

5 Konstruktorientierte Verfahren der Personalauswahl

von Heinz Schuler und Stefan Höft

Inhaltsübersicht

1 Grundlagen	94
1.1 Berufseignung und Eignungsdiagnostik	94
1.2 Drei Ansätze der Eignungsdiagnostik	95
2 Beschreibung ausgewählter Konstrukte und ihrer Erhebungsverfahren	96
2.1 Tests als klassische Meßverfahren	96
2.2 Eine Auswahl konstruktorientierter Ansätze	97
2.2.1 Allgemeine kognitive Fähigkeiten	97
2.2.1.1 Berliner Intelligenzstrukturmodell (BIS-Modell)	97
2.2.1.2 Intelligenz und Beruf	99
2.2.1.3 Validität kognitiver Verfahren für berufseignungsdiagnostische Zwecke	102
2.2.1.4 Allgemeine kognitive Fähigkeit vs. spezifische Fähigkeiten	106
2.2.2 Spezielle kognitive Fähigkeiten	108
2.2.2.1 Allgemeines Wissen und Fachwissen	108
2.2.2.2 Aufmerksamkeit und Konzentration	109
2.2.2.3 Sonstige Leistungsverfahren	110
2.2.3 Allgemeine Persönlichkeitskonstrukte	112
2.2.3.1 Fünf-Faktoren-Modell der Persönlichkeit	112
2.2.3.2 Validität allgemeiner persönlichkeitsorientierter Verfahren	113
2.2.3.3 Kriteriumsabhängige Validität von Persönlichkeits- eigenschaften	115
2.2.4 Spezifische Persönlichkeitseigenschaften: locus of control	117
2.2.5 Motivation und Interessen	118
3 Problemfelder konstruktorientierter Verfahren	122
3.1 Kombiniertes Einsatz von Verfahren	122
3.2 Das „bandwidth-fidelity“-Dilemma	123
3.3 Verfälschungstendenzen	124
3.4 Papier-Bleistift- vs. computerbasierte Testung	126
Zusammenfassung	128
Weiterführende Literatur	129
Literatur	129

1 Grundlagen

In diesem und den beiden folgenden Kapiteln werden die Grundlagen der Berufseignungsdiagnostik und die wichtigsten Methoden der Personalauswahl dargelegt. Die beiden einleitenden Abschnitte geben deshalb als Einführung eine Gegenstandsbeschreibung und stellen die Klassifikation vor, die den eignungsdiagnostischen Verfahrenstypen zugrunde liegt.

1.1 Berufseignung und Eignungsdiagnostik

Berufseignungsdiagnostik bezeichnet die Methodologie der Entwicklung, Prüfung und Anwendung psychologischer Verfahren zum Zwecke eignungsbezogener Erfolgsprognosen und Entscheidungshilfen im beruflichen Kontext. Ihre wissenschaftliche Basis sind vor allem Theorien der Anforderungen, Fähigkeiten und Leistungen sowie Methoden zu deren Messung und Modelle der Klassifikation. Praktische Einsatzbereiche sind alle Arten berufsbezogener Einschätzung, Beratung, Selektion und Zuordnung mit dem Ziel der individuellen Berufs-, Organisations- und Arbeitsplatzwahl, der institutionellen Berufsberatung sowie der Auswahl und Entwicklung von Mitarbeitern in Wirtschafts- und Verwaltungsorganisationen.

Wie im Kapitel über Anforderungsanalyse (vgl. Kapitel 3) bereits angesprochen, heißt Berufseignung, verstanden als Erfolgswahrscheinlichkeit, im Prinzip immer „Eignung wessen wofür“. Das heißt, die Zielgruppen oder Individuen sind ebenso zu bestimmen wie die Zielpositionen. Die Anforderungen sind zu ermitteln, die diese Tätigkeiten oder Positionen an ihre Inhaber stellen, und es sind daraus die erforderlichen Eignungsmerkmale der Positionsinhaber abzuleiten. Um ihre Ausprägung im Einzelfall zu messen und zu vergleichen, stehen eignungsdiagnostische Verfahren zur Verfügung oder werden konstruiert bzw. adaptiert, wenn der Bedarf ein sehr spezifischer ist. Neben der Leistung können auch andere Kriterien beruflichen Erfolgs, namentlich Berufs- und Arbeitszufriedenheit, Gesundheit und Wohlbefinden oder das Verbleiben in der Organisation als Diagnoseziele formuliert werden. Dementsprechend ist in Ergänzung der Anforderungsermittlung auch das Befriedigungspotential der fraglichen Tätigkeit zu bestimmen und mit den Interessen und Bedürfnissen der Personen zu vergleichen.

Zusätzlich zur Bestimmung derzeitiger Anforderungen kann abzuschätzen versucht werden, welche Anforderungsänderungen zu erwarten sind; überdies ist mit einem nicht bestimmaren Anteil an Veränderungen zu rechnen. Das erwünschte Entwicklungspotential einer Person sollte sowohl den absehbaren Veränderungen entsprechen als auch darüber hinaus ausreichende Wahrscheinlichkeit bieten, daß sie auch künftigen Entwicklungen ungewisser Art gewachsen sein wird. Die Prognose ist in diesem Bereich schwierig, aber es haben sich doch einige Eigenschaften als erfolgsrelevant in unterschiedlichstem Berufskontext herausgestellt.

Daraus ergibt sich die Notwendigkeit, Person und Berufstätigkeit auf drei Ebenen zu vergleichen, jener der tätigkeitsspezifischen Anforderungen und Qualifikationen, der tätigkeitsübergreifenden Qualifikationen sowie auf der Ebene von Befriedigungsangebot und -bedarf. In Abbildung 1 erkennt man die auf den Aspekt der Berufseignung reduzierte Abbildung 1 aus Kapitel 3 (Anforderungsanalyse) wieder.

Tätigkeit	Person
Anforderungen	Fähigkeiten, Fertigkeiten und Kenntnisse
Befriedigungspotential	Interessen, Bedürfnisse und Werthaltungen
Veränderung	Entwicklungspotential und allgemein erfolgsrelevante Merkmale

Abbildung 1:

Vergleich zwischen Tätigkeit und Person bei berufsbezogenen Entscheidungen (aus Schuler, 1996)

Berufseignungsdiagnostik: Methodologie der Entwicklung, Prüfung und Anwendung psychologischer Verfahren zur eignungsbezogenen Erfolgsprognose und als Entscheidungshilfe im beruflichen Bereich

Person und Berufstätigkeit müssen für Eignungsuntersuchungen auf drei Ebenen verglichen werden.

Für die Untersuchung der Tätigkeit stehen verschiedene Methoden der Arbeits- und Anforderungsanalyse zur Verfügung, wie an anderer Stelle in diesem Lehrbuch detailliert erörtert wird (vgl. Kapitel 3).

1.2 Drei Ansätze der Eignungsdiagnostik

Zur Klassifikation eignungsdiagnostischer Instrumente ist die Unterscheidung dreier grundsätzlicher methodischer Ansätze hilfreich – des *Eigen-schaftsansatzes*, des *Simulationsansatzes* und des *biographischen Ansatzes* (vgl. Abbildung 2). Jeder dieser Ansätze verfolgt eine teilweise eigenständige Validierungslogik und ist mit speziellen Methoden der Merkmalerfassung verbunden. Mit dem Eigenschafts- oder Konstruktansatz werden Merkmale erfaßt, die als relativ stabil angenommen werden, also z. B. Gewissenhaftigkeit und sprachgebundene Intelligenz; als typische Meßverfahren, diese Merkmale in möglichst homogener Form zu erfassen, werden *psychologische Tests* eingesetzt. Kriterium dafür, wie weit dies gelungen ist, ist die *Konstruktvalidität*. Zielsetzung des Simulationsansatzes ist die Erfassung solchen Verhaltens, das in ähnlicher Form am Arbeitsplatz gefordert wird; das zugehörige Validitätsprinzip ist das der inhaltlichen Repräsentativität oder *Kontentvalidität*, typische Erhebungsform ist die *Arbeitsprobe*. Als drittes diagnostisches Prinzip läßt sich, wenngleich nicht durchgehend trennscharf, der *biographische Ansatz* unterscheiden; typische Erfassungsmethode sind *biographische Fragen*, die in schriftlicher Form im Fragebogen, in mündlicher Form im Interview gestellt werden können. Die Validitätslogik des biographischen Ansatzes ist die der *prognostischen Validität*.

Eigenschafts-, Simulations- und Biographieansatz verfolgen eine eigenständige Validierungslogik und setzen spezifische Methoden zur Merkmalerfassung ein.

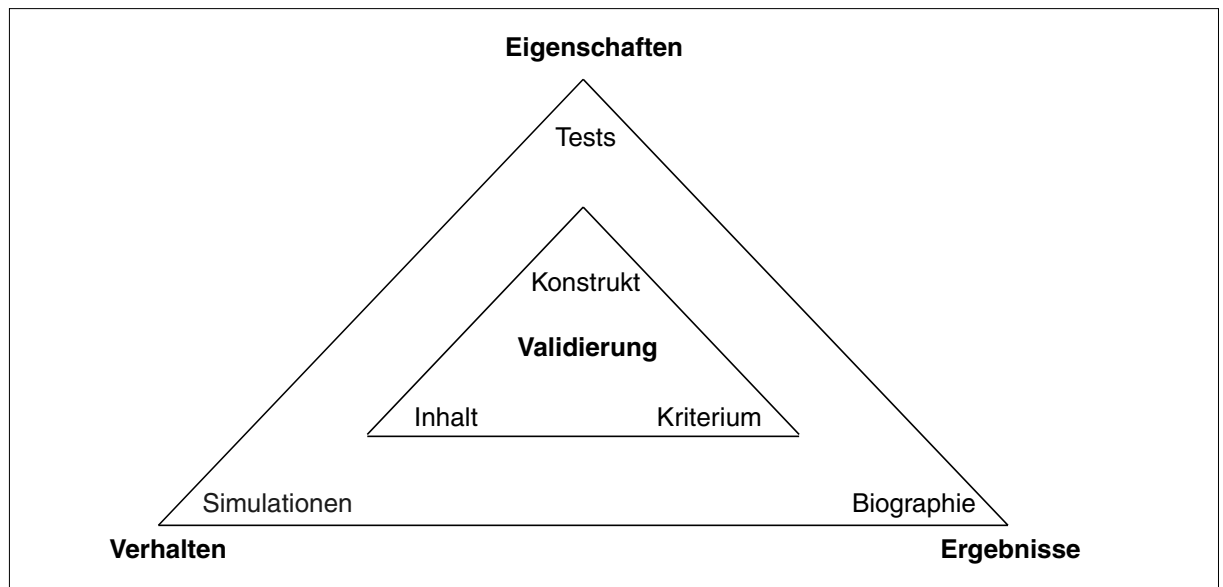


Abbildung 2:

Die Ansätze der Eignungsdiagnostik und ihre Validierungslogik

Für die praktische Berufseignungsdiagnostik bedeutet dies, daß für komplexe Anforderungssituationen zumeist ein multiples Verfahren angemessen ist, denn die Ergebnisse psychologischer Diagnosen sind zum Teil auch von der Erfassungsmethode abhängig: Es kann durch die strukturellen Spezifika der Verfahren keine vollständige Äquivalenz geben, wenn das gleiche Merkmal (z. B. „Kundenorientierung“ oder „Initiative“) mit unterschiedlichen Methoden überprüft wird. Erst durch die systematische Kombination unterschiedlicher Verfahren wird sichergestellt, daß alle relevanten Facetten des interessierenden Merkmalsbereichs abgedeckt werden. Durch das multimethodale Verfahrensrepertoire ist eine große Spannweite möglicher Erhebungsformen abgedeckt, beispielsweise vergangenheitsbezogene wie gegenwärtige Ver-

Die systematische Kombination unterschiedlicher Verfahren gewährleistet, daß alle relevanten Facetten des Merkmalsbereichs abgedeckt werden.

Die Qualität der Potentialdiagnose leitet sich nicht nur aus der Konstruktion der Einzelverfahren, sondern auch aus der theoriegestützten Systematik des Gesamtverfahrens ab.

Tests sind standardisierte, routinemäßig anwendbare Verfahren zur Messung individueller Verhaltensmerkmale, aus denen Schlüsse auf Eigenschaften der betreffenden Person oder auf ihr Verhalten in anderen Situationen gezogen werden können.

haltensbeschreibung, Selbsteinschätzung und Fremdeinschätzung sowie typisches und maximales Verhalten. Durch schriftliche und mündliche Aufgabenvorgabe werden wichtige Stimulusformen variiert, durch die Erfassung mündlicher, schriftlicher und verhaltensbasierter Reaktionen wird ein breiter Responsebereich abgedeckt. Ein multimethodales Auswahlverfahren gewährleistet eine erhöhte Generalisierbarkeit der eignungsdiagnostischen Befunde auf zukünftiges berufliches Verhalten. Darüber hinaus hat die Messung eines Merkmals mit mehreren Methoden den Vorteil höherer Reliabilität, also Verlässlichkeit der Meßwerte. Hohe Qualität der Potentialdiagnose ergibt sich also nicht allein aus der Konstruktion der Einzelverfahren, sondern auch aus der theoriegestützten Systematik des Gesamtverfahrens.

2 Beschreibung ausgewählter Konstrukte und ihrer Erhebungsverfahren

2.1 Tests als klassische Meßverfahren

Die klassische Methode, psychologische Konstrukte zu messen, ist der Test. Psychologische Tests stellen gewissermaßen den Prototyp konstruktorientierter Diagnoseverfahren dar. Sie können mit Brandstätter (1979, S. 82) als standardisierte, routinemäßig anwendbare Verfahren zur Messung individueller Verhaltensmerkmale aufgefaßt werden, aus denen Schlüsse auf Eigenschaften der betreffenden Person oder auf ihr Verhalten in anderen Situationen gezogen werden können. In der wissenschaftlich kontrollierten Eignungsdiagnostik sind Tests die am häufigsten verwendeten Instrumente; ihr Einsatz erstreckt sich auf praktisch alle Berufsbereiche und beruflich relevanten Fähigkeiten. Die wichtigsten Arten von Testverfahren, die in der Berufseignungsdiagnostik eingesetzt werden, sind

- Allgemeine Intelligenztests;
- Tests spezifischer kognitiver Fähigkeiten;
- Tests der Aufmerksamkeit und Konzentration;
- Tests sensorischer und motorischer Leistung;
- Sonstige Leistungstests;
- Allgemeine Persönlichkeitstests;
- Spezifische Persönlichkeitstests;
- Einstellungs-, Motivations- und Interessentests.

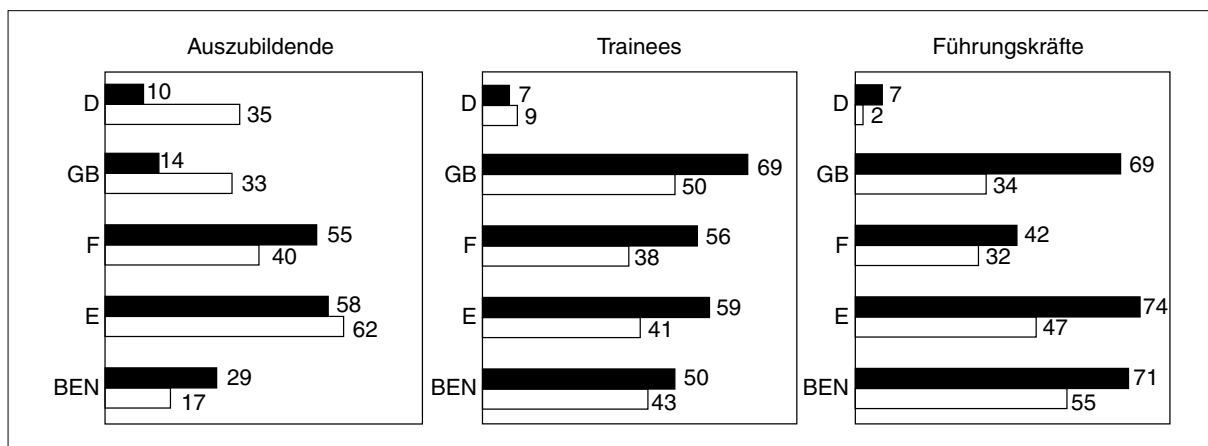


Abbildung 3:

Einsatz psychologischer Personalauswahlverfahren im europäischen Vergleich (aus Schuler, 2000)

Anmerkung: Dargestellt ist die Einsatzhäufigkeit von Persönlichkeitstests (■) sowie Leistungs- und Intelligenztests (□) in Prozent der Unternehmen.

Bei vielen Tests handelt es sich um publizierte und über den Testhandel erhältliche Verfahren. In Großorganisationen (z. B. Arbeitsamt, Unternehmen, Militär) werden aber auch Verfahren eingesetzt, die speziell für den betreffenden Diagnosezweck ausgearbeitet wurden. Die Einsatzhäufigkeiten bestimmter Verfahren sind abhängig von der betreffenden Berufsgruppe. Die Abbildung 3 (Daten von 1990) zeigt, daß sich der Prozentsatz von Großunternehmen, in denen Testverfahren eingesetzt werden, in verschiedenen europäischen Ländern sehr stark unterscheidet (vgl. alternativ auch Ryan, Baron & Page, 1999).

2.2 Eine Auswahl konstruktorientierter Ansätze

Im folgenden soll eine Auswahl der gängigsten konstruktorientierten Ansätze in der beruflichen Eignungsdiagnostik vorgestellt werden. Hierfür wird zunächst genauer auf das Konstrukt, seine Herleitung und seine wichtigsten Definitionsmerkmale eingegangen. Danach werden mögliche Verfahren zur Konstrukterfassung exemplarisch in Auszügen vorgestellt. Soweit dies möglich und sinnvoll ist, schließen sich daran vorliegende Validitätsbefunde an.

Die Darstellung folgt einer groben Zweiteilung: Zuerst werden in dem thematischen Block „kognitive Fähigkeiten“ allgemeine und spezielle kognitive Konstrukte und Verfahren gemeinsam vorgestellt, danach folgt eine Zusammenstellung von allgemeinen und speziellen Persönlichkeitsansätzen sowie Motivations- und Interessenverfahren.

2.2.1 Allgemeine kognitive Fähigkeiten

„Kognitive Fähigkeiten“ werden im weiteren synonym zum Begriff „Intelligenz“ gebraucht, der in seiner allgemeinsten Form die Gesamtheit aller kognitiven oder geistigen Fähigkeiten umschreibt. Gemeinsamer Kern praktisch aller Intelligenzkonzepte ist die Qualität und Geschwindigkeit der Lösung neuartiger (also nicht routinebestimmter) Aufgaben. Die verschiedenen Theorien legen allerdings unterschiedliche Schwerpunkte auf die beteiligten Strukturelemente und Prozesse. Umstritten ist zudem, wie viele und welche Intelligenzfaktoren unterschieden werden sollen und in welchem Maße diese sich zu einem Generalfaktor (dem sogenannten „g“-Faktor) der allgemeinen Intelligenz zusammenfügen. Deshalb soll an dieser Stelle eher exemplarisch das „Berliner Intelligenzstrukturmodell (BIS)“ von Jäger (1984) vorgestellt werden, das als eine Art „Kompromißmodell“ eine Synthese vorhergehender Ansätze anstrebt (vgl. auch den ähnlichen Ansatz von Carroll, 1993).

2.2.1.1 Berliner Intelligenzstrukturmodell (BIS-Modell)

Jäger selbst (1984, S. 25) versteht seinen Ansatz als deskriptives Modell und nicht als Theorie. Es dient als Integrationsversuch einer Vielzahl vorhergehender Strukturmodelle, darunter z. B. der Ansätze von Thurstone, Vernon, Guilford, Meili und Cattell. Es steht somit in der Tradition faktorenanalytisch abgeleiteter Modelle, bei denen Prozeßaspekte der Intelligenz weitgehend ausgeblendet bleiben. Dem Modell liegen folgende Kernannahmen zugrunde (vgl. Jäger, Süß & Beauducel, 1997, S. 4):

- Intelligenzleistungen können hinsichtlich verschiedener Aspekte klassifiziert werden. Im BIS-Modell geschieht dies in Form einer bimodalen Klassifikation (Operationen und Inhalte).
- An jeder Intelligenzleistung sind mit jeweils spezifischer Gewichtung immer alle intellektuellen Fähigkeiten (neben anderen Bedingungen) beteiligt.

In Abbildung 4 ist die z. Zt. gültige Fassung des BIS-Modells dargestellt.

Kern praktisch aller Intelligenzkonzepte: Qualität und Geschwindigkeit der Lösung neuartiger Aufgaben.

Das „Berliner Intelligenzstrukturmodell“ stellt eine Synthese vorhergehender Strukturansätze dar.

