

7 Biographieorientierte Verfahren der Personalauswahl

von Heinz Schuler und Bernd Marcus

Inhaltsübersicht

1 Einleitung	176
2 Bewerbungsunterlagen und Referenzen	176
2.1 Grundlagen	176
2.2 Elemente und Inhalte der Bewerbungsunterlagen	177
2.2.1 Der biographische Ansatz	180
2.2.2 Elemente mit biographischem Inhalt	180
3 Biographische Fragebogen	185
3.1 Grundlagen	185
3.2 Itemselektion und Skalenbildung	187
3.2.1 Der empirische Ansatz	187
3.2.2 Der rationale Ansatz und Mischformen	190
3.2.3 Ein typenbildender Ansatz: Subgrouping	192
3.3 Zusammenfassung und Bewertung biographischer Fragebogen	193
4 Das Interview	196
4.1 Definition und Zwecksetzung	196
4.2 Verbreitung und Akzeptanz	196
4.3 Historische Entwicklung	197
4.4 Methodische Verbesserungen	198
4.5 Komponenten der Interviewstruktur	199
4.6 Typen strukturierter Interviews	200
4.6.1 Das Behavior Description Interview	202
4.6.2 Das situative Interview	203
4.6.3 Das Multimodale Interview	203
4.7 Empirische Evidenz	205
Zusammenfassung	207
Weiterführende Literatur	207
Literatur	208

Nicht alle Auswahlverfahren lassen sich dem eigenschafts- oder simulationsorientierten Konzept zuordnen.

Ein dritter Grundansatz schließt von vergangenem auf zukünftiges Verhalten.

Informationen aus der Vergangenheit des Bewerbers

Bewerbungsunterlagen sind das verbreitetste Auswahlinstrument.

1 Einleitung

In den beiden vorangegangenen Kapiteln wurden zwei Grundansätze der Eignungsdiagnostik vorgestellt. Die direkte Simulation beruflichen Verhaltens (vgl. Kapitel 6) und die eher indirekte Erfassung berufsrelevanter Konstrukte (vgl. Kapitel 5) stellen die beiden Zugänge zur Prognose des Berufserfolgs dar, aus denen sich die meisten validen Einzelverfahren der Personalauswahl ableiten. Dennoch lassen sich nicht alle Eignungsdiagnostika sinnvoll diesen zwei Gruppen zuordnen. Insbesondere gilt dies nicht für die beiden mit Abstand verbreitetsten Auswahlinstrumente, deren Bedeutung die Darstellung in einem eigenen Kapitel erfordert: die Auswertung von Bewerbungsunterlagen und das Einstellungsgespräch bzw. Interview.

Externe Personalauswahl beginnt fast immer damit, daß ein Personalverantwortlicher sich durch einen – heutzutage meist umfangreichen – Stapel von Mappen liest, in denen Bewerber sich selbst und ihre Vorstellungen über die zu besetzende Stelle darstellen. Solche *Bewerbungsunterlagen* enthalten Informationen unterschiedlicher Qualität und Bedeutung, es wird jedoch zu zeigen sein, daß ihre (potentiell) validesten Elemente vor allem über die Vergangenheit des Bewerbers informieren. Der direkte Schluß von vergangenem auf zukünftiges Verhalten zählt zu den ältesten und prognostisch fruchtbarsten Ansätzen der Eignungsdiagnostik. Wir werden ihn hier als dritten Grundansatz neben Simulationen und Eigenschaftsmessungen vorstellen, obwohl sich darin Elemente beider Grundformen wiederfinden. Dabei wird sich zeigen, daß das Potential dieses „biographischen“ Ansatzes mit der Durchsicht von Lebenslauf und Zeugnissen noch keineswegs ausgeschöpft ist, sondern entscheidend von psychometrischer Fundierung und Standardisierung profitiert, was uns über die in der Praxis verbreiteten Personalfragebogen direkt zu einer hierzulande noch unterschätzten Form psychologischer Tests führt: dem *biographischen Fragebogen*.

Ähnlich universell verbreitet wie Bewerbungsunterlagen, jedoch i.d.R. eher am Ende des Auswahlprozesses stehend, ist das *Interview*. Der eigenständige Charakter von Bewerbungsgesprächen als Instrument der Personalauswahl ist dabei weniger inhaltlicher als methodischer Natur; hier – und im Rahmen der Auswahlprozedur häufig nur hier – treten Bewerber und Personalverantwortlicher miteinander in direkte, verbale Interaktion. Dies hat für beide Seiten vielfältige Implikationen, auf die im folgenden einzugehen sein wird (vgl. dazu auch Kapitel 16). Inhaltlich lassen sich im Interview im Prinzip alle Grundformen der Eignungsdiagnostik realisieren. Auch hier zeigt sich jedoch wieder, daß ein nicht zu gering ausgeprägtes Maß an Strukturierung der Validität erheblich zugute kommt.

2 Bewerbungsunterlagen und Referenzen

2.1. Grundlagen

Je nach Berufsgruppe geben zwischen 80 und 99 % der deutschen Unternehmen an, Bewerbungsunterlagen zur externen Personalauswahl heranzuziehen (Schuler, Frier & Kauffmann, 1993). Kein anderes Auswahlinstrument erreicht diese Verbreitung, und zu keinem anderen Verfahren liegt eine solche Flut von Veröffentlichungen in Publikumszeitschriften und Ratgebern vor, aus denen sich potentielle Bewerber über die gewünschten und (vermeintlich) erfolversprechenden Gestaltungsmerkmale informieren können. An dieser Stelle soll der Vielzahl derartiger Publikationen nicht noch eine weitere hinzugefügt werden. Beschäftigten sollen uns vielmehr der Charakter und die Eigenarten von Bewerbungsunterlagen als Personalauswahlverfahren, welche Informationen Personalfachleute in der Praxis daraus gewinnen, und insbesondere, welche Elemente einen tatsächlichen eignungsdiagnostischen Nutzen, also in erster Linie prognostische Validität versprechen.

So häufig Bewerbungsunterlagen in Auswahl-situationen herangezogen werden, so selten dienen sie als *alleinige* Entscheidungsgrundlage zur *Besetzung* einer Stelle. Personalauswahl erfolgt in der Regel in einem mehrstufigen Prozeß, an dessen *Anfang* die Auswertung der Bewerbermappe steht. In diesem Stadium stellen die Unterlagen den ersten Kontakt zwischen Bewerber und Arbeitgeber her und dienen diesem zunächst zur Überprüfung der formalen Voraussetzungen (Gesellenbrief, Hochschulstudium etc.) für die angestrebte Position. Endgültige Entscheidungen aufgrund der Bewerbungsunterlagen sind in aller Regel *Negativentscheidungen* (m.a.W. Ablehnungen); sie dienen zunächst der Reduzierung der Bewerberzahl auf ein handhabbares Maß. Angesichts des gegenwärtigen Bewerberüberhangs trifft dies z.Z. sicher für einen Großteil der Bewerbungen zu, was aus *Bewerbersicht* bedeutet, daß die Stellensuche oft bereits auf der Grundlage ihrer Unterlagen scheitert. Aufgrund dieser Tatsache ist es nicht verwunderlich, daß sich hier dem Stand der Berater ein einträglicher Markt eröffnet hat.

Aus *Arbeitgebersicht* hat die Funktion der „Datenreduktion“ eine ganz andere Implikation. Es lassen sich zwei Arten von Fehlentscheidungen im Auswahlprozeß unterscheiden: die Einstellung ungeeigneter Bewerber (false positives) und die Ablehnung geeigneter Kandidaten (false negatives). Nicht ohne Grund werden bei der endgültigen Entscheidung über die Stellenbesetzung Fehler der ersten Art (false positives) i.d.R. als gravierender empfunden; sie fallen im Gegensatz zu false negatives früher oder später auf (und dann mglw. auf den Verantwortlichen zurück) und lassen sich, bei der gegenwärtigen Rechtslage in Deutschland, auch kaum ohne erhebliche Kosten und Mühen revidieren. In einem frühen Stadium des Auswahlprozesses stellt sich die Sache jedoch anders dar. Hier lassen sich false positives auf nachfolgenden Stufen noch korrigieren, insbesondere wenn dort valide Auswahlinstrumente zur Verfügung stehen; false negatives sind dagegen endgültig. Sie reduzieren das zur Verfügung stehende Potential geeigneter Bewerber. Für die Beurteilung der Validität als wichtigstem Maßstab zur Einschätzung eines Auswahlinstrumentes spielt dieser Umstand keine Rolle (Validität reduziert Fehler *beider* Arten), wohl aber für die Wahl der Selektionsquote (vgl. Kapitel 22), des Prozentsatzes an Bewerbungen also, die weiterhin in Betracht gezogen werden. Eine zu niedrig gesetzte Selektionsquote auf dieser Stufe ist besonders dann kritisch, wenn die Bewerbungsunterlagen selbst keine oder nur sehr geringe Validität besitzen (wobei sich das *Validitätskriterium* beim „screening out“ durchaus von dem bei der Selektion verwendeten unterscheiden kann). Wir werden uns mit der Frage nach der Validität noch beschäftigen.

Eine weitere Eigenart von Bewerbungsunterlagen ist deren Zusammensetzung aus vielen, teils recht *unterschiedlichen Elementen*. Die Auswertung dürfte dabei in der Praxis eher intuitiv als systematisch erfolgen, wobei erfahrene Experten in der Form eines Wechsels zwischen gestalthafter Gesamtbetrachtung und Detailanalyse vorgehen (Mell, 1988). Der Frage, welche Bedeutung einzelnen Elementen und daraus gebildeten Konfigurationen für die (Vor-) Auswahlentscheidung zukommt, gingen Machwirth, Schuler und Moser (1996) mit Hilfe eines Analyseverfahrens nach, das es erlaubt, implizite Regeln der beteiligten Personalreferenten auf statistischem Wege zu rekonstruieren. Dabei zeigte sich, daß die Entscheidung, Bewerber im weiteren Auswahlprozeß *nicht* mehr zu berücksichtigen (Negativselektion), vor allem aufgrund formaler Aspekte wie Mängel in der Bewerbung getroffen wurde, während für die Identifizierung *geeigneter* Bewerber (Positivselektion für weitere Auswahl-schritte) der Inhalt ihrer bisherigen Tätigkeiten eine größere Rolle spielte.

2.2 Elemente und Inhalte der Bewerbungsunterlagen

Im folgenden werden wir uns zunächst vor allem mit zwei Fragen beschäftigen: Welche Informationen *sind* in den Bewerbungsunterlagen üblicherweise enthalten und welche dieser Informationen *sollten* für die Entscheidung berücksichtigt werden? Die erste Frage ist im wesentlichen deskriptiver Na-

Die Analyse von Bewerbungsunterlagen steht in einem i.d.R. mehrstufigen Auswahlprozeß am Beginn.

Unbegründete Ablehnungen sind beim ersten Screening besonders kritisch.

Formale Aspekte der Bewerbung werden in der Praxis von den Inhalten unterschieden.

Eine Vielzahl von Aspekten ist bei der Analyse von Bewerbungsunterlagen zu beachten.

tur und läßt sich relativ leicht beantworten. Tabelle 1 gibt einen Überblick über mögliche Gegenstände der Dokumentenanalyse, die allerdings in der Praxis angesichts einer durchschnittlichen Dauer von zehn Minuten je Auswertungsvorgang (Kleinmann & Seibt, 1990) kaum jemals vollständig abgearbeitet werden dürften.

Tabelle 1:

Auswertung von Bewerbungsunterlagen (aus Schuler, 1998a, S. 80)

1. Formale Aspekte
<ul style="list-style-type: none"> – Ist die Bewerbung ordentlich und übersichtlich angelegt? – Ist sie fehlerfrei und vollständig? – Sind Art und Umfang der Bewerbung der zu besetzenden Position angemessen? – Lichtbild
2. Anschreiben und Lebenslauf
<ul style="list-style-type: none"> – Sind Anschreiben und ausführlicher oder tabellarischer Lebenslauf enthalten? – Simulationsorientierte und eigenschaftsorientierte Auswertung
3. Erforderliche Ausbildung
<ul style="list-style-type: none"> – Zeugnisse – Praktikumsnachweise – sonstige Bescheinigungen – Ausbildungsbedingter Auslandsaufenthalt
4. Erforderliche Spezialkenntnisse
<ul style="list-style-type: none"> – Sprachen – EDV-Kenntnisse – sonstige Zusatzausbildungen, Lehrgänge etc.
5. Übereinstimmung Lebenslauf/Belege
<ul style="list-style-type: none"> – Lückenlosigkeit – Zeitfolgeanalyse
6. Plausibilität des Stellenwechsels
<ul style="list-style-type: none"> – Abfolge der Positionen – Nachvollziehbarkeit der Arbeitgeberwechsel
7. Schulnoten
<ul style="list-style-type: none"> – gut geeignet zur Prognose weiterer Ausbildungsleistungen – weniger geeignet zur Prognose des Berufserfolgs
8. Studienleistungen
<ul style="list-style-type: none"> – falls bekannt, Notenniveau von Hochschule und Studienfach berücksichtigen – Qualität der Diplomarbeit ist wichtiger als das Thema
9. Arbeitszeugnisse und Referenzen
<ul style="list-style-type: none"> – meist nur verlässlich, wenn von Fachleuten ausgestellt – persönliche (mündliche) Referenzen meist aussagekräftiger als schriftliche
10. Ergänzende anforderungsspezifische Aspekte
<ul style="list-style-type: none"> – Berufserfahrung – Mobilität usw.
11. Offengebliebene Fragen werden für das Gespräch vorgemerkt

Die zweite Frage bezieht sich vor allem auf die Validität, und hier besteht vielfach noch erheblicher Klärungsbedarf, zumal bezüglich der Intensität und Qualität der vorliegenden Forschung erhebliche Unterschiede zwischen den einzelnen Bestandteilen der Bewerbungsunterlagen bestehen. Es erscheint sinnvoll, hierbei zunächst die *Elemente* von den darin enthaltenen Informationen oder *Inhalten* zu unterscheiden. Üblicherweise enthält die Bewerbermappe folgende Einzelunterlagen, wobei, besonders bei Berufsanfängern, nicht immer alle Elemente vollständig vorhanden sein müssen:

- Lichtbild
- Anschreiben
- Schul-, Examens- und/oder Ausbildungszeugnisse (Noten)
- Lebenslauf
- Arbeitszeugnisse und Referenzen
- in manchen Fällen: vorab verschickter Personalfragebogen

Inhaltlich lassen sich dabei zumindest drei Gesichtspunkte unterscheiden, wobei wir auf den letzten ausführlicher eingehen werden. Lichtbild und Anschreiben lassen sich, ebenso wie das gesamte Erscheinungsbild der Unterlagen, zunächst auf *formale Aspekte* überprüfen. Im Falle des *Lichtbilds* sind dies, neben Format und Qualität (Fotostudio vs. Automat, Farbe vs. SW) des Fotos, vor allem das äußere Erscheinungsbild des Bewerbers (Frisur, Kleidung) und dessen *physische Attraktivität*. Daß attraktivere Bewerber bei der Vorauswahl und in geringerem Maße auch bei der endgültigen Auswahlentscheidung bevorzugt werden, darf aufgrund übereinstimmender Ergebnisse entsprechender Forschung (z. B. Marlowe, Schneider & Nelson, 1996; Schuler & Berger, 1979) als relativ gesichert gelten. Ob dieses Merkmal dagegen Zusammenhänge mit dem Berufserfolg aufweist, wurde u.W. bislang noch nicht systematisch untersucht. Die einschlägige Forschung setzt größtenteils implizit voraus, daß Attraktivität, häufig mit anderen Merkmalen unter der Kategorie „erster Eindruck“ zusammengefaßt, einen Störfaktor (englisch: bias) darstellt und somit die Validität angenommenermaßen *vermindert*. Diese ungeprüfte Annahme widerspricht allerdings neueren Erkenntnissen, denen zufolge – allerdings auf einer breiteren Informationsbasis gewonnene – erste Eindrücke durchaus positive Validität für die berufliche Leistung besitzen können (Borkenau, 1991). Die Frage nach der Validität muß hier also als ebenso ungeklärt gelten wie bei formalen Aspekten wie *Fehlerfreiheit, Übllichkeit, Vollständigkeit* u.ä., die dem Gesamtbild der Bewerbungsunterlagen und innerhalb der Elemente besonders dem *Anschreiben* entnommen werden können. Hier muß, u.a. als Folge der oben erwähnten Verbreitung hilfreicher Gestaltungsempfehlungen in den Medien, in vielen Berufsfeldern mit einer gewissen Angleichung und entsprechend geringem Informationsgehalt gerechnet werden. Eine Abhebung ist hier vielfach nur in negativer Hinsicht möglich und dient vielen Betrieben als – angesichts hoher Bewerberzahlen willkommenes – „K.O.-Kriterium“ zu einer ersten Vorauslese. Die Berechtigung dieser nicht unplausiblen Vorgehensweise wurde freilich u.W. noch von niemandem systematisch untersucht.

Insbesondere das Anschreiben wird aber noch unter einem anderen inhaltlichen Aspekt interpretiert, den bspw. Hollmann und Reitzig (1995) meinen, wenn dort vom Analogieschluß die Rede ist. Aus Merkmalen wie Stil, Prägnanz und Struktur werden erste Schlüsse auf den Arbeitsstil des Bewerbers gezogen, was im Prinzip der Vorgehensweise bei einer *Arbeitsprobe* entspricht. In der diesem Lehrbuch zugrundeliegenden Diktion wäre diese Interpretation dem simulationsorientierten Prognosekonzept zuzurechnen. Zumindest implizit dürfte jedoch häufig auch auf Fähigkeiten und Persönlichkeit (eigenschaftsorientiert) geschlossen werden. Auch zur Validität dieser Auswertungen liegen u.W. derzeit keine kontrollierten Studien vor. Obwohl damit zu rechnen ist, daß erfahrene Personalpraktiker dem Inhalt des Anschreibens wertvolle Einsichten entnehmen können, zeigt andererseits die Forschung zur klinischen Urteilsbildung, daß solche freien Interpretationen mit erheblichen Reliabilitätsmängeln behaftet sind, was auch für die Validität eher bescheidene Werte erwarten läßt.

Bestandteile der Bewerbungsunterlagen

Formale Aspekte dienen oft als erstes Selektionskriterium.

Die Validität klinischer Schlußfolgerungen bei der Auswahl ist weitgehend ungeklärt.

Bewerbungsunterlagen informieren vor allem über die Vergangenheit eines Bewerbers.

Der biographische Ansatz der Eignungsdiagnostik beruht auf der Stabilität menschlichen Verhaltens.

