel Psychology*, 41, 315-327.
- Dunnette, M.D., McCartney, J., Carlson, H.C. & Kirchner, W.K. (1962). A study of faking behavior on a forced-choice self-description Checklist. *Personnel Psychology*, 15, 13-24.
- Eckardt, H.-H. (1991). Die Unterstützung individueller beruflicher Entscheidungen durch den Einsatz von Tests und Selbsterkundungsverfahren. In K.H. Ingenkamp & R.S. Jäger (Hrsg.), *Tests und Trends* 9 (S. 109-150). Weinheim: Beltz.
- Eder, R.W. & Ferris, G.R. (Eds.). (1989). *The employment interview: Theory, research, and practice*. Beverly Hills: Sage.
- Eder, R.W., Kacmar, K.M. & Ferris, G.R. (1989). Employment interview research: History and Synthesis. In R.W. Eder & G.R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 17-31). Beverly Hills: Sage.
- Eichel, E. & Bender, H.E. (Eds.). (1984). *Performance appraisal: A study of current techniques*. New York: American Management Association.
- Engelbrecht, W. (1991). Maschinelle berufsbezogene Testbefundinterpretation bei der Bundesanstalt für Arbeit. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 88-90). Göttingen: Hogrefe/Verlag für Armewandte Psychologie.
- England, G.W. (1971). *Development and use of weighted application blanks* (rev. ed.). Minneapolis: Industrial Relations Center, University of Minnesota.
- Facaoaru, C. & Frieling, E. (1985). Verfahren zur Ermittlung informatorischer Belastungen. Teil 1: Theoretische und konzeptionelle Grundlagen. Teil 2: Aufbau und Darstellung eines Verfahrensentwurfs. *Zeitschrift für Arbeitswissenschaft*, 39, 65-72 und 90-96.
- Farr, J.L. (1993). Informal performance feedback: Seeking and giving. In H. Schuler, J.L. Farr & M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 163-180). Hillsdale, NJ: Erlbaum.
- Feldman, J.M. (1981). Beyond attribution theory: Cognitive processes in Performance appraisal. *Journal of Applied Psychology*, 66, 127-148.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Flanagan, J.C. & Burns, R.K. (1955). The employee performance record: A new appraisal and development tool. *Harvard Business Review*, 33, 95-102.
- Fleishman, E.D. & Ouaintance, M.K. (1984). *Taxonomies of human Performance*. Orlando: Academic Press.
- Frieling, E. (1977). Die Arbeitsplatzanalyse als Grundlage der Eignungsdiagnostik. In J.K. Triebe & E. Ulich (Hrsg.), *Beiträge zur Eignungsdiagnostik* (S. 20-90). Bern: Huber.
- Frieling, E., Facaoaru, J., Benedix, J., Pfaus, H. & Sonntag, K. (1990). *Tätigkeits-Analyse-Inventar*. Kassel: Universität Gesamthochschule, Fachgebiet Arbeitswissenschaft.
- Frieling, E. & Hoyos, C. Graf (1978). *Fragebogen zur Arbeitsanalyse*. Bern: Huber.
- Fruhner, R., Schuler, H., Funke, U. & Moser, K. (1991). Einige Determinanten der Bewertung von Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 35, 170-178.

- Funke, J. (1986). *Komplexes Problemlösen*. Berlin: Springer.
- Funke, U. (1986). Die Validität verschiedener eignungsdiagnostischer Verfahren bei Lehrstellenbewerbern. *Psychologie und Praxis. Zeitschrift für Arbeits- und Organisationspsychologie*, 30, 92-97.
- Funke, U. (1989). Die Wirksamkeit von Führungsstrainings. Eine Bestandsaufnahme in zwei Übersichtsstudien. *Zeitschrift für Arbeits- und Organisationspsychologie*, 33, 51-52.
- Funke, U. (1991). Die Validität einer computergestützten Systemsimulation zur Diagnose von Problemlösekompetenz. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 114-122). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Funke, U. (in Vorbereitung). *Arbeitsanalyse und Eignungsdiagnostik: Konzepte und Anwendungen am Beispiel industrieller Forschung und Entwicklung*. Diss., Universität Hohenheim, Stuttgart.