Die Teilkomponenten des Berliner Intelligenzstrukturmodells

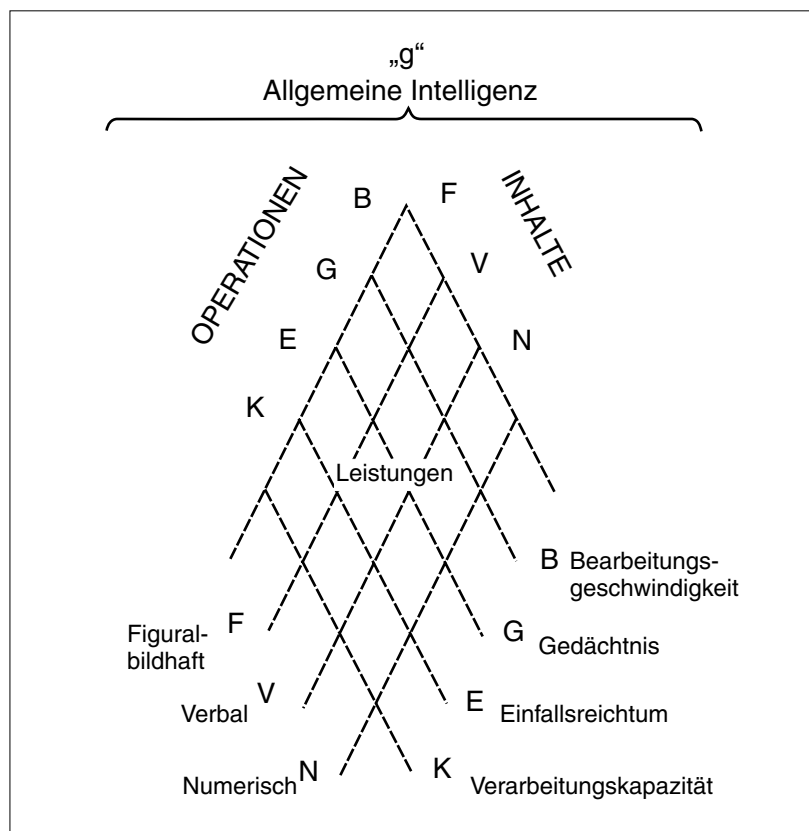


Abbildung 4:

Das Berliner Intelligenzstrukturmodell (aus Jäger et al., 1997, S. 5)

Jede Intelligenzleistung wird als Indikator von drei Fähigkeitsfeldern verstanden: der operativen, inhaltsgebundenen und Allgemeinen Intelligenz.

Jede Intelligenzleistung wird damit als Indikator für drei Fähigkeitsfelder verstanden: der operativen, inhaltsgebundenen und Allgemeinen Intelligenz.

Das *operative Fähigkeitskonstrukt* wird weitergehend in vier Facetten untergliedert (Verarbeitungskapazität, Einfallsreichtum, Merkfähigkeit, Bearbeitungsgeschwindigkeit), das *inhaltsgebundene Fähigkeitskonstrukt* in drei Facetten (sprachgebundenes Denken, zahlengebundenes Denken, anschauungsgebundenes Denken). Durch Verknüpfung dieser operativen und inhaltlichen Facetten entsteht die in Abbildung 4 dargestellte Matrix mit 12 Zellen. Die 45 Subtests des von der Berliner Arbeitsgruppe entwickelten BIS-Tests sind jeweils zu 3er- bis 5er-Gruppen einer dieser Fähigkeitsverknüpfungszellen zugewiesen. In Kasten 1 ist beispielhaft jeweils ein Ausschnitt aus den Subtests RD (Rechnerisches Denken) und ZS (Zahlen-Symbol-Test) angegeben.

Jeder der 45 Subtests wird bei der Auswertung dreifach verwertet: Zur Bestimmung der jeweiligen Operations- und Inhaltsfacettenausprägung, indem die Tests der zugehörigen Zellen aggregiert werden, sowie zur Bestimmung der Allgemeinen Intelligenz, für die über alle Subtests aggregiert wird. Im BIS-Modell wird Allgemeine Intelligenz also eher operational als Zusammenfassung der gesamten Spannbreite spezifischer Fähigkeiten verstanden.

Das BIS-Modell wird hier nur beispielhaft dargestellt, da in der (eignungs-) diagnostischen Praxis eine Vielzahl von intelligenzbezogenen Verfahren eingesetzt wird. Wenn einem Verfahren allerdings ein faktorenthoretisches Modell zugrunde liegt, so sollte dieses im Sinne einer Vor- bzw. Subform des BIS-Modells reinterpretiert werden können. Zumindest auf hochaggregiertem Niveau können so allgemeingültige intelligenzbezogene Validitätsbefunde abgeleitet werden.

Kasten 1:

Auszüge aus dem BIS-Test (Jäger et al., 1997)

Modifiziertes Beispiel aus dem Subtest RD:

Ein Schüler geht mit einem Betrag von 7,30 DM in einen Schreibwarenladen. Dort kostet ein Buntstift 2,50 DM. Wie viele Buntstifte kann er bezahlen und wieviel Geld behält er übrig?

Lösung: Er kann ____ Buntstifte bezahlen und behält ____ DM übrig.

Der Subtest RD ist der Fähigkeitsverknüpfung „Verarbeitungskapazität“ x „zahlengebundenes Denken“ zugeordnet.

Modifiziertes Beispiel aus dem Subtest ZS:

Tragen Sie die den jeweiligen Zahlen zugeordneten Symbole in die unteren Kästchen ein. Lassen Sie kein Kästchen aus und halten Sie sich an die vorgegebene Reihenfolge. Arbeiten Sie so rasch wie möglich!

Beispiel:

1	2	3	4	5	6		3	6	1	2
—	/	□	L	□	○					

Der Subtest ZS ist der Fähigkeitsverknüpfung „Bearbeitungsgeschwindigkeit“ x „figural-bildhaftes Denken“ zugeordnet.

Auszüge aus dem BIS-Test**2.2.1.2 Intelligenz und Beruf**

Kognitive Verfahren wurden bereits frühzeitig für eignungsdiagnostische Zwecke eingesetzt. So wurden die beiden Verfahren der U.S. Army zur Erfassung allgemeiner kognitiver Fähigkeiten, der „Army Alpha“-Test sowie die sprachungebundene „Army Beta“-Parallelform, während des Ersten Weltkriegs zunächst noch unstrukturiert und im Zweiten Weltkrieg systematisch für die Eignungsdiagnose von Hunderttausenden von Rekruten eingesetzt (vgl. für einen kurzen historischen Überblick Kanfer, Ackerman, Murtha & Goff, 1995).

Bereits auf einer einfachen deskriptiven Ebene läßt sich die Bedeutsamkeit der (allgemeinen) Intelligenz für das Berufsleben herleiten. In Kasten 2 sind beispielhaft nach Engelbrecht (1994) die Mittelwerte des „Allgemeinen intellektuellen Leistungsvermögens“ (ermittelt über Teilfacetten der „Eignungs-Untersuchungsserie für die Berufsberatung (EUB)“ der Bundesanstalt für Arbeit) für Schülerinnen und Schüler gesammelt, die vier bis sieben Jahre später den aufgelisteten 92 Ausbildungsberufen angehörten und zufrieden mit ihrem Beruf waren.

Kriterium der Bedeutsamkeit der (allgemeinen) Intelligenz für das Berufsleben kann auch die Zufriedenheit sein.

Kasten 2:

Mittelwerte des „Allgemeinen intellektuellen Leistungsvermögens“ verschiedener Berufsgruppen (nach Engelbrecht, 1994, Tab. 2)

Altenpfleger 46	Apothekenhelfer 51
Arzthelfer 53	Augenoptiker 58
Bäcker 43	Bankkaufmann 59
Bauschlosser 44	Bauzeichner 58
Bekleidungsfertiger 46	Bekleidungsschneider 49
Beton- u. Stahlbetonbauer 47	Betriebsschlosser 49
Blechschröcker 47	Bürogehilfe 53
Bürokaufmann 53	Büromaschinenmechaniker 55
Chemielaborant 59	Dachdecker 46
Damenschneider 49	Dienstleistungsfachkraft im Postbetrieb 46
Dreher 48	Drogist 52

Drucker 51	Druckvorlagenhersteller 55
Einzelhandelskaufmann 50	Elektroanlageninstallateur 51
Elektroinstallateur 50	Elektromaschinenbauer 51
Elektromechaniker 55	Energieanlagenelektroniker 56
Energiegeräteelektroniker 56	Erzieher 55
Fachhilfe in steuerber. Berufen 56	Feinblechner 48
Feingeräteelektroniker 59	Feinmechaniker 52
Fernmeldeelektroniker 59	Fernmeldehandwerker 56
Fleischer 43	Fliesenleger 43
Florist 49	Flugzeugmechaniker 54
Friseur 44	Funkelektroniker 58
Gärtner 49	Gas- u. Wasserinstallateur 45
Glaser 46	Hauswirtschaftler 47
Holzmechaniker 48	Hotelfachmann 52
Industrie Kaufmann 56	Informationselektroniker 58
Karosseriebauer 46	Kaufmann im Groß- u. Außenhandel 53
Kinderpfleger 46	Koch 46
Konditor 47	Kraftfahrzeugelektriker 51
Kraftfahrzeugmechaniker 47	Krankenpflegehelfer 44
Krankenpfleger 53	Kunststoffformgeber 51
Landmaschinenmechaniker 45	Landwirt 48
Maler u. Lackierer 42	Maschinenbauer 50
Maschinenschlosser 51	Masseur 50
Maurer 43	Mechaniker 50
Nachrichtengerätemechaniker 58	Radio- u. Fernsehtechniker 57
Raumausstatter 47	Rechtsanwaltsgehilfe 53
Schauwerbegestalter 55	Schlosser 46
Schriftsetzer 55	Sozialversicherungsfachangestellter 57
Speditionskaufmann 54	Stahlbauschlosser 46
Stahlformenbauer 51	Technischer Zeichner 57
Tischler 47	Verkäufer 46
Verkäufer im Nahrungshandwerk 45	Vermessungstechniker 58
Versicherungskaufmann 55	Werkzeugmacher 53
Zahnarzthelfer 52	Zahntechniker 55
Zentralheizungs- und Lüftungsbauer 46	Zimmerer 46

Anmerkung: N = 30477. T-Wertnormierung (GesamtMW=50); gleichgeschlechtliche Normen; Normgruppen bestehen aus 50% Haupt- und 50% Realschülern/-innen (Berufsbezeichnung nur männlich). Die Werte wurden bei männlichen und weiblichen Haupt- und Realschülern gewonnen, die als 14- bis 16jährige an EUB-Untersuchungen teilgenommen und 4-7 Jahre später in einer follow up-Untersuchung angegeben hatten, im betreffenden Beruf die Ausbildung erfolgreich abgeschlossen zu haben und mit dem Beruf „sehr zufrieden“ oder „zufrieden“ zu sein. Standardabweichungen 8 oder 9 T-Werte, vereinzelt 7 oder 10 T-Wert-Punkte.

Massenerhebungen zeigen Intelligenzunterschiede zwischen Angehörigen verschiedener Berufssparten. Innerhalb der Berufe wird die Streuung kleiner, je höher die Qualitätsanforderungen des Berufs sind.

Deutlich wird eine große Abweichung der Intelligenzwerte zwischen den Berufen: Bei der vorliegenden Untersuchung streuen die Werte immerhin über zwei Standardabweichungen. Zusätzlich ergibt sich ein weiterer Effekt, der über das retrospektive follow up-Erhebungsdesign von Engelbrecht (1994), bei dem Gymnasiasten bewußt ausgeklammert wurden, leider nicht deutlich wird: Innerhalb der Berufe ist eine schrumpfende Streuung zu beobachten, je höher die Qualifikationsanforderungen des jeweiligen Berufs sind (vgl. Ackerman & Humphreys, 1991, pp. 227-229). Sehr deutlich wird dieser Umstand z. B. bei der graphischen Darstellung der Intelligenzdaten, die während des 2. Weltkriegs von der U.S. Army zur Vorbeurteilung der Rekruten erhoben wurden (vgl. Abbildung 5).

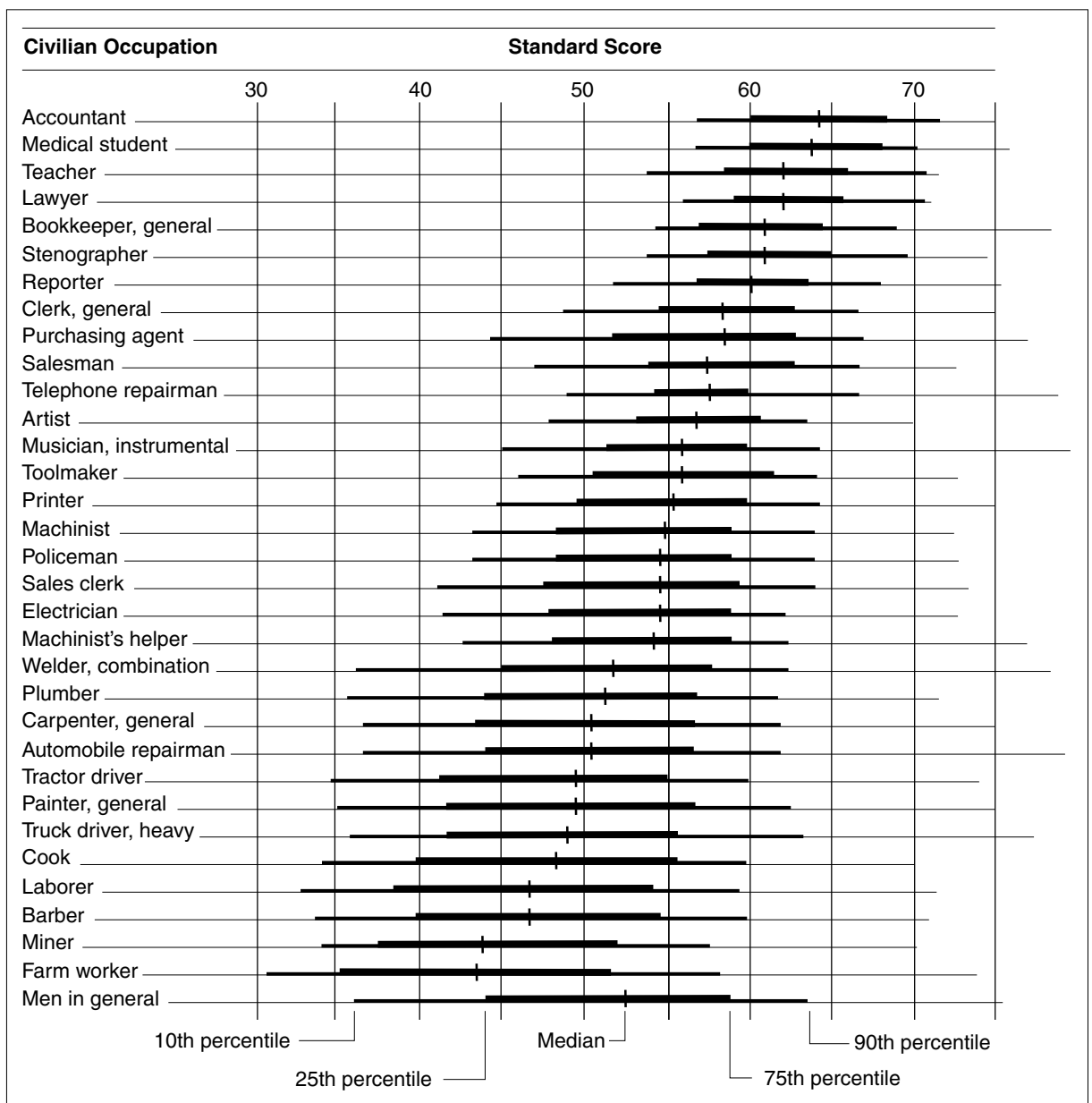


Abbildung 5:

Standardisierte Werte des „Army General Classification Test“ für unterschiedliche Berufsgruppen (Cronbach, 1960, zitiert nach Ackerman & Humphreys, 1991)

Die schwächste Gruppe stellen danach die Landarbeiter (farm worker) mit einem Median von ca. 44 Standardwerten. Ihr Interquartilbereich (Differenz aus Prozentrang 75 und Prozentrang 25) beträgt ca. 22 Standardwerte. Buchhalter (accountants) erzielen hingegen die besten Werte. Der Median liegt bei ca. 64 Standardwerten, ihr Interquartilabstand beträgt nur noch ca. 16 Standardwerte.

Die Erklärungsgrundlage für diese deskriptiven Unterschiede ist kontrovers. Die Testintelligenz weist hohe Korrelationen mit dem Schulabschluß und weitergehenden sozioökonomischen Statusvariablen auf. Beide Variablenkomplexe (kognitive und gesellschaftliche Variablen) wirken sich auf die Berufswahl, -ausbildung und -ausübung einer Person aus und tragen zum deskriptiven Befund bei:

- Eine Berufstätigkeit stellt für ihre erfolgreiche Erfüllung bestimmte Anforderungen an die kognitiven Fähigkeiten der ausübenden Person. Dies können Mindestanforderungen i.S. eines Schwellenmodells („eine be-

Als Erklärung für die berufsspezifischen Intelligenzunterschiede werden unterschiedliche kognitive und gesellschaftliche Ursachen herangezogen.

Die Erklärung der Intelligenzunterschiede zwischen sozialen Schichten ist politisch umstritten.

Die in den sechziger Jahren aufgekommene Skepsis zum Einsatz von kognitiven Verfahren in der Eignungsdiagnostik resultierte aus enttäuschenden empirischen Befunden und konzeptionellen Zweifeln.

stimmte Fähigkeitsausprägung muß mindestens vorhanden sein“) oder einer Linearitätshypothese („je höher die kognitiven Fähigkeiten, desto besser die Berufsausübung“) sein. Wenn eine Person die Mindestanforderungen nicht erfüllt, wird sie dauerhaft nicht erfolgreich sein und der (Selbst-)Selektion unterliegen. Da die Mindestanforderungen mit dem Berufsniveau steigen, ist damit die Streuungsverkleinerung zu erklären.

- Viele Berufe unterliegen formalisierten Aufnahmebeschränkungen, die stark von intelligenzgesättigten Variablen abhängen (Schulabschluß, „numerus clausus“-Zulassungsbeschränkungen für Hochschulstudiengänge usw.). Zusätzlich wirken neben diesen Bildungsbarrieren auch gesellschaftliche Klassifizierungen (z. B. Zugehörigkeit zu einer ethnischen Majorität) oder Sozialisationseffekte (z. B. Bildungsniveau und sozialer Status der Eltern) als „gesellschaftliche Türöffner“ für Prestigeberufe.

In den USA wurde die Debatte zur Alltagsrelevanz von Intelligenz und Intelligenztestergebnissen Mitte der Neunziger Jahre mit der Veröffentlichung von „The Bell Curve“ (Herrnstein & Murray, 1994) neu entfacht. Die Autoren versuchen unter Bezug auf bestehende Untersuchungen und eigene Längsschnittanalysen nachzuweisen, daß die über entsprechende Tests erfaßte Intelligenz eine wichtige Vorhersage für den allgemeinen Lebenserfolg einer Person (in Schule, Beruf, Familie, hinsichtlich Wohlstand, Gesundheit usw.) leistet. Von einer starken Heredität der Intelligenz ausgehend plädieren sie dafür, bestehende Unterschiede zwischen Personen und Personengruppen (z. B. ethnischen Gruppen) als gegeben zu akzeptieren und offensiv damit umzugehen (z. B. stärkere intelligenzbezogene Entscheidungskriterien, Förderung der bestehenden kognitiven Elite). Das Buch hat z. T. heftigen Widerspruch erregt und führte u. a. zu einer durch die American Psychological Association unterstützten wissenschaftlichen Projektgruppe zur Sichtung der empirischen Befundlage zu Intelligenz (Neisser et al., 1996). Das Ergebnis ist wie bereits beschrieben: Intelligenzunterschiede bestehen; die Erklärungsgrundlage ist unklar; empirische Befunde werden immer auch gleich hinsichtlich ihrer politischen Implikationen beurteilt.

2.2.1.3 Validität kognitiver Verfahren für berufseignungsdiagnostische Zwecke

Der Einsatz von kognitiven Verfahren in der Eignungsdiagnostik ist historisch gesehen durch einen tiefen Einschnitt gekennzeichnet, der sich auch bei persönlichkeitsorientierten Verfahren nachweisen läßt (vgl. Abschnitt 2.2.3). Bis 1960 wurden kognitive Verfahren mit großer Häufigkeit eingesetzt. Die narrativen Zusammenfassungen von Dorcus und Jones (1950) sowie Super und Crites (1962) zu mehr als 500 Einzelstudien weisen ein positives Ergebnis hinsichtlich der erzielten Validität aus. Die danach ausgebrochene Skepsis führt z. B. Cook (1993) im wesentlichen auf zwei parallele Vorgänge zurück:

- Die ersten systematischeren Analysen von Ghiselli (1973) zeigten, daß die kriteriumsbezogene Validität für die meisten kognitiven Verfahren im Bereich zwischen .2 und .3 lag, was als äußerst unzureichend beurteilt wurde.
- Parallel dazu wuchsen neben den ideologischen Vorbehalten gegen eignungsdiagnostische Verfahren auch die Einwände gegen eigenschaftsorientierte Verfahren an sich. Nach der Entdeckung, daß wesentliche Studienergebnisse des bedeutenden Vererbungsforschers Burt beschönigt bzw. gefälscht waren (vgl. Jensen, 1969), wurde von vielen ein gesamter Forschungszweig disqualifiziert. Zusätzlich veröffentlichte Mischel 1968 sein einflußreiches Buch „Personality and assessment“, in dem er versucht nachzuweisen, daß aufgrund einer prinzipiell falschen Konzeption die Interkorrelationen von eigenschaftsorientierten Verfahren mit anderen Kriterien praktisch nie die magische „3 Schranke“ durchbrechen können. Dies führte zu einer weitreichenden konzeptionellen Umorientierung vie-

ler Wissenschaftler in diesem Bereich. So postulierte z. B. McClelland (1973, vgl. hierzu auch Barrett & Depinet, 1991) alternativ zum „Intelligenz“-einen sog. „Kompetenz“-Ansatz, der durchaus als „geistiger Urverter“ der simulationsorientierten Verfahren (vgl. Kapitel 6) angesehen werden kann.