2.2.1 Der biographische Ansatz

Auch die restlichen vier Elemente aus obiger Aufstellung dürften zunächst unter formalen Aspekten analysiert werden. Weitaus bedeutsamer sind jedoch die daraus entnehmbaren inhaltlichen Informationen, die überwiegend den Charakter „harter“, überprüfbarer Fakten haben. Das gemeinsame Merkmal von Schul- und Ausbildungszeugnissen, Lebenslauf, Referenzen früherer Arbeitgeber und großteils auch Personalfragebogen ist, daß sie über die *Vergangenheit des Bewerbers* informieren. Darin spiegeln sich Ereignisse, Verhalten und Leistungen – im Sinne erzielter Resultate (zum Leistungsbegriff vgl. Kapitel 15) – wider, aus denen eignungsdiagnostische Prognosen *zukünftigen Verhaltens* abgeleitet werden. Dieser Ansatz ist theoretisch weniger gut untermauert als das simulations- und besonders das eigenschaftsorientierte Prognosekonzept; er beruht im wesentlichen auf der Annahme, daß menschliches Verhalten in gewissem Ausmaß über lange Zeiträume stabil bleibt. Obwohl die Stabilitätshypothese keineswegs im Sinne einer absoluten Unveränderlichkeit interpretiert werden sollte, spricht für sie in fast einzigartigem Maße die empirische Evidenz, was sich vielleicht am besten in einem Satz von Lord Byron zusammenfassen läßt: „The best prophet of the future is the past.“

Wir möchten den auf der empirischen Beobachtung der Verhaltensstabilität beruhenden *biographischen Ansatz* an dieser Stelle als das dritte Grundkonzept der Eignungsdiagnostik einführen, da die im folgenden zu schildernden Auswahlinstrumente im wesentlichen auf diesem Axiom beruhen. Stabilität ist auch ein wichtiges Element des konstruktorientierten Prognosekonzepts, das auf der angenommenen Existenz stabiler individueller Differenzen beruht. Anders als dort geht es beim biographischen Ansatz meist (zu Ausnahmen siehe unten) jedoch nicht um dem Verhalten zugrundeliegende abgrenzbare (homogene) Konstrukte, sondern zukünftiges Verhalten soll *direkt* aus vergangenem Verhalten bzw. äußeren Ereignissen vorhergesagt werden, die in der Regel multipel bedingt und daher in sich weniger homogen sind. Diese eher direkte Form der Schlußfolgerung teilt das biographische Konzept wiederum mit dem simulationsorientierten Ansatz. Allerdings werden hier nicht zukünftige Arbeitsaufgaben simuliert, und es findet üblicherweise auch keine direkte Verhaltensbeobachtung statt. An deren Stelle treten, wie bei den nun darzustellenden Komponenten der Bewerbungsunterlagen, externe Nachweise bzw. Selbst- und Fremddarstellungen von begrenzter Systematisierung oder, wie bei dem nachfolgend zu schildernden biographischen Fragebogen, vollstandardisierte Verfahren, die sich in technischer Hinsicht nur wenig von konstruktorientierten Tests unterscheiden.

Nach dieser etwas abstrakten Einführung wenden wir uns jetzt wieder den Verfahren zu. Zu einigen der folgenden Bestandteile von Bewerbungsunterlagen liegen durchaus zahlreiche Validitätsstudien, teilweise sogar metaanalytische Zusammenfassungen dieser Arbeiten vor, die eine etwas eingehendere Beschäftigung lohnen.

2.2.2 Elemente mit biographischem Inhalt

Schul-, Ausbildungs- und Studienleistungen genießen in der Öffentlichkeit und teilweise auch in der wissenschaftlichen Diskussion (Althoff, 1986) nicht unbedingt den Ruf eines validen Prädiktors für Berufserfolg. Noch Hollman und Reitzig (1995) kommen zu dem Schluß, daß „... Studien zur prädiktiven Validität von Zeugnisnoten ... zwar gelegentlich erstaunlich hohe Korrelationen aufweisen, das Gesamtbild jedoch insgesamt zu widersprüchlich erscheint, als daß ihre Verwendung empfohlen werden könnte“ (S. 467). Zur Auflösung solcher Widersprüche wurde die Technik der Metaanalyse (vgl. Kapitel 5) entwickelt, deren Ergebnisse oftmals ganz andere Schlußfolgerungen nahelegen. Zur Validität von Zeugnisnoten liegen inzwischen mehrere in- und ausländische Metaanalysen vor, die bei Schuler (1998b) zusammengefaßt sind. Dabei zeigen sich die „erstaunlich hohen Korrelationen“ zumin-

dest für die Beziehung zwischen Schulnoten und *Ausbildungserfolg* durchaus generalisierbar; im Mittel errechnet sich für den deutschsprachigen Raum ein um statistische Artefakte bereinigter Wert von $r = .46$ für den Studienerfolg bzw. $.41$ für die Abschlußnote von Berufsausbildungen. Auch hier kommt das statistische Prinzip zum Tragen, demzufolge eine Kombination mehrerer Einzelprädiktoren reliabler (und deshalb auch valider) ist als jeder Prädiktor für sich: Die Gesamtnote sagte den Ausbildungserfolg jeweils besser voraus als jede Einzelnote. Bester Einzelprädiktor war in beiden Fällen die Mathematiknote. Allerdings fand sich auch eine deutliche Abwärtstendenz der Validität von Schulabschlüssen jüngerer Datums (mittlere Korrelation in neueren Studien: $r = .26$).

Nun geht es bei Schul- und Ausbildungsabschlüssen aber jeweils um die Erhebung weitgehend theoretischer (Lern-)Leistungen, deren konzeptionelle Ähnlichkeit eine substantielle Korrelation erwarten läßt. Berufliche Eignungsdiagnostik beschäftigt sich jedoch mit der Prognose des Berufserfolgs, und an diesem Kriterium muß die Effektivität von Noten als Auswahlinstrument letztlich gemessen werden. Hierzu liegen fast nur nordamerikanische Studien vor. Dort kommt die jüngste Metaanalyse (Roth, BeVier, Switzer & Schippmann, 1996) aber immer noch auf den beachtlichen Wert von $r = .32$, der deutlich über den Angaben aus früheren quantitativen Zusammenfassungen liegt und darüber hinaus auch generalisierbar (also nur in geringem Maße situationsabhängig) war. Allerdings zeigte sich eine eindeutige Tendenz *abnehmender* Korrelationen bei *zunehmendem* Zeitintervall zwischen den Meßzeitpunkten (Ausbildungsabschluß und Erhebung der Berufsleistung). Hierin wird exemplarisch eine Schwäche des biographischen Ansatzes deutlich, daß nämlich die Verhaltens- bzw. Leistungsstabilität – und in der Folge auch die Prognosequalität – im Laufe der Zeit abnimmt. Dennoch sprechen die berichteten Ergebnisse eindeutig dafür, daß Zeugnisnoten, entgegen verbreiteter Annahmen, zu den validesten Einzelkomponenten der Bewerbungsunterlagen zählen.

Zu den Standardunterlagen zählt auch ein tabellarischer, seltener ein ausführlicher *Lebenslauf*. Die darin enthaltenen biographischen Daten geben Auskunft über die schulische und berufliche Qualifikation und Entwicklung einschließlich Aufgaben- und Stellenwechseln, aber auch über den familiären Hintergrund und über das außerberufliche Engagement. Die Interpretation erfolgt dabei in zweifacher Hinsicht. Zunächst geht aus dem Lebenslauf eine sehr wesentliche Information hervor, die eine eher simulationsorientierte Prognose erlaubt: das Ausmaß an *Berufserfahrung*. Berufliche Erfahrung spielt eine wichtige Rolle in neueren pfadanalytischen Kausalmodellen zur Erklärung des Berufserfolgs (Schmidt & Hunter, 1992), wobei sich insbesondere kognitive Fähigkeiten teilweise nur indirekt, über die dadurch bedingten Unterschiede bei der Aufnahme und Verarbeitung berufsrelevanter Informationen (also die *genutzte* Erfahrung), auf die berufliche Leistung auswirken. Gleichwohl fand die jüngste Metaanalyse zum direkten Zusammenhang zwischen Berufserfahrung und -erfolg (Quiñones, Ford & Teachout, 1995) lediglich eine artefaktkorrigierte Validität von $r = .27$, ein Wert, der, gemessen an der intuitiven Bedeutungseinschätzung vieler Praktiker, eher enttäuschend ausfällt. Bemerkenswert ist dabei, daß die *Dauer* und *Art* der Erfahrung – Informationen also, die sich direkt dem Lebenslauf entnehmen lassen – mit Korrelationen von $.27$ bzw. $.21$ wesentlich geringere Validität besitzen als deren *Menge* (gemessen als numerische *Anzahl* bspw. der Ausführungen einer bestimmten Aufgabe oder der bekleideten Positionen) mit einem mittleren Wert von $.43$. Gleichfalls zur Erhöhung der Validität trug es bei, wenn die Erfahrung mit *spezifischen Aufgaben* eher globalen berufs- oder organisationsbezogenen Erfahrungsmaßen gegenübergestellt wurde. Auch hierüber findet sich i.d.R. wenig im Lebenslauf, so daß die Validität der Erfahrungskomponente aus dieser Unterlage eher im unteren Bereich der von Quiñones et al. (1995) berichteten typischen Bandbreite von $.09$ bis $.46$ ¹ vermutet werden dürfte.

Gemessen an der Validität sind Schulnoten besser als ihr Ruf.

Aus dem Lebenslauf ersichtliche Elemente der Berufserfahrung sind nur mäßig valide.

¹ Anmerkung für interessierte Leser: eigene Berechnung; der Originalartikel enthält an dieser Stelle eine etwas mißverständliche Angabe

Praktiker ziehen aus den im Lebenslauf vorhandenen Daten Schlüsse auf kognitive Fähigkeiten, Motivation und soziale Kompetenz von Bewerbern ohne daß deren Validität geklärt wäre.

Die Graphologie zählt aus Sicht der Bewerber zu den am wenigsten geschätzten Auswahlverfahren.

Lebensläufe werden allerdings nicht nur anhand der daraus ersichtlichen Berufserfahrung ausgewertet. Brown und Campion (1994) etwa konnten zeigen, daß mit der Personalauswahl betraute Praktiker aus den im Lebenslauf enthaltenen biographischen Daten Schlüsse sowohl auf kognitive Fähigkeiten des Bewerbers als auch auf dessen Motivation, soziale und Führungskompetenz etc. ziehen, Merkmale also, die eindeutig dem konstruktorientierten Prognosekonzept zuzurechnen wären. Über die Validität solcher Schlußfolgerungen und die des Lebenslaufs insgesamt existieren freilich eher indirekte Hinweise. Interessanterweise finden sich diese in Untersuchungen zu einer dritten, gänzlich andersartigen Verwertungsmöglichkeit von Lebensläufen, deren Betrachtung einen kurzen Exkurs erfordert.

Exkurs: Graphologie

Gelegentlich kann man in Stellenanzeigen lesen, daß das betreffende Unternehmen von Bewerbern fordert, den Lebenslauf in *handschriftlicher* Form anzufertigen. In solchen Fällen ist damit zu rechnen, daß über den Kandidaten ein *graphologisches Gutachten* erstellt werden soll. Die Verbreitung der Handschriftenanalyse als Auswahlverfahren differiert erheblich zwischen verschiedenen Staaten, insbesondere in französischsprachigen Ländern zählt sie nach wie vor zu den meistgenutzten Instrumenten (Lévy-Leboyer, 1994), während sie hierzulande allenfalls noch zur Auswahl von Führungskräften eine gewisse Rolle spielt (Schuler et al., 1993). Die Urteilsbildung bei diesem Verfahren ist dabei sehr indirekter Natur und in dieser Hinsicht mit projektiven Tests vergleichbar (vgl. Driver, Buckley & Frink, 1996, zu einer Gegenposition zu dieser Einordnung sowie zu den folgenden Ausführungen): Aus objektiv meßbaren Merkmalen wie Neigung, Höhe und Weite und aus qualitativen Kennzeichen wie Regelmäßigkeit und Harmonie einer Handschrift werden stabile Eigenschaften des Schreibers, insbesondere dessen Persönlichkeit erschlossen, die wiederum zu einer konstruktorientierten Erfolgsprognose herangezogen werden. Durch diese etwas umständliche und für den Laien vollkommen undurchsichtige Form der Analyse soll, wie bei projektiven Tests, u.a. eine willentliche Verfälschung des Analysematerials verhindert werden. Zwar gelingt dies bei der Graphologie zweifellos, dieser Vorzug wird jedoch durch erhebliche Mängel erkauft, die diesen mit Sicherheit weit überkompensieren.

Nicht nur führt die mangelnde Augenscheinigkeit dazu, daß die Graphologie zu den von Bewerberseite am wenigsten geschätzten Auswahlverfahren zählt (siehe z.B. Steiner & Gilliland, 1996), auch Validitätsstudien kommen hier i.d.R. zu niederschmetternden Ergebnissen. Neter und Ben-Shakhar (1989) fassen die wenigen Untersuchungen metaanalytisch zusammen und schätzen zwar die Validität graphologischer Gutachten auf immerhin $r = .20$; Nicht-Graphologen (Psychologen und Laien) kommen aber aufgrund der gleichen Vorlagen fast durchweg zu *valideren* Schlüssen als Graphologen. Werden neutrale (inhaltsfreie) Schriftproben benutzt, versagen graphologische Gutachten vollkommen ($r = -.01$). Dies läßt nur den Schluß zu, daß nützliche Erkenntnisse aus Lebensläufen lediglich auf der Grundlage ihrer *Inhalte* gewonnen werden können, für deren Validität Werte knapp über $.20$ eine brauchbare Schätzung darstellen dürften. Das aber ist keine Graphologie, und es spielt auch keine Rolle, ob der Lebenslauf in handschriftlicher oder gedruckter Form vorliegt.

Dafür, daß die Ursache der mangelnden prognostischen Validität von Handschriftenanalysen auf der Ebene der *Konstruktvalidität* zu suchen ist, findet sich ein Hinweis bei Bushnell (1996). Er fand keinerlei Zusammenhang zwischen Selbst- und Partneereinschätzungen bestimmter Persönlichkeitsmerkmale auf der einen Seite und graphologischen Beurteilungen der gleichen Merkmale auf der anderen Seite, während die Ergebnisse eines schriftlichen Persönlichkeitstests gut mit den Selbst- und Fremdeinschätzungen übereinstimmten. Es scheint so, als wäre die Graphologie nicht in

der Lage, die Eigenschaften zu messen, die sie zu messen beabsichtigt. Es ist nicht auszuschließen, daß hier mit Hilfe computergestützter Verfahren in Zukunft Fortschritte erzielt werden; so lange aber ist von der Verwendung der Graphologie als Auswahlverfahren abzuraten.

Während die im Lebenslauf enthaltenen Informationen im wesentlichen quantitativer bzw. deskriptiver Natur sind, enthalten qualifizierte *Arbeitszeugnisse* in gewissem Umfang *Wertungen* des Verhaltens in früheren beruflichen Stellungen, im Grunde also eine Art Leistungsbeurteilung. Kein anderes Element der Bewerbungsunterlagen kommt den simulationsorientierten Auswahlverfahren so nahe. Allerdings steht die Aussagekraft von Arbeitszeugnissen unter der Einschränkung, daß aufgrund des geltenden Arbeitsrechts kaum ausdrücklich Nachteiliges über den Mitarbeiter ausgesagt wird, weshalb sich eine „Zeugnissprache“ entwickelt hat, auf deren Fragwürdigkeit seit langem hingewiesen wird. Werden andererseits nicht die Regeln dieser Zeugnissprache benutzt, ist die Interpretation weniger eindeutig, insbesondere wenn Zeugnisse nicht von Fachleuten ausgestellt oder ausgewertet werden. Auch ist nicht auszuschließen, daß ein Zeugnis vom Beurteilten selbst verfaßt ist (laut Weuster, 1994, in 17 % der Fälle) oder daß der Ausstellende vorrangig das Ziel verfolgt, sich vom Beurteilten konfliktfrei zu trennen und deshalb wahrheitswidrige Angaben macht. Dies bedeutet allerdings nicht, daß Arbeitszeugnisse gänzlich zu verwerfen wären. Quellen perfekter Validität stehen uns nicht zur Verfügung, und im Rahmen einer sequentiellen Strategie der Informationssammlung kommt auch dem Arbeitszeugnis ein gewisser Wert zu. Weuster (1994) hat den interessanten Versuch unternommen, ein einheitlich formuliertes Arbeitszeugnis mit der Bitte um Analyse an eine größere Zahl von Personalberatern zu schicken. Die Zusammenstellung ergibt, daß die Zeugnisaussagen teilweise recht unterschiedlich interpretiert wurden; insgesamt dürften aber die Diskrepanzen nicht größer sein als bei der Beurteilung beobachteten Verhaltens, etwa im Assessment Center.

Arbeitszeugnisse sind bei gehobenen Positionen von größerer Bedeutung. In diesen Fällen werden oft zusätzliche *Referenzen* eingeholt, die oftmals in mündlicher Form informativer, weil offener sind als in schriftlicher. Die vor allem von Personalberatern gelegentlich praktizierte Überprüfung der Referenzinformation kann ergänzenden Aufschluß liefern. Mit *reference check* wird im Englischen neben der Überprüfung auch das aktive Einholen zusätzlicher Beurteilungsinformationen von früheren Arbeitgebern bezeichnet. Nach Hunter und Hunter (1984) kommt diesem Verfahren mit $r = .26$ eine durchaus brauchbare Validität zu.

Alle bisher beschriebenen Komponenten der Bewerbungsunterlagen leiden unter einem Standardisierungsdefizit, das sich nicht immer auf die Unterlagen selbst bezieht (bspw. können Lebenslauf und von Fachleuten ausgestellte Arbeitszeugnisse gewissermaßen als teilstandardisiert gelten), wohl aber auf deren Auswertung. Das Problem dieser bereits angesprochenen „klinischen Urteilsbildung“ besteht vor allem in deren mangelnder Reliabilität, daneben allerdings auch in der geringen Nachvollziehbarkeit und Ökonomie; gleiche Inhalte werden von Beurteiler zu Beurteiler, von Fall zu Fall und von Zeitpunkt zu Zeitpunkt unterschiedlich eingeschätzt. Es ist ein meßtheoretisches Gesetz, daß diese Ungenauigkeit auch auf die Validität negativ durchschlägt. Eine konsequente Standardisierung der Erhebung biographischer Daten führt zwangsläufig zu der Form des biographischen Fragebogens (nicht unbedingt aber in Papier-und-Bleistiftform), mit der wir uns im folgenden Abschnitt ausführlich beschäftigen.