- Funke, U. & Barthel, E. (1990). Nutzen und Kosten von Personalentscheidungen. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 647-658). Göttingen: Hogrefe.
- Funke, U., Krauß, J., Schuler, H. & Stapf, K.H. (1987). Zur Prognostizierbarkeit wissenschaftlich-technischer Leistungen mittels Personvariablen: Eine Metaanalyse der Validität diagnostischer Verfahren im Bereich Forschung und Entwicklung. *Gruppendynamik*, 18, 407-428.
- Funke, U. & Schuler, H. (1986). Personalauswahl im Bereich industrieller Forschung und Entwicklung - Forschungsstand und Perspektiven. *Psychologie und Praxis. Zeitschrift für Arbeits- und Organisationspsychologie*, 30, 34-51.
- Funke, U. & Schuler, H. (1990). Weiterentwicklung biographischer Fragbogen durch Konstruktaufklärung: Grundlagen und erste empirische Ergebnisse. In H. Schuler & W. Stehle (Hrsg.), *Biographische Fragebogen als Methode der Personalauswahl* (2. Aufl., S. 114-139). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Ghiselli, E.E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, 26, 461-477.
- Giese, F. (1924). Die Arbeitsprobe in der Psychodiagnostik. *Zeitschrift für Angewandte Psychologie*, 23, 3-4.
- Guion, R.M. (1976). Recruiting, selection and job placement. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 777-828). New York: Wiley.
- Hacker, W. (1983). Psychische Anforderungen. In F. Stoll (Hrsg.), *Arbeit und Beruf: Kindlers Psychologie des 20. Jahrhunderts* (S. 86-109). Weinheim: Beltz.
- Heinze, B. (1990). Graphologie. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 368-371). Göttingen: Hogrefe.
- Hirsh, H.R., Schmidt, E.L. & Hunter, J.E. (1986). Estimation of employment validities by less experienced judges. *Personnel Psychology*, 39, 337-344.
- Holling, H. & Leippold, W. (1991). Zur sozialen Validität von Assessment Centern. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 305-312). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Hollmann, H. (1991). *Validität in der Eignungsdiagnostik. Neue Ansätze einer sachgerechten Bewertung und effizienten Verbesserung*. Göttingen: Hogrefe.
- Horn, R. (1986). *Alle wichtigen Tests zur Auswahl von Bewerbern*. München: Heyne.
- Hough, L.M. (1984). Development and evaluation of the <Accomplishment Record> method of selecting and promoting professionals. *Journal of Applied Psychology*, 69, 135-146.
- Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D. & McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75, 581-595.
- Hunter, J.E. & Hirsh, H.R. (1987). Applications of meta-analysis. In C.L. Cooper & I.T. Robertson (Eds.), *International Review of industrial and organizational psychology* (pp. 321-357). New York: Wiley.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and Utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hunter, J.E. & Schmidt, E.L. (1982). Fitting people to jobs: The impact of personnel selection on national productivity. In M.D. Dunnette & E.A. Fleishman (Eds.), *Human Performance and productivity* (Vol. 1, pp. 233-284). Hillsdale, NJ: Erlbaum.
- Hunter, J.E. & Schmidt, E.L. (1990). *Methods of meta-analysis. Correcting error and bias in research findings*. Newbury Park: Sage.
- Husted, H. (1991). Dezentrales Testvorgabe- und Testauswertungssystem im Psychologischen Dienst der Bundesanstalt für Arbeit (DELTA): Ziele, Realisierungen, Entwicklungsvorhaben. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 132-134). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Ilgen, D.R. & Feldman, J.M. (1983). Performance appraisal: A process focus. In B.M. Staw & L.L. Cummings (Eds.), *Research in organizational psychology* (pp. 141-197). Greenwich: JAI Press.
- Ilgen, D.R. & Schneider, J. (1991). Performance measurement: A multi-discipline view. In C.L. Cooper & I.T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology 1991* (Vol. 6, pp. 71-108). Chichester: Wiley.