Erst durch die systematisierte Analyse und Bewertung einzelner Validitätsstudien mit neueren metaanalytischen Techniken konnte eine Relativierung der Negativbefunde erreicht werden. Im Kasten 3 werden die Grundzüge der für die Eignungsdiagnostik bedeutendsten Metaanalyse-Technik beschrieben.

Kasten 3:

Metaanalytische Validitätsgeneralisierung

In der Organisationspsychologie hat von den existierenden Metaanalyseschulen nur die Validitätsgeneralisierung nach Hunter und Schmidt (Hunter & Schmidt, 1990) größere Bedeutung erlangt. Dies liegt sicher daran, daß Hunter und Schmidt als originäre Organisationspsychologen ihre Metaanalysetechnik auf die Belange des Fachs praktisch maßgeschneidert haben (vgl. für einen allgemeinen Schulenüberblick z. B. Bangert-Drowns, 1986). Der Ansatz basiert im wesentlichen auf einer systematischen Zusammenfassung (= Sekundärstudie) von unterschiedlichen empirischen Einzelstudien zur gleichen Fragestellung (= Primärstudien) mit einer Kontrolle möglicher Artefaktquellen. Die generelle Vorgehensweise soll anhand eines kleinen, inhaltlich stark vereinfachten Beispiels erläutert werden.

Hypothetische Ausgangssituation:

Die kriteriumsbezogene Validität eines Personalauswahlverfahrens (z. B. eines kognitiven Fähigkeitstests) soll überprüft werden. Eine Literaturstudie soll ergeben haben, daß bisher vier vergleichbare Validierungsstudien durchgeführt wurden. Bei allen Studien wurden jeweils die Ergebnisse der Bewerber in dem entsprechenden Verfahren erfaßt und mit der globalen Leistungsbeurteilung der Vorgesetzten sechs Monate nach der Einstellung korreliert. In der nachfolgenden Tabelle sind die (frei erfundenen) Ergebnisse (r_{TC}) der vier Studien zusammen mit der jeweils zugrundeliegenden Stichprobengröße (n) angegeben.

Studie	1	2	3	4
r_{TC}	.30	.15	.15	.20
n	150	100	100	200

a) Kontrolle des Meßfehlers: Attenuationskorrektur

Eine erste Korrektur der Korrelationen kann vorgenommen werden, indem der Meßfehler der Kriteriumsmessung berücksichtigt wird. Die hierfür benötigten Informationen (die Reliabilitätsangaben der eingesetzten Beurteilungsverfahren) sind entweder bei den berücksichtigten Primärstudien mit angegeben, müssen zusätzlich über die üblichen Schätzmethoden erfaßt oder aus der allgemeinen empirischen Befundlage geschlossen werden. Die Metaanalyse von Viswesvaran, Ones & Schmidt (1996) zeigt, daß für unser globales Leistungsurteil eine Reliabilität von .52 durchaus realistisch erscheint. In der nachfolgenden Tabelle sind die attenuationskorrigierten Korrelationen $r_{TC(korr1)}$ angegeben (vgl. für diese und alle weiteren rechnerischen Durchführungen z. B. Hunter & Schmidt, 1990):

Studie	1	2	3	4
r_{TC}	.30	.15	.15	.20
$r_{TC(korr1)}$.42	.21	.21	.28
n	150	100	100	200

Die metaanalytische Validitätsgeneralisierung faßt systematisch Befunde von bestehenden Studien zusammen und korrigiert methodisch bedingte Mängel.

Die Grundzüge der metaanalytischen Validitätsgeneralisierung veranschaulicht an einem vereinfachten Beispiel.

Eine Attenuationskorrektur der Prädiktorwerte wäre übrigens nur angemessen, wenn die Validität des Konstrukts geprüft würde, nicht aber die des Auswahlverfahrens.

b) Kontrolle von Varianzeinschränkungen

Eine zusätzliche Verzerrung tritt bei den berücksichtigten Studien durch das eingesetzte Erhebungsdesign auf: Vorgesetztenbeurteilungen liegen nur für die angenommenen Bewerber vor. Die auf der Grundlage des Verfahrens abgelehnten Bewerber bleiben bei der Zusammenhangsanalyse unberücksichtigt. Diese Varianzeinschränkung der Testwerte führt zu einer Unterschätzung des wahren Zusammenhangs. Für eine Korrektur der Validitätskoeffizienten muß lediglich die Selektionsquote (Anteil der Ausgewählten an den Bewerbern) bekannt sein. Für unser Beispiel nehmen wir vereinfachend an, daß bei allen Studien eine Selektionsquote von 50% vorlag. Die korrigierten Korrelationskoeffizienten $r_{TC(korr2)}$ sind in der folgenden Tabelle angegeben.

Studie	1	2	3	4
r_{TC}	.30	.15	.15	.20
$r_{TC(korr1)}$.42	.21	.21	.28
$r_{TC(korr2)}$.46	.25	.25	.33
n	150	100	100	200

c) Kontrolle des Stichprobenfehlers

Um die Stichprobenschwankungen zu kontrollieren, wird ein „bester Schätzer“ für den wahren Zusammenhang zwischen Testwert und dem Berufserfolgskriterium gebildet, indem ein nach der Stichprobengröße gewichteter Mittelwert der Einzelkorrelationen gebildet wird. (Hunter & Schmidt, 1990, pp. 213-218, sprechen sich übrigens gegen eine vorherige Fisher's z-Transformation der Koeffizienten aus.). Für unser Beispiel resultiert ein dreifach korrigierter Validitätskoeffizient von $r_{TC(korr3)} = .34$, der wesentlich über den ursprünglichen Koeffizienten liegt. Auf seiner Grundlage würde man dem Verfahren eine akzeptable Validität attestieren. Sollten nach Berücksichtigung der üblichen Artefaktkorrekturen noch größere Unterschiede zwischen den nunmehr korrigierten Koeffizienten der Primärstudien bestehen bleiben, so kann eine nachfolgende *Moderatorenanalyse* sinnvoll sein. Hierbei wird überprüft, ob Besonderheiten in den Kontextmerkmalen der Einzelstudien (z. B. Merkmale der Versuchspersonengruppe, unterschiedliche Zielsetzungen der Studien, Unterschiede in der Durchführung, besondere Kriterienmerkmale usw.) diese Restvarianz zwischen den Studien aufklären können.

Auf der Basis dieser Technik reanalysierten Schmidt und Hunter in mehreren Studien (z. B. Schmidt & Hunter, 1977; Hunter & Hunter, 1984) die vorliegenden Einzelstudien unter Einschluß der Datenbasis von Ghiselli (1973). Hunter und Hunter (1984) ermittelten so z. B. einen durchschnittlichen Validitätskoeffizienten für allgemeine Intelligenz von durchschnittlich .53, der durch kein anderes Verfahren übertroffen wurde.

Seit diesen ersten Arbeiten ist die Zahl der einschlägigen Metaanalysen stetig gewachsen, in denen eine Vielzahl unterschiedlicher Verfahren, Personenstichproben und Berufserfolgskriterien berücksichtigt wurden. In Kasten 4 ist hierzu eine vorläufige Zusammenfassung der Ergebnisse dargestellt.

Metaanalytische Befundlage zur Validität unterschiedlicher eignungsdiagnostischer Verfahren: Allgemeine kognitive Fähigkeitstests werden von Schmidt und Hunter als beste Standardverfahren zur Vorhersage von Berufserfolg ausgewiesen.

Kasten 4:

Zusammenfassung der metaanalytischen Befundlage zur Validität eignungsdiagnostischer Verfahren (verkürzt nach Schmidt & Hunter, 1998, S. 22)

Prädiktor	Val	mR	inkrV	%Zuw	Quelle
<i>allgemeine kognitive Fähigkeitstests</i>	.51				Abs. 2.2.1
Arbeitsproben	.54	.63	.12	24%	Kap. 6
„Integrity“-Tests	.41	.65	.14	27%	Abs. 3.1
Gewissenhaftigkeitstests	.31	.60	.09	18%	Abs. 2.2.3
strukturiertes Einstellungsgespräch	.51	.63	.12	24%	Kap. 7
unstrukturiertes Einstellungsgespräch	.38	.55	.04	8%	Kap. 7
Fachkenntnistests	.48	.58	.07	14%	Abs. 2.2.2
Probezeit	.44	.58	.07	14%	
Biographische Daten	.35	.52	.01	2%	Kap. 7
Assessment Center	.37	.53	.02	4%	Kap. 6
Interessen	.10	.52	.01	2%	Abs. 2.2.5
Graphologie	.02	.51	.00	0%	Kap. 7

Als Artefakte wurden die Kriteriumsreliabilität und Varianzeinschränkungen korrigiert.

Folgende Abkürzungen werden verwendet:

Val: kriteriumsbezogene Validität (Kriterium: Leistungsbeurteilung)

mR: multiple Korrelation mit allgemeiner kognitiver Fähigkeit als erster und mit dem jeweiligen Prädiktor als zweiter Variable

inkrV: inkrementelle Validität, d.h. der Validitätszuwachs durch Hinzunahme des zweiten Prädiktors

%Zuw: Prozentualer Zuwachs der Validität durch Hinzunahme des zweiten Prädiktors

Quelle: Hier findet sich ein Hinweis, in welchem Kapitel (Kap.) des Lehrbuchs bzw. in welchem Abschnitt innerhalb dieses Kapitels (Abs.) das entsprechende Verfahren vorgestellt wird.

Hinweis:

Diese Validitätszusammenstellung von Schmidt und Hunter (1998) kann nicht unkommentiert bleiben:

- Bedingt durch andere gesetzliche Rahmenbedingungen, andere Berufsordnungen usw. sind nicht alle Befunde direkt auf deutsche Verhältnisse übertragbar. So gibt es z. B. in Deutschland keinen regulär in der Praxis eingesetzten „Integrity“-Test.
- Die Autoren haben bei konkurrierenden metaanalytischen Befunden eine „wohlwollende Auswahl“ getroffen, indem sie die jeweils höchste gefundene Validität in die Übersicht aufgenommen haben (vgl. als Beispiel Abschnitt 2.2.3.2 zu persönlichkeitsorientierten Verfahren).
- Die Koeffizienten stammen aus unterschiedlichen Metaanalysen. Dadurch konnten einige Randvariable nicht parallelisiert werden, z. B. die Kriterien, nach denen Primärstudien in die Metaanalyse einbezogen (oder verworfen) wurden.
- Als Berufserfolgskriterium diente in den meisten Fällen eine globale Leistungsbeurteilung durch den Vorgesetzten. Wie Kapitel 15 zeigt, wird damit nicht der gesamte Leistungsbereich abgedeckt.
- Generell dienen die angegebenen Koeffizienten nur als grobe Orientierungsgröße. Zu den Validitätsbefunden nicht-kognitiver Verfahren sei explizit auf die restlichen Abschnitte in diesem Kapitel sowie auf die Kapitel 6 und 7 verwiesen. Die hohe Validität kognitiver Fähigkeitstests setzt voraus, daß keine starke Vorselektion in bezug auf Intelligenz existiert.

Zusammenfassung der aktuellen metaanalytischen Befundlage zur Validität eignungsdiagnostischer Verfahren

Die Ergebnisübersicht von Schmidt und Hunter (1998) sollte nicht unreflektiert übernommen werden.

Allgemeine kognitive Verfahren sind bedingungsunabhängig valide. Spezifische kognitive Verfahren müssen für den jeweiligen Zweck maßgeschneidert werden.

Die Vorhersagekraft von Intelligenztestergebnissen hängt immer auch von den spezifischen Randbedingungen der Auswahlsituation ab.

Schmidt und Hunter (1998) gehen davon aus, daß es praktisch keinen Beruf gibt, für den allgemeine Fähigkeitstests keine kriteriumsbezogene Validität für berufliche Leistungsvariablen aufweisen. Da diese Verfahrensgruppe häufig der beste Prädiktor für Berufserfolg ist, legen sie in ihrer Übersicht besonderen Wert auf den Informationszuwachs, den andere Verfahren in Ergänzung liefern können.

Diese Argumentationsstrategie soll zunächst anhand der Gegenüberstellung allgemeiner und spezifischer Fähigkeitstests diskutiert werden (vgl. Abschnitt 2.2.1.4). Nach der Darstellung der nicht-kognitiven Ansätze wird dieses Thema noch einmal in verallgemeinerter Form aufgegriffen (vgl. Abschnitt 3.1).

2.2.1.4 Allgemeine kognitive Fähigkeit vs. spezifische Fähigkeiten

Zeitgleich zur Entwicklung von standardisierten Testbatterien für allgemeine kognitive Fähigkeiten fand auch die Konstruktion maßgeschneiderter Verfahren für bestimmte Berufsgruppen statt, bei denen auf die Ergebnisse einer vorhergehenden detaillierten Arbeitsanalyse (vgl. Kapitel 3) zurückgegriffen wird. Beispielsweise wäre für die Auswahl von Fluglotsen ein (computergestütztes) Verfahren denkbar, in dem die wesentlichen Anforderungselemente des Arbeitsplatzes (z. B. Streßbelastbarkeit, räumliches Vorstellungsvermögen usw.) abgebildet werden. Abhängig von der Breite der definierten Berufssparte führt diese Herangehensweise allerdings dazu, daß kostenaufwendige, schwer miteinander vergleichbare Spezialverfahren konstruiert werden. Da sich das Berufsbild z. B. durch technische Neuerungen schnell verändern kann, muß die Gültigkeit der durchgeführten Arbeitsanalyse und der Verfahrenskomponenten regelmäßig überprüft werden.

Die Diskussion, welche dieser Herangehensweisen (Erfassung fundamentaler Globalfähigkeiten vs. berufsrelevante Spezialfähigkeiten) adäquater sei, wurde durch die Ergebnisse der metaanalytischen Validitätsgeneralisierung und ihre offensive Propagierung durch die Schmidt-Hunter-Arbeitsgruppe noch weiter forciert. Wie bereits angedeutet, stellen Schmidt, Hunter, Pearlman und Hirsh (1985) in einer dokumentierten Debatte zum Einsatz der Validitätsgeneralisierung fest, daß allgemeine Intelligenz der eignungsdiagnostische Prädiktor der Wahl sein muß für alle Berufe, in denen übliche Anforderungen hinsichtlich verbaler, räumlicher und mathematischer Fähigkeiten gestellt werden. Spezifische Fähigkeitstests sollten danach nur zusätzlich eingesetzt werden, um inkrementelle Validität zu erzielen. Unterstützt wird diese Argumentation auch durch einen statistischen Umstand: Häufig können in der betrieblichen Praxis nur kleine Stichproben (unter 300 Personen) erhoben werden. Im direkten Vergleich weist hier ein allgemeiner Fähigkeitssummenscore eine höhere (Kreuz-)Validität auf als eine optimierte Regressionsgleichung der zugehörigen Untertests. Er wäre aus statistischen Gründen also wegen seiner größeren Robustheit zu bevorzugen (vgl. Ackerman & Humphreys, 1991).

Allerdings muß beachtet werden, daß die metaanalytische Validitätsgeneralisierung, auf der die Favorisierung der allgemeinen Intelligenz im wesentlichen fußt, teilweise mit vereinfachenden Annahmen arbeitet. So hängt die Vorhersagekraft von Intelligenz immer auch von den spezifischen Randbedingungen der Auswahlsituation ab. Wenn Bewerbergruppen in bestimmten Berufssparten z. B. aus spezialisierten Fachakademikern bestehen, so sind diese Gruppen bereits durch ihren Ausbildungsweg stark nach kognitiven Fähigkeiten vorselektiert. Ein Beispiel für diesen Umstand liefert die Metaanalyse von Funke, Krauss, Schuler und Stapf (1987) zur Prognose wissenschaftlich-technischer Leistungen von Mitarbeitern im Forschungs- und Entwicklungsbereich von Großunternehmen: Für allgemeine kognitive Fähigkeiten ergab sich nur eine hinsichtlich Stichprobenfehler und Kriteriumsreliabilität korrigierte mittlere Validität von $r = .16$. Nichtkognitive Variablen wiesen hier durch die geringere Vorselektion höhere Validitäten auf.

Ein weiterer Befund ist, daß allgemeine kognitive Fähigkeiten bessere Be-

rufserfolgsprädiktoren für komplexe Berufe als für weniger komplexe Berufe sind. Erstere zeichnen sich durch heterogene, über standardisierte Arbeitsanalysen nur schwer erfassbare Anforderungen aus, für die allgemeine Fähigkeiten per definitionem durch ihre Universalität beste Vorhersagen liefern sollten. In einfacheren oder hochspezialisierten Berufen erlangen (kombinierte) spezielle Fähigkeiten eine höhere Bedeutung. So ergab die Metaanalyse von Hunter und Hunter (1984) eine kriteriumsbezogene Validität der allgemeinen Intelligenz für Führungspositionen von .58, für Arbeitsplätze mit geringer Vorbildung („semi-skilled jobs“) eine Validität von .40 und für Arbeitsplätze ohne Vorbildung („unskilled jobs“) nur noch von .23.

Mit etwas Distanz betrachtet, erscheint die im Laufe der wissenschaftlichen Diskussion aufgeworfene Dichotomie zudem reichlich überspitzt. Murphy (1996) argumentiert beispielsweise, daß der Einsatz von allgemeinen kognitiven Fähigkeitsmessungen besonders durch die Einfachheit der Anwendung, die Generalisierbarkeit der Ergebnisse und die Parsimonität der Erklärung überzeugt. Allerdings hat „Berufserfolg“ als vorherzusagendes Kriterium viele, teilweise nur mäßig interkorrelierte Facetten (vgl. Kapitel 15), für die allgemeine Fähigkeiten unterschiedlich gute Vorhersagen leisten.

Zudem erscheint eine strikte Trennung zwischen „g-Faktor“ und spezifischen Fähigkeiten sinnlos, da sich in heute allgemein anerkannten hierarchischen Intelligenzmodellen diese beiden Komponenten mit faktorenanalytischen Methoden schwerlich trennen lassen. Sehr deutlich wird dies z. B. im BIS-Modell: Hier wird weitgehend offengelassen, ob der „g-Faktor“ der Intelligenz eher als Sammelvariable spezifischer Fähigkeiten oder als eigenständige Größe verstanden wird.

Zu beachten ist schließlich auch, daß durch den Blick auf die Prognose (=kriteriumsbezogene Validität) die Erklärung (=Konstruktvalidität) nicht vernachlässigt werden darf. Zur Illustration kann ein Beispiel von Schmidt (1992; der Reanalyse liegt eine Studie von Schmidt, Hunter & Outerbridge, 1986, zugrunde) dienen, in dem die Wirkungen von Intelligenz und Berufserfahrung auf Berufskennnisse, die praktische Leistung in Arbeitsproben und schließlich auf die Leistungsbeurteilung untersucht wurde. Als Analysegrundlage für das in Abbildung 6 dargestellte Pfaddiagramm diente eine

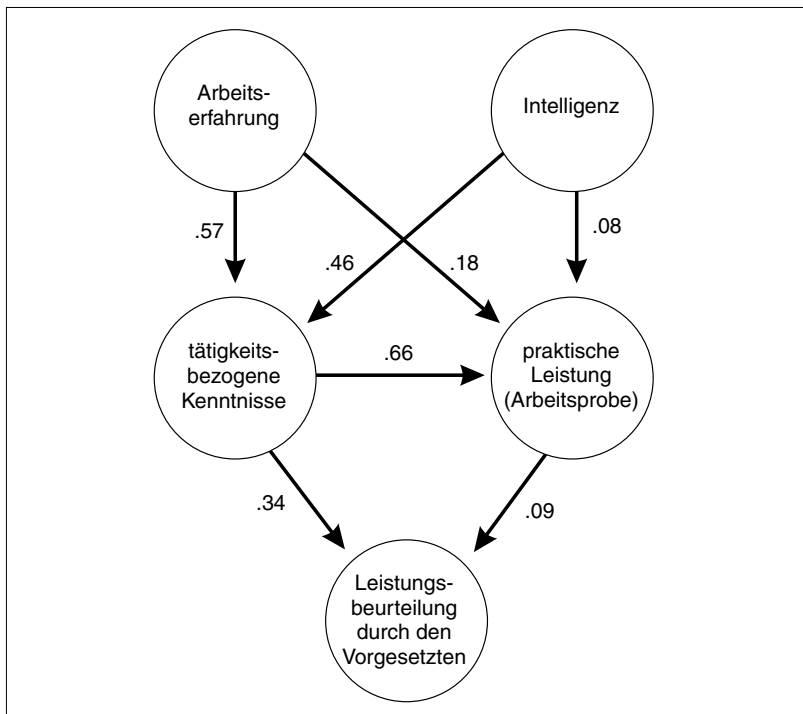


Abbildung 6:

Ein pfadanalytisches Modell der Leistung und Leistungsbeurteilung (nach Schmidt, 1992, p. 1179)

Die strenge Dichotomie „Erfassung fundamentaler Globalfähigkeiten“ vs. „berufsrelevante Spezialfähigkeiten“ ist letztlich nicht haltbar.

Die Prognosequalität eines Prädiktors gibt ohne weitergehende Untersuchungen wenig Aufschluß über Ursachenzusammenhänge.

Nachweis für die Auswirkung der Intelligenz auf die Vorgesetztenbeurteilung

Fachkenntnisse haben großen Einfluß auf die Urteilsbildung des Vorgesetzten.

Es gibt drei Varianten von Wissenstests: Fachkenntnistests, Tests des praktischen oder impliziten Wissens sowie situative Aufgaben.

metaanalytisch generierte Korrelationsmatrix der fünf Variablen. Die Pfadkoeffizienten zeigen, daß der Haupteinfluß der Intelligenz auf die Arbeitsleistung nur indirekt zustande kommen kann, nämlich über den besseren Erwerb tätigkeitsbezogener Kenntnisse. Zusätzlich basiert nach diesem Pfadmodell die Leistungsbeurteilung durch die Vorgesetzten in stärkerem Maße auf den Berufskennnissen der Mitarbeiter als auf ihren praktischen Leistungen.