Einen ersten Schritt in diese Richtung unternehmen viele Firmen bereits, indem sie ihre Bewerber auffordern, den Unterlagen zuvor ausgefüllte *Personalfragebogen* beizulegen². Gelegentlich werden diese auch erst an Ort und Stelle vorgelegt. In jedem Fall dienen sie der Erhebung vergleichbarer

Graphologie ist ein für die Personalauswahl z.Z. untaugliches Instrument.

Valide Schlüsse aus Arbeitszeugnissen erfordern, daß früherer und künftiger Arbeitgeber die gleiche „Sprache“ sprechen bzw. verstehen.

² Tatsächlich liegt hier ein historischer Ursprung des biographischen Fragebogens, indem man begann, einzelne Fragen auf ihre empirische Validität zur Prognose beruflicher Leistung zu überprüfen (vgl. unten).

Personalfragebogen stellen einen ersten Schritt in Richtung standardisierter Eignungsdiagnostik dar. Sie sind mitbestimmungspflichtig.

Informationen über die Bewerber und sind insofern eine erste Quelle standardisierter Messung im Auswahlprozeß. Die Auswahl der Fragen erfolgt jedoch i.d.R. betriebsspezifisch nach dem „gesunden Menschenverstand“ bzw. nach der personalbezogenen Entscheidungsstruktur im Unternehmen und ist eigentlich nicht der psychologischen Eignungsdiagnostik i.e.S. zuzurechnen. Auch der Zweck von Personalfragebogen ist nicht ausschließlich eignungsdiagnostisch; sie werden nach erfolgter Einstellung archiviert und dienen als einheitliche Datenbasis für Personalverwaltung und -planung, bei geeigneter Gestaltung können auch Personalentwicklungsmaßnahmen darauf aufbauen. Eine u.a. daraus abgeleitete Besonderheit von Personalfragebogen besteht in deren relativ eindeutiger *rechtlicher Regelung* durch Gesetz und Rechtsprechung. Nach § 94 Abs.1 Betriebsverfassungsgesetz bedürfen Personalfragebogen im Geltungsbereich dieses Gesetzes der Zustimmung durch den Betriebsrat, was sich nicht auf die grundsätzliche Entscheidung über die Einführung (kein Initiativrecht des Betriebsrats), wohl aber auf die inhaltliche Ausgestaltung erstreckt. In Zweifelsfällen entscheidet eine paritätisch besetzte Einigungsstelle (§ 76 BetrVG). Darüber hinaus hat sich zur grundsätzlichen Zulässigkeit bestimmter Fragenbereiche eine reichhaltige Rechtsprechung entwickelt. Tabelle 2 faßt wesentliche Fragenbereiche mit Beispielen für deren rechtliche Zulässigkeit zusammen.

Tabelle 2:

Fragenbereiche im Personalfragebogen und ihre rechtliche Zulässigkeit (aus Schuler, 1998a, S. 94)

Fragenbereich	zulässig, Beispiel	unzulässig, Beispiel
Allg. identifizierende Merkmale	Name, Anschrift, Geburtsdatum	
Rechtsstatus	Staatsangehörigkeit, Kranken- und Rentenversicherung	
Familie	Familienstand, Geburtsdaten von Ehepartner und Kindern	Heiratsabsichten, intime Beziehungen
Bildungsweg	Ausbildung, Zeugnisnoten, Lehrgänge	
Bisherige Berufstätigkeit	Arbeitsplätze, Arbeitgeber, Kontinuität	
Fachliche Qualifikation	Kenntnisse, Fertigkeiten, Erfahrung	
Einkommen	letztes Einkommen	falls ohne Bezug zur Eignung
Vermögensverhältnisse	bei leitenden Angestellten und in Vertrauensstellungen	falls ohne Bezug zur Position
Wettbewerbsverbot	diesbezgl. besteht sogar Offenbarungspflicht des Bewerbers	
Urlaubsanspruch	gewährter oder abgegoltener Urlaub	
Wehr-/Zivildienst	zulässig	
Vorstrafen	bei Einschlägigkeit; z. B. Unterschlagung bei Kassierer	bei mangelnder Einschlägigkeit
Religions- und Parteizugehörigkeit	in Tendenzbetrieben, z. B. konfess. Einrichtungen, pol. Parteien	grundsätzlich unzulässig
Gewerkschaftszugehörigkeit	in Tendenzbetrieben, etwa Arbeitgeberverbänden; bei Interessenkollision (Itd. Angestellte); wegen Tarifbindung; bei betrieblichem Beitragseinzug	grundsätzlich unzulässig
Schwangerschaft	in Sonderfällen (z. B. Infektionsgefahr, Tänzerin)	nach europäischem Recht grundsätzlich unzulässig
Gesundheitszustand	bei Bedeutung für die Arbeitstätigkeit und Leistungsfähigkeit; gefährliche Infektionskrankheiten (z. B. Tbc, Aids)	allgemein gehaltene Fragen über frühere und derzeitige Krankheiten
Schwerbehinderung	zulässig	

Abschließend zu diesem Abschnitt bliebe zu klären, welche Kriteriumsvalidität die Bewerbungsunterlagen *insgesamt* für den Berufserfolg besitzen. Erstaunlicherweise scheint diese Frage von der personalpsychologischen Forschung bislang weitgehend ignoriert worden zu sein; einschlägige Arbeiten beziehen sich praktisch immer auf einzelne Elemente. Wie wir gesehen haben, sind dabei insbesondere solche Komponenten zu beachten, die über die Vergangenheit des Bewerbers informieren. Deren zum Teil durchaus respektable Einzelvalidität ließe erwarten, daß eine sinnvoll gewichtete Kombination einzelner Elemente einen substantiellen Beitrag zur Erfolgsprognose leisten könnte. Was einer solchen Integration relevanter Informationen entgegensteht, ist – wie erwähnt – die intuitive Erhebung und Auswertung biographischer Daten. Im nächsten Abschnitt werden wir uns daher ausführlich einem Auswahlverfahren zuwenden, mit dessen Hilfe biographische Informationen mit den diesbezüglichen Vorzügen vollstandardisierter Tests kombiniert werden können.

3 Biographische Fragebogen

3.1 Grundlagen

Die Idee, biographische Daten von Stellenbewerbern zur Prognose beruflichen Erfolgs zu verwenden, läßt sich bis ins letzte Jahrhundert zurückverfolgen (vgl. zur Geschichte biographischer Fragebogen Stehle, 1990). Ein zunächst dominierendes Anwendungsfeld war die Auswahl von Versicherungsvertretern; spätestens seit der Mitte des 20. Jahrhunderts erfolgte jedoch eine Ausweitung auf die Bereiche Militär und Industrie einschließlich der Diagnose des Führungspotentials, so daß man heute von einer universellen Anwendung des biographischen Fragebogens sprechen kann.

Sehr früh wurde in diesem Zusammenhang bereits das methodische Prinzip angewandt, einzelne Items anhand ihrer empirisch ermittelten Validität zur Prognose eines Außenkriteriums auszuwählen und zu gewichten. Wir werden uns damit im nächsten Abschnitt etwas eingehender beschäftigen. Eine Folge dieser streng (manche sagen „blind“) empirizistischen Vorgehensweise war, daß sich die Methode „biographischer Fragebogen“ von dem Grundaxiom des biographischen Ansatzes, daß nämlich der beste Prädiktor zukünftigen Verhaltens das Verhalten in der Vergangenheit sei, teilweise löste. Items, die inhaltlich eher einer bewertenden Selbsteinschätzung entsprachen, wie sie für Persönlichkeitstests typisch ist, wurden in die Fragebogen aufgenommen, sofern sie nur empirische Zusammenhänge mit dem Kriterium zeigten. Dies führte wiederum dazu, daß bis heute der Begriff „biographischer Fragebogen“ vielfach weniger mit dem biographischen Grundkonzept verbunden wird als vielmehr mit dem rein technischen Prinzip der empirischen Itemselektion. Die Kriteriumsvalidität solchermaßen konstruierter Instrumente vermag durchaus zu befriedigen, *was* damit aber gemessen wird, bleibt oftmals weitgehend im Dunkeln.

Es blieb nicht aus, daß die ungeklärte Konstruktvalidität der meisten frühen biographischen Fragebogen eine Gegenbewegung zum strikten Empirizismus auf den Plan rief (vgl. Stokes, 1994). Konzeptionelle Pionierarbeiten auf diesem Gebiet sind insbesondere der Forschergruppe um William S. Owens zu verdanken, deren entwicklungspsychologisch orientierter Ansatz gleichfalls im folgenden Abschnitt näher dargestellt wird. In den letzten Jahren zeichnet sich immer deutlicher eine Rückbesinnung auf das ursprüngliche Konzept des biographischen Ansatzes und, damit verbunden, eine Revitalisierung der Forschungsbemühungen ab. Die Methode „biographischer Fragebogen“ beginnt sich wieder von der Kennzeichnung durch das rein technische Itemselektionsprinzip zu emanzipieren, die durch eine inhaltliche Definition dessen verdrängt wird, was ein biographisches *Item* als solches qualifiziert und von anderen Items unterscheidet.

Nach anfänglich eingeschränkter Anwendung biographischer Daten auf die Auswahl von Versicherungsvertretern werden biographische Fragebogen heute universell eingesetzt.

Biographischer Ansatz und das technische Prinzip empirischer Itemselektion werden beim biographischen Fragebogen oft vermengt.

**Zehn mögliche Attribute
kennzeichnen biographische
Items.**

Ausgangspunkt dieser Überlegungen ist, daß die Konstruktion eines Fragebogens gewöhnlich nicht mit der Skalenbildung beginnt, sondern mit der Formulierung einzelner Fragen und Antwortalternativen, die die Bausteine der Skala bilden. Es liegt nahe, mit der inhaltlichen Definition auf dieser basalen Ebene anzusetzen. Eine wichtige konzeptionelle Arbeit auf diesem Gebiet stammt von Mael (1991), der sich mit der Systematisierung einiger schon seit längerem diskutierter Itemcharakteristika beschäftigt hat. Mael unterscheidet zehn *mögliche* Attribute zur *Kennzeichnung biographischer Items* (siehe Tabelle 3).

Tabelle 3:

Attribute zur Kennzeichnung biographischer Items (Mael, 1991)

1. historisch (vergangenheitsbezogen)	vs. hypothetisch (zukunftsbezogen; Absichten oder Erwartungen)
2. external (äußere, beobachtbare Ereignisse)	vs. internal (innere Vorgänge wie Gedanken, Gefühle; nicht beobachtbar)
3. objektiv (faktisch; nicht wertend)	vs. subjektiv (interpretativ; bewertend)
4. aus erster Hand (eigene Beobachtungen der Befragten)	vs. aus zweiter Hand (Vermutungen darüber, was andere beobachtet haben)
5. diskret (einzelne, abgrenzbare Ereignisse)	vs. summarisch (Zusammenfassungen mehrerer Ereignisse)
6. verifizierbar (durch Zeugen, Dokumente o.ä. prinzipiell nachprüfbar)	vs. nicht verifizierbar (Aussage des Befragten ist die einzig mögliche Quelle)
7. beeinflussbar (Verhalten oder Ereignis steht unter der Kontrolle des Befragten)	vs. nicht beeinflussbar (vom Befragten nicht kontrollierbares Ereignis wie bspw. Geschwisterzahl)
8. allgemein zugänglich (jeder Befragte könnte Verhalten gezeigt haben)	vs. ungleich zugänglich (Zugänglichkeit hängt von äußeren Faktoren ab, wie z.B. ausgeübte Sportarten von den Angeboten vor Ort)
9. offensichtlich berufsrelevant (auch für Laien augenscheinlicher Berufsbezug)	vs. nicht offensichtlich berufsrelevant (Bezug zur Tätigkeit nur indirekt nachvollziehbar)
10. non-invasiv (die Privatsphäre respektierend)	vs. invasiv (die Privatsphäre verletzend)

Einige dieser Attribute sind wünschenswert, weil sie durch einen gegenüber Persönlichkeitsitems verringerten Bewertungsspielraum geeignet sind, die psychometrische Qualität zu verbessern (*external*, *objektiv*, *aus erster Hand*); allerdings fanden Lefkowitz, Gebbia, Balsam & Dunn, 1999, eine höhere kriterienbezogene Validität für Items *aus zweiter Hand*), leichter zu erinnern (*diskret*) oder weniger täuschungsanfällig (*verifizierbar*) sind. Eine zweite Gruppe von Attributen dient der Verbesserung von Fairneß (*beeinflussbar*,

allgemein zugänglich) und/oder rechtlicher Zulässigkeit (*offensichtlich berufsrelevant, non-invasiv*). In allen diesen Fällen sind aber auch mögliche Nachteile zu beachten. Die Kernaussage von Maels Arbeit besteht jedoch darin, daß nur ein *einziges* Attribut zur definitorischen Abgrenzung unerlässlich ist: *Biographische Items sind historisch*. Sie beziehen sich auf konkrete Ereignisse in der Vergangenheit. Verhaltensabsichten oder hypothetische Fragen (Was würden Sie tun, wenn...) sind *nicht* biographisch – in den meisten Fällen, aber nicht zwingend, sollten zusätzlich die Attribute 2 bis 4 vorliegen. Mit dieser denkbar einfachen Definition läßt sich auch eine Abgrenzung von anderen Fragetypen nachvollziehen, wie sie bspw. von Knoblauch (1990, siehe Kasten 1) exemplarisch herausgearbeitet wurde.

Kasten 1:

Biographisches Item im Vergleich zu anderen Fragetypen (aus Knoblauch, 1990)

Fragetypen
<p><i>Biographisches Item</i> Wie viele Bücher über Antibiotika haben Sie in den letzten zwei Jahren gelesen?</p>
<p><i>Persönlichkeitsitem</i> Würden Sie lieber allein neue Antibiotika erforschen oder lieber beim Vertrieb dieser Präparate mitarbeiten?</p>
<p><i>Interessenitem</i> Interessieren Sie sich für die Bedeutung und Wirkungsweise von Antibiotika?</p>
<p><i>Meinungs-/Bewertungsitem</i> Antibiotika werden häufig unnötigerweise verschrieben.</p>
<p><i>Wissensitem</i> Welches Antibiotikum hat Sir Alexander Fleming entdeckt?</p>

Maels taxonomischer Ansatz und andere theoretische Arbeiten haben die jüngere Entwicklung des biographischen Fragebogens, zumindest soweit sie sich in anspruchsvolleren Publikationen niederschlägt, erheblich beeinflußt. Da er selbst allerdings noch relativ neu ist, wird in den folgenden Abschnitten auch von Instrumenten zu sprechen sein, deren Items nicht immer dem Kriterium der Historizität genügen. Sind die Items erst einmal generiert, bestehen die nächsten Schritte i.d.R. darin, aus dem vorliegenden Itempool eine Auswahl zu treffen und diese Fragen zu einer oder mehreren Skalen zusammenzufassen, in deren Gesamtwert die Items mit einheitlichem oder unterschiedlichem Gewicht eingehen. Ob die Phasen der Itemselektion, -gewichtung und Skalenbildung voneinander abgrenzbar sind und in welcher Reihenfolge sie stattfinden, hängt u.a. von der methodischen Vorgehensweise ab. Mit den sehr verschiedenen, zum Teil sogar gegensätzlichen Ansätzen hierzu beschäftigt sich der folgende Abschnitt.

3.2 Itemselektion und Skalenbildung

3.2.1 Der empirische Ansatz

Aus der Zeit des ersten Weltkriegs datieren frühe Ansätze, die intuitive Auswertung von Personalfragebogen durch eine streng empirische Systematik zu ersetzen, deren Überlegenheit erstmals bei Goldsmith (1922) dokumentiert ist. Dabei wird jedes einzelne Item daraufhin untersucht, ob es statistisch bedeutsame Zusammenhänge zu einem Erfolgskriterium aufweist. Häufig, aber nicht zwingend, besteht die Methode darin, zu untersuchen, ob zwei oder mehr zuvor festgelegte Gruppen, die sich hinsichtlich ihres beruflichen Erfolgs unterscheiden, auch in der Beantwortung des Items systemati-

Vergangenheitsbezug ist definitorisches Merkmal biographischer Items.

Biographische Items lassen sich von anderen Fragetypen abgrenzen.

Kontrastgruppenmethode

Die vertikale Prozentmethode der Itemgewichtung hat sich bei Einsatz von Kontrastgruppen als Verfahren bewährt.

sche Unterschiede aufweisen. Ist dies empirisch nachweisbar, so geht das betreffende Item in den endgültigen Testwert ein, andernfalls wird es aus der Skala entfernt bzw. im Itemschlüssel nicht berücksichtigt. Diese als *Kontrastgruppenmethode* bekannt gewordene Vorgehensweise ist dabei vollkommen unabhängig vom *Inhalt* des Items anwendbar und keineswegs auf biographische Fragebogen oder auch nur die Berufseignungsdiagnostik beschränkt (ein prominentes Beispiel aus der klinischen Psychologie sind die Skalen des Minnesota Multiphasic Personality Inventory [MMPI]). Die Methode wurde jedoch lange Zeit so universell zur Entwicklung biographischer Fragebogen verwendet, daß sie schließlich vielfach mit diesem gleichgesetzt wurde mit der Folge, daß viele derart entstandene Fragebogen sich aus Items zusammensetzen, die inhaltlich eindeutig Persönlichkeits- oder Interessentests zuzuordnen sind.

Es existieren zahlreiche Varianten des empirischen Ansatzes, die teilweise anstelle von Kontrastgruppen auch kontinuierlich skalierte Kriterien und statt der üblichen Multiple-Choice-Items intervallskalierte Prädiktoren verwenden (für einen Überblick siehe Hogan, 1994). In diesem Fall kann bspw. die multiple Regression bzw. Einzelkorrelation zwischen Item und Kriterium zur Selektion und Gewichtung der Items herangezogen werden. Das bei Einsatz von Kontrastgruppen verbreitetste Einzelverfahren ist die *vertikale Prozentmethode* (England, 1971, nach ihrem Urheber auch Englandverfahren genannt), die in Kasten 2 dargestellt wird. Vergleichsstudien zwischen verschiedenen Methoden (z. B. Devlin, Abrahams & Edwards, 1992) zeigen dabei, daß elaborierte Verfahren mit differentieller Gewichtung der Items bzw. Antwortalternativen relativ einfachen Vorgehensweisen wie dem Englandverfahren keineswegs überlegen sind. Grundsätzlich wird bei allen empirischen Methoden ähnlich vorgegangen, und auch die Probleme unterscheiden sich nur graduell.