- Ivancevich, J.M. & Donnelly, J.M. (1971). Job offers acceptance behavior and reinforcement. *Journal of Applied Psychology*, 55, 119-122.
- Jäger, R.S. (1982). Diagnostische Urteilsbildung. In K.J. Groffmann & L. Michel (Hrsg.), *Grundlagenpsychologischer Diagnostik. Enzyklopädie der Psychologie, B/II/1* (S. 295-375). Göttingen: Hogrefe.
- Jäger, R.S. (1988). Biographische Daten. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (S. 292-303). München: Psychologie Verlags Union.
- Jäger, R.S. (1990). Computardiagnostik-ein oberblick. *Diagnostica*, 36, 96-114.
- Janke, W. (1982). Klassenzuordnung. In K.J. Groffmann & L. Michel (Hrsg.), *Grundlagen psychologischer Diagnostik. Enzyklopädie der Psychologie, B/II/1* (S. 376-466). Göttingen: Hogrefe.
- Janz, J.T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology*, 67, 577-580.
- Kisser, R. (1988). Adaptive Strategien. In R.S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (S. 123-130). München: Psychologie Verlags Union.
- Klieme, E. & Stumpf, H. (1990). Computereinsatz in der pädagogisch-psychologischen Diagnostik. In K. Ingenkamp & R.S. Jäger (Hrsg.), *Tests und Trends 8* (S. 13-63). Weinheim: Beltz.
- Klimoski, R.J. & Raffaelli, A. (1983). Inferring personality qualities through handwriting analysis. *Journal of Occupational Psychology*, 56, 191-202.
- Kluwe, R. (1990). Computergestützte Systemsimulation

- nen. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 458-463). Göttingen: Hogrefe.
- Kompa, A. (1984). Personalbeschaffung und Personalauswahl. Stuttgart: Enke.
- Landy, F. (1988). Selection procedure development and usage. In S. Gael (Ed.), *The job analysis handbook for business, industry, and government* (Vol. 1, pp. 271-287). New York: Wiley.
- Landy, F.J. & Farr, J.L. (Eds.). (1983). *The measurement of work performance*. Orlando: Academic Press.
- Latham, G.P. (1986). Job Performance and appraisal. In C.L. Cooper & I.T. Robertson (Eds.), *International review of industrial and organizational psychology 1986* (pp. 117-155). Chichester: Wiley.
- Latham, G.P. & Finnegan, B. (1993). The perceived practicality of unstructured, patterned and situational interviews. In H. Schuler, J.L. Farr & M. Smith, (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 41-55). Hillsdale, NJ: Erlbaum.
- Latham, G.P. Saari, L.M., Pursell, E.D. & Campion, M.A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.
- Latham, G.P. & Wexley, K.N. (1977). Behavioral observation scales for performance appraisal purposes. *Personnel Psychology*, 30, 255-268.
- Lattmann, C. (Hrsg.) (1989). *Das Assessment-Center-Verfahren der Eignungsbeurteilung*. Heidelberg: Physica.
- Lazer, R.I. & Wikstrom, W.S. (Eds.). (1977). *Appraising managerial performance: Current practices and future directions*. New York: Conference Board.
- Locke, E.A. & Latham, G.P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice-Hall.
- Lord, R.G. (1985). Accuracy in behavioral measurement: An alternative definition based on raters' cognitive schema and signal detection theory. *Journal of Applied Psychology*, 70, 66-71.
- Lorenzo, R.V. (1984). Effects of assessorship on manager's proficiency in acquiring, evaluating, and communicating information about people. *Personnel Psychology*, 37, 617-634.
- McCormick, E.J. & Jeanneret, P.R. (1988). Position analysis questionnaire (PAQ). In S. Gael (Ed.), *The job analysis handbook for business, industry, and government* (Vol. 2, pp. 825-842). New York: Wiley.
- McDaniel, M.A., Whetzel, D.L., Schmidt, F.O., Hunter, J.E., Maurer, S. & Russel, J. (1986). *The validity of employment interviews: A review and meta-analysis*. Unpublished manuscript, US Office of Personnel, Washington DC.