2.2.2 Spezielle kognitive Fähigkeiten

2.2.2.1 Allgemeines Wissen und Fachwissen

Fachkenntnisse, verstanden als Wissen über Fakten, Abläufe und Zusammenhänge innerhalb der eigenen Berufstätigkeit, gehören zu den wichtigsten Voraussetzungen beruflichen Erfolgs und deshalb auch zu seinen validesten Prädiktoren. Diesem Sachverhalt entspricht das hohe Gewicht, das den Fachkenntnissen für die Einschätzung der Mitarbeiterleistung durch Vorgesetzte zukommt. Das in Abbildung 6 wiedergegebene pfadanalytische Modell weist den Fachkenntnissen große Bedeutung für die Urteilsbildung des Vorgesetzten zu: ihr Gewicht übertrifft das der Leistung in praktischen Arbeitsproben bei weitem. Wie auch Borman, Hanson, Oppler, Pulakos und White (1993) zeigen konnten, nehmen die Fachkenntnisse eine Mediatorfunktion zwischen Intelligenz und Leistungsbeurteilung ein. Während Fachkenntnisse wie auch allgemeines Wissen in den vorausgegangenen Jahrzehnten eher gering geschätzt wurden, finden sie in neuerer Zeit wieder Zuwendung.

Wie Salgado (1999) darlegt, können drei Arten von Meßverfahren als Tests des Fachwissens angesehen werden: *Fachkenntnistests* (job knowledge tests) im engeren Sinne, *Tests des praktischen oder impliziten Wissens* (tacit knowledge tests) sowie *situative Aufgaben* (situational judgement tests). Für schriftliche Fachkenntnistests errechneten Dye, Reck und McDaniel (1993) durchschnittliche korrigierte Validitäten von $r = .45$ (Kriterium Berufsleistung) und $r = .47$ (Kriterium Trainingsleistung). Ihrer Metaanalyse liegen 502 Validitätskoeffizienten und ein Stichprobenumfang von 363528 Personen zugrunde. Als Moderatoren der Validität erwiesen sich Aufgabenkomplexität und die Ähnlichkeit zwischen Test und Arbeitstätigkeit. Bei geringer Komplexität beträgt die Validität $r = .39$ für Berufsleistung und $r = .46$ für Trainingsleistung, während sie bei hoher Komplexität auf $r = .57$ für beide Kriterien steigt. Stärker noch ist die Auswirkung des Moderators Ähnlichkeit: Bei geringer Ähnlichkeit zwischen Arbeitstätigkeit und Test beläuft sich die Validität auf $r = .35$ (Berufsleistung) bzw. $r = .46$ (Trainingsleistung), bei hoher Ähnlichkeit dagegen auf $r = .62$ und $.72$. Die Methode der Wahl scheint also die Entwicklung von Fachkenntnistests zu sein, die der Arbeitstätigkeit sehr ähnlich sind.

Tests des impliziten Wissens oder auch der „praktischen Intelligenz“ wurden vor allem von Sternberg und Wagner (1993) konzipiert und propagiert. Sternberg und Wagner sehen darin eine der „theoretischen“ oder „akademischen“ Intelligenz gegenüberzustellende Art der Befähigung, die sich dadurch auszeichnen soll, daß sie prozeduraler Natur ist, den Zielen entspricht, die Menschen erreichen wollen, und ohne wesentliche Hilfe durch andere zustande kommt. Die Aufgaben sehen typischerweise so aus, daß arbeitsbezogene Situationen geschildert und jeweils eine Reihe von Fragen dazu gestellt werden. Nach Sternberg (1997) liegt die Validität zwischen .20 und .40. In einer Studie von Sue-Chan, Latham und Evans (1997, zitiert nach Salgado, 1999) hatte *tacit knowledge* bei der Prognose des Ausbildungserfolgs von Krankenschwestern keine inkrementelle Validität über die allgemeine Intelligenz hinaus.

Die dritte Variante von berufsbezogenen Wissenstests sind situative Aufgaben, ähnlich den situativen Interviewfragen (Latham, Saari, Pursell & Campion, 1980), allerdings in schriftlicher Vorgabe. Motowidlo und Tippins (1993), von denen dieser Verfahrenstyp als *low fidelity simulation* bezeich-

net wird, fanden eine Validität von $r = .28$. Pulakos und Schmitt (1996) fanden Korrelationen situativer Aufgaben von $r = .20$ mit Intelligenztests, wohingegen diese von McDaniel, Finnegan, Morgeson, Campion und Braveman (1997, zitiert nach Salgado, 1999) auf $r = .53$ beziffert wird. Als Moderator der Korrelationshöhe machten sie die Durchführung einer Arbeitsanalyse aus: Während arbeitsanalysebasierte Aufgaben eine Korrelation von $r = .68$ zu Intelligenztests aufwiesen, lag diese bei Aufgaben ohne Arbeitsanalyse lediglich bei $.29$. Damit haben wir das merkwürdige Phänomen vor uns, daß ausgezeichnet arbeitsplatzähnliche Testaufgaben zu einem höheren Anteil ihrer Validität auf allgemeiner Intelligenz basieren und folglich geringere Aussicht auf inkrementelle Validität aufweisen. Hier hat die weitere Forschung noch für Konstruktaufklärung zu sorgen.

Aus den bisherigen Erörterungen sollte erkennbar geworden sein, daß Intelligenz die Basis des Fachwissens darstellt, etwa in dem Sinne, daß sie das Potential zum Erwerb von Kenntnissen bereitstellt. Dies gilt natürlich nicht nur für Fachkenntnisse, sondern für Kenntnisse oder Wissen ganz allgemein. Vermutlich liegt hierin auch der wichtigste Grund dafür, daß der Wortschatz eines Menschen einen passablen Schätzwert für seine allgemeine Intelligenz abgibt, sowie für die hohe Korrelation ($r = .50$) zwischen Intelligenz und Schulleistungen. Auch allgemeine Wissenstests sind deshalb zu den kognitiven Fähigkeitstests zu rechnen. Naheliegenderweise ist ihr Bezug zur Berufsleistung geringer als der von Fachkenntnistests. Ein Beispiel ist der Differentielle Wissenstest (DWT) von Jäger (1969). Er besteht aus 11 Subtests, die sich auf verschiedene Wissensgebiete beziehen (Sport, Geldwesen, Politik, Technik, Physik/Chemie, Biologie, Erdkunde, Geschichte, Literatur, Kunst, Musik). Das Profil eines Testteilnehmers gibt Aufschluß über dessen Kenntnisschwerpunkte (und -lücken) bzw. über die Ausprägung seiner Allgemeinbildung. Die Testergebnisse enthalten sowohl Tätigkeits- wie Interessenkomponenten und eignen sich deshalb relativ gut zu Zwecken der Berufsberatung. Andere Tests des Allgemeinwissens sind direkt als Subtests in Intelligenztests integriert, beispielsweise im Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE-R, Tewes, 1991).

2.2.2.2 Aufmerksamkeit und Konzentration

Das psychologische Konstrukt „Aufmerksamkeit“ beschreibt „die auf die Beachtung eines Objekts (Vorgangs, Gegenstands, Idee usw.) gerichtete Bewußtseinshaltung, durch die das Beobachtungsobjekt apperzipiert [d.h. ‘bewußt’ wahrgenommen] wird“ (Dorsch, Häcker & Stapf, 1994, S. 69). „Konzentration“ bezeichnet hingegen eine leistungsbezogene, kontinuierliche Reizselektion, bei der die Aufmerksamkeit auf relevante Stimuli beschränkt wird, während irrelevante Stimuli abgeschirmt werden (Brickenkamp, 1994). Eine herausgehobene eignungsdiagnostische Bedeutung erhalten die beiden Merkmale bei beruflichen Tätigkeiten mit hochgewichteten Sicherheits- und Prüfbedingungen. Dazu zählen z. B. Tätigkeiten mit kursorischer Korrekturlesung (z. B. Korrektor im Verlagswesen) oder Datenprüfungen (z. B. EDV-Operatoren in Atomkraftwerken). Genauso zählen hierzu aber auch Fließbandarbeiten, bei denen eine Qualitätskontrolle von industriellen Produkten unter Zeitdruck erfolgen muß. Ein großes Anwendungsgebiet sind zusätzlich Verkehrstauglichkeitsprüfungen.

Das im deutschen Sprachraum am häufigsten eingesetzte Meßinstrument für Aufmerksamkeit und Konzentration dürfte der „Aufmerksamkeits-Belastungs-Test d2“ von Brickenkamp (1994; vgl. aber auch Moosbrugger & Heyden, 1998) sein. Er erfaßt Tempo und Sorgfalt des Arbeitsverhaltens bei der Unterscheidung ähnlicher visueller Reize. In Kasten 5 ist ein Testauszug dargestellt.

Intelligenz stellt die Basis des Fachwissens dar, indem sie das Potential zum Erwerb von Kenntnissen bereitstellt.

Allgemeine Wissenstests sind zur Gruppe der kognitiven Fähigkeitstests zu zählen.

Definitionen von Aufmerksamkeit und Konzentration

Auszug aus dem Aufmerksamkeits-Belastungs-Test d2

Konzentration ist eine notwendige, aber keine hinreichende Bedingung für gute Leistungen in Fähigkeitstests.

Spezielle kognitive Fähigkeiten: Teilkonstrukte der Intelligenz oder Mischungen aus Grundfähigkeit und Fertigkeit oder Wissen?

Der ABAT-R: Eine Mischform zwischen Konstrukt- und Simulationsansatz

Kasten 5:

Auszug aus dem Aufmerksamkeits-Belastungs-Test d2 von Brickenkamp (1994)

Testmaterial (Auszug):

" " ' " " ' ' " " ' ' "
 d p d d d d p d d p d d d d p p
 ' ' " " " ' ' " " "

Instruktion (verkürzt):

Aus einer Reihe ähnlicher Zeichen ist innerhalb eines knapp bemessenen Zeitrahmens (20 Sekunden) jedes d, das mit insgesamt zwei Strichen versehen ist, „so schnell wie möglich – aber natürlich auch ohne Fehler“ durchzustreichen.

Nach einigen Kontroversen zur besten Erfassung der Testleistung (vgl. zusammenfassend Leitner, 1998) stellte Brickenkamp 1994 in der 8. Auflage des Verfahrens den sog. KL-Wert vor (Differenz aus der Anzahl richtiger Kennzeichnungen und der Anzahl falscher Kennzeichnungen), der seines Erachtens ein ausgewogenes und reliables Maß für Sorgfalt und Schnelligkeit darstellt.

Der Zusammenhang des d2-Ergebnisses mit dem Ergebnis der Fahrprüfung konnte in mehreren Studien zur Verkehrstauglichkeit von Kraftfahrern nachgewiesen werden (vgl. für einen Überblick zu den Validitätsstudien Brickenkamp, 1994). Andere berufliche eignungsdiagnostische Untersuchungen sind selten und durch den beruflichen Wandel zum Teil veraltet (Untersuchungen zu Fernsehprüfern, Locherinnen und Datentypistinnen).

Trotz der großen Anwendungshäufigkeit (vgl. Schorr, 1995) ist allerdings letztlich noch immer nicht geklärt, aus welchen basalen Fähigkeiten sich die d2-Leistung zusammensetzt. Die in diesem Zusammenhang relativ niedrige Korrelation von maximal .4 mit kognitiven Fähigkeitstests (vgl. den Überblick bei Brickenkamp, 1994) stützt die These, daß Konzentration anscheinend eine notwendige, aber keine hinreichende Bedingung für gute Leistungen in Fähigkeitstests ist. In neueren Untersuchungen wird versucht, das durch den Konzentrationstest erfaßte Konstrukt durch Modifikation der Reizvorlage (Schmidt-Atzert & Bühner, 1997) und genauere Analyse der auftretenden Fehlertypen (Schmidt-Atzert & Bühner, 1998) weiter aufzuklären.

2.2.2.3 Sonstige Leistungsverfahren

Unter speziellen kognitiven Fähigkeiten werden zum einen Teilkonstrukte der Allgemeinen Intelligenz verstanden, also beispielsweise sprachliche Fähigkeiten oder räumliches Vorstellungsvermögen. Zum anderen werden mit diesem Begriff Fähigkeiten bezeichnet, wie sie in bezug auf Tätigkeitsanforderungen in Erscheinung treten. Dabei handelt es sich oft um eine Mischung aus Grundfähigkeit und Fertigkeit oder Wissen. Die zweite Begriffsverwendung ist besonders dort verbreitet, wo das Meßverfahren und dessen praktische Verwendung im Vordergrund steht. Ein Beispiel hierfür ist der „Mannheimer Test zur Erfassung des physikalisch-technischen Problemlösens (MTP)“ von Conrad, Baumann und Mohr (1980), der zur Auswahl von Auszubildenden in technisch-handwerklichen Berufen eingesetzt werden kann.

Ein Beispiel aus dem Bereich der kaufmännischen Berufe ist der Allgemeine Büroarbeitstest (ABAT-R) von Lienert und Schuler (1994). Der ABAT-R verbindet den Konstruktansatz mit dem Simulationsansatz der Eignungsdiagnostik, er kann also sowohl als Test wie auch als Arbeitsprobe angesehen werden. Auf der „Oberfläche“ der Messung werden Fertigkeiten erfaßt, wie sie für Bürotätigkeiten erforderlich sind (z. B. Kategorisierungen vornehmen). Sie entsprechen den Grundanforderungen in Büroberufen – wie Sekretariats-, Verwaltungs- und kaufmännische Tätigkeit –, die rasche und

qualitätsorientierte Arbeitsweise vorsehen, wobei ein Großteil der Arbeitszeit mit dem Sichten, Vergleichen, Kategorisieren und Kontrollieren von Information in verbaler und numerischer Form verbracht wird. „Hinter“ diesen Prozessen stehen generelle kognitive Fähigkeiten, wie sie auch mit klassischen Intelligenztests gemessen werden. So konnten insbesondere die Validitätsstudien von Schmidt (1993) am Berliner Arbeitsamt vier Faktoren nachweisen, die im Berliner Intelligenzstrukturmodell (Jäger, 1984; vgl. Abschnitt 2.2.1.1) zu den Kernfaktoren geistiger Leistung zählen: Verarbeitungskapazität, Bearbeitungsgeschwindigkeit, verbale Fähigkeiten und numerische Fähigkeiten.

Der ABAT-R besteht aus folgenden sechs Subtests:

- Kundenbriefe-Sortieren
- Adressen-Prüfen
- Summen-Prüfen
- Rechtschreibung-Korrigieren
- Textaufgaben-Lösen
- Zeichen-Setzen

Ein Auszug aus dem Subtest *Kundenbriefe-Sortieren* wird in Abbildung 7 vorgestellt. Die Aufgabe besteht in einer Mehrfachkategorisierung, die unter Zeitbegrenzung vorzunehmen ist.

Der ABAT-R als Beispiel für einen Leistungstest aus dem Bereich kaufmännischer Berufe

Kundenbriefe-Sortieren									
1. Kartei: Alphabet									
Aa-AI 1	Am-Az 2	Ba-Bi 3	Bj-Br 4	Bs-Bz 5	Ca-Cz 6	Da-Dk 7	DI-Dz 8	Ea-Ek 9	El-Ez 10
Fa-Fr 11	Fs-Fz 12	Ga-Go 13	Gp-Gz 14	Ha-Hj 15	Hk-Hz 16	Ia-Iz 17	Ja-Jz 18	Ka-Ko 19	Kp-Kz 20
La-Le 21	Lf-Lz 22	Ma-Me 23	Mf-Mo 24	Mp-Mz 25	Na-Nk 26	Ni-Nz 27	Oa-Oz 28	Pa-Pq 29	Pr-Pz 30
Qua-Quz 31	Ra-Rz 32	Sa-Si 33	Sj-Ss 34	St-Sz 35	Ta-Ti 36	Tj-Tz 37	U-V 38	Wa-Wz 39	X-Y-Z 40
2. Kartei: Branche									
Lebensmittel L			Technik T				Behörden B		
3. Kartei: Datum									
1. Vierteljahr I		2. Vierteljahr II			3. Vierteljahr III		4. Vierteljahr IV		
	1.	2.	3.		1.	2.	3.		1.
1. Fuester, Vermessungsamt, 15.09.				19. Kloeters, Klimaanlagen, 10.02					
2. Hack, Flugzeugbau, 04.01.				20. Luetgens, Oberschulamt, 01.08.					
3. Timmermann, Feinkost, 21.12.				21. Meyerhoot, Metzgerei, 18.10.					
...									

Abbildung 7:

Auszug aus dem Revidierten Allgemeinen Büroarbeitstest (ABAT-R) von Lienert und Schuler (1994)

Für den ABAT-R wurden möglicherweise aufgrund seines „Doppelcharakters“ relativ hohe Validitätskoeffizienten ermittelt; für die Vorgängerversion ABAT errechnete Lienert – allerdings an teilweise kleinen Stichproben – Validitäten bis $r = .71$.

Parallel- und Retestreliabilität des ABAT-R betragen $r = .79$ bzw. $r = .83$. Im Rahmen der Retestdurchführung nach einer Zeitspanne von acht Wochen wurde auch der Übungsgewinn festgehalten, den die Testteilnehmer erzielten. Er betrug ca. $.4$ bzw. $.7$ Standardabweichungen bei Vorgabe der Parallelversion bzw. der gleichen Testversion. Bei wiederholter Testanwendung ergibt sich daraus eine Spanne der Unsicherheit, wenn nicht bekannt ist, ob der Proband schon einmal Testteilnehmer war. Dies ist allerdings ein Problem sämtlicher Tests und auch aller übrigen eignungsdiagnostischen Verfahren. Da es selten untersucht wird, fehlen entsprechende Veränderungswerte für wiederholte Teilnahme gewöhnlich in den Testbeschreibungen.

2.2.3 Allgemeine Persönlichkeitskonstrukte

Zunächst soll wieder ein allgemeines Modell vorgestellt werden. Im Gegensatz zum Intelligenzbereich ist diese Darstellung allerdings nicht als exemplarisch zu bezeichnen, da das vorgestellte Fünf-Faktoren-Modell der Persönlichkeit im Bereich der strukturtheoretischen Ansätze im Laufe des letzten Jahrzehnts eine Hegemonialstellung erreicht hat.

2.2.3.1 Fünf-Faktoren-Modell der Persönlichkeit

Der Streit, über wie viele und welche Faktoren die nicht-kognitive Persönlichkeitsstruktur eines Menschen ausreichend beschrieben wird, wird schon lange ausgetragen (vgl. als Beispiel die Kommentierung des Artikels von Goldberg, 1993, im *American Psychologist*). In neueren Studien konnten allerdings gehäuft fünf abstrakte Faktoren nachgewiesen werden, die eine hinreichende Robustheit aufzuweisen scheinen. Bemerkenswert ist die Konvergenz ganz unterschiedlicher Forschungsansätze: So fanden sich ähnliche Befunde in lexikalischen Ansätzen, in faktorenanalytischen Reanalysen und in interkulturellen Vergleichsstudien (vgl. als Übersicht Digman, 1990, oder Goldberg, 1993). In Kasten 6 ist eine grobe Charakterisierung der fünf Faktoren (auch kurz „Big5“ genannt) wiedergegeben.

Kasten 6:

Charakterisierung der Faktoren des Big5-Modells (in Anlehnung an Schuler, 1996)

Extraversion
gesellig, gesprächig, großzügig, bestimmt, dominant, aktiv, impulsiv
Neurotizismus (Gegenpol: <i>Emotionale Stabilität</i>)
ängstlich, deprimiert, verlegen, emotional, leicht verärgert, besorgt, unsicher
Verträglichkeit
freundlich, höflich, flexibel, vertrauensvoll, kooperativ, tolerant, versöhnlich, weichherzig
Gewissenhaftigkeit
verlässlich, sorgfältig, verantwortungsbewußt, planvoll, organisiert, leistungsorientiert, ausdauernd
Offenheit für Erfahrungen (auch <i>Intellekt</i> oder <i>Kultiviertheit</i> genannt)
einfallsreich, kultiviert, originell, vielseitig, intellektuell, aufgeschlossen, ästhetikbetont

Das Fünf-Faktoren-Modell der Persönlichkeit hat eine Hegemonialstellung im Bereich der allgemeinen Persönlichkeitseigenschaften erreicht.

Charakterisierung der Faktoren des Big5-Modells

Allerdings ist auch diese sehr abstrakte Faktorencharakterisierung nicht allgemein in der Fachwelt akzeptiert. So beschäftigen sich beispielsweise alle Beiträge in einem Sonderheft des *European Journal of Personality* (Heft 4, Band 8, 1994) einzig mit der Charakterisierung des Faktors „Offenheit für Erfahrungen“. Zum jetzigen Zeitpunkt kann das Big5-Modell deshalb nur als grober Orientierungsrahmen für angewandte Forschung dienen.

Während im angloamerikanischen Sprachraum bereits einige Big5-Verfahren entwickelt wurden, ist bisher (Stand: Frühjahr 2000) nur eine Kurzform des NEO-PI-R von Costa und McCrae (1992), der NEO-FFI, noch ohne adäquate Normierung ins Deutsche übersetzt und regulär veröffentlicht worden (Borkenau & Ostendorf, 1993). In Kasten 7 sind einige Items des NEO-FFI wiedergegeben.