Kasten 2:

Vertikale Prozentmethode der Itemgewichtung (England, 1971)

Im folgenden Beispiel (zur Übung sei empfohlen, das Item danach einzuschätzen, welche Attribute aus der Taxonomie von Mael darauf zutreffen) wird gemäß der Englandmethode jede *Antwortalternative* separat gewichtet, obwohl das Responseformat zumindest Ordinalskalenniveau erreicht. Für jede Alternative wird zunächst der Prozentsatz der jeweiligen Kontrastgruppe berechnet, der die Option angekreuzt hat. Aus den beiden Prozentwerten wird anschließend die Differenz gebildet. Diesen Differenzen werden dann die von Strong tabellierten (abgedruckt bei England, 1971) positiven oder negativen Nettogewichte zugeordnet, die eine Aussage über die *Richtung* und das *Ausmaß* treffen, in dem die Antwortalternative zwischen den beiden Gruppen differenziert. England empfiehlt nun weiter, auf die letzte Information (Ausmaß) zu verzichten, um den Itemschlüssel zu vereinfachen und nicht zu stark an Zufallseinflüsse bei der gegebenen Stichprobe anzupassen. Daher werden nur bedeutsame Nettogewichte berücksichtigt und in einheitliche Grobgewichte (+1: -1) konvertiert; die übrigen Alternativen werden, sofern ihre Eliminierung aus Plausibilitätsgründen nicht opportun erscheint, mit Null gewichtet. Der Gesamtscore ergibt sich dann durch einfache Addition der Gewichte der angekreuzten Optionen.

Wie wichtig war	<input type="radio"/>	sehr großer Einfluß
Unabhängigkeit	<input type="radio"/>	großer Einfluß
als Grundlage für	<input type="radio"/>	gewisser Einfluß
Ihre Berufswahl?	<input type="radio"/>	geringer Einfluß
	<input type="radio"/>	gar kein Einfluß

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

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

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

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

Der technische Grundansatz aller empirischen Methoden besteht darin, daß jedes Item, in der Regel – bspw. bei kategorialen Responseformaten (Multiple Choice) oder bei Verfahren wie der Englandmethode – sogar jede Antwortalternative, als *Einzelprädiktor* des Kriteriums angesehen wird. Dies entspricht im Prinzip, läßt man das Skalenniveau dabei außer acht, einem *regressionsanalytischen Ansatz*³. In einer Itemanalysestichprobe wird dann zunächst für jeden Prädiktor anhand der Prognosegüte ein empirisches Gewicht ermittelt, das im einfachsten Fall (siehe Kasten 2) positiv (+1), negativ (-1) oder neutral (0) ist. Diese empirische Itemgewichtung führt in der *dafür eingesetzten Stichprobe* zwangsläufig zu einer maximalen Kriteriumsvalidität; Werte um $r = .80$ sind hier nicht selten. Leider fangen die Probleme danach erst an. Die „blind“ empirizistische Vorgehensweise führt dazu, daß der so entstandene Itemschlüssel vollständig auf die Besonderheiten der Stichprobe zugeschnitten ist mit der Folge, daß die Gewichtung kaum auf einen anderen Kontext übertragbar ist (*Situationsspezifität*) und die *Höhe* der Korrelation eine absolut unrealistische Überschätzung darstellt; sie erfordert eine unabhängige Überprüfung. Idealerweise sollte dies durch eine Kreuzvalidierung an einer zweiten, unabhängigen Stichprobe erfolgen; üblich ist jedoch die Aufspaltung des Itemanalysesamples in zwei Teile, wobei für eine Teilstichprobe die Itemgewichte ermittelt werden, deren Validität anschließend an den „zurückbehaltenen“ Versuchspersonen überprüft wird (und vice versa: sog. double-cross validation). Auf diese Weise läßt sich allerdings nur der durch den Stichprobenumfang bestimmte Zufallsfehler eliminieren, nicht aber systematische Besonderheiten der Stichprobe insgesamt. Eine Alternative zur Kreuzvalidierung stellt die Anwendung von Korrekturformeln (Kennedy, 1988) dar, die allerdings beim empirischen Ansatz zu einer Validitätsüberschätzung führen (Mitchell & Klimoski, 1986). Das Problem der Spezifität oder Überanpassung stellt sich im übrigen nicht nur hinsichtlich der Stichprobe, sondern auch bezüglich des gewählten *Kriteriums*, was gegebenenfalls eine *Kreuzvalidierung* an alternativen Kriterien erforderlich machen kann.

In jedem Fall erfordert der empirische Ansatz sehr *große Stichproben* zur Ermittlung eines stabilen Itemschlüssels. Zu kleine Stichproben führen zum Zusammenbruch der Validität in der Kreuzvalidierung, wobei der erforderliche Umfang zusätzlich von der Anzahl der Items oder Antwortalternativen abhängt. Je mehr Items in der Gewichtungsstichprobe „ausprobiert“ werden, desto mehr erscheinen allein aufgrund des Zufallsfehlers empirisch valide (*capitalization on chance*). Differentielle Gewichte (z. B. aus der multiplen Regression) sind hierfür anfälliger als Einheitsgewichte (Englandverfahren).

³ Dem mit der Technik der multiplen Regression vertrauten Leser werden die augenfälligen Parallelen in der folgenden Diskussion nicht entgehen.

Beim empirischen Ansatz werden Items einzeln aufgrund ihrer Kriteriumsvalidität ausgesucht und ggf. gewichtet.

Kreuzvalidierungen oder äquivalente Methoden sind beim empirischen Vorgehen unerlässlich.

Empirisch skalierte biographische Fragebogen messen i.d.R. keine homogenen Konstrukte.

Das empirische Vorgehen gewährleistet eine gewisse Validität.

Hogan (1994) empfiehlt, daß die Stichprobengröße bei differentieller Gewichtung mindestens dem Fünf- bis Zehnfachen der Item- bzw. Antwortalternativenzahl entsprechen sollte. Das bedeutet, daß für einen Multiple-Choice-Fragebogen mit 50 Items und jeweils vier Antwortalternativen bereits eine vierstellige Stichprobengröße nur zur Ermittlung der Gewichte (ohne Kreuzvalidierung) erforderlich wäre.

Ein weiterer Nachteil erwächst aus der „blinden“ Vorgehensweise beim empirischen Ansatz: die derart entstandenen Skalen sind inhaltlich oft kaum zu interpretieren, was keineswegs nur den Theoretiker stört, sondern u.a. auch zu ernstesten Akzeptanzproblemen und rechtlicher Angreifbarkeit führen kann. Besonders die Gewichtung einzelner Antwortalternativen führt nicht selten zu kontraintuitiven Befunden, wenn sich bspw., wie exemplarisch bei dem Item in Kasten 2, bei einem kontinuierlichen Merkmal nichtlineare Zusammenhänge zum Kriterium ergeben. Solche schwer erklärbaren Zusammenhänge erweisen sich häufig auch als besonders instabil. Auf der Ebene der Gesamtskala führt außerdem die Orientierung am regressionsanalytischen Ansatz dazu, daß, im krassen Gegensatz zur klassischen Testtheorie, zur Vermeidung von Multikollinearitätsproblemen (vgl. z. B. Bortz, 1993) möglichst *geringe* Korrelationen zwischen den einzelnen Items gefordert sind. Dies wiederum äußert sich in einer äußerst geringen, manchmal negativen internen Konsistenz (Homogenität) der Skalen, was die Etablierung von Konstruktvalidität für biographische Fragebogen zu einem schwierigen Unterfangen macht.

Der Hauptvorteil des empirischen Ansatzes liegt in der nicht zu bezweifelnden Kriteriumsvalidität der so entstandenen Instrumente. Die mittlere Kreuzvalidität für eine Vielzahl von Kriterien liegt üblicherweise zwischen $r = .20$ und $.50$ (Zusammenfassungen verschiedener Metaanalysen finden sich bei Bliesener, 1995, sowie Schmidt & Rothstein, 1994). Darüber hinaus konnten Rothstein, Schmidt, Erwin, Owens und Sparks (1990) in einer originellen Anwendung metaanalytischer Techniken zeigen, daß eine Vorauswahl der Items aufgrund von Anforderungsanalysen in Verbindung mit extremen Stichprobengrößen ($N = 11.000$) dazu führt, daß empirische Itemschlüssel und die damit erzielten Validitäten auf unterschiedliche Stichproben *und* Kriterien übertragbar (generalisierbar) sind. Die dort zitierten Itembeispiele lassen darauf schließen, daß sich dieser Befund eher auf die empirische Methode als auf biographische Inhalte bezieht. Zusammenfassend läßt sich zum empirischen Ansatz konstatieren, daß sich mit dieser Methode, genügend große Stichproben vorausgesetzt, hochvalide Instrumente für spezifische Stichproben konstruieren lassen. Das Problem geringer Generalisierbarkeit kann durch die Verwendung extremer Stichprobenumfänge reduziert werden, wofür die Voraussetzungen in der Praxis allerdings selten gegeben sein dürften. Die geringe Inhalts- und Konstruktvalidität liegt allerdings in der Methode begründet und läßt sich auch durch Massenuntersuchungen kaum verbessern.

3.2.2 Der rationale Ansatz und Mischformen

Die unbefriedigende Interpretierbarkeit itemweise validierter Skalen führte insbesondere in den letzten Jahren zu einer Gegenposition, die sich als konstruktorientierte Eignungsdiagnostik mit den Mitteln des biographischen Fragebogens beschreiben läßt. In reinsten Form wird sie durch die *deduktive Methode* der Fragebogenkonstruktion repräsentiert. Hier sind, theoriegeleitet und/oder auf der Grundlage von Anforderungsanalysen, *a priori* Konstrukte zu definieren, die annahmegemäß mit beruflichem Erfolg zusammenhängen. Solche Hypothesen setzen selbstverständlich das Vorhandensein einer Theorie, zumindest aber die genaue Kenntnis der jeweiligen Auswahl-situation voraus. Zur Messung dieser Konstrukte werden anschließend Items formuliert, die, um als biographisch zu gelten, zumindest dem Kriterium der Historizität genügen sollten. Die Festlegung des Itemschlüssels und die Zusammenfassung zu Skalen erfolgt gleichfalls *vor* Erhebung empirischer Da-

ten. Im Gegensatz zum empirischen Ansatz liegt der Schwerpunkt hier also nicht auf der Berechnung von Itemgewichten, sondern auf der *Itemgenerierung* (eine relativ elaborierte Methode hierzu wird bei Mumford, Costanza, Connelly & Johnson, 1996, beschrieben). Ziel ist zunächst eine Maximierung der Inhaltsvalidität und Interpretierbarkeit der Skalen. Neben diesen qualitativen Maßstäben erfolgt, dies selbstverständlich auf empirischem Weg, regelmäßig eine Überprüfung der testtheoretischen Gütekriterien Reliabilität und Validität. Besonders wichtig ist dabei auch die Berücksichtigung der Konstruktvalidität, um dem Hauptvorwurf gegenüber der deduktiven Methode zu begegnen, sie verlasse sich zu sehr auf ungeprüfte Vorannahmen (*Subjektivität*). Da die Skalenbildung a priori erfolgt, entfällt das Problem der Überanpassung vollständig und damit auch die *Notwendigkeit* einer Kreuzvalidierung – obwohl unabhängige Replikationen natürlich immer wünschenswert sind.

Eine nicht ganz so radikal antiempirizistische, aber gleichwohl streng konstruktorientierte Variante des rationalen Ansatzes ist die *induktive Methode*. Der Schwerpunkt liegt hier auf der Analyse der *internen Struktur* eines Fragebogens mit Hilfe der Faktorenanalyse. Ziel ist es, innerhalb des Itempools homogene (intern konsistente) Konstrukte bzw. latente Variablen (Faktoren) zu identifizieren, die eindeutig interpretierbar sein und gleichzeitig einen möglichst großen Teil der gesamten im Fragebogen enthaltenen Information (Varianz) auf eine möglichst sparsame Struktur (geringe Faktorenzahl) reduzieren sollten. Die Skalierung erfolgt anschließend durch die Zusammenfassung von Items, die die zuvor identifizierten Faktoren möglichst gut repräsentieren. Die Werte einzelner Personen auf diesen Faktoren werden schließlich zu Validitätskriterien in Beziehung gesetzt.

Induktiv konstruierte Fragebogen stellen zweifellos, gemessen an der Interpretierbarkeit und Konstruktvalidität, gegenüber empirischen Itemvalidierungen einen erheblichen Fortschritt dar und können gleichzeitig im Vergleich zu deduktiv gebildeten Skalen für sich den Vorzug geringerer Subjektivität in Anspruch nehmen. Dennoch: Während faktorenanalytische Persönlichkeitstests derzeit eine bemerkenswerte Renaissance erleben, gilt dies für biographische Fragebogen in weit geringerem Maße. Der Grund liegt in der Heterogenität biographischer Daten. Faktorenanalysen biographischer Itempools fanden i.d.R. eine große Zahl einzelner Komponenten (etwa 10 bis 30 Faktoren; Hough & Paullin, 1994), die zusammen jedoch nur einen geringen Prozentsatz der gesamten Varianz aufklärten (zwischen 20 und 50 %). Beides ist im Sinne des induktiven Ansatzes eher unerwünscht. Große Faktorenzahlen bedeuten eine nur mäßige Datenreduktion, und ein hoher Anteil nicht erklärter Varianz führt zum Verlust an Information. Inhaltlich fanden sich zudem viele Faktoren nur in einzelnen Untersuchungen. Zu den am häufigsten replizierten (generalisierbarsten) Komponenten zählen Persönlichkeitskonstrukte wie emotionale Stabilität und Extraversion, Indikatoren für Interessen wie kulturelles oder sportliches Engagement, aber auch Leistungsindikatoren wie schulischer Erfolg oder Karriereentwicklung (Mumford & Owens, 1987). Es ist vielleicht der wichtigste Beitrag der faktorenanalytischen Schule, einen Einblick in die Vielgestaltigkeit biographischer Daten vermittelt zu haben.

In jüngster Zeit sind einige Ansätze zu beobachten, die versuchen, einen Mittelweg zwischen den rein empirischen und rationalen Vorgehensweisen zu beschreiten. Zwei solcher Vorschläge wurden von Mael & Hirsch (1993) untersucht. Bei der *quasi-rationalen Methode* handelt es sich um den Versuch, die Messung von Persönlichkeitskonstrukten mit biographischen Daten nachzubilden. Die Verbindung zum empirischen Ansatz besteht hier darin, daß die Items gleichfalls auf der Grundlage empirischer Zusammenhänge ausgewählt werden, wobei als Kriterium jedoch nicht der berufliche Erfolg dient, sondern die Korrelation mit entsprechenden Persönlichkeitsskalen. Die zweite Methode steht dem rein empirischen Vorgehen näher, wurde von Mael (1991) jedoch – in Abhebung vom blinden „dustbowl empiricism“ – als „rainforest empiricism“ bezeichnet. Hier wird die empirische Itemselektion um eine theoretische Begründung zur *Auswahl* der Items ergänzt, wobei zu-

Biographische Fragebogen lassen sich auch rational (theoriegeleitet) konstruieren, was eine spätere Validierung jedoch nicht erspart.

Faktorenanalytische Ansätze waren für biographische Fragebogen bislang weniger erfolgreich.

Neuerdings wurden auch rational-empirische Mischformen vorgeschlagen.

Subgrouping ist ein entwicklungspsychologisch fundierter Ansatz zur Unterscheidung von Personengruppen.

sätzlich Richtlinien existieren, die die Aufnahme kontraintuitiver Gewichtungen in den Itemschlüssel ausschließen. Ziel der quasi-rationalen Methode ist die Objektivierung und Verminderung der Verfälschbarkeit klassischer Persönlichkeitsmessung, während der „rainforest empiricism“ zu einer besseren Interpretierbarkeit der Skalen gegenüber dem blind empirizistischen Vorgehen führen soll. Mael und Hirsch (1993) konnten in einer großen Stichprobe die Brauchbarkeit beider Methoden auch zur Prognose eines Leistungskriteriums (Führungserfolg) demonstrieren.

3.2.3 Ein typenbildender Ansatz: Subgrouping

Wir haben mit der faktorenanalytischen Methode bereits einen Ansatz zur Strukturierung biographischer Daten vorgestellt, zu dem die Arbeitsgruppe um William Owens wesentliche Beiträge geliefert hat. Aus der gleichen Gruppe (Owens, 1976; Owens & Schoenfeldt, 1979) stammen Arbeiten, die sich durch ihren theoretischen Ansatz und das statistische Werkzeug zur Datenstrukturierung von den bisher vorgestellten Methoden unterscheiden. Der theoretische Ansatz beim *Subgrouping* ist entwicklungspsychologisch; es geht um die Identifizierung stabiler Muster von Lebensereignissen, die sich zur generalisierbaren Beschreibung typischer Lebensgeschichten eignen. Diese in objektiven Ereignissen ausgedrückten Lebensgeschichten werden als Persönlichkeitstypen interpretiert. Die Subgrouping-Methode stellt den bislang ambitioniertesten Versuch dar, den biographischen Ansatz als eigenständiges theoretisches Konzept in der Eignungsdiagnostik zu etablieren (siehe Mumford, Reiter-Palmon & Snell, 1994, für eine ausführliche Beschreibung des theoretischen Hintergrunds).

Das statistische Werkzeug hierzu ist die Clusteranalyse, bei der *nicht* Items oder Variablen nach ihrer Ähnlichkeit zusammengefaßt, sondern *Personen*, die sich durch eine ähnliche Lebensgeschichte auszeichnen, zu Typen oder „Klumpen“ aggregiert werden. Statistisch stabile und theoretisch bedeutsame clusteranalytische Lösungen setzen große Stichproben und einen umfassenden Itempool voraus, der keine wesentlichen Lebensbereiche ausspart. Owens und Schoenfeldt (1979) fanden unter Verwendung von Persönlichkeitsskalen und biographischen Daten eine Lösung mit 23 männlichen und 15 weiblichen Subgruppen, die sich in späteren Untersuchungen als weitgehend replizierbar erwies (Mumford et al., 1994). In Tabelle 4 sind beispielhaft einige dieser Gruppen beschrieben.