- McHenry, J., Hough, L.M., Toquam, J.L., Hanson, M.A. & Ashworth, S. (1990). Projekt A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43, 335-354.
- Meyer, H., Kay, E. & French, J.R.P. (1965). Split roles in performance appraisal. *Harvard Business Review*, 43, 123-129.
- Miner, J.B. (1988). Development and application of the rated ranking technique in Performance appraisal. *Journal of Occupational Psychology*, 61, 291-305.
- Moser, K. (1992). Einführung und Überblick. In K. Moser, W. Stehle & H. Schuler (Hrsg.), *Personalmarketing* (S.1-18). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Münsterberg, H. (1912). *Psychologie und Wirtschaftsleben*. Leipzig: Barth.
- Murphy, K.R., Martin, C. & Garcia, M. (1982). Do behavioral observation scales measure observation? *Journal of Applied Psychology*, 67, 562-567.
- Neuberger, O. (1974). Das Einstellungsgespräch. *Psychologie und Praxis*, 18, 8-22.
- Neuberger, O. (1990). *Miteinander arbeiten - miteinander reden!* (12. Aufl.). München: Bayerisches Staatsministerium für Soziales und Gesundheit.
- Neuberger, O. (1991). *Organisationspsychologie: Eine Disziplin auf der Suche nach ihrem Gegenstand*. Augsburg: Beiträge zu Organisationspsychologie und Personalwesen, Heft 12.
- Odiome, G.S. (Ed.). (1965). *Management by objectives: A system of managerial leadership*. Belmont: Pitman.
- Owens, W.A. (1976). Background data. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 609-644). Chicago: Rand McNally.
- Peterson, N.G. & Bownas, D.A. (1982). Skill, task structure, and performance acquisition. In M.D. Dunnette & E.A. Fleishman (Eds.), *Human performance and productivity: Human capability assessment* (Vol. 1, pp. 49-105). Hillsdale, NJ: Erlbaum.
- Porter, L.W., Lawler, E.E.III & Hackman, J.R. (1975). *Behavior in Organizations*. New York: McGraw Hill.
- Powell, G.N., Kiman, J., Farley, J.A. & Geisinger, K.F. (1989). The relationship between recruiting source, applicant quality, and hire performance: An analysis by sex, ethnicity, and age. *Personnel Psychology*, 42, 293-308.
- Premack, S.L. & Wanous, J.P. (1985). A meta-analysis of realistic job preview experiments. *Journal of Applied Psychology*, 70, 706-719.
- Pulling, K.K. (1990). Relevante Personengruppen und deren Rekrutierung. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 55-64). Göttingen: Hogrefe.
- Pursell, E.D., Campion, M.A. & Gaylord, S.R. (1980). Structured interviewing: Avoiding selection problems. *Personnel Journal*, 11, 907-912.
- Putz-Osterloh, W. (1991). Computergestützte Eignungsdiagnostik: Warum Strategien informativer als Leistungen sein können. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 97-102). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Reilly, R.R. & Chao, G.T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, 35, 1-62.
- Robertson, I.T. & Downs, S. (1989). Work-sample tests of trainability: A meta-analysis. *Journal of Applied Psychology*, 74, 402-410.
- Robertson, I.T. & Kandola, R.S. (1982). Work sample tests: Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology*, 55, 171-183.
- Roest, F. & Horn, R. (1990). Mailbox-90: Computergestützte Diagnostik im Assessment Center. *Diagnostica*, 36, 213-219.
- Rynes, S.L. (1993). When recruitment fails to attract: Individual expectations meet organizational realities in recruitment. In H. Schuler, J.L. Farr & M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 27-40). Hillsdale, NJ: Erlbaum.
- Sackett, P.R. & Wilson, M.A. (1982). Factors affecting the consensus judgement process in managerial assessment centers. *Journal of Applied Psychology*, 67, 10-17.
- Sackett, P.R., Zedeck, S. & Fogli, R. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, 73, 482-486.

- Samson, G.E., Graue, E.M., Weinstein, T. & Walberg, H.J. (1984). Academic and occupational performance: A quantitative Synthesis. *American Educational Research Journal*, 21, 311-321.