Kasten 7:

Einige Items des NEO-FFI (Borkenau & Ostendorf, 1993)

Extraversion
<ul style="list-style-type: none"> • Ich habe gerne viele Leute um mich herum. • Ich habe oft das Gefühl, vor Energie überzuschäumen.
Neurotizismus
<ul style="list-style-type: none"> • Ich fühle mich anderen oft unterlegen. • Ich empfinde selten Furcht oder Angst. (negativ gepolt)
Verträglichkeit
<ul style="list-style-type: none"> • Ich versuche zu jedem, dem ich begegne, freundlich zu sein. • Manche Leute halten mich für selbstsüchtig und selbstgefällig. (negativ gepolt)
Gewissenhaftigkeit
<ul style="list-style-type: none"> • Ich halte meine Sachen ordentlich und sauber. • Ich vertrödele eine Menge Zeit, bevor ich mit einer Arbeit beginne. (negativ gepolt)
Offenheit für Erfahrungen
<ul style="list-style-type: none"> • Ich probiere oft neue und fremde Speisen aus. • Wenn ich Literatur lese oder ein Kunstwerk betrachte, empfinde ich manchmal ein Frösteln oder eine Welle der Begeisterung.

Einige Items des NEO-FFI

2.2.3.2 Validität allgemeiner persönlichkeitsorientierter Verfahren

Beim Einsatz von persönlichkeitsorientierten Verfahren in der Eignungsdiagnostik sind historische Parallelen zu den intelligenzbezogenen Verfahren erkennbar (vgl. Abschnitt 2.2.1.3): Nach anfänglich regem Einsatz zeigten erste systematische Sekundärstudien (z. B. Guion & Gottier, 1965) nur Validitätskoeffizienten bis maximal .3, was von den Autoren als äußerst unbefriedigend angesehen wurde. Bei allen älteren Metaanalysen fallen allerdings methodische Mängel auf:

- *Mängel der durchgeführten Studien und der verwendeten Stichproben*
Sehr viele Primärstudien wiesen Mängel auf hinsichtlich des eingesetzten Validierungsdesigns und der Stichprobenwahl. Neben einer nur teilweise repräsentativen Personenauswahl ist hier auch die Vermengung ziviler und militärischer Testanwendungen sowie eine mangelhafte Trennung hinsichtlich des ursprünglichen Zwecks der Verfahrensanwendung (Auswahl vs. Screening) zu nennen.
- *Mängel der erhobenen Konstrukte*
In der Zusammenschau wurden allgemeine Konstrukte, die über ursprünglich für psychopathologische Zielgruppen optimierte Meßverfahren erhoben wurden, gleichgestellt mit sehr spezifischen, berufsbezogenen Konstrukten,

Ältere Metaanalysen zur Validität von Persönlichkeitseigenschaften weisen methodische Mängel auf.

die sich nur auf schmale, relativ eindeutig definierte Verhaltensausschnitte beziehen.

- *Wahl der Kriterien*

Bei den verwendeten Kriterien wurde nicht zwischen globalen und spezifischen Variablen unterschieden; eine Abstufung nach ihrer Zentralität für Berufserfolg fand nicht statt.

- *mangelhafte theoretische Verbindung zwischen Konstrukten und Kriterien*

In vielen Studien wurden die Persönlichkeitsverfahren im Sinne einer „Breitbandanwendung“ auch zu inhaltlich eindeutig unabhängigen Kriterien in Beziehung gesetzt und die gefundenen Nullkorrelationen dann fälschlicherweise als aussagekräftig interpretiert.

- *Schwächen der angewendeten metaanalytischen Technik*

Bei den eingesetzten frühen metaanalytischen Verfahren wurde häufig über Persönlichkeitsdimensionen, Berufe und Kriterien ohne gleichzeitige Moderatoranalysen generalisiert.

In der Forschungsliteratur zur Eignungsdiagnostik ist spätestens ab 1973 ein nachlassendes Interesse an Persönlichkeitsverfahren zu verzeichnen, was neben der enttäuschenden Befundlage auch mit der Interaktionismusdebatte in der Differentiellen Psychologie (wie bereits erwähnt veröffentlichte Mischel 1968 sein Buch „*Personality and assessment*“), der „civil rights act“-Festschreibung (vgl. z. B. Arvey, 1979) und der zunehmenden Sensibilisierung für Eingriffe in die Privatsphäre (Anastasi, 1985) erklärt werden kann. Wie bereits angedeutet, fokussierte sich u. a. als Konsequenz daraus das Interesse in der beruflichen Eignungsdiagnostik zunehmend auf simulationsorientierte Verfahren, prototypisch umgesetzt im „Assessment Center“-Ansatz (vgl. Kapitel 6).

Das neu erwachte Interesse Anfang der Neunziger Jahre läßt sich zum einen direkt mit der Rehabilitation strukturtheoretischer Persönlichkeitsmodelle in Gestalt des „Big5-Modells“ erklären. Zum anderen wiesen methodisch vorbildlich durchgeführte Einzelstudien die inkrementelle Validität von Persönlichkeitsvariablen für bestimmte Leistungskriterien nach (hier ist an erster Stelle die Untersuchung von Hough, Eaton, Dunnette, Kamp & McCloy, 1990, beim amerikanischen Militär zu nennen).

Neuere Metaanalysen (Barrick & Mount, 1991; Salgado, 1997; Tett, Jackson & Rothstein, 1991) nehmen direkt Bezug auf das Fünf-Faktoren-Modell. Die Big5-Struktur dient bei diesen Studien als konzeptionelles Schema, in das die Primärstudien, deren Konstruktdefinitionen von denen der „Big 5“ abweichen, post hoc integriert wurden. Allerdings ergeben sich im direkten Vergleich sehr widersprüchliche Befunde. Eine Ergebnisübersicht ist in Tabelle 1 zusammengestellt. Zu berücksichtigen ist dabei, daß die Analysen auf unterschiedlichen Primärstudienstichproben basieren: Barrick und Mount (1991) berücksichtigten praktisch alle Studien, derer sie habhaft werden konnten. Tett et al. (1991) beschränkten sich auf Studien, die eine hypothesenprüfende Zielstellung verfolgten, und ließen rein explorative Studien (Korrelation von Testbatterie A mit Kriteriumssammlung B und post hoc-Interpretation der Signifikanzen) aus. Salgado (1997) schließlich berücksichtigte nur Studien, die in Staaten der Europäischen Union durchgeführt wurden.

Festzuhalten bleibt hier also, daß unterschiedliche Metaanalysen zu demselben Thema zu ganz verschiedenen Ergebnissen führen können. Die Details der Studien helfen hier wesentlich weiter (vgl. auch Hossiep, Paschen & Mühlhaus, 2000):

- Die post hoc-Einordnung der ursprünglichen Konstrukte in das Big5-Schema ist wegen der Unschärfe der Zordnungsregeln und der damit verbundenen geringen Konvergenz der Einzelmessungen als eher kritisch zu betrachten.
- Nachgewiesene Moderatoren für die gefundenen Validitätskoeffizienten sind neben den Persönlichkeitsfaktoren zusätzlich die Art der berücksich-

Neuere Metaanalysen nehmen direkten Bezug auf das Fünf-Faktoren-Modell.

Nachgewiesene Validitätsmoderatoren: Art der Eigenschaften, des Berufs, des Kriteriums

tigten Leistungskriterien sowie die untersuchten Berufe. Alle drei Analysen können hier nur mit exemplarischen Befunden aufwarten, die große Unterschiede aufzeigen.

- Auch nach einer detaillierten Moderatorenanalyse (besonders ausführlich z. B. bei Tett et al., 1991) verbleibt noch eine große unaufgeklärte Restvarianz. Dies weist daraufhin, daß die Heterogenität der verwendeten Primärstudien einen zusätzlichen Einfluß auf die gefundenen Validitätskoeffizienten ausübt.

Tabelle 1:

Ergebnisse neuerer Metaanalysen zur kriteriumsbezogenen Validität von allgemeinen Persönlichkeitsvariablen

Studien	Barrick & Mount (1991)	Tett et al. (1991)	Salgado (1997)
Big5-Dimension	korrigierte Val ^a	korrigierte Val ^b	korrigierte Val ^c
Extraversion	.13	.16	.08
Neurotizismus	-.08	-.22	-.13
Verträglichkeit	.07	.33	.01
Gewissenhaftigkeit	.22	.18	.15
Offenheit für Erfahrungen	.04	.27	.06

a: Analyse basiert auf Tabelle 3 aus Barrick & Mount (1991): Berücksichtigt wurden zwischen 91 und 141 Koeffizienten basierend auf einer Gesamtstichprobe zwischen 15939 und 31211 Personen. Die Koeffizienten wurden hinsichtlich Varianzeinschränkungen, Prädiktor- und Kriteriumsreliabilität korrigiert.

b: Darstellung basiert auf Tabelle 5 aus Tett et al. (1991). Berücksichtigt wurden zwischen 4 bis 15 Koeffizienten basierend auf einer Gesamtstichprobe zwischen 280 und 2302 Personen. Die Koeffizienten wurden nur hinsichtlich Prädiktor- und Kriteriumsreliabilität korrigiert.

c: Darstellung basiert auf Tabelle 1 in Salgado (1997). Berücksichtigt wurden zwischen 18 und 32 Koeffizienten basierend auf einer Gesamtstichprobe zwischen 2722 und 3877 Personen. Die Koeffizienten wurden hinsichtlich Varianzeinschränkungen, Prädiktor- und Kriteriumsreliabilität korrigiert.

Ergebnis dreier konkurrierender Metaanalysen zur kriteriumsbezogenen Validität von allgemeinen Persönlichkeitsvariablen

Die vorliegenden Metaanalysen scheinen zum jetzigen Zeitpunkt noch zu stark über vorhandene systematische Einflüsse zu generalisieren. Möglicherweise müssen noch einige Jahre vergehen, bis genügend Primärstudien mit direkten Big5-Messungen für eine metaanalytische Aggregation vorhanden sind. So könnte zumindest ein großer Unsicherheitsfaktor (die post hoc-Zuordnung von Persönlichkeitsskalen) ausgeschaltet werden.

Als Abschluß soll beispielhaft die Studie von Stewart (1996) berichtet werden, die sehr anschaulich die Kriterienabhängigkeit der Validität ganz bestimmter Persönlichkeitsmerkmale aufzeigt.

2.2.3.3 Kriteriumsabhängige Validität von Persönlichkeitseigenschaften

Stewart (1996) greift für seine Studie auf die Theorie von Gray (1973) zurück (vgl. auch Asendorpf, 1996). Dieser wiederum führt die von Eysenck (1967) aufgestellte ARAS-Theorie weiter. Danach unterscheiden sich extravertierte und introvertierte Personen nach dem Ausprägungsgrad ihres Verhaltensaktivierungssystems („behavioral activation system“): Extravertierte weisen ein niedriges internes Aktivitätsniveau auf und suchen ihre Umgebung aktiv nach Reizen ab, die dieses Niveau steigern können. Sie sollten also wesentlich empfänglicher für externale Bekräftigungen sein als introvertierte Personen.

Die vorliegenden Metaanalysen scheinen noch zu stark über vorhandene systematische Einflüsse zu generalisieren.

Wird der Zusammenhang zwischen Extraversion und Berufserfolg moderiert durch die Art des bestehenden Gratifikationssystems?

Eine einfache Inspektion der Korrelationsmatrix zeigt keinen Zusammenhang zwischen Extraversion und den Berufserfolgsvariablen.

Erst eine moderierte Regressionsanalyse offenbart die bestehende Interaktion zwischen Extraversion und der Art des Belohnungssystems.

Stewart leitet aus diesen grundsätzlichen Annahmen die Hypothese ab, daß der Zusammenhang zwischen Extraversion und Berufserfolg unter anderem moderiert wird durch die Art des bestehenden Gratifikationssystems des Unternehmens für berufliche Leistungen: Es sollte sich nur ein Zusammenhang zwischen Extraversion und beruflichen Leistungsdaten ergeben, wenn diese Leistungselemente explizit belohnt werden. Kein bzw. nur ein geringer Zusammenhang sollte hingegen mit nichtbelohnten Leistungsaspekten bestehen. Die Persönlichkeitseigenschaft „Gewissenhaftigkeit“ sollte im Kontrast dazu unabhängig vom Belohnungssystem einen durchweg positiven Zusammenhang mit Berufserfolgsvariablen aufweisen.

Zur Überprüfung dieser Hypothesen untersucht Stewart 152 Außendienstmitarbeiter einer politischen Interessenvertretung für Unternehmen. Aufgabe dieser Mitarbeiter ist zum einen die Akquisition neuer Mitglieder für den Dachverband, zum anderen aber auch die Erneuerung laufender Mitgliedsverträge. 77 Personen gehörten Abteilungen an, in denen die Akquisition von neuen Mitgliedern belohnt wurde, während 75 Personen zu Abteilungen gehörten, in denen die Kundenbewahrung explizit gratifiziert wurde.

Die Ausprägung von Extraversion und Gewissenhaftigkeit wurde über entsprechende Skalen des NEO-PI-R erfaßt. Als Leistungsdaten wurde der Prozentsatz der betreuten Mitgliedschaftserneuerungen sowie die Anzahl der angeworbenen Neumitgliedschaften innerhalb einer 9-Monats-Phase erhoben. Durch eine geeignete Stichprobenziehung und durch die Berücksichtigung der Mitgliedsanzahl im jeweiligen Bezirk des Außendienstmitarbeiters wurden geographische Einflüsse kontrolliert. In Tabelle 2 ist die Interkorrelationsmatrix der angesprochenen Variablen dargestellt.

Tabelle 2:

Interkorrelation der Variablen bei Stewart (1996) (verkürzt nach Stewart, 1996, table 1)

	Variablen	1	2	3	4	5
1	% Erneuerungen ^a	1				
2	Neue Mitglieder ^b	-.14*	1			
3	Extraversion	.10	.01	1		
4	Gewissenhaftigkeit	.11	.17*	.35**	1	
5	Belohnungssystem ^c	.01	.07	-.14*	.00	1

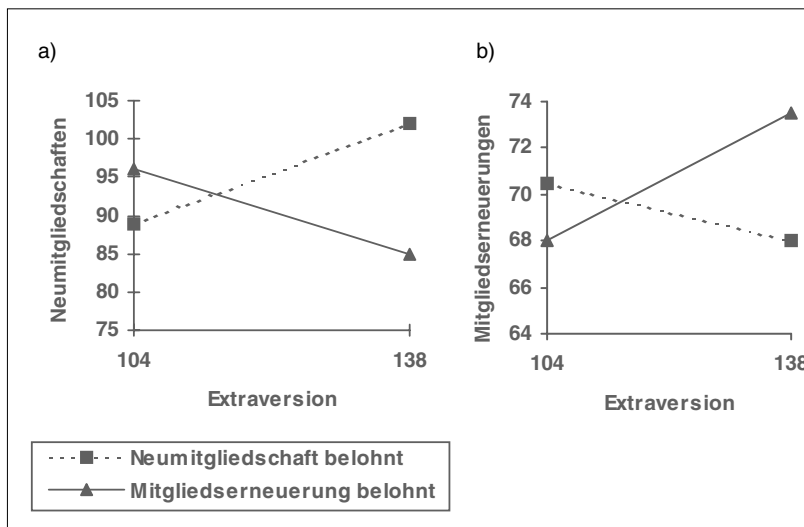
* $p < .05$ (einseitig); ** $p < .01$ (einseitig)

a: Prozentsatz von Mitgliedschaftserneuerungen

b: Anzahl neugeworbener Mitglieder

c: Dummykodierung der jeweils gratifizierten Leistungsvariable

Bei alleiniger Analyse der Korrelationsmatrix scheint Extraversion mit keiner der Leistungsvariablen zusammenzuhängen (nichtsignifikante $r = .10$ und $r = .10$), während Gewissenhaftigkeit zumindest für die Anzahl neuakquirierter Mitglieder einen signifikanten positiven Zusammenhang aufweist ($r = .10$). Die hypothetisierte Interaktion zwischen Extraversion und Art des Belohnungssystems läßt sich erst über eine moderierte Regressionsanalyse überprüfen (vgl. Stewart, 1996, table 2; vgl. für die Technik der moderierten Regressionsanalyse z. B. Jaccard, Turrisi & Wan, 1993): Hier zeigt sich, daß bei Extraversion nur der Interaktionsterm „Extraversion x Belohnungssystem“ eine substantielle Varianzaufklärung bei der Vorhersage beider Berufserfolgsvariablen leistet. Als Interpretationshilfe sind in Abbildung 8 die Regressionsdiagramme dargestellt.

**Abbildung 8:**

Regressionsdiagramme zu den abhängigen Variablen in der Stewart (1996)-Studie

Es zeigt sich also: Der Zusammenhang zwischen einer Persönlichkeitsvariable (Extraversion) und Berufserfolgsindikatoren (Neumitgliedschaften, Mitgliedererneuerungen) wird moderiert durch eine Kontextvariable (organisationales Belohnungssystem). Dieser komplexere Zusammenhang würde bei einer einfachen Korrelationsinspektion verborgen bleiben.

Diese Darstellung ist nur exemplarisch, da das organisationale Belohnungssystem sicherlich nicht der einzige (und vielleicht auch nicht der bedeutendste) Moderator des Extraversion-Leistungs-Zusammenhangs ist. Es wird aber deutlich, was die möglichen Ursachen für die heterogenen Zusammenhangsbefunde in unterschiedlichen Berufssparten sein können.

2.2.4 Spezifische Persönlichkeitseigenschaften: locus of control

So interessant es ist, psychische Eigenschaften relativ alters- und kulturinvariant zu fünf globalen Merkmalen gruppieren zu können, und so wichtig es ist, Instrumente zur Messung dieser Merkmale als Referenzverfahren zur Verfügung zu haben – in vielen Fällen wird man sich entscheiden, differenziertere Konstrukte erfassen zu wollen. Ein Beispiel hierfür ist der 16-Persönlichkeits-Faktoren-Test nach Cattell (Schneewind & Graf, 1998), mit dem, wie der Name sagt, 16 Eigenschaften unterschieden werden (die sich auf einer „Ebene zweiter Ordnung“ zu ebenfalls fünf Faktoren zusammenfassen lassen). Ein Beispiel dafür, daß selbst innerhalb eines breiter aufgefaßten Merkmals noch weiter differenziert werden kann, ist das Leistungsmotivationsinventar (LMI, Schuler & Prochaska, 2000), mit dem 17 Teilaspekte der Leistungsmotivation unterschieden werden (die sich allerdings nur zu drei orthogonalen Faktoren gruppieren).

Aber nicht nur, um Konstrukte genauer zu beleuchten, auch im Sinne der kriteriumsbezogenen Validität kann eine Aufschlüsselung von Nutzen sein. Beispielsweise zeigte die bereits erwähnte Metaanalyse von Funke et al. (1987) zum Bereich Forschung und Entwicklung, daß zur Prognose wissenschaftlich-technischer Leistungen spezielle Persönlichkeitstests tauglicher sind – also höhere Validitäten erreichen – als allgemeine Persönlichkeitstests.

Zu den spezifischen Persönlichkeitseigenschaften, die viel Aufmerksamkeit auf sich gezogen haben, gehört das Merkmal *Kontrollüberzeugung*. Es geht unter dem Terminus *locus of control* auf die soziale Lerntheorie von Rotter (1966) zurück und bezeichnet generalisierte Erwartungen von Individuen, inwieweit ihre Verstärkung internaler (also eigener) oder externaler

Spezifische Eigenschaften erfassen möglicherweise zusätzliche Validitätsinformationen, die bei globalen Eigenschaftskonstrukten verlorengehen.

Beispielitems aus dem Fragebogen zu Kompetenz- und Kontrollüberzeugungen

Kontrollüberzeugung und Leistungsmotivation sind eher spezifische Perspektiven auf Verhaltensorientierungen als spezifische Eigenschaften.

Motivation: Probleme der Begriffsklärung und Messung

Kontrolle unterliegt. Eine deutschsprachige Testfassung wurde von Krampen vorgelegt, der das Merkmal *Kontrollüberzeugung* in den Rahmen eines handlungstheoretischen Modells selbstbezogener Kognitionen stellt. Neben Kontrollüberzeugungen mißt dieser Test auch *Kompetenzüberzeugungen* (FKK, Krampen, 1991). In Kasten 8 ist jeweils ein Item für jede der vier Skalen wiedergegeben, die der Test umfaßt.

Kasten 8:

Beispielitems aus dem Fragebogen zu Kompetenz- und Kontrollüberzeugungen (Krampen, 1991)

Skala „Selbstkonzept eigener Fähigkeiten“
Ich komme mir manchmal taten- und ideenlos vor.
Skala „Internalität“
Ob ich einen Unfall habe oder nicht, hängt alleine von mir und meinem Verhalten ab.
Skala „Soziale Externalität“
Ich habe das Gefühl, daß vieles von dem, was in meinem Leben passiert, von anderen Menschen abhängt.
Skala „Fatalistische Externalität“
Zufällige Geschehnisse bestimmen einen großen Teil meines Lebens und Alltags.

Zusammenhänge der vier Primärskalen des FKK (die sich auch zu zwei Sekundärskalen sowie einer Tertiärskala zusammenfassen lassen) werden vor allem zu klinischen Variablen wie Depression und psychosomatischen Beschwerden aufgezeigt. Deutliche Zusammenhänge finden sich auch mit handlungsbezogenen Parametern wie dem sozialen Engagement in der Gemeinde oder Alltagsaktivitäten älterer Menschen. Direkte Validitätsdaten zu berufsbezogenen Außenkriterien werden von Krampen nicht vorgelegt, aber es können doch erhebliche Profilunterschiede zwischen Führungskräften und Angestellten mit unselbständiger Tätigkeit sowie, etwas weniger deutlich, zwischen Arbeitern und Vorarbeitern aufgezeigt werden.