Tabelle 4:

Biographische Subgruppen (nach Owens & Schoenfeldt, 1979)

Beispiele für weibliche Subgruppen
<p><i>materialistisch-autoritäre Extravertierte</i> (beschrieben u.a. durch: engere Beziehung zur Mutter als zum Vater; künstlerisch desinteressiert; religiöse und ökonomische Werte; Wirtschaftsstudium; hohe Extraversion; geringe Kreativität; schlechte Schulnoten)</p> <p><i>unkonventionelle Hochbegabte</i> (beschrieben u.a. durch: kulturelle und wissenschaftliche Interessen; gute Zeugnisnoten; religiöser Nonkonformismus; kognitive Werte; Ausdruck negativer Emotionen)</p> <p><i>unpopuläre, einfältige Introvertierte</i> (beschrieben u.a. durch: sozial unreif; unbeliebt beim anderen Geschlecht; musisch interessiert; geringe kognitive Komplexität; durchschnittliche Noten)</p>
Beispiele für männliche Subgruppen
<p><i>traditionelle, wissenschaftlich orientierte Führer</i> (beschrieben u.a. durch: gute Noten; naturwissenschaftliche Interessen;</p>

religiöser Konformismus; Engagement in Schüler- und Studentenvertretungen; sozial eher introvertiert)

unabhängige Analytiker

(beschrieben u.a. durch: wissenschaftliche Interessen; antiautoritäre Erziehung; langfristige Zielsetzung; kognitiv komplex; Drogenerfahrung)

„Bauernjungen“

(beschrieben u.a. durch: geringer sozioökonomischer Status; geringer Radikalismus; landwirtschaftliches Studieninteresse; kurzfristige Zielorientierung; Schulwechsler)

Aus eignungsdiagnostischer Perspektive unterscheidet sich die Subgruppenmethode erheblich von den zuvor beschriebenen Skalierungsverfahren. Das individuelle Testergebnis ist kein Wert auf einer kontinuierlichen Skala (Score), sondern die Einordnung des Probanden in eine festgelegte Typologie anhand der Ähnlichkeit zu einer Referenzgruppe. Diese Information ist eignungsdiagnostisch wertlos, solange keine Daten zur differentiellen Eignung dieser Gruppe für unterschiedliche Positionen vorliegen. Sind solche Daten allerdings erhoben worden oder können zumindest durch anforderungsanalytisch begründete Hypothesen angenähert werden, können mit einem typologischen Verfahren mehrere Kriterien simultan berücksichtigt werden (Hein & Wesley, 1994), so daß sich die Subgruppenmethode nicht nur für Auswahl-, sondern auch für Plazierungsentscheidungen (Auswahl zwischen mehreren Positionen) und Personalentwicklungsmaßnahmen sowie zur Berufsberatung eignet.

Das Subgrouping-Konzept ist eher theoretisch orientiert. Praktische Anwendungen erfordern Zusatzanalysen.

3.3 Zusammenfassung und Bewertung biographischer Fragebogen

In Tabelle 5 sind noch einmal einige kennzeichnende Merkmale der vier wichtigsten Ansätze zur Konstruktion biographischer Fragebogen zusammengestellt.

Vergleich der wichtigsten Konstruktionsprinzipien

Tabelle 5:

Merkmale unterschiedlicher Konstruktionsprinzipien für biographische Fragebogen

	empirisch	deduktiv	induktiv	Subgrouping
konzeptio- nelles Prinzip:	• Selektion und Gewichtung einzelner Items oder Antwortalternativen gemäß ihrer Validität bzgl. eines Erfolgskriteriums	• theoriegeleitete oder anforderungsanalytisch fundierte Itemgenerierung und Skalenbildung vor Erhebung von Validitätsdaten	• Bildung homogener Subskalen auf der Grundlage von Inter-Item-Korrelationen (kein Außenkriterium)	• Identifizierung homogener Gruppen von Personen anhand empirisch vorgefundener Antwortmuster
methodisches Prinzip:	• Regressionsanalyse	• bei Skalenbildung keines (später Überprüfung auf testtheoretische Gütekriterien)	• Faktorenanalyse	• Clusteranalyse
Schwerpunkt der Zielsetzung:	• Kriteriumsvalidität	• Inhaltsvalidität; Interpretierbarkeit	• Konstruktvalidität; Homogenität (interne Konsistenz)	• Typologisierung (Homogenität innerhalb, Heterogenität zwischen den Gruppen)
Testergebnis:	• vollstandardisiert erhobener Testwert	• vollstandardisierte Gesamt- und/oder Teilskalenwerte	• vollstandardisierte Gesamt- und/oder Teilskalenwerte	• standardisierte Zuordnung von Personen zu Gruppen

Fortsetzung Tabelle 5

	empirisch	deduktiv	induktiv	Subgrouping
Vorteile:	<ul style="list-style-type: none"> • im spezifischen Bereich sehr gute Kriteriumsvalidität • geeignet zur Exploration unbekannter Situationen, Positionen 	<ul style="list-style-type: none"> • intuitiv unmittelbar zugängliche Ergebnisse • kein oder geringer Validitätsverlust bei Übertragung auf andere Situationen bzw. Replikation 	<ul style="list-style-type: none"> • erleichtert Interpretation bei Exploration unbekannter Situationen • homogene, abgrenzbare Konstrukte 	<ul style="list-style-type: none"> • bei Kenntnis der Anforderungen und Eignungsprognose für die Subgruppen auch für Platzierungsentscheidungen geeignet • theoretische Fundierung des biographischen Ansatzes möglich
Probleme:	<ul style="list-style-type: none"> • Ergebnisse situativ instabil (auf Stichprobe und Kriterium zugeschnitten) • schlecht interpretierbare Skalen • große Stichproben erforderlich • Kreuzvalidierung zwingend 	<ul style="list-style-type: none"> • subjektive Vorentscheidungen • ungeeignet bei schlecht definierten oder wenig erforschten Situationen • großer konzeptioneller Aufwand 	<ul style="list-style-type: none"> • Heterogenität biographischer Daten führt zu einer Vielzahl von Faktoren • Informationsverluste bei Verwendung von Faktorskalen 	<ul style="list-style-type: none"> • große Stichproben erforderlich • praktische Umsetzung erfordert umfangreiche zusätzliche Forschung

Die Validitätsunterschiede der verschiedenen Konstruktionsprinzipien sind gering.

Die *Kriteriumsvalidität* biographischer Fragebogen wurde bereits im Abschnitt zum empirischen Ansatz angesprochen. Die dort zitierten Ergebnisse beziehen sich auf Überblicksarbeiten, die aus den achtziger Jahren oder früher datieren und zum größten Teil die Validität empirisch skalierten Fragebogen widerspiegeln. In einer jüngeren Metaanalyse (Hough & Paullin, 1994) wurden systematisch empirischer, deduktiver und induktiver Ansatz einander gegenübergestellt⁴. Die Arbeit bezog sich ausschließlich auf Studien, in denen zumindest zwei Ansätze direkt miteinander verglichen wurden. Hough und Paullin fanden eine minimal höhere Validität des deduktiven Ansatzes gegenüber empirischen Skalen (Validitätsunterschied [r_{diff}]: .01) und des empirischen gegenüber dem induktiven Ansatz ($r_{\text{diff}} = .02$). Etwas größer ($r_{\text{diff}} = .05$) war der Vorsprung deduktiver Skalen im Vergleich zu induktiven Faktorskalen. Die Ergebnisse sind insbesondere dann verblüffend, wenn man sich vor Augen führt, daß das Hauptziel der empirischen Methode die Maximierung der Kriteriumsvalidität ist; sie entsprechen übrigens aber vollkommen Burischs (1984) Befunden zu Persönlichkeitstests. Offenbar funktionieren zumindest deduktiv konstruierte Skalen, gemessen an diesem Kriterium, wenigstens genauso gut. Angesichts der angesprochenen Probleme des empirischen Ansatzes spricht somit nur wenig für dessen Anwendung in einigermaßen explorierten Feldern, und tatsächlich findet der reine, „blinde“ Empirizismus in der veröffentlichten Literatur der letzten Jahre kaum noch Unterstützung. Diese Entwicklung läßt sich, mit der gebotenen Vorsicht, durchaus als Paradigmenwechsel in der Erforschung biographischer Daten werten.

Etwas detaillierter mit methodischen und inhaltlichen Moderatoren der Validität biographischer Fragebogen setzt sich eine Metaanalyse von Bliesener (1992; 1995) auseinander. Bliesener fand eine mittlere Validität von $r =$

⁴ Die Qualität der Subgrouping-Methode läßt sich aus konzeptionellen Gründen kaum mit der anderer Verfahren vergleichen und wurde bislang auch wenig erforscht (für Ausnahmen siehe Hein & Wesley, 1994). Sie bleibt daher im folgenden ausgespart.

.30, ein Wert, der etwas unter dem früherer Zusammenfassungen liegt, was teilweise sicher dem Umstand zuzuschreiben ist, daß der Autor keine Korrekturen für Artefakte wie mangelnde Reliabilität oder Streuungseinschränkung vornahm. Die Untersuchung *methodischer* Moderatoren ergab im wesentlichen, daß konservativere Vorgehensweisen bei Konstruktion und Validierung zu niedrigeren Validitätsschätzungen führen. So sind die Korrelationen aus konkurrenten Erhebungen höher als in Längsschnittstudien, spezifisch entwickelte Itemschlüssel produzieren höhere Validitäten als die Verwendung genereller Standardschlüssel, und ähnliches gilt für die Kreuzvalidierung durch Aufspaltung der Stichprobe gegenüber der Verwendung eines zweiten, unabhängigen Sample. Die Ergänzung durch nicht-biographische Items führt zur Erhöhung der Validität. Diese Befunde ließen sich schon aufgrund von Plausibilitätsüberlegungen erwarten; größere Überraschungen ergab die Untersuchung *inhaltlicher* (situativer) Moderatoren. Ohne auf die absolute Höhe der Korrelationen einzugehen⁵, zeigt sich bspw. für weibliche Teilnehmer eine deutlich höhere Validität als bei männlichen Probanden. Noch unerwarteter ist Blieseners Befund, daß die „klassische“ Berufsgruppe in der biographischen Eignungsdiagnostik (Verkäufer) eher am unteren Ende der Validitätsskala steht und daß sich objektive Leistungskriterien besser prognostizieren lassen als Vorgesetztenbeurteilungen.

Abschließend sei noch auf zwei weitere Merkmale zur Beurteilung der Qualität biographischer Fragebogen eingegangen. In den wenigen vergleichenden Untersuchungen zur *Akzeptabilität* verschiedener Auswahlinstrumente schnitten biographische Fragebogen gegenüber anderen Tests in den USA etwas besser (Steiner & Gilliland, 1996), in Frankreich gleich (Steiner & Gilliland, 1996) und in Deutschland etwas schlechter (Marcus, 1999) ab. Alle Tests finden sich jedoch regelmäßig im Mittelfeld platziert, wobei an der Spitze der Akzeptanzskala Interviews und Arbeitsproben rangieren und am Ende, deutlich abgeschlagen, die Graphologie und ähnlich schwer nachvollziehbare Instrumente landen. Die größte Diskrepanz zwischen einzelnen Untersuchungen fand sich für die Einschätzung der Bewerbungsunterlagen, insbesondere des Lebenslaufs. Eine wachsende Zahl von Studien beschäftigt sich inzwischen mit der Tendenz zur *Verfälschung* (faking) biographischer Daten. Dabei zeigte sich bspw., daß die Einhaltung der ersten sechs Kriterien aus der in Abschnitt 3.1 vorgestellten Liste von Mael (1991) zur Reduzierung von Täuschungstendenzen (operationalisiert als Mittelwertschiebung) beiträgt (Becker & Colquitt, 1992). Das gleiche gilt für die explizite *Warnung*, daß die Antworten gegebenenfalls verifiziert bzw. durch eine Lügenskala überprüft werden, allerdings nur, sofern die Items durchschaubar sind (Kluger & Colella, 1993; siehe dort auch zum Effekt auf die *Varianz* von Items und Skalen). Es sei angefügt, daß nicht durchschaubare Fragen und empirisch gewichtete Itemschlüssel schon aufgrund logischer Überlegungen als relativ resistent gegenüber Faking gelten dürfen. Die oben zitierten Untersuchungen zeigen jedoch, daß auch für die in dieser Hinsicht als besonders anfällig geltenden deduktiven Skalen Strategien existieren, die zu einer erheblichen Reduzierung von Täuschungstendenzen beitragen.

Biographische Fragebogen sind die am stärksten formalisierte und wohl auch am intensivsten erforschte Methode zur Erhebung biographischer Daten, deren wohldokumentierte Effektivität in der psychologischen Eignungsdiagnostik den relativ breiten Raum rechtfertigt, den ihre Darstellung in diesem Kapitel einnimmt. Wir werden im folgenden Abschnitt jedoch u.a. sehen, daß sich biographische Inhalte auch im mündlichen Gespräch so erheben lassen, daß dadurch Aussagen von hoher prognostischer Validität ermöglicht werden.

Insgesamt zählen biographische Fragebogen zu den validesten Instrumenten der Eignungsdiagnostik, wobei eine Reihe von Moderatorvariablen zu beachten ist.

Bewerber schätzen biographische Instrumente sehr unterschiedlich, im Mittel eher neutral ein.

Es existieren Methoden, mit denen sich die Verfälschung biographischer Angaben reduzieren läßt.

⁵ Die berichteten Werte sind um den Einfluß methodischer Moderatoren bereinigt, d.h. nach *unten* korrigiert. Wir werden auf diese etwas ungewöhnliche und methodisch nicht unumstrittene Praxis hier nicht näher eingehen.

Das Einstellungsinterview bietet Gelegenheit zum Austausch bewerbungsrelevanter person-, arbeits- und organisationsbezogener Information.

Das Interview kann in verschiedenen Phasen des Auswahlprozesses eingesetzt werden.

4 Das Interview

4.1. Definition und Zwecksetzung

Unter einem Interview als Methode der Personalauswahl ist eine Gesprächssituation zwischen zwei oder mehreren Personen – Repräsentanten der auswählenden Organisation einerseits und Stellenbewerber andererseits – zu verstehen, die Gelegenheit zum Austausch bewerbungsrelevanter person-, arbeits- und organisationsbezogener Information bietet und damit als Grundlage für Auswahlentscheidungen seitens der Organisation und der Organisationswahl seitens der Bewerber dient. Im Regelfall handelt es sich um eine direkte, unreduzierte, also vis-à-vis-Interaktion, aber auch ein Telefongespräch kann den Charakter eines Auswahlgesprächs haben. Demgegenüber kommt einer Befragung in schriftlicher Form, auch wenn sie über elektronische Medien ohne Zeitverzug durchgeführt wird, nicht die Eigenheit eines persönlichen Gesprächs zu, weshalb wir sie hier nicht dem Begriff Interview zurechnen wollen. Wir sehen die Bezeichnungen Interview, Einstellungsinterview, Auswahl- und Vorstellungsgespräch hier als annähernde Synonyme an, wobei gewisse Akzente darin liegen mögen, von einem Interview oder Einstellungsinterview dann zu sprechen, wenn das Verfahren durch eine Systematik gekennzeichnet ist, während von Vorstellungsgespräch eher dann die Rede ist, wenn die Seite des Bewerbers betont wird.

Ebenso wie andere Auswahlverfahren dient das Interview dem Stellenanbieter als Quelle der Information über Fähigkeiten, Fertigkeiten, Kenntnisse, Interessen und andere personbezogene Information über die Kandidaten. Es wird teilweise zur Vorselektion eingesetzt, um überhaupt erst formelle Bewerbungen einzuleiten, als frühe Stufe eines Auswahlprozesses, innerhalb dessen weitere, häufig aufwendigere Schritte folgen, als integrierter Teil eines umfangreicheren Verfahrens (z. B. eines Assessment Centers) oder auch als einzige Methode der Eindrucksbildung und Entscheidungsfindung.

Der Durchführungsmodus von Einstellungsinterviews reicht von der völlig freien Gesprächsform über teilstrukturierte bis zu vollstrukturierten Varianten mit standardisierten Abläufen und Fragestellungen. Die gestellten Fragen beziehen sich insbesondere auf Berufserfahrung und Berufsausbildung, auf Aspekte des Lebenslaufs und deren subjektive Verarbeitung, gelegentlich auch auf persönliche Bereiche wie den des familiären Hintergrunds. In erweiterten Gesprächsformen sind auch arbeitsprobenartige Komponenten wie kleine Rollenspiele oder Präsentationen integriert. Das Interview stellt damit ein Auswahlverfahren dar, das potentiell alle drei in diesem Band vorgestellten Zugangsweisen der Eignungsdiagnostik repräsentiert – den eigenschaftsorientierten, den simulationsorientierten und den biographieorientierten Ansatz. Die Antworten des Bewerbers wie auch weitere Eindrücke aus dem Gesprächsverlauf, beispielsweise nonverbales Verhalten betreffend, werden traditionell zu einem „klinischen“ Urteil – in intuitiver Kombination und Gewichtung –, bei neueren Interviewformen auch nach einem vorgegebenen Algorithmus zusammengefaßt.

4.2 Verbreitung und Akzeptanz

Einstellungsinterviews sind nach der Auswertung der Bewerbungsunterlagen die verbreitetste Methode der Personalauswahl; dies gilt für deutsche Unternehmen ebenso wie für Firmen in anderen europäischen Ländern (Schuler et al., 1993). Das Gespräch als eignungsdiagnostische Methode genießt hohe Wertschätzung nicht nur seitens der Verwender, sondern auch von seiten der Bewerber. Beide sehen das Gespräch als taugliche – also valide – Methode an, sich ein Bild vom Gegenüber zu machen und relevante Information als Entscheidungsgrundlage zu gewinnen. Beide schätzen an der Gesprächssituation die Möglichkeit des direkten persönlichen Austauschs und halten die dabei zu gewinnenden Hinweise für glaubhafter als die aus schrift-

lich übermittelter Information. Interviewer nutzen darüber hinaus Auswahlgespräche als Gelegenheit zur Öffentlichkeitsarbeit (Anderson & Shackleton, 1993), Bewerber schätzen die Möglichkeiten der Situationskontrolle höher ein als bei anderen Auswahlverfahren und nutzen – soweit das Interview als transparentes Verfahren gedacht ist und über die Anforderungen informiert – die Situation als Gelegenheit, Feedback zu gewinnen, sowie als Grundlage ihrer Selbstselektion. Schließlich bietet das Interview als einziges Auswahlverfahren den Rahmen, Vereinbarungen über den weiteren Auswahlprozeß und über die Bedingungen der Zusammenarbeit zu treffen. Dementsprechend gehört das Interview nicht nur zu den von den Verwendern am besten bewerteten Verfahren (Schuler et al., 1993), sondern wird von den Bewerbern sogar vor allen anderen Auswahlmethoden präferiert (Fruhner, Schuler, Funke & Moser, 1991; Steiner & Gilliland, 1996).