- Schein, E.A. (1965). *Organizational psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Schmidt, E.L., Caplan, J.R., Bemis, S.E., Decuir, R., Dunn, L. & Antone, L. (1979). *The behavioral consistency method of unassembled examining* (Technical publication 79-21). Washington DC: Office of Personnel Management, Personnel Research and Development Center.
- Schmidt, E.L. & Hunter, J.E. (1981). Employment test-retest. Old theories and new research findings. *American Psychologist*, 36, 1128-1137.
- Schmidt, F.L., Hunter, J.E., McKenzie, R.C. & Muldrow, T.W. (1979). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology*, 64, 609-626.
- Schmitt, N. (1976). Social situational determinants of interview decisions: Implications for the employment interview. *Personnel Psychology*, 29, 79-101.
- Schmitt, N. & Coyle, B. (1976). Applicant decisions in the employment interview. *Journal of Applied Psychology*, 61, 184-192.
- Schmitt, N., Gooding, R.Z., Noe, R.D. & Kirsch, M. (1984). Metaanalysis of validity studies published between 1974 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422.
- Schmitt, N. & Ostroff, C. (1986). Operationalizing the «behavioral consistency» approach: Selection test development based on a content-oriented strategy. *Personnel Psychology*, 39, 91-108.
- Schneider, B. & Schmitt, N. (1986). *Staffing organizations*. Glenview: Scott, Foresman and Company.
- Schneider, W., Heim, H. & Wacker, P.A. (1975). *Tätigkeitsspezifische Eignungstests*. Göttingen: Schwarz.
- Scholz, G. & Schuler, H. (1993). Das nomologische Netzwerk des Assessment Centers: eine Metaanalyse. *Zeitschrift für Arbeits- und Organisationspsychologie* 37, 73-95.
- Schor, A. (1991). Diagnostische Praxis in der Arbeits- und Organisationspsychologie. Teilergebnisse aus einer repräsentativen Umfrage zur diagnostischen Praxis. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 6-14). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Schule, H. (1989a). Construct validity of a multimodal employment interview. In B.J. Fallon, H.P. Pfister & J. Brebner (Eds.), *Advances in Industrial Organizational Psychology* (pp. 343-354). North Holland: Elsevier.
- Schuler, H. (1989b). Leistungsbeurteilung. In E. Roth (Hrsg.), *Organisationspsychologie. Enzyklopädie der Psychologie D/III/3* (S. 399-430). Göttingen: Hogrefe.
- Schuler, H. (1990). Personalauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 34, 184-191.
- Schuler, H. (1991a). Auswahl von Mitarbeitern. In L. v. Rosenstiel, E. Regnet & M. Domsch (Hrsg.), *Führung von Mitarbeitern* (S. 100-125). Stuttgart: Schäffer.
- Schuler, H. (1991b). Leistungsbeurteilung- Funktionen, Formen und Wirkungen. In H. Schuler (Hrsg.), *Beurteilung und Förderung beruflicher Leistung* (S. 11-40). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Schuler, H. (1992). Das Multimodale Einstellungsinterview. *Diagnostica*, 38, 1-20.
- Schuler, H., Farr, J.L. & Smith, M. (Eds.). (1993). *Personnel selection and assessment: Individual and organizational perspectives*. Hillsdale, NJ: Erlbaum.
- Schuler, H., Frier, D. & Kauffmann, M. (1993). *Personalauswahl im europäischen Vergleich*. Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Schuler, H. & Funke, U. (1989). The interview as a multimodal procedure. In R.W. Eder & G.R. Ferris (Eds.), *The employment interview* (pp. 183-192). Newbury Park: Sage.
- Schuler, H. & Funke, U. (Hrsg.). (1991). *Eignungsdiagnostik in Forschung und Praxis. Psychologische Information für Auswahl, Beratung und Förderung von Mitarbeitern*. Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Schuler, H., Funke, U., Moser, K., Donat, M. & Barthelme, D. (1994). *Personalauswahl in Forschung und Entwicklung*. Göttingen: Hogrefe.