Wir sollten nicht versäumen, uns zum Abschluß dieses Abschnitts die Frage zu stellen, inwieweit es gerechtfertigt ist, Merkmale wie Leistungsmotivation und Kontrollüberzeugung als *spezifische Persönlichkeitseigenschaften* zu bezeichnen. Sie sind spezifisch in dem Sinne, daß nicht der Anspruch erhoben wird, mit einer geringen Zahl von Merkmalen „die ganze Persönlichkeit“ zu erfassen (wie mit einem Verfahren, das die „Big Five“ mißt). Ob sie spezifisch allerdings auch in dem Sinne genannt werden können, daß sie ein „kleineres“, enger abgegrenztes Merkmal umfassen, scheint dagegen fraglich: So haben die Konstruktanalysen zum LMI so vielfältige Beziehungen der Leistungsmotivation zu anderen Persönlichkeitsmerkmalen gezeigt, daß ein Verständnis der Leistungsmotivation als Ausrichtung eines großen Teils der Gesamtperson auf die Leistungsthematik nahegelegt wird. Ähnliches gilt für die Merkmale Kompetenz- und Kontrollüberzeugungen. Der wesentliche Unterschied zwischen spezifischen und allgemeinen Merkmalen und den entsprechenden Testverfahren scheint also vor allem darin zu liegen, daß wir im ersteren Fall eine spezifische Perspektive einnehmen, den Fokus auf eine bestimmte Verhaltensorientierung richten, während wir in letzterem Fall an einer möglichst „ökonomischen“, breiteren Beschreibung der Persönlichkeit interessiert sind.

2.2.5 Motivation und Interessen

Die Begriffe „Motiv“ und „Motivation“ spielen zur Erklärung des Verhaltens in der Persönlichkeitspsychologie eine ebenso große Rolle wie im Alltagsleben. Alle Versuche, Motive zu messen, haben sich jedoch mit dem Problem auseinanderzusetzen, daß wir es hier mit „hypothetischen Konstruk-

ten“ zu tun haben – mit Erklärungsbegriffen, die ein Agens als Erklärung für beobachtetes Verhalten oder Verhaltensergebnisse postulieren. Diese treibende Kraft ist als solche nicht beobachtbar, und in manchen psychologischen Theorien wird sogar angenommen, sie sei auch für das handelnde Individuum nicht erkennbar. Dies schließt eine Messung selbstverständlich nicht aus, erfordert aber jeweils den Nachweis eines (kausalen) Zusammenhangs zwischen der postulierten Triebkraft und dem Verhaltensereignis. Die Erklärung sollte möglichst nichttrivial sein – trivial wäre etwa die Erklärung sportlicher Aktivitäten durch ein Sportmotiv, nichttrivial die durch ein Leistungs-, Anerkennungs- oder Aggressionsmotiv. Naturwissenschaftliche Stringenz – wie etwa bei der Rückführung des Appetits nach einem Apfel auf einen objektiv bestimmbaren Vitaminmangelzustand – ist auch auf diesem Feld in der Psychologie schwer zu erreichen.

Lange Zeit maßgeblich für die Motivmessung war der Versuch, durch mehr oder weniger unbestimmte, aber interpretierbare Reizvorlagen die dominierenden Strebungen erkennbar zu machen. Einflußreiches Beispiel hierfür ist der Thematische Apperzeptionstest (TAT) von Murray (1943). Aus den Geschichten, die von den Probanden zu den einzelnen Bildern erzählt werden, hofft man Zugang zu Motiven und Handlungstendenzen der Person zu finden, insbesondere zu solchen Beweggründen des Handelns oder auch unerfüllten Bedürfnissen, die der Selbstzensur unterliegen. Die psychologische Funktionstheorie eines solchen *operanten* Meßverfahrens – im Unterschied zu einem *respondenten* Verfahren, etwa einem Fragebogen – ist bei Murray und seinen Nachfolgern die der *Projektion*. Ohne dieses schwierige tiefenpsychologische Konzept versuchen heutige Interpretationen auszukommen, die das Entsprechungsverhältnis von Bedürfnis und Aufforderungscharakter der Situation in den Vordergrund stellen. Schmalt und Sokolowski (1996) sprechen von „subjektiven Anreiz-Landschaften“, die in den erzählten Geschichten entstehen. Ein Beispiel für ein Testverfahren, das nach diesem Prinzip konstruiert wurde, ist der von Heckhausen in Anlehnung an McClelland erarbeitete Leistungsmotivationstest (Heckhausen, 1963).

Legt man übliche psychometrische Qualitätskriterien an, so hat sich die These nicht bestätigt, daß Motive besser mittels operanter als mittels respondenten Verfahren meßbar seien. Ganz im Gegenteil haben sich in bezug sowohl auf die Konstruktvalidierung als auch auf die Validierung an Außenkriterien Fragebogenverfahren auch zur Messung der Motivation ganz eindeutig als überlegen erwiesen (Prochaska, 1998). Deshalb finden sich, um beim gewählten Beispielkonstrukt zu bleiben, unter den heute eingesetzten Tests zur Erfassung der Leistungsmotivation nahezu ausschließlich Verfahren vom Fragebogentypus. Beispiele hierfür sind der *Leistungs Motivations Test (LMT)* von Hermans, Petermann und Zielinski (1978), der als Teilaspekte *Leistungsstreben, Ausdauer und Fleiß, Leistungshemmende Prüfungsangst* und *Leistungsfördernde Prüfungsangst* unterscheidet, sowie das *Leistungsmotivationsinventar (LMI)* (Schuler & Prochaska, 2000), hinter dem nicht die Auffassung der Leistungsmotivation als enges separierbares Persönlichkeitsmerkmal steht, sondern die einer Orientierung großer Teilbereiche der Persönlichkeit auf die Leistungsthematik hin. Der Versuch, mit einem „semi-projektiven“ Verfahren den operanten mit dem respondenten Ansatz der Messung zu verknüpfen – es handelt sich de facto um ein projektives Verfahren mit Multiple-Choice-Vorgabe – wurde von Schmalt (1999) mittels der sogenannten „Gittertechnik“ unternommen.

Notorisch unterschätzt wird die Bedeutung von *Interessentests* bei der Personalauswahl. Üblicherweise hält man ihr Anwendungsgebiet für beschränkt auf die Berufsberatung – wo sie tatsächlich eine besonders wichtige Funktion haben –, übersieht dabei aber, daß Interessen vor allen anderen Parametern die Bestimmungsgrößen der Selbstselektion sind. Daß sie auch gute Prädiktoren der Berufszufriedenheit, Fluktuation und ähnlicher Kriterien sind, muß beim Stand des Wissens leider eher als Plausibilität denn als Faktum dargestellt werden.

Eine Ursache für die Vernachlässigung von Interessentests dürfte in der Vermutung liegen, sie seien für vorsätzliche Verfälschung noch anfälliger

Zur Meßproblematik von Motiven

Interessen sind vorrangig Bestimmungsgrößen der Selbstselektion, möglicherweise auch Prädiktoren der Berufszufriedenheit, der Fluktuation und ähnlicher Kriterien.

Zur Meßproblematik von Interessen

als andere Fragebogenverfahren. Das könnte (muß aber nicht) zutreffen auf *Einstufungen* als Skalierungsformat. Wenn die Frage vorgelegt wird, „Wie groß ist Ihr Interesse an der genauen, gewissenhaften Prüfung eingehender Rechnungen?“, so mag es verlockend scheinen, auf einer Skala von 1 bis 5 den höchsten Wert anzukreuzen, wenn man annehmen kann, eben dieses Verhalten werde von einem Mitarbeiter verlangt. Allerdings müssen die Items keineswegs zwangsläufig mit einer Einstufungsskalierung verbunden sein. Eine *Rangordnung* beispielsweise zwischen Items, die auf durchschnittlich gleiche Attraktivität vorgeprüft sind, zwingt zur Auswahl und Differenzierung. Eine interessante Variante der Rangordnung ist die sogenannte *Forced-choice-Skalierung*, nach deren Muster der *Berufs-Interessen-Test* (BIT II, Irlé & Allehoff, 1988) aufgebaut ist. Bei diesem Verfahren werden dem Probanden immer gleichzeitig vier Aussagen vorgelegt, die Aufgaben aus verschiedenen Tätigkeitsbereichen benennen und von denen jeweils die am stärksten bevorzugte und die am meisten abgelehnte Aufgabe gekennzeichnet werden muß. Eine Gesamtzahl von 81 Vergleichen ergibt schließlich ein recht aussagekräftiges Profil der Interessen in bezug auf neun Berufsfelder. Der Berufs-Interessentest ermöglicht durch seine Konstruktionsweise die Einschätzung der relativen Interessen oder Berufspräferenzen. Was andererseits in Kauf genommen werden muß, ist das Fehlen einer Aussage über die absolute Höhe der Interessen. Eine solche Aussage erlaubt die *Generelle Interessen-Skala* (GIS, Brickenkamp, 1990). In diesem Verfahren werden sechzehn Interessenbereiche unterschieden. Anders als beim BIT II ist allerdings kein unmittelbarer Bezug zu beruflichen Tätigkeitsfeldern gegeben, auch wird hier nicht qua Skalierungsverfahren Vorsorge gegen beschönigende Selbstdarstellung getroffen.

Auf der Basis allgemeiner Grundhaltungen oder Lebensorientierungen arbeitet der *Allgemeine Interessen-Struktur-Test* von Holland (deutsche Fassung von Bergmann & Eder, 1992). Die dahinterstehende Theorie sieht Interessen als grundlegende Persönlichkeitsorientierungen an und damit die Berufswahl als bedingt durch allgemeine Wesensmerkmale der Person (Holland, 1997). Holland folgt hier einer alten Typologie von Spranger (1913), ohne daß dies in der Testpublikation zum Ausdruck gebracht würde (siehe Kasten 9).

Kasten 9:

Alter Wein in neuen Schläuchen? Hollands Typenschema (aus Seifert, 1977, S. 209f)

Hollands RIASEC-Theorie als Theorie allgemeiner beruflicher Interessen

Typen oder Orientierungen nach Holland

Typen oder Orientierungen nach Holland

1. *Realistische Orientierung*

Personen dieses Typus sind „männlich“, aktiv, aggressiv; sie sind an physischer Aktivität interessiert und motorisch befähigt; sie bevorzugen konkrete Gegebenheiten gegenüber abstrakten Problemen; sie haben eher konventionelle Werthaltungen – sowohl im politischen wie im ökonomischen Bereich. Bei der Berufswahl tendieren sie am meisten zu handwerklichen und technischen sowie zu land- und forstwirtschaftlichen Berufen.

2. *Intellektuelle Orientierung*

Dieser Typus ist aufgabenorientiert und in gewissem Sinne „asozial“; er trachtet danach, Probleme intellektuell zu bewältigen, „durch Manipulation von Ideen, Worten und Symbolen“, nicht dagegen durch physische oder soziale Aktivität. Er hat ein starkes Bedürfnis, Zusammenhänge zu verstehen und besitzt eher unkonventionelle Wertvorstellungen und Einstellungen. Menschen dieses Typus werden vor allem in naturwissenschaftlichen und mathematischen Berufen gefunden.

3. *Soziale Orientierung*

Menschen dieses Typus sind sozial orientiert und von sozialer Verantwortung erfüllt. Sie haben ein starkes Bedürfnis nach Beachtung und sozialer Interaktion und verfügen über gute verbale und soziale Fähigkeiten. Sie

tendieren dazu, Probleme eher emotional und durch soziale Aktivität zu bewältigen als intellektuell. Typische Berufe in diesem Bereich sind: pädagogische und sonderpädagogische Berufe, Sozialarbeiter, klinischer Psychologe, Berufsberater.

4. Konventionelle Orientierung

Dieser Typus bevorzugt weitgehend strukturierte verbale und numerische Aktivitäten und Untergebenenrollen; er ist konformistisch („extracptive“) eingestellt und vermeidet unklare Situationen sowie Probleme, die soziale Aktivität oder ausgeprägte physische Fähigkeiten erfordern; er identifiziert sich mit Machtpositionen und schätzt materiellen Besitz und Status. Bei der Berufswahl und der beruflichen Ausbildung werden z. B. folgende Berufe gewählt: Buchhalter, Rechnungsprüfer, Bankangestellter, Statistiker, EDV-Operator.

5. Unternehmerische Orientierung

Menschen dieser Art verstehen sich als starke, männliche Führerpersönlichkeiten. Sie besitzen ausgeprägte verbale Fertigkeiten und fühlen sich wohl, wenn sie anderen etwas verkaufen oder mit ihnen in Konkurrenz treten können. Sie vermeiden jedoch klar definierte verbale Situationen sowie Aufgaben, die einen längeren, angestregten intellektuellen Einsatz erfordern. Die beruflichen Präferenzen betreffen u. a. folgende Berufe: Hotelier, Unternehmer, Industrieberater, Immobilienhändler, Wahlkampfmanager, Versicherungsvertreter und dergleichen.

6. Künstlerische Orientierung

Dieser Typus ähnelt dem intellektuellen hinsichtlich seiner „intrazeptiven“ und eher „asozialen“ Ausrichtung, er unterscheidet sich jedoch von ihm durch sein Bedürfnis nach Selbst-Ausdruck mit Hilfe künstlerischer Medien. Er meidet hochgradig strukturierte Probleme und Aufgaben, die grobmotorische Fertigkeiten erfordern. Menschen dieser Art haben ferner eine geringere Ichstärke, sind eher feminin und leiden häufiger unter emotionalen Störungen. Sie tendieren natürlich vor allem zu künstlerischen oder mit dem Kultur- und Kunstleben befaßten Berufen.

Eine Besonderheit des Hollandschen Ansatzes, der ihn für die Berufsberatung, aber auch für Maßnahmen beruflicher Klassifikation und Entwicklung besonders interessant macht, ist die Einschätzung von Berufen nach den gleichen Kategorien. Die theoretische Annahme ist, daß Menschen nach (beruflichen) Umwelten suchen, die es ihnen ermöglichen, ihre Fähigkeiten bestmöglich anzuwenden und ihre Werthaltungen zu verwirklichen; dies ist ihnen am besten dort möglich, wo hohe *Kongruenz* zwischen den Orientierungen der Person und den Anforderungen der Umwelt vorliegt. Das Ausmaß an Kongruenz ergibt sich nicht allein durch die Übereinstimmung oder Nichtübereinstimmung der dominierenden Werthaltungen, sondern durch die Prüfung der Orientierungsmuster von Person und Umwelt auf Übereinstimmung. Mit dem in Abbildung 9 wiedergegebenen hexagonalen Modell soll dies möglich sein. Die räumliche Nähe im Modell symbolisiert die Ähnlichkeit der Orientierungen.

Bergmann und Eder (1992, S. 11) erläutern, daß es nach dem Modell von Holland vier Abstufungen von Kongruenzen gibt:

„Wenn z. B. eine praktisch-technisch orientierte Person (R) einen praktisch-technischen Beruf (R) ergreift, dann liegt maximale Person-Umwelt-Kongruenz vor; ergreift dieselbe Person (R) einen intellektuell-forschenden (I) oder konventionellen (C) Beruf, dann ergibt sich eine mittlere und bei einem künstlerisch-sprachlichen (A) oder unternehmerischen (E) Beruf niedrige Kongruenz. Ergreift ein R-Typ einen sozialen Beruf (S), so handelt es sich um eine inkongruente Wahl.“

Außer der Kongruenz wird zur Bestimmung, wie gut Person und Umwelt zusammenpassen, auch das Niveau der *Differenziertheit* bestimmt, das durch den Grad der Eindeutigkeit des Person- oder Umweltprofils definiert ist.

Eine Besonderheit des Hollandschen Ansatzes besteht in der Analyse der Kongruenz zwischen den Orientierungen der Person und den Anforderungen der Umwelt.

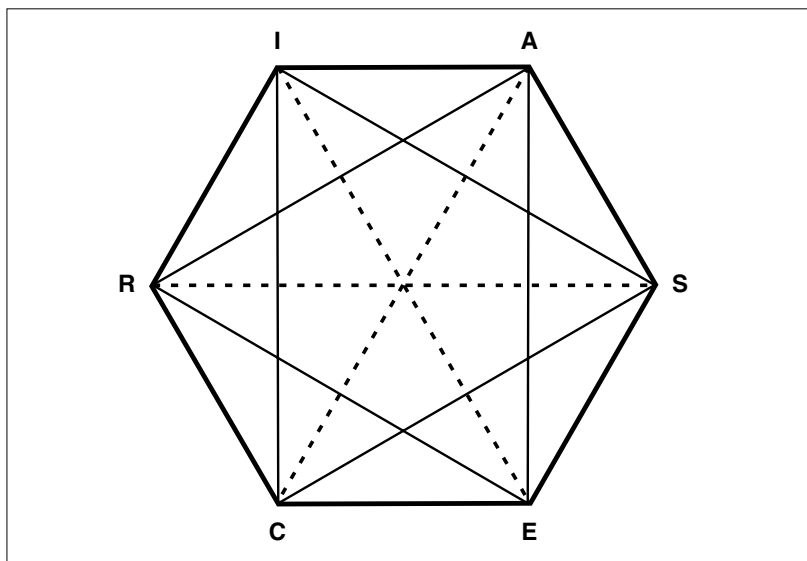


Abbildung 9: Hexagonales Modell zur Bestimmung der Ähnlichkeit zwischen Persönlichkeitstypen, Umwelttypen und deren Beziehungen (aus Bergmann & Eder, 1992, S. 11)

Während die Typologie der Berufsorientierungen nach Holland eine ergiebige Heuristik für die Eignungsdiagnostik zu sein scheint, ist bis heute ungewiß, welcher Nutzen ihr für die Entscheidungsfindung in der Praxis zukommt.

Während die Typologie der Berufsorientierungen nach Holland eine ergiebige Heuristik für die Eignungsdiagnostik zu sein scheint, ist bis heute ungewiß, welcher Nutzen ihr für die Entscheidungsfindung in der Praxis zukommt. Dies liegt zum einen, wie eingangs dieses Abschnitts erwähnt, an der seltenen Verwendung von Interessenkonzepten in der Berufseignungsdiagnostik, zum anderen aber auch an der mangelnden Eindeutigkeit der Auswertung des Umwelt-Struktur-Tests, vor allem hinsichtlich der Bestimmung von „Kongruenz“ und „Differenziertheit“. Auch wäre noch zu belegen, ob das hexagonale Modell weitergehende und verlässlichere Information liefert als ein einfacher ipsativer Rangreihenvergleich zwischen Person und Umwelt, wie es etwa das *Verhaltensrangprofil* zur Leistungsbeurteilung als Konkordanzmaß vorsieht (vgl. Kapitel 15). Zu den Versuchen, das Hollandsche Interessenmodell für die Berufseignungsdiagnostik nutzbar zu machen, gehört die beratungsorientierte Anwendung an prospektiven Unternehmensgründern (Schuler & Rolfs, 2000).

3 Problemfelder konstruktorientierter Verfahren

Im folgenden Abschnitt sollen zunächst einige allgemeine Problemsituationen diskutiert werden, die auftreten, wenn mehrere (eigenschaftsorientierte) Verfahren in Kombination zur Kriteriumsvorhersage eingesetzt werden. Danach wird auf die Problematik von Verfälschungstendenzen speziell bei persönlichkeitsorientierten Verfahren eingegangen. Den Abschluß bildet eine eher „technisch“ orientierte Darstellung zum Computereinsatz in der konstruktorientierten Eignungsdiagnostik.

3.1 Kombiniertes Einsatz von Verfahren

In den meisten eignungsdiagnostischen Anwendungsfällen werden mehrere Verfahren gleichzeitig eingesetzt, die ganz unterschiedliche Konstrukte erfassen sollen. Es wird bei dieser Vorgehensweise angenommen, daß die Ver-

fahrenskombination eine bessere Vorhersage liefert als die isolierten Einzelkomponenten. Insofern ist die bisherige Berichterstattung zu den Validitätseinzelbefunden der unterschiedlichen Verfahren irreführend, da hier nur Hinweise zu *Einzelzusammenhängen* mit Berufserfolgskriterien gegeben werden. Beim kombinierten Verfahrenseinsatz gelten aber die statistischen Grundprinzipien der Multiplen Regression: Vorher in der Einzelprüfung als hoch kriteriumsvalide nachgewiesene Verfahren können sich plötzlich als redundant herausstellen, da sie nur indirekt über einen zweiten Prädiktor mit dem Kriterium zusammenhängen.

Üblicherweise können solche Redundanzeffekte (aber auch Suppressions-effekte) erst bei einer Evaluation (vgl. Kapitel 22) einer bestehenden Verfahrenskombination erfaßt werden, wenn für die ersten Absolventen der Testreihe sinnvolle Kriteriumswerte vorliegen. Hilfestellung für eine rationale Konstruktion gibt aber auch die Übersicht von Schmidt und Hunter (1998), wiedergegeben in Kasten 4 (Abschnitt 2.2.1.3). Ihr Plädoyer ist zunächst, ein Verfahren zur Erfassung allgemeiner kognitiver Fähigkeiten einzusetzen (vgl. zur Diskussion dieses Ansatzes Abschnitt 2.2.1.4), da für diese Verfahrensklasse normalerweise die höchste kriteriumsbezogene Validität zu erwarten ist. Weitere Verfahren sind zum einen nach den Validitätseinzelbefunden, zum anderen aber auch nach dem durch sie erzielten Validitätszuwachs (inkrementelle Validität) zu beurteilen.