4.3 Historische Entwicklung

Die Erforschung des Interviews als Methode der Personalauswahl umfaßt nun bald ein ganzes Jahrhundert. Im Jahre 1915 prüfte Walter Scott die Beurteilerübereinstimmung von Auswahlgesprächen und kam zu dem Ergebnis, daß sie sehr gering war; beispielsweise wurde von einem der beurteilenden Manager einer der 36 Bewerber auf den ersten, von einem anderen auf den 32. Rangplatz gesetzt (Scott, 1915). In der amerikanischen Versicherungsbranche waren diese Ergebnisse ein Impuls zur Entwicklung biographischer Fragebogen, die wenige Jahre später die Überlegenheit psychometrischer Prinzipien in der Diagnostik demonstrieren sollten (Goldsmith, 1922). In den darauffolgenden Jahrzehnten folgten weitere Studien, so daß der erste umfassende Überblick zu diesem Thema (Wagner, 1949) bereits 106 Arbeiten berücksichtigen konnte. Als mittlere Beurteilerübereinstimmung wurde $r = .57$ errechnet (sie lag höher für die Merkmale „Intelligenz“ und „Soziabilität“), als durchschnittliche Validität $r = .27$.

Die nächste große Übersicht wurde von Mayfield (1964) zusammengestellt. Sie konnte bereits von den „McGill Studies“ profitieren, einer Serie von Arbeiten zur Urteilsbildung und Entscheidungsfindung im Interview, die unter Leitung von Edward Webster an der McGill-Universität in den 50er und 60er Jahren durchgeführt wurden, desgleichen von der sozialpsychologischen Forschung zur Eindrucksbildung sowie der zur Frage der „klinischen versus statistischen“ Informationsintegration in der Diagnostik. Die wichtigsten von Mayfields Schlußfolgerungen können wie in Tabelle 6 zusammengefaßt werden:

Tabelle 6:

Schlußfolgerungen aus der Interviewforschung nach Mayfield (1964)

1. Interviews sind von befriedigender Wiederholungsreliabilität, aber von unbefriedigender Objektivität.
2. Interviewer verhalten sich relativ konsistent in ihren Interviews.
3. Die relevanten Gesprächsbereiche werden nicht vollständig abgedeckt.
4. Interviewer unterscheiden sich in ihrer Informationsverwertung.
5. Strukturierte Interviews weisen höhere Objektivität auf als unstrukturierte.
6. Die Ergebnisse verschiedener strukturierter Interviews weisen erhebliche Unterschiede auf.
7. Die Validität von Interviews ist gering.
8. Die Ergänzung valider Testinformation durch ein Interview führt meist nicht zu erhöhter, manchmal zu verringerter Validität.
9. Intelligenz scheint eher als andere Merkmale reliabel und valide eingeschätzt werden zu können.
10. Die Form der Frage beeinflusst die erhaltene Antwort.

Das Interview ist die verbreitetste und die am meisten geschätzte Methode der Personalauswahl.

Lange Zeit wurden nur unbefriedigende Ergebnisse für das Interview als Verfahren der Personalauswahl berichtet.

Die empirische Interviewforschung konnte eine Vielzahl von Ursachen für die Unzulänglichkeit herkömmlicher Auswahlgespräche finden.

Die Person des Interviewers ist wichtig.

Methodische Weiterentwicklungen führten zu einer wesentlichen Verbesserung der prognostischen Validität.

11. Die Einstellung des Interviewers beeinflusst seine Interpretation der Antwort.
12. Im unstrukturierten Interview spricht der Interviewer mehr als der Interviewte.
13. Negative Information hat stärkeren Einfluß als positive.
14. Im unstrukturierten Interview treffen Interviewer ihre Entscheidung zu einem sehr frühen Zeitpunkt.
15. Äußere Erscheinung, Gesichtsausdruck und Benehmen eines Bewerbers spielen wahrscheinlich eine größere Rolle als das, was er sagt.

Die Einsichten von Mayfield und anderen Eignungsdiagnostikern belebten die weitere Interviewforschung, verschafften sie doch Einblick in die Prozesse der Interaktion und Urteilsbildung im Auswahlgespräch und boten damit Ansatzpunkte für Weiterentwicklungen. Schmitt (1976) und andere Autoren schlugen „Modelle“ des Interviewprozesses vor, die Einflußgröße, Prozeßvariablen und Ergebnisse in plausible Zusammenhangsdarstellungen zu bringen suchten. Die Kausalitätsheuristiken dieser Zeit hatten allerdings noch nicht den Charakter heutiger Strukturgleichungsmodelle, die zur empirischen Prüfung des gesamten Gefüges an Annahmen aufgestellt werden. Eine gründliche Darstellung des Interviewprozesses in allen seinen Komponenten findet man beispielsweise bei Dipboye (1992).

Schmitt (1976) und insbesondere Arvey (1979) befaßten sich auch mit Rechtsproblemen und Fragen der Fairneß von Auswahlentscheidungen gegenüber ethnischen Gruppen, Behinderten, Frauen und älteren Personen. Damit wandte sich das Interesse auch der Reaktion der Bewerber auf Auswahl-situationen zu. Schmitt & Coyle (1976) fanden, daß die Entscheidung qualifizierter Bewerber, ein Einstellungsangebot anzunehmen oder abzulehnen, maßgeblich von der Person des Interviewers abhängt. Rynes, Heneman & Schwab (1980) stellten Untersuchungen zusammen, die solche Einflüsse auch für andere Charakteristika der Organisation und des Auswahlprozesses nachweisen. Schuler und Stehle (1983) schlugen mit dem Begriff „soziale Validität“ eine Sammelbezeichnung derjenigen Situationsmerkmale vor, von denen es abhängt, ob Personalauswahl-situationen von Bewerbern als akzeptable soziale Situationen erlebt werden. Als bedeutsam für das Einstellungs-interview sollte sich diese Arbeitsrichtung insofern erweisen, als das Gespräch sich unter allen eignungsdiagnostischen Verfahren als dasjenige herausstellte, das am ehesten in der Lage ist, den Erwartungen der Bewerber zu entsprechen und positiv erlebt zu werden. Daraus resultierte ein Anstoß, der die Mühe lohnenswert erscheinen ließ, das Interview angesichts niedriger Validitätswerte weiterzuentwickeln, um sowohl den methodischen als auch den Akzeptanzanforderungen gerecht zu werden.

4.4 Methodische Verbesserungen

Für solche Weiterentwicklungen wurden verschiedene Ansatzpunkte gefunden, Interviews methodisch so zu verbessern, daß sie zu verlässlichen Auswahlinstrumenten werden. Dies ist nicht ohne methodischen Aufwand zu erreichen, ermöglicht aber Validitätswerte, die bis zur Höhe der besten sonstigen Auswahlverfahren reichen (Eder & Ferris, 1989). Vor allem Anforderungsbezug und Struktur haben sich als wirksame Moderatoren der Validität erwiesen. McDaniel, Whetzel, Schmidt, Hunter, Maurer und Russel (1986, publ. 1994) errechneten metaanalytisch Werte von $r = .30$ für anforderungsbezogene und $r = .21$ für traditionelle psychologische Interviews. Die Metaanalyse von Wiesner & Cronshaw (1988) resultierte in einem mittleren Validitätskoeffizienten von $r = .13$ für unstrukturierte und von $r = .40$ für strukturierte Interviews. Huffcutt und Arthur (1994) konnten 114 einzelne Studien reanalysieren, die sie in vier Grade der Strukturierung klassifizierten. Der Strukturierungsgrad zeigte einen starken Moderatoreffekt auf die Validität, die von $r = .20$ beim unstrukturierten Verfahren bis zu $r = .56$ beim höchststrukturierten Interview reichte.

Aus den Defiziten konventioneller Einstellungsgespräche und aus der Interviewforschung der siebziger und achtziger Jahre ließen sich Prinzipien und Maßnahmen ableiten, die eine substantielle Verbesserung des Interviews als Auswahlmethode erwarten lassen (Tabelle 7).

Tabelle 7:

Maßnahmen zur methodischen Verbesserung des Interviews (nach Schuler, 1989)

1. Anforderungsbezogene Gestaltung
2. Beschränkung auf diejenigen Merkmale, die nicht anderweitig zuverlässiger gesammelt werden können (z.B. Zeugnisnoten)
3. Durchführung in strukturierter bzw. (teil-) standardisierter Form
4. Verwendung geprüfter und verankerter Skalen
5. Empirische Prüfung (Itemanalyse, Validierung) von Einzelfragen
6. Bei geringem Standardisierungsgrad Einsatz mehrerer Interviewer
7. Integration von Verfahrenskomponenten aus dem Assessment Center
8. Trennung von Information und Entscheidung
9. Standardisierung der Gewichtungs- und Entscheidungsprozedur
10. Vorbereitung der Interviewer durch ein verfahrensbezogenes Training

Die in Tabelle 7 aufgeführten Maßnahmen sind als „technische Hilfe“ gleichzeitig geeignet, den großen – allerdings bislang nur unzureichend erforschten – Unterschieden in der Urteilskompetenz der Interviewer zu begegnen und auch bei schwächeren Interviewern ausreichende Ergebnisqualität sicherzustellen.

4.5 Komponenten der Interviewstruktur

Wenn von „strukturierten Interviews“ die Rede ist, so handelt es sich bei näherem Besehen keineswegs um eine homogene und eindeutig definierte Kategorie. Manche Autoren sprechen dann von strukturiert, wenn allen Kandidaten die gleichen Fragen in der gleichen Reihenfolge gestellt werden, andere verzichten auf eine dieser beiden Bestimmungen, dritte wieder ziehen die Art der Antwortverwertung oder Entscheidungsfindung als Definitionsgrößen heran. Campion, Palmer und Campion (1997) hatten bereits so viele Publikationen über die Effekte von Strukturierungsmaßnahmen zur Verfügung, daß sie 15 Strukturkomponenten, unterteilt in die Kategorien „Inhalt“ und „Bewertung“, zusammenstellen konnten, die sich tatsächlich oder mutmaßlich auf die Qualität des Interviews auswirken. Tabelle 8 gibt ihre Zusammenstellung der 15 Komponenten in ihrem Einfluß auf verschiedene Kennwerte der Reliabilität, der Validität und der Reaktion von Verwendern und Bewerbern wieder.

Trotz aller vorliegenden Daten darf nicht übersehen werden, daß manche Aussage, die aus der Auflistung der Tabelle 8 folgt, derzeit noch mehr der Plausibilität als generalisierbaren Belegen entspringt. Überdies ist das Zusammenwirken der verschiedenen Strukturierungsmaßnahmen noch weit davon entfernt, in allen Kombinationsmöglichkeiten geprüft worden zu sein. Gleichwohl kommt dieser Zusammenstellung hohe Anregungsfunktion für Forschung und Praxis zu. Campion et al. äußern jedenfalls die Überzeugung, jedes Interview könne dadurch gewinnen, daß zumindest einige der fünfzehn Komponenten genutzt würden. Einige Interviewmethoden, in denen dies realisiert ist, werden im nächsten Abschnitt vorgestellt.

Verschiedene Maßnahmen zur methodischen Verbesserung von Interviews sind umsetzbar.

„Strukturierung“ von Interviews kann aus einer Vielzahl von Einzelmaßnahmen bestehen.

Im Laufe der Zeit sind verschiedene konkurrierende Verfahrensformen strukturierter Interviews entstanden.

Biographische Interviewfragen sind vergangenheitsbezogen.

4.6 Typen strukturierter Interviews

Nachdem sich strukturierte Interviews der freien Gesprächsform hinsichtlich der Validität als grundsätzlich überlegen erwiesen haben, entstanden verschiedene konkurrierende Verfahrensformen. Einige davon konzentrieren sich ganz auf einen bestimmten Fragetypus, andere kombinieren unterschiedliche Arten von Fragen zu einem Interviewsystem. Im folgenden werden einige typische Vertreter beider Formen vorgestellt. Wir werden dabei sehen, daß eine wichtige Gemeinsamkeit mit dem Begriff „Struktur“ nicht gut erfaßt ist, aber gleichwohl eine große Rolle bei der methodischen Verbesserung des Interviews spielt.

Eines der erfolgsleitenden Prinzipien der Berufseignungsdiagnostik besteht in der Erkenntnis, daß Schlüsse über berufliche Qualifikation im großen und ganzen um so besser gelingen, je ähnlicher das geprüfte Verhalten dem Zielverhalten ist. „Der beste Prädiktor ist das Kriterium“ lautet ein altes Bonmot, das diesen Umstand auf die Spitze treibt. Während man in der eigenschaftsorientierten Diagnostik gewissermaßen die gebündelten Kräfte hinter dem Verhalten sucht, ist es in der simulationsorientierten Diagnostik die Verhaltensbeobachtung selbst, die Gewähr bieten soll, durch direkte Analogiebildung die Risiken eigenschaftsbezogener Spekulation zu vermeiden. Dieser Gedanke ist so neu nicht, spielte er doch in der Diagnostik der „Psychotechnik“ (Giese, 1924) schon eine große Rolle, bei der allerdings beide Prinzipien kombiniert wurden.

Der gleiche Grundsatz – möglichst große Ähnlichkeit zum Zielverhalten anzustreben – läßt sich auch beim biographieorientierten Diagnoseansatz verfolgen, indem man beispielsweise nicht nach der Geschwisterkonstellation fragt, um über relativ komplizierte (und häufig genug unbewiesene) gedankliche Konstruktionen Schlüsse über Kooperation und Durchsetzung im Berufsleben zu ziehen, sondern möglichst direkt das Verhalten in der Vergangenheit anspricht, um das es einem in der Zukunft geht. Für manche mag darin eine gewisse Entzauberung oder Banalisierung der Diagnostik liegen, aber die Erfolge geben dieser Vorgehensweise unzweifelhaft recht.

4.6.1 Das Behavior Description Interview

Vergangenes Verhalten, lautet der Grundsatz der biographieorientierten Diagnostik, ist der beste Prädiktor künftigen Verhaltens. Mit dem Behavior Description Interview (BDI) bzw. Patterned Behavior Description Interview (PBDI) wurde von Janz, Hellervik und Gilmore (1986) ein Verfahren zur Verfügung gestellt, das diesem Grundsatz klarer verpflichtet ist als alle publizierten Interviewverfahren davor. Viele später erarbeitete Verfahren orientieren sich an der Methode des BDI.

Ausgangspunkt der Fragen im BDI ist die Sammlung *kritischer Ereignisse* im Sinne der *Critical Incident Technique* von Flanagan (1954; vgl. Kapitel 3). Der Vorzug dieses Ausgangsmaterials ist, daß es sich um reale Ereignisse handelt, nicht um Meinungen oder andere unbelegte Äußerungen. Die Sammlung dieser Ereignisse in der gegebenen Organisation erhöht die Anpassung des Verfahrens an die dort gegebenen Bedingungen, nutzt das Wissen der Beschäftigten und erzielt gleichzeitig einen Partizipationseffekt. Critical Incidents können von Stelleninhabern, Vorgesetzten und auch von Kunden kommen, und wie zumeist, so ist auch in diesem Zusammenhang die Nutzung mehrerer Quellen einer einzelnen Informationsquelle vorzuziehen.

Critical Incidents beschreiben positives oder negatives Verhalten. Ein Beispiel für ein negatives Vorkommnis (Janz et al. 1986, S. 46) lautet folgendermaßen (eigene Übersetzung):

Zur Hauptverkehrszeit äußerte ein frustrierter Passagier, nachdem er endlich an den Check-in vorgerückt war, eine sarkastische Bemerkung über den Service der Fluglinie. Der Mitarbeiter am Schalter antwortete mit einer Beleidigung, was den Fluggast weiter aufbrachte. Kurz danach waren mehrere Passagiere und der Mitarbeiter in eine lautstarke Auseinandersetzung verwickelt.

Tabelle 8: Maßnahmen der Interviewstrukturierung und ihre Auswirkung auf Reliabilität, Validität und Verwenderreaktionen (Campion et al., 1997, S. 657, eigene Übersetzung)

Inhalt	Reliabilität						Validität			Verwenderreaktionen		
	Test-retest	Inter-rater	Kandid. Konsist.	Interaktion zw. Kandidaten	Interne Konsistenz	Beurteiler-übereinstimmung	Anforderungsbezug	Reduz. Defizienz	Reduz. Kontamination	Reduz. Benachteilig. best. Gruppen	Akzeptanz seitens Kandidaten	Akzeptanz seitens Interviewer
1. Anforderungsanalyse							+	+	+	+	+	+
2. Gleiche Fragen	+	+	+	+			+	+	+	+	+	+
3. Verzicht auf Hilfen und Nachfragen	+	+	+	+				-	+	+	-	-
4. Bessere Fragen	+	+	+		+		+	+	+	+		+
5. Längere Interviews	+	+			+			+			-	-
6. Verzicht auf ergänzende Information	+	+					-	+	+	+	-	-
7. Keine Fragen der Kandidaten	+	+	+	+			-	+	+		-	-
Auswertung												
8. Beurteilung der einzelnen Antworten oder Verwendung von Einstufungsdimensionen	+	+			+			+	+	+		+
9. Verankerte Einstufungsskalen	+	+					+	+	+	+		-
10. Detaillierte Aufzeichnungen	+	+					+	+	+	+		
11. Mehrere Interviewer	+	+		+	+			+	+	+	-	
12. Gleiche Interviewer	+			+								
13. Kein Meinungsaustausch zw. Interviewern	+	-						+	+	+		-
14. Training	+	+	+	+	-		+	+	+	+	+	+
15. Statistische Prognose	+	+			+		+	+	+	+		+

Anmerkungen: „+“ bedeutet positiver Effekt, „-“ bedeutet negativer Effekt.

Die kritischen Ereignisse werden zu 5 bis 10 Leistungsdimensionen gruppiert, die sich im wesentlichen durch Aufgabenbereiche ergeben. Zu jeder Dimension werden zumindest 10 bis 20 Ereignisse gesammelt, die gegebenenfalls um Items aus einer bereits bestehenden Sammlung ergänzt werden. Im Interview werden mehrere Fragen zu jeder dieser Leistungsdimensionen gestellt, die Antworten werden protokolliert. Ein Beispiel für eine Interviewfrage an einen Computerprogrammierer mit Folgefragen findet sich in Kasten 3.