- Schuler, H. & Moser, K. (1992). Entscheidung von Bewerbern. In K. Moser, W. Stehle & H. Schuler (Hrsg.), *Personalmarketing* (S. 51-75). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Schuler, H. & Schmitt, N. (1987). Multimodale Messung in der Personalpsychologie. *Diagnostica*, 32, 259-271.
- Schuler, H. & Stehle, W. (Hrsg.). (1990). *Biographische Fragebogen als Methode der Personalauswahl* (2. Aufl.). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Schuler, H. & Stehle, W. (Hrsg.). (1992). *Assessment Center als Methode der Personalentwicklung* (2. Aufl.). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Schwaab, M.O. & Schuler, H. (1991). Die Attraktivität der deutschen Kreditinstitute bei Hochschulabsolventen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 35, 105-114.
- Schwoerer, C. & Rosen, B. (1989). Effects of employment-at-will policies and compensation policies on corporate image and job pursuit intentions. *Journal of Applied Psychology*, 74, 653-656.
- Scott, W.D. (1916). Selection of employees by means of quantitative determinations. *Annals of the American Academy of Political and Social Science*. 65.
- Seibt, H. & Kleinmann, M. (1991). Personalvorauswahl von Bewerbern: Derzeitiger Stand und Alternativen. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 174-177). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Seifert, K.H. (1982). *Die Bedeutung der Beschäftigungsaussichten im Rahmen des Berufswahlprozesses* (Beitr. AB 67). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit.
- Smith, P.C. & Kendall, L.M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47, 149-155.
- Staffelbach, B. (1986). Personal-Marketing. In I.E. Rührli & H.P. Wehrli (Hrsg.), *Strategisches Marketing und Management* (S. 124-143). Bern: Haupt.
- Steeger, F.W. & Aschenbrenner, H. (1991). Computergestützte Diagnostik im Psychologischen Dienst der Bundeswehr. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 123-126). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Streufer, S., Pogash, R. & Piasecki, M. (1988). Simulation-based assessment of managerial competence:

- Reliability and validity. *Personnel Psychology*, 41, 537-557.
- Taylor, H.C. & Russel, J.F. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 23, 565-578.
- Thorndike, R.L. (1949). *Personnel selection*. New York: Wiley.
- Thornton, G.C. & Byham, W.C. (1982). *Assessment Centers and managerial performance*. New York: Academic Press.
- Thornton, G.C. III, Gaugler, B.B., Rosenthal, D.B. & Bentson, C. (1992). Die prädiktive Validität des Assessment Centers - eine Metaanalyse. In H. Schuler & W. Stehle (Hrsg.), *Assessment Center als Methode der Personalentwicklung* (2. Aufl., S. 36-77). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Trost, G. (1986). Die Bedeutung des Interviews für die Diagnose der Studieneignung. Darstellung der internationalen Forschungsergebnisse. In R. Lohölter, K. Hinrichsen, G. Trost & S. Drolshagen (Hrsg.), *Das Interview bei der Zulassung zum Medizinstudium* (S. 49-80). Stuttgart: Schattauer.
- Wanous, J.P. (1980). *Organizational entry*. Reading, MA: Addison-Wesley.
- Wanous, J.P. (1989). Impression management at organizational entry. In R.A. Giacalone & P. Rosenfeld (Eds.), *Impression management in the organization* (pp. 253-267). Hillsdale, NJ: Erlbaum.
- Webster, E.C. (1982). *The employment interview*. Schamberg, Ontario: SIP Publications.
- Weuster, A. (1991). Das Arbeitszeugnis als Instrument der Personalauswahl. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 177-181). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Wherry, R.J. & Bartlett, C.J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology*, 35, 521-551.
- 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.
- Wildgrube, W. (1990). Computergestützte Diagnostik in einer Großorganisation. *Diagnostica*, 36, 127-147.
- Wottawa, H. & Hossiep, R. (1987). *Grundlugen psychologischer Diagnostik*. Göttingen: Hogrefe.
- Wunderer, R. (1991). Personalmarketing. *Die Unternehmung*, 45, 119-131.