So weisen beispielsweise „Integrity“-Tests eine hohe kriteriumsbezogene Validität von $r=.41$ auf, die weitgehend unabhängig von Intelligenz zu sein scheint: In Kombination mit einem allgemeinen kognitiven Fähigkeitstest eingesetzt, erzielen sie immerhin einen Validitätszuwachs von 27%. Gleichzeitig wird anhand dieses Beispiels auch das Grundproblem einer solchen theoriefreien Verfahrenselektion (Auswahl nach Inkrements gegenüber einem Intelligenzverfahren) offensichtlich: „Integrity“-Tests sind eine relativ heterogene Verfahrensgruppe, die ursprünglich zur Vorhersage von kontraproduktivem Verhalten (Diebstahl, Sabotage, Absentismus usw.) konstruiert wurde. Eher „unbeabsichtigt“ ergibt sich auch die dokumentierte Vorhersagequalität für allgemeine berufliche Leistungskriterien (Ones, Schmidt & Viswesvaran, 1993). Bis heute ist aber noch nicht völlig geklärt, was genau durch diese Verfahren erfaßt wird und wie sie konzeptionell in das Ensemble konstruktorientierter Verfahren einzuordnen sind (vgl. für einen Überblick Marcus, 2000).

Bei dieser Art der Konstruktions- bzw. Anwendungsstrategie mit dem alleinigen Ziel der Vorhersageoptimierung würde also wieder die Erklärung vernachlässigt (vgl. analog Abschnitt 2.2.1.4), die durch einen Rückgriff auf die Ergebnisse einer Arbeits- und Anforderungsanalyse (vgl. Kapitel 3) bei der Verfahrensauswahl wesentlich eher gegeben wäre.

3.2 Das „bandwidth-fidelity“-Dilemma

Der Einsatz von eignungsdiagnostischen Verfahren unterliegt in der Praxis immer einem Balanceakt: Zum einen kann eine Vielzahl von Verfahren eingesetzt werden, um einen möglichst großen Überblick über das Fähigkeits- und Eigenschaftsprofil eines Bewerbers zu erhalten. Dieser Überblick geht allerdings auf Kosten der Detailtiefe, da spezifische Einzelinformationen gewungenermaßen nicht erfaßt werden können. Wenn andererseits bei einer eignungsdiagnostischen Untersuchung der Schwerpunkt auf einer Detailerfassung liegt, geht damit zwangsläufig der große Überblick zum allgemeinen Fähigkeits- und Eigenschaftsprofil verloren.

Cronbach (z. B. in Cronbach, 1990, p. 208) hat diese Abwägungsproblematik unter dem Stichwort „bandwidth-fidelity-dilemma“ (die deutsche Direktübersetzung „Bandbreiten-Vertrauens-Problem“ konnte sich bisher noch nicht durchsetzen) publik gemacht. Mit „bandwidth“ wird in diesem Zusammenhang die Bandbreite, Komplexität oder Menge der Informationen bezeichnet, die ein bestimmtes eignungsdiagnostisches Verfahren liefert. Mit „fidelity“ wird die Vertrauenswürdigkeit bzw. Akkuratheit dieser Informa-

Informationen zu Einzelvaliditäten sagen wenig über die mögliche Validitätsverbesserung durch Verfahrenskombinationen aus.

Das „bandwidth-fidelity“-Dilemma beschreibt den Konflikt zwischen möglichst breiter Informationsausbeute und Detailtiefe der erhobenen Informationen.

Hilfestellung durch die Symmetriehypothese: Das Abstraktionsniveau von Prädiktor und Kriterium muß vergleichbar sein.

Besonders bei respondenten eigenschaftsorientierten Verfahren sowie bei den biographiebezogenen Methoden, Fragebogen und Interviews wird angenommen, ihre auf typisches Verhalten fokussierende Meßtechnik würde zu Verzerrungen der wahren Werte führen.

tionen beschrieben; der Begriff stellt damit eine eher umgangssprachliche Vermengung der testtheoretischen Konzepte „Validität“ und „Reliabilität“ dar. Prinzipiell bezeichnen „bandwidth“ und „fidelity“ zwei unabhängig voneinander variierbare Dimensionen. In der Realität werden (wie bereits angedeutet) zwei entgegengesetzte Paare allerdings gehäuft auftreten: Die Kombination „niedrige bandwidth, hohe fidelity“ stellt gewissermaßen das in der Psychometrie angestrebte Ideal dar, während „hohe bandwidth, niedrige fidelity“ z. B. bei eignungsdiagnostischen Screenings anzutreffen ist.

Eine Spielart des Abwägungskonflikts wurde bereits bei der Erfassung von allgemeinen vs. spezifischen kognitiven Fähigkeiten diskutiert (Abschnitt 2.2.1.4). In neuerer Zeit ist die Diskussion durch die Renaissance des am Fünf-Faktoren-Modell der Persönlichkeit orientierten eignungsdiagnostischen Ansatzes (Abschnitt 2.2.3.1) neu entflammt. Einige Organisationspsychologen (z. B. Hough & Schneider, 1996) vertreten die Ansicht, daß die Big5-Faktoren zu breit definiert seien und konzeptionell sinnvollerweise zu trennende Subdimensionen konfundieren (i. S. des Dilemmas liegt also eine zu große „bandwidth“ vor). Besser sei der Rückgriff auf enger umgrenzte Eigenschaften (wie z. B. Kontrollüberzeugung, vgl. Abschnitt 2.2.4), die eine höhere kriteriumsbezogene Validität erwarten ließen.

Ausgangspunkt der „bandwidth-fidelity-Überlegungen“ (z. B. Hogan & Roberts, 1996; Ones & Viswesvaran, 1996; Schneider, Hough & Dunnette, 1996) ist immer die Beziehung zwischen dem eignungsdiagnostischen Prädiktor (hier z. B. eine Persönlichkeitsvariable) und dem vorherzusagenden Berufserfolgskriterium. Unbestritten ist hier die Symmetriehypothese von Cronbach und Gleser (1965): Wenn ein eng umgrenztes Kriterium vorhergesagt werden soll, so muß auch ein spezifischer Prädiktor eingesetzt werden. Ist hingegen ein allgemeines Globalmaß das Kriterium, so muß auch der Prädiktor ein vergleichbares Abstraktionsniveau aufweisen (vgl. Wittmann & Matt, 1986).

Im eignungsdiagnostischen Vorhersagemodell (vgl. Kapitel 22) können ganz unterschiedliche Berufserfolgsmaße herangezogen werden: In der Praxis wird meist nur ein globales Leistungsmaß erhoben. Die neuere Forschung zur beruflichen Leistung zeigt allerdings, daß beim Leistungskonzept verschiedene, nur mäßig interkorrelierte Facetten unterschieden werden können (vgl. Kapitel 15). Wenn also ein spezifisches Leistungskriterium vorhergesagt werden soll (z. B. „Krankenstand im Kalenderjahr“ als Negativindikator von Leistung), so sollte hierfür auch ein spezifischer Prädiktor eingesetzt werden (z. B. „Selbstdisziplin“ als Subfacette der Big5-Dimension „Gewissenhaftigkeit“).

Das „bandwidth-fidelity-Dilemma“ wird von dieser Warte aus also aufgelöst durch eine systematisierte Vorgehensweise: Erst nach einer Konkretisierung des gewünschten Zielkriteriums erfolgt eine entsprechend parallelisierte Prädiktorauswahl.

3.3 Verfälschungstendenzen

Bei der Anwendung eignungsdiagnostischer Verfahren wird häufig befürchtet, die Ergebnisse könnten aufgrund der Wirksamkeit bestimmter Tendenzen zur Selbstdarstellung auf systematische Weise verzerrt sein. Dies wird besonders für Personalauswahlsituationen – weniger für Beratungssituationen – angenommen, wo man beim Bewerber die Neigung erwartet, sich so zu verhalten, daß es seinem Ziel, ein Stellenangebot zu erhalten, entgegenkommt. Bei Leistungstests stellt dies insofern kein Problem dar, als sie darauf ausgelegt sind, *maximales Verhalten* zu erfassen, also die „Bestleistung“ jeder teilnehmenden Person; allenfalls Leistungsunterschätzungen, die durch mangelnde Motivation zustande kommen, könnten dort zu Fehlschlüssen führen. Auch bei simulationsorientierten Verfahren wie Gruppendiskussionen oder Rollenspielen wird im wesentlichen ein Maximalleistungscharakter angenommen. Besonders bei respondenten eigenschaftsorientierten Verfahren sowie bei den biographiebezogenen Methoden Fragebogen und Inter-

view wird dagegen angenommen, ihre auf *typisches Verhalten* fokussierende Meßtechnik würde zu Verzerrungen der wahren Werte führen.

Diese Annahme wurde in mehreren Untersuchungen geprüft. In manchen Studien erfolgte diese Prüfung auf dem Weg, daß man Versuchspersonen in einer Forschungssituation aufforderte, sich so positiv wie möglich darzustellen. Für viele Persönlichkeitsskalen sowie für biographische Fragebögen erhielt man dabei Werte, die erheblich von den Normwerten abwichen. Hieraus den Schluß zu ziehen, diese Instrumente würden in Auswahl-situationen verfälschte Werte liefern, wäre allerdings übereilt, denn nicht zwangsläufig muß dort die Tendenz zu positiver Selbstdarstellung andere Verhaltensorientierungen überwiegen. Beispielsweise könnte die Erwartung, man müsse in der späteren Tätigkeit in irgendeiner Form den Nachweis für die jetzt gemachten Angaben erbringen, von einer allzu forcierten Verfälschung abhalten. Tatsächlich ergeben sich für Auswahl-situationen durchschnittlich „bessere“ Werte als für Forschungssituationen, die aber von diesen bei weitem nicht so stark abweichen wie diejenigen, deren Grundlage die anweisungsgemäße Beschönigung war (Ryan & Sackett, 1987).

Wichtiger noch als die Frage, ob in Auswahl-situationen beschönigende Selbstdarstellung erfolgt, ist das Problem, ob dadurch validitätsschädliche Effekte entstehen. Ones und Viswesvaran (1998) führten hierzu eine Metaanalyse durch. An einer sehr großen Gesamtstichprobe (N zwischen 39 000 und 143 000 Personen) konnten sie zunächst zeigen, daß Skalen zur Messung von *Social Desirability* substantiell mit drei der fünf großen Persönlichkeitsmerkmale korreliert sind, nämlich mit Emotionaler Stabilität als Gegenpol von Neurotizismus ($r = .37$), Gewissenhaftigkeit ($r = .20$) und Verträglichkeit ($r = .14$). Zwei der drei Beziehungen zeigten sich in geringerer Stärke auch dann, wenn den Werten nicht Selbsteurteilungen – wie bei Fragebogentests üblich –, sondern Fremdbeurteilungen zugrunde lagen. Dies legt die Annahme nahe, daß die Neigung zu positiver Selbstdarstellung zunächst nicht ausschließlich zur Verfälschung anderer Persönlichkeitsmaße führt, sondern teilweise zur Substanz dieser Merkmale gehört. An einer geringeren, aber immer noch aussagekräftigen Stichprobengröße berechneten Ones & Viswesvaran (1998) die Beziehung zu Kriterien des beruflichen und Ausbildungserfolgs. Dabei ergab sich einzig zu Trainingserfolg eine positive Korrelation, nicht aber zu Ausbildungs- und Berufserfolg. Dementsprechend hatte die Auspartialisierung von *Social Desirability* aus der Korrelation zwischen Persönlichkeitswerten und Berufserfolg keinerlei Effekt, d.h. die Validität von Persönlichkeitsmaßen wird durch die Selbstdarstellung *nicht* beeinträchtigt.

Beide Ergebnisse zusammen unterstützen die Annahme, daß die Neigung zu vorteilhafter Selbstdarstellung nicht primär als Quelle der Verfälschung in Auswahl-situationen betrachtet werden sollte, sondern als Orientierung an den Erwartungen anderer und damit auch als Komponente sozialer Anpassung. Gerade in Kontaktberufen ist die erfolgreiche Gestaltung von Beziehungen immer auch mit der Art der Selbstpräsentation, also dem Bemühen verbunden, „einen guten Eindruck“ auf andere zu machen. Diese Auffassung lag einer Studie von Diemand und Schuler (1991) zugrunde, in der mehrere Selbstdarstellungsskalen mit dem Abschneiden in einem Assessment Center verglichen wurden. Die Ergebnisse bestätigen die Unterscheidung von „assertiver“ und „defensiver“ Selbstdarstellung, wie sie von Lennox und Wolfe (1984) für das Konstrukt *Self-monitoring* getroffen wurde. Während assertive Selbstdarstellung für diejenigen Teilnehmer charakteristisch war, die im Auswahlverfahren insgesamt positiv abschnitten, traf das Gegenteil auf Personen mit hoher Ausprägung der defensiven Komponente zu. Nicht mit dieser Differenzierung, wohl aber mit der positiven Beziehung des *Self-monitoring*-Gesamtwerts zu Beurteilungen durch Vorgesetzte beschäftigte sich eine Nachfolgestudie im Rahmen der prädiktiven Validierung einer Potentialanalyse (Diemand & Schuler, 1998). Auch in diesem Kontext war die Validität der Prädiktoren nicht durch Selbstdarstellung beeinträchtigt. Angesichts dieser Ergebnisse ist durchaus fraglich, ob man die psychometrischen Möglichkeiten nutzen sollte, die Neigung zu positiver Selbstdarstellung bei der Konstruktion berufsseignungsdiagnostischer Instrumente dadurch zu eliminieren,

Tatsächlich ergeben sich für Auswahl-situationen durchschnittlich „bessere“ Werte als für Forschungssituationen, die aber von diesen bei weitem nicht so stark abweichen wie diejenigen, deren Grundlage die anweisungsgemäße Beschönigung war.

Die Auspartialisierung von Social Desirability aus der Korrelation zwischen Persönlichkeitswerten und Berufserfolg hat keinerlei Effekt, d.h. die Validität von Persönlichkeitsmaßen wird durch die Selbstdarstellung nicht beeinträchtigt.

Positive Selbstdarstellung sollte nicht primär als Quelle der Verfälschung in Auswahl-situationen betrachtet werden, sondern als Orientierung an den Erwartungen anderer und damit als Komponente sozialer Anpassung.

daß Items, die hoch mit Darstellungswerten korreliert sind, aus dem weiteren Konstruktionsprozeß des eignungsdiagnostischen Verfahrens ausgeschlossen werden. Von Nutzen ist allerdings in jedem Fall, diese Komponente mitzunormieren, d.h. die Testnormen sollten in realen Auswahl-situationen gewonnen werden.

Das Problem der Verfälschung von Testwerten durch vorteilhafte Selbstdarstellung stellt sich also, bezogen auf Durchschnittswerte, als geringer dar als oft befürchtet. Dies löst freilich nicht vollständig das Problem der interindividuell unterschiedlichen Beschönigungstendenz, die sehr wohl zu unangemessener Bevorzugung einzelner Kandidaten führen kann (Rosse, Stecher, Miller & Levin, 1998). Die früher gelegentlich praktizierte Korrektur mittels „Lügenskalen“ stellt jedenfalls keine adäquate Lösung des Problems dar. Eine bessere, allerdings aufwendigere Möglichkeit ist die Überprüfung der Testwerte durch weitere, insbesondere interaktive Verfahren (vgl. Kapitel 6) im Rahmen eines multimethodalen Vorgehens. Stark verfälschte Testwerte sollten durch einen deutlichen Leistungsabfall zwischen den schriftlichen und den interaktiven Verfahren in Erscheinung treten.

3.4 Papier-Bleistift- vs. computerbasierte Testung

Der Einsatz von Computern ist praktisch in allen Phasen einer Testentwicklung, -durchführung, -auswertung, bei der Urteilsbildung und auch bei der Testevaluation möglich. Als häufiger Vorteil wird z. B. der hohe Standardisierungsgrad der Durchführung und Auswertung bei Einsatz eines Computerverfahrens angeführt. So werden z. B. alle Teilnehmer identisch instruiert, die Zeitvorgabe für Items und Verfahrensteile kann genau kontrolliert werden usw. Zudem wird der gesamte Erhebungsprozeß rationalisiert: Da eine fehleranfällige separate Dateneingabe entfällt, sind die Ergebnisse wesentlich schneller verfügbar. Kubinger (1993) diskutiert anhand diverser Testgütekriterien die grundsätzlichen Vor- und Nachteile von Computertests gegenüber Papier-Bleistift-Verfahren.

Die z.Zt. praktizierte Computerdiagnostik kann grob in drei Gruppen eingeteilt werden.

- *Computerversionen von Papier-Bleistift-Testverfahren*

Zu den meisten bekannten Papier-Bleistift-Verfahren liegen EDV-Programme vor, die eine äquivalente Testdurchführung am Bildschirm gewährleisten sollen. Weitergehende Vorteile der Computerdurchführung gegenüber der traditionellen Testvorgabe werden dabei nicht genutzt. Ziel ist eine möglichst äquivalente Erhebung.

- *Adaptive Testung*

Unter adaptivem Testen versteht man Strategien der Testvorgabe, bei denen sich die Auswahl der Testitems am Leistungsniveau der betreffenden Person orientiert (Kisser, 1992, S. 164). Da auf der Grundlage der bisherigen Itemantworten sofort auf die Fähigkeitsausprägung des Probanden geschlossen werden muß und sich danach die Auswahl der folgenden Items richtet, ist die praktische Umsetzung einer solchen Testung eigentlich nur über Computer denkbar.

In der Eignungsdiagnostik wird adaptives Testen derzeit noch kaum eingesetzt. Ein Grund hierfür ist sicherlich der große Konstruktionsaufwand dieser Verfahren. Zudem greift diese Verfahrensklasse auf neuere testtheoretische Ansätze zurück (Teilbereiche der Item-Response-Theorie, vgl. z. B. Rost, 1996), die sich stark von der klassischen Testtheorie unterscheiden.

- *Komplexe Simulationen*

Hierunter fallen Computerszenarien, in denen komplexe Systeme simuliert werden und in die der Proband regelhaft eingreifen soll. Diese Verfahren sind der simulationsorientierten Eignungsdiagnostik zuzuordnen. Sie werden in Kapitel 6 ausführlich behandelt.

Computerbasierte Testung kann als Ersatz für eine Papier-Bleistift-Erhebung oder als Erweiterung traditioneller Erhebungsformen realisiert werden.

Da die erste Gruppe (Computerumsetzung von traditionellen Papier-Bleistift-Verfahren) bis dato den mit Abstand größten Anwendungsbereich darstellt, soll auf die Äquivalenzproblematik der traditionellen und der neueren Verfahrensform ausführlicher eingegangen werden.

Überprüfung der Äquivalenz von Papier-Bleistift- und Computertestformen
Als psychometrischer Äquivalenzkennwert wird üblicherweise (Mead & Drasgow, 1993) die Korrelation zwischen den beiden parallel durchgeführten Testformen herangezogen (hohe Korrelation = gleiche Rangabfolge der Probanden = äquivalente Erhebungen). Zusätzlich werden die Mittelwerte der Testformen verglichen (gleiche Mittelwerte = gleiche Schwierigkeit = äquivalente Erhebungen). Miles und King (1998) weisen darauf hin, daß diese Analysen für eine Überprüfung der Konstruktäquivalenz nicht ausreichen. Sie plädieren dafür, zumindest auch einen Faktorstrukturvergleich und eine Analyse des Faktorladungsmusters z.B. mit Hilfe einer konfirmatorischen Faktorenanalyse durchzuführen. Nur durch eine möglichst große Informationssammlung können potentielle Ursachen von Ungleichheiten gefunden werden. Klinck (1998, S. 62) nennt in diesem Zusammenhang drei Bereiche, die Einfluß auf Veränderungen bei der Verfahrensadaptation haben können:

- *Merkmale der Testsituation:* Statt in einer Gruppe mit einem Versuchsleiter kann der Test in Einzelsitzungen ohne menschliche Anleitung durchgeführt werden.
- *Merkmale des Mediums:* Statt Papier und Bleistift stehen Maus, Tastatur und Bildschirm zur Verfügung.
- *Merkmale der realisierten Testumsetzung:* Hierunter fallen die Besonderheiten der gewählten Computerversion. Werden z.B. die Items einzeln oder in Gruppen dargeboten, kann der Proband Items auslassen und später bearbeiten, sind graphische Darstellungen deckungsgleich mit ihren Papiervorlagen?

Die bisherigen Ergebnisse deuten darauf hin, daß die Äquivalenzproblematik besonders bei kognitiven Fähigkeitstests auftritt, die unter Zeitdruck bearbeitet werden müssen (speed-Bedingung). So fanden Mead und Drasgow (1993) in ihrer Metaanalyse zur Computerumsetzung von kognitiven Verfahren bei dieser Verfahrensklasse nur eine korrigierte Verfahrenskorrelation von .72, während sich bei Verfahren ohne Zeitdruck (power-Bedingung) ein Übereinstimmungskoeffizient von .97 ergab. Mead und Drasgow vermuten, daß die unterschiedlichen motorischen Anforderungen der Verfahrensformen bei den speed-Verfahren besonders stark ins Gewicht fallen und für die Unterschiede verantwortlich sind. Einen interessanten Hinweis gibt hier die Arbeit von Overton, Rodgers Taylor, Zickar und Harms (1996). Sie setzten neben der traditionellen Papier-Bleistift- und einer Computeradaptation mit Tastatureingabe eine dritte Verfahrensform ein, bei der die Probanden die ihres Erachtens richtigen Antworten mit Hilfe eines Lichtgriffels direkt am Bildschirm markierten. Sie zeigte erwartungsgemäß eine größere Ähnlichkeit mit der Papier-Bleistift-Form als die Tastatur-Version. Wesentlich dünner ist die Befundlage bei den nicht-kognitiven Verfahren. In einer Studie mit vier solcher Verfahren konnten zumindest Miles und King (1998) keine Unterschiede zwischen Papier-Bleistift- und Computerversion finden.