Kasten 3:

Interviewfragen aus dem BDI (Janz et al. 1986, S. 134, eigene Übersetzung)

1. Berichten Sie über eines Ihrer Projekte, das erheblich zu spät zum Abschluß kam.
2. Was war das für ein Projekt?
3. Welches waren die größten Schwierigkeiten, denen Sie sich gegenübersehen?
4. Wie haben Sie versucht, diese Schwierigkeiten zu überwinden?
5. Was war das Ergebnis Ihrer Bemühungen?
6. Welche Auswirkung hatte die Verzögerung für denjenigen, der dieses Programm angefordert hatte?
7. Was haben Sie unternommen, um künftige Verzögerungen dieser Art zu verhindern?
8. Wie oft gerieten (Ihre) Projekte im letzten Jahr in Verzug?

Im Behavior Description Interview wird konkreten Sachverhalten gründlich nachgegangen, um die tatsächliche Handlungsweise des Bewerbers zu erfassen.

Das Beispiel in Kasten 3 zeigt, daß der Interviewer im BDI einem Sachverhalt gründlich nachgeht. Hierdurch wird zum einen angestrebt, sich ein konkretes Bild von der Handlungsweise des Bewerbers zu machen. Zum anderen wird dadurch verhindert, daß der Bewerber allgemeine Verhaltensregeln oder Lehrbuchwissen anstelle des eigenen Vorgehens äußert. Der Gefahr, sozial erwünschte Antworten anstelle realer Verhaltensschilderungen zu bekommen, kann also durch die konkreten Nachfragen entgegengewirkt werden. Gleichzeitig dürfte aus diesem Beispiel auch deutlich werden, daß die Formulierung der Fragen selbst und ebenso die Bewertung der Antworten intensive Kenntnis der Tätigkeit und ihrer Anforderungen voraussetzt.

4.6.2 Das situative Interview

Von Latham wurde ein Fragentypus vorgeschlagen, der heute weite Verbreitung gefunden hat, die situative Frage (Latham, Saari, Pursell & Campion, 1980). Ihre Gemeinsamkeit mit den biographischen Fragen des Behavior Description Interviews besteht in der Herkunft aus kritischen Ereignissen. Anders als im BDI werden allerdings die gesammelten Ereignisse nicht notwendigerweise zu Dimensionen zusammengefaßt, um dann den Stoff für Fragen zu ergeben, die ihrerseits häufig nur noch in lockerer Verbindung mit den kritischen Ereignissen stehen; situative Fragen werden demgegenüber so eng wie möglich an die kritischen Ereignisse angelehnt. Anstelle des – positiven oder negativen – Ausgangs des Falls steht allerdings die Frage: „Was würden Sie in dieser Situation tun?“. Dies kennzeichnet den größten Unterschied zwischen den beiden Fragearten: die biographiebezogene Frage richtet sich auf die Vergangenheit, die situative auf die Zukunft; sie erfaßt gewissermaßen Verhaltensvorsätze oder zumindest kognizierte Verhaltensmöglichkeiten. Latham (1989) bezieht sich dabei ausdrücklich auf die Zielsetzungstheorie (Locke & Latham, 1984), die Intentionen oder Ziele als unmittelbare Vorläufer des Verhaltens einer Person ansieht. Ein weiterer wichtiger Unterschied ist, daß zur Bewertung der Antworten zu jeder Frage eine verhaltensverankerte Einstufungsskala konstruiert wird, die Verhaltensverankerungen an den Extrempunkten und am mittleren Wert aufweist. Ein Beispiel für eine situative Frage gibt Kasten 4.

Situative Interviewfragen sind zukunftsbezogen.

Den Kandidaten werden selbstverständlich die Antwortverankerungen nicht gezeigt, ebenso wird ihnen nicht mitgeteilt, welche Anforderungsdimension gegebenenfalls damit erfaßt wird. Soweit ein Bewerber eine Antwort gibt, deren Bewertung sich nicht eindeutig aus dem Bewertungsschlüssel ergibt, ist auch bei diesem Fragetyp die Kompetenz des Interviewers zur eigenständigen Beurteilung erforderlich.

Situative Interviews haben den Vorzug, Kandidaten über Anforderungen, insbesondere über möglicherweise auftretende Schwierigkeiten zu informieren. Sie werden von den Bewerbern geschätzt, wenn auch nicht in gleichem Maße wie frei geführte Gespräche (Latham & Finnegan, 1993), vermutlich aufgrund der verminderten Möglichkeit zur Situationskontrolle. Verschiedentlich durchgeführte Versuche, situative Interviews nicht in Gesprächsform, sondern via Bildschirm durchzuführen, verlieren dadurch alle von Bewerbern besonders gutgeheißenen Charakteristika von Vorstellungsgesprächen und gehören eigentlich in eine andere Kategorie von Auswahlinstrumenten.

Kasten 4:

Beispiel für eine situative Frage mit Antwortbewertung (aus Latham, 1989, S. 172, eigene Übersetzung)

Sie sind der Vorgesetzte von Lastwagenfahrern in Philadelphia. Ihr Kollege ist der Vorgesetzte von Lastwagenfahrern 800 Meilen entfernt in Atlanta. Sie haben beide den gleichen Chef. Ihr Gehalt und Bonus resultiert zu 100 % aus Ihren Kosten. Ihr Kollege benötigt dringend einen Ihrer Lastwagen. Wenn Sie nein sagen, bleiben Ihre Kosten niedrig, und Ihr Team wird in diesem Quartal wahrscheinlich den Goldenen Flügel gewinnen. Wenn Sie ja sagen, wird wahrscheinlich das Atlanta-Team diesen angesehenen Preis gewinnen, weil sie der Firma erheblichen Gewinn bringen. Ihr Chef predigt Kosten, Kosten, Kosten ebenso wie Kooperation mit den Kollegen. Ihr Chef hat keine Kontrolle über die Buchhalter, die die Schiedsrichter sind. Ihr Chef ist sehr wettbewerbsorientiert, er belohnt Sieger. Sie selbst sind genauso wettbewerbsorientiert, Sie sind ein Siegertyp! Was würden Sie tun?

Antwortprotokoll:

Bewertungsanleitung

- (1) Ich würde mich um den Preis bemühen. Ich würde meinem Kollegen die Umstände erklären und sein Verständnis dafür gewinnen.
- (3) Ich würde den Rat meines Vorgesetzten einholen.
- (5) Ich würde meinem Kollegen den Lastwagen leihen. Ich würde die Anerkennung meines Vorgesetzten wie meines Kollegen dafür gewinnen, daß ich meinen guten Abschluß dem ihren geopfert habe. Dann würde ich meinen Leuten den Sinn der Sache erklären.

Die Information der Kandidaten über Tätigkeitsanforderungen dient der Selbstselektion.

Situative Fragen lehnen sich eng an kritische Ereignisse an und erfassen kognizierte Verhaltensmöglichkeiten.

4.6.3 Das Multimodale Interview

Es wurden verschiedene Konzepte entwickelt, mehrere Fragenarten zu einer komplexeren Interviewform zu kombinieren. Pursell, Campion und Gaylord (1980) stellten eine Interviewstruktur vor, die neben situativen Fragen und Kenntnisfragen auch kleinere Arbeitsproben enthält. Dieses Interview eignet sich damit vor allem zur Auswahl berufserfahrener Bewerber. Das von Westhoff und Kluck (1994) entwickelte *Entscheidungsorientierte Gespräch* wur-

Kombination mehrerer Fragetypen im Multimodalen Interview

Acht Bestandteile kennzeichnen das Multimodale Interview.

Kombination der drei Ansätze in der Eignungsdiagnostik

de nicht speziell als Einstellungsinterview konzipiert, dürfte allerdings auch für diesen Zweck verwendbar sein. In Form eines Leitfadens gibt es Regeln für die Vorbereitung, Durchführung und Auswertung sowie die schriftliche Darstellung der Gesprächsergebnisse in Gutachterform.

In unmittelbarer Reaktion auf die Defizite konventioneller Auswahlgespräche und als Versuch, die durch die Interviewforschung aufgezeigten Möglichkeiten der methodischen Verbesserung (s. Tabelle 7) umzusetzen, wurde das *Multimodale Interview (MMI)* konzipiert (Schuler, 1992). Es besteht aus einer relativ invarianten Abfolge von acht Gesprächskomponenten. Fünf dieser Komponenten dienen zur diagnostischen Urteilsbildung, also zur Bewertung der Antworten und des Verhaltens der Interviewten; die übrigen drei Gesprächskomponenten erfüllen vor allem die Funktion, dem Interaktionsprozeß einen natürlichen Gesprächsablauf zu geben sowie die Kandidaten über die Tätigkeitsanforderungen zu informieren. Einen Überblick über den Aufbau des *MMI* gibt Kasten 5.

Kasten 5:

Der Aufbau des Multimodalen Interviews (nach Schuler, 1992)

1. *Gesprächsbeginn.* Kurze informelle Unterhaltung. Bemühen um angenehme und offene Atmosphäre. Skizzieren des Verfahrensablaufs. Keine Beurteilung.
2. *Selbstvorstellung des Bewerbers.* Der Bewerber spricht einige Minuten über seinen persönlichen und beruflichen Hintergrund, seine derzeitige Situation und seine Erwartungen für die Zukunft. Sein Verhalten wird im Hinblick auf anforderungsbezogene Urteilsdimensionen eingestuft.
3. *Berufsorientierung und Organisationswahl.* Es werden standardisierte Fragen zu Berufswahl, Berufsinteressen, Organisationswahl und Bewerbung, bei berufserfahrenen Bewerbern auch zum Fachwissen gestellt. Die Antwortbewertung erfolgt auf verhaltensverankerten Skalen.
4. *Freier Gesprächsteil.* Der Interviewer stellt offene Fragen in Anknüpfung an Selbstvorstellung und Bewerbungsunterlagen. Die Bewertung erfolgt summarisch.
5. *Biographiebezogene Fragen.* Biographische (oder „Erfahrungs-“) Fragen werden aus Anforderungsanalysen dimensionsbezogen abgeleitet oder als validierte Fragen aus biographischen Fragebogen übernommen. Die Antworten werden auf verhaltensverankerten Einstufungsskalen beurteilt.
6. *Realistische Tätigkeitsinformationen.* Der Interviewer gibt dem Bewerber ausgewogene, bedarfsgerechte Information über Tätigkeit, Arbeitsplatz und Unternehmen. Überleitung zu situativen Fragen. Keine Beurteilung.
7. *Situative Fragen.* Auf critical incident-Basis konstruierte situative Fragen werden gestellt, die Antworten werden auf verhaltensverankerten Einstufungsskalen beurteilt.
8. *Gesprächsabschluß.* Fragen des Bewerbers werden beantwortet. Das weitere Vorgehen wird besprochen. Gegebenenfalls werden bereits Vereinbarungen getroffen.

Charakteristisch für den Aufbau des *MMI* ist, daß es sowohl dem Erfordernis der Verfahrensstandardisierung gerecht wird als auch durch die freien Gesprächsteile einen von Bewerbern bevorzugten natürlichen Gesprächsablauf und die Möglichkeit der Situationskontrolle durch die Bewerber zuläßt. In den verschiedenen Bewertungsteilen sind konstruktorientierte, simulationsorientierte und biographieorientierte diagnostische Ansätze repräsentiert; das *MMI* stellt damit gewissermaßen einen Mikrokosmos multimethodaler Eignungsdiagnostik dar. In der Gesprächskomponente „realistische Tätigkeitsinformation“ wird der Einsicht Rechnung getragen, daß eine Darstel-

lung von Tätigkeit und Organisation, die auch negative Seiten nicht verschweigt, der Selbstselektion dienlich ist und späteren Enttäuschungen sowie unerwünschter Fluktuation vorbeugt (Wanous, 1992); überdies ist sie ein Gebot der Fairneß gegenüber Bewerbern, von denen ihrerseits veridikale Information erwartet wird. Um zu verhindern, daß Interviewer während des Gesprächs ein frühes Globalurteil bilden, wird bei den meisten Gesprächsteilen jede Antwort unmittelbar bewertet, bei einigen auch eine summarische Einschätzung vorgenommen. Die Auswertung ergibt sich anschließend sowohl dimensionsbezogen als auch als Gesamtwert. Letzterer ist – gegebenenfalls in normierter Form – gewöhnlich die Grundlage einer Auswahlentscheidung, dimensionsbezogene Werte sind zumeist als Grundlage von Personalentwicklungsmaßnahmen vorzuziehen.

4.7 Empirische Evidenz

Die Zahl der für Interviews durchgeführten Validierungsstudien ist so groß, daß auch bereits eine ganze Reihe von Metaanalysen vorliegt. Ihre Ergebnisse sind nicht ganz einheitlich, was damit zusammenhängt, daß teilweise verschiedene Einzelstudien aufgenommen und unterschiedlich klassifiziert wurden. Die Einzelstudien wiederum beziehen sich auf unterschiedliche Bewerbergruppen und verwenden unterschiedliche Kriterien. Auch die Frage der angemessenen Korrektur der Validitätskoeffizienten wird nicht einheitlich beantwortet: Während viele Metaanalytiker, insbesondere jene aus der Schmidt/Hunter-Schule, Korrekturen bezüglich mangelnder Reliabilität des Kriteriums und Streuungseinschränkung im Prädiktor für selbstverständlich ansehen (und für konstruktbezogene Fragestellungen zusätzlich die Reliabilitätskorrektur des Prädiktors), ziehen andere – namentlich Bobko, Roth und Potosky (1999) und Schmitt, Rogers, Chan, Sheppard und Jennings (1997) – es vor, sich auch für metaanalytische Berechnungen am praktischen Gebrauch der eignungsdiagnostischen Verfahren zu orientieren und auf Korrekturen vollständig zu verzichten.

Welchen Einfluß Reliabilitäts- und Streuungskorrekturen haben, zeigt der Vergleich korrigierter Validitätskoeffizienten bei McDaniel et al. (1994) – $r = .44$ für strukturierte und $.33$ für unstrukturierte Verfahren – mit unkorrigierten Werten, die sich auf $r = .24$ und $.18$ belaufen. Bemerkenswerterweise hatten frühe Metaanalysen wie die von Hunter & Hunter (1984) für Interviews insgesamt eine korrigierte Validität von $r = .14$ mitgeteilt, was zum Boom der Bemühungen um Verfahrensverbesserungen beigetragen hat. Es könnte sein, daß mittlerweile auch das, was unter „unstrukturiertem Interview“ gemeint ist, validitätsfördernde Verbesserungen erfahren hat. Bei der bereits berichteten Metaanalyse von Huffcutt und Arthur (1994) – es wurden vier Stufen der Strukturierung unterschieden, was in $r = .20$ für die niedrigste und $r = .56$ für die höchste Strukturierungsstufe resultierte – handelt es sich ebenfalls um korrigierte Werte. Man tut jedenfalls gut daran, im Einzelfall zu prüfen, worauf sich Validitätsangaben beziehen.

Beim Vergleich verschiedener Fragetypen fanden Campion, Campion und Hudson (1994) in einer Einzelstudie für vergangenheitsbezogene (biographische) Fragen höhere Validität als für zukunftsbezogene, also situative ($r = .51$ gegenüber $.39$, korrigiert). Die prädiktive Validität des MMI-Gesamtwerts wurde in verschiedenen Einzelstudien bezogen auf Kriterien des Ausbildungs- und Berufserfolgs größtenteils zwischen $r = .30$ und $r = .50$ (unkorrigiert) errechnet (Schuler & Moser, 1995). Campion et al. (1994) errechneten für vergangenheitsbezogene Fragen inkrementelle Validität über zukunftsbezogene Fragen hinaus, nicht aber umgekehrt. Beide Fragetypen hatten inkrementelle Validität über einen Intelligenztest hinaus. Inkrementelle Validität gegenüber Intelligenztests wurde auch für das Multimodale Interview in bezug auf verschiedene Kriterien des Ausbildungserfolgs nachgewiesen (Schuler, Moser, Diemand & Funke, 1995). Schmidt und Hunter (1998) untersuchten 18 verschiedene diagnostische Verfahren und andere Datenquellen daraufhin, welche inkrementelle Validität sie über Intelligenztests hinaus zu-

Beim Multimodalen Interview wird verhindert, daß der Interviewer sich ein frühes Globalurteil über den Bewerber bildet.

Korrigierte und unkorrigierte Validitätsangaben unterscheiden sich erheblich.

Inkrementelle Validität

Konstruktvalidität

stande bringen. Neuen Ergebnissen zufolge liegen strukturierte Interviews gemeinsam mit Arbeitsproben an zweiter Stelle (Validitätsgewinn $r = .12$), übertroffen nur durch Integrity Tests mit $r = .14$. In Übereinstimmung damit halten Rindermann und Oubaid (1999) eine Kombination von Abiturnoten, Intelligenztests und MMI für den aussichtsreichsten Weg zur Auswahl von Studierenden durch die aufnehmenden Universitäten.

Bezüglich der Konstruktvalidität des Einstellungsinterviews waren Hunter und Hirsh (1987) noch auf Spekulationen angewiesen. Ihre Vermutung war, daß mit konventionellen Interviews soziale Kompetenzen erfaßt würden, mit strukturierten Interviews dagegen Intelligenz. Schuler und Funke (1989) zeigten demgegenüber, daß mittels des MMI mehrere psychologische Konstrukte, darunter soziale Kompetenz, psychische Stabilität und Leistungsmotivation, erfaßt werden können, wobei das Wesentliche ist, daß die Zielmerkmale nicht zwangsläufig mit der Interviewstruktur verbunden sind, sondern anforderungsgemäß bestimmbar sind. Huffcutt, Roth und McDaniel (1996) fanden metaanalytisch eine mittlere korrigierte Korrelation zwischen Interview und Intelligenz von $r = .40$ (unkorrigiert $.25$), wobei geringer strukturierte Interviews die höhere Korrelation mit dem Ergebnis von Intelligenztests aufwiesen als hochstrukturierte Verfahren. (Bei genauerem Besehen stellt sich allerdings heraus, daß eine große Teilstichprobe bei Huffcutt et al. aus einer MMI-Studie bestand – Schuler & Funke, 1989 –, deren Daten aus einem Verfahren stammten, das als zusätzliches Auswahlverfahren auf möglichst geringe Korrelation mit Intelligenz – und damit auf hohe inkrementelle Validität – angelegt war. Ohne diese Teilstichprobe wäre die Differenz zwischen hoch- und mittelhochstrukturierten Verfahren hinfällig.)

Salgado und Moscoso (in press) vergleichen metaanalytisch die Konstruktvalidität konventioneller und strukturierter verhaltensbezogener Interviews. Sie kommen zum Ergebnis, daß mit konventionellen Interviews Intelligenz, globale Persönlichkeitsmerkmale (u.a. Extraversion und Neurotizismus) sowie soziale Kompetenz erfaßt wird, während strukturierte Interviews vornehmlich Fachkenntnisse, Berufserfahrung und ebenfalls soziale Kompetenz messen (wobei letzteres wiederum vor allem auf die erwähnte MMI-Studie zurückgeht).