Generell ist zu fragen, was größeren Einfluß auf die Äquivalenz von Verfahrensformen hat: der generelle Methodenunterschied zwischen den grundverschiedenen Medien oder die Realisierungsbesonderheiten der gewählten EDV-Version. So kann Klinck (1998) die von ihr gefundenen Unterschiede bei der Adaptation von Teilaufgaben des Leistungsprüfsystems von Horn (1983) recht plausibel mit Abweichungen in der Aufgabenvorgabe erklären.

Qualitativ unterschiedliche Adaptationen können auch mit ein Grund für die heterogene Befundlage zur Akzeptanz, Tendenz zur sozialen Erwünschtheit und Anfälligkeit für „impression management“ sein (vgl. Booth-Kewley, Edwards & Rosenfeld, 1992; Richman, Kiesler, Weisband & Drasgow, 1999).

Die einfache Computerumsetzung von traditionellen Papier-Bleistift-Verfahren stellt bis dato den größten computergestützten Anwendungsbereich dar.

Die bisherigen Ergebnisse deuten darauf hin, daß die Äquivalenzproblematik besonders bei kognitiven Fähigkeitstests auftritt, die unter Zeitdruck bearbeitet werden müssen.

Generell ist zu fragen, was größeren Einfluß auf die Äquivalenz von Verfahrensformen hat: der generelle Methodenunterschied zwischen den grundverschiedenen Medien oder die Realisierungsbesonderheiten der gewählten EDV-Version.

Da gesicherte generelle Aussagen zur (In-)Äquivalenz zum jetzigen Zeitpunkt nicht getroffen werden können, muß die Äquivalenz der Computer- mit der entsprechenden Papier-Bleistift-Version für jedes Einzelverfahren nachgewiesen werden.

Zusammenfassung

Letztlich bleibt nur die Forderung, die Äquivalenz der Computer- mit der entsprechenden Papier-Bleistift-Version für jede spezifische Realisierung zu zeigen, da gesicherte generelle Aussagen zur (In-)Äquivalenz zum jetzigen Zeitpunkt nicht getroffen werden können.

Allerdings soll hier auch auf ein Grunddilemma der Computardiagnostik verwiesen werden (vgl. Miles & King, 1998; Klinck, 1998): Wenn die Vorteile der Computerisierung wirklich effektiv genutzt werden sollen (z. B. durch Implementierung von Multimediaelementen in das Verfahren), so muß sich die Computerversion zwangsläufig von den Vorgaben des Papier-Bleistift-„Vorbilds“ trennen. Statt die Parallelität zu prüfen, muß dann „nur noch“ gezeigt werden, daß beide Verfahren konstruktvalide sind.

Zusammenfassung

Wie in der Einführung zu diesem Kapitel dargestellt wurde, können die vorherrschenden Verfahren in der beruflichen Eignungsdiagnostik grob in drei Gruppen unterteilt werden, die ganz unterschiedlichen (Validierungs-)Ansätzen folgen. Entsprechend dieser Unterscheidung stellt das vorliegende Kapitel „Konstruktorientierte Verfahren der Personalauswahl“ den ersten Teil einer Kapiteltrilogie zur Personalauswahl dar. Die beiden folgenden Kapitel widmen sich dem simulations- (Kapitel 6) und dem biographieorientierten Ansatz (Kapitel 7).

Für die Darstellung von eignungsdiagnostisch relevanten Konstrukten und ihren Erhebungsverfahren wurde zunächst zwischen kognitiven und nicht-kognitiven Verfahren unterschieden. Im Bereich der kognitiven Fähigkeitstests dominieren sicherlich die allgemeinen Verfahren. Die Diskussion zu spezifischen Verfahren sowie die Darstellung von spezifischen Fähigkeitstests sollte aber gezeigt haben, daß unterschiedliche Zielsetzungen (letztlich Vorhersage vs. Erklärung) beim Verfahrenseinsatz unbedingt berücksichtigt werden müssen.

Auch bei den nicht-kognitiven Verfahren wurde zunächst zwischen globalen und spezifischen Persönlichkeitsvariablen getrennt. Anhand der Befunde zum Fünf-Faktoren-Modell der Persönlichkeit wurde diskutiert, ob die Technik der metaanalytischen Validitätsgeneralisierung mit ihrer Aggregation über unterschiedliche Verfahren, Berufe und Leistungskriterien in diesem Bereich bisher adäquat eingesetzt wurde. Anhand eines Beispiels wurde die Kriteriumsabhängigkeit zumindest einer Persönlichkeitsvariable (Extraversion) veranschaulicht.

Ergänzend wurden weitere spezifische Persönlichkeitsvariablen (Kontrollüberzeugung, Motivation und Interessen), allerdings ohne eignungsdiagnostische Validitätsinformationen, dargestellt.

Zum Abschluß wurden einige ausgewählte Problemfelder der (konstruktorientierten) Eignungsdiagnostik behandelt. Zum einen wurde noch einmal die Diskussion zum kombinierten Verfahrenseinsatz („Man nehme einen allgemeinen kognitiven Fähigkeitstest und zusätzlich ...“) aufgegriffen und mit dem „bandwidth-fidelity“-Dilemma der Kriteriumsvorhersage verknüpft. Zum anderen wurde die für nicht-kognitive Verfahren relevante Problematik von Verfälschungstendenzen untersucht. Schließlich wurde die konventionelle Papier-und-Bleistift-Form der Testdurchführung mit computergestützten Möglichkeiten der Testanwendung verglichen, deren Verbreitung künftig zunehmen wird.

Ein großer Vorteil, den der konstruktorientierte Ansatz gegenüber den noch folgenden Ansätzen aufweist, ist der Rückgriff auf „wohldefinierte“ Konstrukte aus der Grundlagenforschung der Differentiellen und Persönlichkeitspsychologie. Dadurch kann sich dieser Ansatz zumeist auf ein bereits bestehendes, breites nomologisches Netzwerk stützen. Eignungsdiagnostische Befunde können direkt mit Ergebnissen aus anderen Anwendungsbereichen verglichen werden.

Ein Nachteil ist sicherlich der teilweise nicht immer eindeutige Bezug der erfaßten Konstrukte zu Kriterien des Berufserfolgs. Über eine kriteriumsbezogene Validierung kann zumindest die Prognosequalität der Konstrukterhebung nachgewiesen werden. Eine tiefergehende Erklärung kann nur über eine stringente Ableitung der Konstrukte aus den Ergebnissen einer Arbeits- und Anforderungsanalyse sowie einer Detailanalyse der Konstrukt-Kriteriumsbeziehung erfolgen.

Weiterführende Literatur

- Cook, M. (1993). *Personnel selection and productivity*. Chichester, England: John Wiley & Sons.
- Schmidt, F.L. & Hunter, J.E. (1998). Meßbare Personmerkmale: Stabilität, Variabilität und Validität zur Vorhersage zukünftiger Berufsleistung und berufsbezogenen Lernens. In M. Kleinmann & B. Strauss (Hrsg.), *Potentialfeststellung und Personalentwicklung* (S. 16-43). Göttingen: Hogrefe.
- Schuler, H. (1996). *Psychologische Personalauswahl. Einführung in die berufliche Eignungsdiagnostik*. Göttingen: Hogrefe.

Literatur

- Ackerman, P.L. & Humphreys, L.G. (1991). Individual differences theory in industrial and organizational psychology. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 223-282). Palo Alto, CA: Consulting Psychologists Press.
- Anastasi, A. (1985). The use of personality assessment in industry: Methodological and interpretative problems. In H.J. Bernardin & D.A. Bownas (Eds.), *Personality assessment in organizations* (pp. 1-20). New York: Praeger Publishers.
- Arvey, R.D. (1979). *Fairness in selecting employees*. Reading, MA: Addison-Wesley.
- Asendorpf, J.B. (1996). *Psychologie der Persönlichkeit*. Berlin: Springer.
- Bangert-Drowns, R.L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin*, 99, 388-399.
- Barrett, G.V. & Depinet, R.L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist*, 46, 1012-1024.
- Barrick, M.R. & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Bergmann, Ch. & Eder, F. (1992). *Allgemeiner Interessen-Struktur-Test/Umwelt-Struktur-Test (AIST/UST)*. Weinheim: Beltz.
- Booth-Kewley, S., Edward, J.F. & Rosenfeld, P. (1992). Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? *Journal of Applied Psychology*, 77, 562-566.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI)*. Göttingen: Hogrefe.
- Borman, W.C., Hanson, M.A., Oppler, S.H., Pulakos, E.D. & White, L.A. (1993). Role of early supervisory experience in supervisor performance. *Journal of Applied Psychology*, 78, 443-449.
- Brandstätter, H. (1979). Die Ermittlung personaler Eigenschaften kognitiver Art. In G. Reber (Hrsg.), *Personalinformationssysteme* (S. 74-95). Stuttgart: Poeschel.
- Brickenkamp, R. (1990). *Generelle Interessen-Skala (GIS)*. Göttingen: Hogrefe.
- Brickenkamp, R. (1994). *Test d2: Aufmerksamkeits-Belastungs-Test*. Göttingen: Hogrefe.
- Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. Cambridge, UK: Cambridge University Press.

Weiterführende Literatur

Literatur

Fortsetzung Literatur

- Conrad, W., Baumann, E. & Mohr, V. (1980). *Mannheimer Test zur Erfassung des physikalisch-technischen Problemlösens (MTP)*. Göttingen: Hogrefe.
- Cook, M. (1993). *Personnel selection and productivity*. Chichester, England: John Wiley & Sons.
- Costa, P.T. & McCrae, R.R. (1992). *NEO-PI-R*. Odessa, FL: PAR – Psychological Assessment Resources.
- Cronbach, L.J. (1990). *Essentials of psychological testing*. Row: Harper.
- Cronbach, L.J. & Gleser, G.C. (1965). *Psychological tests and personnel decisions*. University of Illinois Press: Urbana.
- Diemand, A. & Schuler, H. (1991). Sozial erwünschtes Verhalten in eignungsdiagnostischen Situationen. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 135-145). Göttingen: Hogrefe.
- Diemand, A. & Schuler, H. (1998). Wirksamkeit von Selbstdarstellungsvariablen im Rahmen der prognostischen Validierung eines Potentialanalyseverfahrens. *Zeitschrift für Arbeits- und Organisationspsychologie*, 42, 134-146.
- Digman, J.M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.
- Dorcus, R.M. & Jones, M.H. (1950). *Handbook of employee selection*. New York: McGraw Hill.
- Dorsch, F., Häcker, H. & Stapf, K.H. (Hrsg.). (1994). *Psychologisches Wörterbuch*. Göttingen: Hogrefe.
- Dye, D.A., Reck, M. & McDaniel, M.A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment*, 1, 153-157.
- Engelbrecht, W. (1994). Computergestützte berufsbezogene Testauswertung im Dienst der Berufsberatung. *Zeitschrift für Arbeits- und Organisationspsychologie*, 38, 175-181.
- Eysenck, H.J. (1967). *The biological basis of personality*. Springfield, IL: Charles C. Thomas.
- 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.
- Ghiselli, E.E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, 26, 461-477.
- Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48, 26-34.
- Gray, J.A. (1973). Causal theories of personality and how to test them. In J.R. Royce (Ed.), *Multivariate analysis and psychological theory* (pp. 409-464). New York: Academic Press.
- Guion, R.M. & Gottier, R.F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18, 135-164.
- Heckhausen, H. (1963). *Hoffnung und Furcht in der Leistungsmotivation*. Meisenheim/Glan: Hain.
- Hermans, H., Petermann, F. & Zielinski, W. (1978). *Leistungsmotivationstest (LMT)*. Amsterdam: Swets & Zeitlinger.
- Herrnstein, R.J. & Murray, C.A. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Hogan, J.C. & Roberts, B.W. (1996). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior*, 17, 627-637.
- Holland, J.L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Odessa, FL: Psychological Assessment Resources.
- Horn, W. (1983). *Leistungsprüfsystem (LPS)*. Göttingen: Hogrefe.
- Hossiep, R., Paschen, M. & Mühlhaus, O. (2000). *Persönlichkeitstests im Personalmanagement*. Göttingen: Hogrefe.
- 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, 95-108.
- Hough, L.M. & Schneider, R.J. (1996). Personality traits, taxonomies, and applications in organizations. In K.R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 3-30). San Francisco, CA: Jossey-Bass.

- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hunter, J.E. & Schmidt, F.L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.
- Irle, M. & Allehoff, W. (1988). *Berufs-Interessen-Test II (BIT II)* (2. Aufl.). Göttingen: Hogrefe.
- Jaccard, J., Turrisi, R. & Wan, C.K. (1993). *Interaction effects in multiple regression*. Newbury Park, CA: Sage.
- Jäger, A.O. (1969). *Differentieller Wissenstest (DWT)*. Göttingen: Hogrefe.
- Jäger, A.O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35, 21-35.
- Jäger, A.O., Süß, H.-M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test (BIS, Form 4)*. Göttingen: Hogrefe.
- Jensen, A.R. (1969). *Genetics and education*. London: Methuen.
- Kanfer, R., Ackerman, P.L., Murtha, T.C. & Goff, M. (1995). Personality and intelligence in industrial and organizational psychology. In D.H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 577-602). New York: Plenum Press.
- Kisser, R. (1992). Adaptive Strategien. In R.S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik* (S. 161-170). Weinheim: Beltz/PVU.
- Klinck, D. (1998). Papier-Bleistift- versus computerunterstützte Administration kognitiver Fähigkeitstests: Eine Studie zur Äquivalenzfrage. *Diagnostica*, 44, 61-70.
- Krampen, G. (1991). *Fragebogen zu Kompetenz- und Kontrollüberzeugungen (FKK)*. Göttingen: Hogrefe.
- Kubinger, K.D. (1993). Testtheoretische Probleme der Computerdiagnostik. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 130-137.
- Latham, G.P., Saari, L.M., Pursell, E.D. & Campion, M.A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.
- Leitner, W.G. (1998). *Konzentrationsleistung und Aufmerksamkeitsverhalten*. WVB: Bamberg.
- Lennox, R.D. & Wolfe, R.D. (1984). Revision of the Self-Monitoring Scale. *Journal of Personality and Social Psychology*, 46, 1349-1364.
- Lienert, G.A. & Schuler, H. (1994). *Revidierter Allgemeiner Büroarbeitstest (ABAT-R)*. Göttingen: Hogrefe.
- Marcus, B. (2000). *Kontraproduktives Verhalten im Betrieb*. Göttingen: Hogrefe.
- McClelland, D.C. (1973). Testing for competence rather than for „intelligence“. *American Psychologist*, 28, 1-14.
- Mead, A.D. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449-458.
- Miles, E.W. & King, W.C. (1998). Gender and administration mode effects when pencil-and-paper personality tests are computerized. *Educational and Psychological Measurement*, 58, 68-76.
- Mischel, W. (1968). *Personality and assessment*. New York: John Wiley & Sons.
- Moosbrugger, H. & Heyden, M. (1998). *Frankfurter Adaptiver Konzentrationsleistungs-Test (FAKT)*. Göttingen: Hogrefe.
- Motowidlo, S.J. & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology*, 66, 337-344.
- Murphy, K.R. (1996). Individual differences and behavior in organizations: Much more than g. In K.R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 3-30). San Francisco: Jossey-Bass.
- Murray, H.A. (1943). *Thematic Apperception Test*. Cambridge, MA: University Press.
- Neisser, U., Boodoo, G.M., Bouchard, T.J., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J. & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77-101.
- Ones, D.S., Schmidt, F.L. & Viswesvaran, C. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78, 679-703.

Fortsetzung Literatur

Fortsetzung Literatur

- Ones, D.S. & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17*, 609-626.
- Ones, D.S. & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11*, 245-269.
- Overton, R.C., Rodgers Taylor, L., Zickar, M.J. & Harms, H.J. (1996). The pen-based computer as an alternative platform for test administration. *Personnel Psychology, 49*, 455-464.
- Prochaska, M. (1998). *Leistungsmotivation: Methoden, soziale Erwünschtheit und das Konstrukt*. Frankfurt/M.: Peter Lang.
- Pulakos, E.D. & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9*, 241-258.
- Richman, W.L., Kiesler, S., Weisband, S. & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84*, 754-775.
- Rosse, J.G., Stecher, M.D., Miller, J.L. & Levin, R.A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634-644.
- Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Bern: Huber.
- Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General & Applied, 80*, 1-28.
- Ryan, A.M., Baron, H. & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52*, 359-391.
- Ryan, A.M. & Sackett, P.R. (1987). Pre-employment honesty testing: Fakability, reactions of test takers, and company image. *Journal of Business and Psychology, 1*, 248-256.
- Salgado, J.F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology, 82*, 30-43.
- Salgado, J.F. (1999). Personnel selection methods. In C.L. Cooper & I.T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol 12, pp. 1-54). London, UK: John Wiley & Sons.
- Schmalt, H.-D. (1999). Assessing the achievement motive using the grid technique. *Journal of Research in Personality, 33*, 109-130.
- Schmalt, H.-D. & Sokolowski, K. (1996). Motivationale Konstrukte. In M. Amelang (Hrsg.), *Temperaments- und Persönlichkeitsunterschiede. Enzyklopädie der Psychologie C/VIII/3* (S. 301-336). Göttingen: Hogrefe.
- Schmidt, F.L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist, 47*, 1173-1181.
- Schmidt, F.L. & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalisation. *Journal of Applied Psychology, 62*, 529-540.
- Schmidt, F.L. & Hunter, J.E. (1998). Meßbare Personmerkmale: Stabilität, Variabilität und Validität zur Vorhersage zukünftiger Berufsleistung und berufsbezogenen Lernens. In M. Kleinmann & B. Strauss (Hrsg.), *Potentialfeststellung und Personalentwicklung* (S. 16-43). Göttingen: Hogrefe.
- Schmidt, F.L., Hunter, J.E. & Outerbridge, A.N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432-439.
- Schmidt, F.L., Hunter, J.E., Pearlman, K. & Hirsh, H.R. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology, 38*, 697-798.
- Schmidt, J.U. (1993). Der Allgemeine Büroarbeitstest (ABAT) – mehr als ein Bürottest? *Diagnostica, 39*, 151-168.
- Schmidt-Atzert, L. & Bühner, M. (1997). Komponenten von Konzentrationstestleistungen: Eine Untersuchung mit systematisch variierten Durchstreichtest-Varianten. *Diagnostica, 43*, 314-326.

- Schmidt-Atzert, L. & Bühner, M. (1998). Fehlertypen im Aufmerksamkeits-Belastungs-Test d2. *Diagnostica*, 44, 142-152.
- Schneewind, K.A. & Graf, J. (1998). *Der 16-Persönlichkeits-Faktoren-Test, revidierte Fassung (16 PF-R)*. Göttingen: Hogrefe.
- Schneider, R.J., Hough, L.M. & Dunnette, M.D. (1996). Broadsided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior*, 17, 639-655.
- Schorr, A. (1995). Stand und Perspektiven diagnostischer Verfahren in der Praxis. Ergebnisse einer repräsentativen Befragung westdeutscher Psychologen. *Diagnostica*, 41, 3-20.
- Schuler, H. (1996). *Psychologische Personalauswahl. Einführung in die berufliche Eignungsdiagnostik*. Göttingen: Hogrefe.
- Schuler, H. (2000). Personalauswahl im europäischen Vergleich. In E. Regnet & L.M. Hoffmann (Hrsg.), *Personalmanagement in Europa*. Göttingen: Hogrefe.
- Schuler, H. & Prochaska, M. (2000). *Leistungsmotivationsinventar (LMI)*. Göttingen: Hogrefe.
- Schuler, H. & Rolfs, H. (2000). Hohenheimer Gründerdiagnose: Konzeption zur eignungsdiagnostischen Untersuchung potentieller Unternehmensgründer. In G.F. Müller (Hrsg.), *Existenzgründung und Führung als unternehmerisches Handeln – Forschung und Förderung* (S. 55-73). Landau: Empirische Pädagogik.
- Seifert, K.H. (1977). Theorien der Berufswahl und der beruflichen Entwicklung. In K.H. Seifert (Hrsg.), *Handbuch der Berufspsychologie* (S. 171-279). Göttingen: Hogrefe.
- Spranger, E. (1913). *Lebensformen*. Halle: Niemeyer.
- Sternberg, R.J. (1997). Tacit knowledge and job success. In N. Anderson & P. Herriot (Eds.), *International handbook of selection and assessment* (pp. 201-213). London: Wiley.
- Sternberg, R.J. & Wagner, R.K. (1993). The g-ocentric view of intelligence and job performance is wrong. *Current Directions in Psychological Science*, 2, 1-5.
- Stewart, G.L. (1996). Reward structure as a moderator of the relationship between extraversion and sales performance. *Journal of Applied Psychology*, 81, 619-627.
- Super, D.E. & Crites, J.O. (1962). *Appraising vocational fitness by means of psychological tests*. New York: Harper and Row.
- Tett, R.P., Jackson, D.N. & Rothstein, M.G. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44, 703-742.
- Tewes, U. (1991). *Hamburg-Wechsler-Intelligenztest für Erwachsene – Revision 1991 (HAWIE-R)*. Göttingen: Hogrefe.
- Viswesvaran, C., Ones, D.S. & Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557-574.
- Wittmann, W.W. & Matt, G.E. (1986). Aggregation und Symmetrie. Grundlagen einer multivariaten Reliabilitäts- und Validitätstheorie, dargestellt am Beispiel der differentiellen Validität des Berliner Intelligenzstrukturmodells. *Diagnostica*, 32, 309-329.

Fortsetzung Literatur