Ein letzter Blick im Zusammenhang mit der Konstruktvalidität von Einstellungsinterviews soll der Frage gelten, inwieweit deren Ergebnisse durch die Neigung zu sozial erwünschtem Antwortverhalten beeinflusst werden. Pennock und Shultz (in press) verglichen dahingehend die Werte aus einem *Patterned Behavior Description Interview* und einem situativen Interview mit den Scores eines gleichzeitig durchgeführten Social Desirability-Tests, des *Paulhus Balanced Inventory of Desirable Responding*. Das Ergebnis waren geringe Korrelationen für alle verglichenen Teilskalen. Für strukturierte Interviews bestätigt sich damit, was wir bei der Erörterung anderer Diagnoseverfahren schon feststellen konnten – daß die Neigung zu beschönigender Selbstdarstellung für gut konstruierte Auswahlverfahren eine geringere Gefahr darstellt als vielfach befürchtet.

Das Interview als Personalauswahlverfahren, läßt sich abschließend feststellen, konnte durch methodische Verbesserungen nicht nur in seiner psychometrischen Qualität verbessert werden, sondern es stellt eine ernstzunehmende Alternative zu anderen, teilweise weit aufwendigeren Verfahren der Personalauswahl dar. Mit der Zuordnung zu den biographieorientierten Verfahren der Eignungsdiagnostik wird strenggenommen nur der Teil des Interviews erfaßt, der sich auf Informationen aus der Vergangenheit des Bewerbers bezieht. Dies dürfte bei den meisten Auswahlgesprächen der überwiegende Teil sein. Fragen zu Einstellungen und Interessen können insoweit noch dem biographiebezogenen Ansatz zugerechnet werden, als sie Berufswahl und Bewerbungsgründe zum Gegenstand haben, die in der Vergangenheit liegen, auch wenn ihre Auswirkungen in die Gegenwart des Gesprächs reichen. Außerhalb dieses Rahmens liegen dagegen situative Fragen; wir können sie gewissermaßen als mentale Arbeitsproben ansehen und damit dem simulationsorientierten Ansatz der Diagnostik zurechnen.

Unabhängig davon müssen wir uns vor Augen halten, daß es sich bei der

Verfälschungstendenzen bei strukturierten Interviews sind relativ gering.

Kategorisierung „Konstrukt – Simulation – Biographie“ nur um eine idealtypische Einteilung handelt; de facto werden wir in den meisten Diagnoseverfahren eine Mischung dieser diagnostischen Ansätze vorfinden und eine Zuordnung nur schwerpunktgemäß vornehmen können.

Bei der Einteilung in konstruktorientierte, simulationsorientierte und biographieorientierte diagnostische Verfahren handelt es sich um eine idealtypische Kategorisierung.

Zusammenfassung

Neben dem simulations- und dem konstruktorientierten Ansatz läßt sich als drittes Prinzip der Eignungsdiagnostik der biographieorientierte Ansatz unterscheiden, der auf Schlüssen aus der Vergangenheit des Bewerbers auf dessen zukünftiges Verhalten im Beruf beruht. Idealtypische Verfahren dieser Kategorie sind insbesondere die Auswertung von Bewerbungsunterlagen und der biographische Fragebogen. Zu ersteren liegen bislang wenig positive Validitätshinweise vor, dennoch erscheinen sie als Instrument zur ersten Vorauswahl unverzichtbar. Als standardisiertes und wesentlich stärker formalisiertes Auswahlverfahren kann dagegen der biographische Fragebogen zu den Instrumenten der Eignungsdiagnostik mit gut bestätigter Validität gezählt werden. Während dieses Verfahren oftmals mit der Methode der empirischen Itemselektion gleichgesetzt wurde, hat sich inzwischen gezeigt, daß mittels rationaler (logischer) und induktiver (faktorenanalytischer) Methoden der Skalenbildung Schlüsse mit vergleichbarer kriteriumsbezogener aber ungleich höherer konstruktbezogener Validität möglich sind.

Ein Verfahren, das sich nicht so eindeutig dem biographieorientierten Ansatz der Eignungsdiagnostik zurechnen läßt wie die beiden erstgenannten Verfahren, ist das Interview. Zum überwiegenden Teil wird aber auch bei dieser Methode Information aus der Vergangenheit des Bewerbers gesammelt. Das Gespräch ist das verbreitetste und auch am meisten geschätzte Auswahlverfahren. Ausreichende psychometrische Qualität vermag es allerdings nur in anforderungsbezogener und strukturierter Form zustandezubringen. Unter den veröffentlichten und empirisch geprüften Varianten strukturierter Interviews haben vor allem Verfahren Verbreitung gefunden, die mit biographischen Fragen oder mit situativen Fragen arbeiten. Mit Verfahren, die mehrere verschiedene Komponenten zu einem Gesamtverfahren kombinieren, beispielsweise dem Multimodalen Interview, wird versucht, die Vorteile mehrerer, darunter auch unstrukturierter, Vorgehensweisen miteinander zu verbinden.

Zusammenfassung

Weiterführende Literatur

- Eder, R. & Ferris, G.R. (Eds.). (1989). *The employment interview: Theory, research, and practice*. Newbury Park, CA: Sage.
- Schuler, H., Farr, J.L. & Smith, M. (Eds.). (1993). *Personnel selection and assessment: Individual and organizational perspectives*. Hillsdale, NJ: Lawrence Erlbaum.
- Stokes, G.S., Mumford, M.D. & Owens, W.A. (Eds.). (1994). *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction*. Palo Alto, CA: Consulting Psychologists Press.

Weiterführende Literatur

Literatur

Literatur

- Althoff, K. (1986). Zur Aussagekraft von Schulzeugnissen im Rahmen der Eignungsdiagnostik. *Psychologie und Praxis. Zeitschrift für Arbeits- und Organisationspsychologie*, 30, 77-85.
- Anderson, N. & Shackleton, S. (1993). *Successful selection interviewing*. Oxford: Blackwell.
- Arvey, R.D. (1979). Unfair discrimination in the employment interview: Legal and psychological aspects. *Psychological Bulletin*, 86 (4), 736-765.
- Becker, T.E. & Colquitt, A.L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. *Personnel Psychology*, 45, 389-406.
- Bliesener, T. (1992). Ist die Validität biographischer Daten ein methodisches Artefakt? Ergebnisse einer meta-analytischen Studie. *Zeitschrift für Arbeits- und Organisationspsychologie*, 36, 12-21.
- Bliesener, T. (1995). Methodological moderators in validating biographical data in personnel selection. *Journal of Occupational and Organizational Psychology*, 69, 107-120.
- Bobko, P., Roth, P.L. & Potosky, D. (1999). Derivation and Implications of a Meta-Analytic Matrix Incorporating Cognitive Ability, Alternative Predictors, and Job Performance. *Personnel Psychology*, 52, 561-589.
- Borkenau, P. (1991). Sind erste Eindrücke valide, und wenn ja, warum? In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis* (S. 252-255). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Brown, B.K. & Campion, M.A. (1994). Biodata phenomenology: Recruiters' perceptions and use of biographical information in resume screening. *Journal of Applied Psychology*, 79, 897-908.
- Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist*, 39, 214-227.
- Bushnell, I.W.R. (1996). A comparison of the validity of handwriting analysis with that of the Cattell 16PF. *International Journal of Selection and Assessment*, 4, 12-17.
- Campion, M.A., Campion, J.E. & Hudson, J.P. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology*, 79, 998-1002.
- Campion, M.A., Palmer, D.K. & Campion, J.E. (1997). A review of structure in the selection interview. *Personnel Psychology*, 50, 655-702.
- Devlin, S.E., Abrahams, N.M. & Edwards, J.E. (1992). Empirical keying of biographical data: Cross-validity as a function of scaling procedure and sample size. *Military Psychology*, 4, 119-136.
- Dipboye, R.L., (1992). *Selection Interviews: Process Perspectives*. Cincinnati, OH: South-Western Publishing Co.
- Driver, R.W., Buckley, M.R. & Frink, D.D. (1996). Should we write off graphology? *International Journal of Selection and Assessment*, 4, 78-86.
- Eder, R. & Ferris, G.R. (Eds.). (1989). *The employment interview: Theory, research, and practice*. Newbury Park, CA: Sage.
- England, G.W. (1971). *Development and use of weighted application blanks* (rev. ed.). Minneapolis, MN: Industrial Relations Center, University of Minnesota.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 52, 327-358.
- Fruhner, R., Schuler, H., Funke, U. & Moser, K. (1991). Einige Determinanten der Bewertung von Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 35, 170-178.
- Giese, F. (1924). Die Arbeitsprobe in der Psychodiagnostik. *Zeitschrift für Angewandte Psychologie*, 23, 3-4.
- Goldsmith, D.B. (1922). The use of the personal history blank as a salesmanship test. *Journal of Applied Psychology*, 6, 149-155.
- Hein, M. & Wesley, S. (1994). Scaling biodata through subgrouping. In G.S. Stokes, M.D. Mumford & W.A. Owens (Eds.), *Biodata handbook: Theory, re-*

- search, and use of biographical information in selection and performance prediction (pp. 171-196). Palo Alto, CA: Consulting Psychologists Press.
- Hogan, J.B. (1994). Empirical keying of background data measures. In G.S. Stokes, M.D. Mumford & W.A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 69-107). Palo Alto, CA: Consulting Psychologists Press.
- Hollmann, H. & Reitzig, G. (1995). Referenzen und Dokumentenanalyse. In W. Sarges (Hrsg.), *Management-Diagnostik* (2. Aufl., S. 463-470). Göttingen: Hogrefe.
- Hough, L.M. & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G.S. Stokes, M.D. Mumford & W.A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 109-145). Palo Alto, CA: Consulting Psychologists Press.
- Huffcutt, A.I. & Arthur Jr., W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79, 184-190.
- Huffcutt, A.I., Roth, P.L., & McDaniel, M.A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology*, 81, 459-474.
- Hunter, J.E. & Hirsh, H. (1987). Applications of meta-analysis. In C.L. Cooper & I.T. Robertson (Eds.), *International review of industrial and organizational psychology 1987* (pp. 321-357). Chichester: Wiley.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Janz, T., Hellervik, L. & Gilmore, D.C. (1986). *Behavior Description Interviewing*. Newton, MA: Allyn & Bacon.
- Kennedy, E. (1988). Estimation of the squared cross-validity coefficient in the context of best subset regression. *Applied Psychological Measurement*, 12, 231-237.
- Kleinmann, M. & Seibt, H. (1990). Personalvorauswahl von Hochschulabsolventen: Derzeitiger Stand und Perspektiven. In M. Methner & A. Gebert (Hrsg.), *Psychologen gestalten die Zukunft: Anforderungen und Perspektiven* (S. 292-304). Bonn: Deutscher Psychologen Verlag.
- Kluger, A.N. & Colella, A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. *Personnel Psychology*, 46, 761-780.
- Knoblauch, R. (1990). Die Auswahl von Außendienstmitarbeitern in der pharmazeutischen Industrie mit Hilfe eines biographischen Fragebogens. In H. Schuler & W. Stehle (Hrsg.), *Biographische Fragebogen als Methode der Personalauswahl* (2. Aufl., S. 91-113). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Latham, G.P. (1989). The Reliability, Validity, and Practicability of the Situational Interview. In R.W. Eder & G.R. Ferris (Eds.), *The Employment Interview: Theory, research, and practice* (pp.169-182). Newbury Park, CA: Sage.
- Latham, G.P. & Finnegan, B.J. (1993). Perceived practicality of unstructured, patterned, and situational interviews. In H. Schuler, J.L. Farr & M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 41-55). Hillsdale, NJ: Lawrence Erlbaum.
- Latham, G.P., Saari, L.M., Pursell, E.D. & Champion, M.A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.
- Lefkowitz, J., Gebbia, M.I., Balsam, T. & Dunn, L. (1999). Dimensions of biodata items and their relationships to item validity. *Journal of Occupational and Organizational Psychology*, 72, 331-350.
- Lévy-Leboyer, C. (1994). Selection and assessment in Europe. In H.C. Triandis, M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 4, 2nd ed., pp. 173-190). Palo Alto, CA: Consulting Psychologists Press.
- Locke, E.A. & Latham, G.P. (1984). *Goal setting: A motivational technique that works*. Englewood Cliffs, NJ: Prentice-Hall.

Fortsetzung Literatur

Fortsetzung Literatur

- Machwirth, U., Schuler, H. & Moser, K. (1996). Entscheidungsprozesse bei der Analyse von Bewerbungsunterlagen. *Diagnostica*, 42, 220-241.
- Mael, F.A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology*, 44, 763-792.
- Mael, F.A. & Hirsch, A.C. (1993). Rainforest empiricism and quasi-rationality: Two approaches to objective biodata. *Personnel Psychology*, 46, 719-738.
- Marcus, B. (1999). *Attitudes toward personnel selection methods: A partial replication and extension in a German sample* (Manuscript submitted for publication). Stuttgart: University of Hohenheim.
- Marlowe, C.M., Schneider, S.L. & Nelson, C.E. (1996). Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology*, 81, 11-21.
- Mayfield, E.C. (1964). The selection interview: A re-evaluation of published research. *Personnel Psychology*, 17, 239-260.
- McDaniel, M.A., Whetzel, D.L., Schmidt, F.O., Hunter, J.E., Maurer, S. & Russel, J. (1994). The validity of employment interviews: A review and meta-analysis. *Journal of Applied Psychologie*, 79, 599- 602 (unpublished manuscript 1986).
- Mell, H. (1988). *Bewerbung auf dem Prüfstand*. Stuttgart: Schäffer.
- Mitchell, T.W. & Klimoski, R.J. (1986). Estimating the validity of cross-validity estimation. *Journal of Applied Psychology*, 71, 311-317.
- Mumford, M.D., Costanza, D.P., Connelly, M.S. & Johnson, J.F. (1996). Item generation procedures and background data scales: Implications for construct and criterion-related validity. *Personnel Psychology*, 49, 361-398.
- Mumford, M.D. & Owens, W.A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied psychological Measurement*, 11, 1-31.
- Mumford, M.D., Reiter-Palmon, R. & Snell, A.F. (1994). Background data and development: Structural issues in the application of life history measures. In G.S. Stokes, M.D. Mumford & W.A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 555-581). Palo Alto, CA: Consulting Psychologists Press.
- Neter, E. & Ben-Shakhar, G. (1989). The predictive validity of graphological inferences: A meta-analytic approach. *Personality and Individual Differences*, 10, 737-745.
- Owens, W.A. (1976). Background data. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (1st ed., pp. 609-644). Chicago: Rand McNally.
- Owens, W.A. & Schoenfeldt, L.F. (1979). Toward a classification of persons [monograph]. *Journal of Applied Psychology*, 64, 569-607.
- Pennock, R.P. & Shultz, K.S. (in press). Fakeability of structured selection interviews: A comparison of situational and past behavior formats. *International Journal of Selection and Assessment*.
- Pursell, E.D., Campion, M.A. & Gaylord, S.R. (1980). Structured interviewing: Avoiding selection problems. *Personnel Journal*, 11, 907-912.
- Quiñones, M.A., Ford, J.K. & Teachout, M.S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology*, 48, 887-910.
- Rindermann, H. & Oubaid, V. (1999). Auswahl von Studienanfängern durch Universitäten – Kriterien, Verfahren und Prognostizierbarkeit des Studienerfolgs. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 20, 172-191.
- Roth, P.L., BeVier, C.A., Switzer, F.S. & Schippmann, J.S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology*, 81, 548-556.
- Rothstein, H.R., Schmidt, F.L., Erwin, F.W., Owens, W.A. & Sparks, C.P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology*, 75, 175-184.
- Rynes, S.L., Heneman, H.G. & Schwab, D.P. (1980). Individual reactions to organizational recruiting: A review. *Personnel Psychology*, 33, 529-542.
- Salgado, J.F. & Moscoso, S. (in press). *Construct validity of employment interviews*.

- Schmidt, F.L. & Hunter, J.E. (1992). Development of a causal model of processes determining job performance. *Current Directions in Psychological Science*, 1, 89-92.
- Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmidt, F.L. & Rothstein, H.R. (1994). Application of validity generalization to biodata scales. In G.S. Stokes, M.D. Mumford & W.A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 237-260). Palo Alto, CA: Consulting Psychologists Press.
- Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *Personnel Psychology*, 29, 79-101.
- Schmitt, N. & Coyle, B.W. (1976). Applicant decisions in the employment interview. *Journal of Applied Psychology*, 61, 184-192.
- Schmitt, N., Rogers, W., Chan, D., Sheppard, L. & Jennings, D. (1997). Adverse impact and predictive efficiency using various predictor combinations. *Journal of Applied Psychology*, 82, 719-730.
- Schuler, H. (1989). Interviews. In S. Greif, H. Holling & N. Nicholson (Hrsg.), *Arbeits- und Organisationspsychologie* (S. 260-265). München: Psychologie Verlags Union.
- Schuler, H. (1992). Das Multimodale Einstellungsinterview. *Diagnostica*, 38, 281-300.
- Schuler, H. (1998a). *Psychologische Personalauswahl: Einführung in die Berufseignungsdiagnostik* (2. Aufl.). Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Schuler, H. (1998b). Noten und Studien- und Berufserfolg. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 370-374). Weinheim: Psychologie Verlags Union.
- Schuler, H. & Berger, W. (1979). Physische Attraktivität als Determinante von Beurteilung und Einstellungsempfehlung. *Psychologie und Praxis*, 23, 59-70.
- Schuler, H., Frier, D. & Kauffmann, M. (1993). *Personalauswahl im europäischen Vergleich*. Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Schuler, H. & Funke, U. (1989). The interview as a multimodal procedure. In R.W. Eder & G.R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 183-192). Newbury Park: Sage.
- Schuler, H. & Moser, K., (1995). Die Validität des Multimodalen Interviews. *Zeitschrift für Arbeits- und Organisationspsychologie*, 39, 2-12.
- Schuler, H., Moser, K., Diemand, A. & Funke, U. (1995). Validität eines Einstellungsinterviews zur Prognose des Ausbildungserfolgs. *Zeitschrift für Pädagogische Psychologie*, 9, 45-54.
- Schuler, H. & Stehle, W. (1983). Neuere Entwicklungen des Assessment-Center-Ansatzes – beurteilt unter dem Aspekt der sozialen Validität. *Psychologie und Praxis. Zeitschrift für Arbeits- und Organisationspsychologie*, 27, 33-44.
- Scott, W.D. (1915). Scientific selection of salesmen. *Advertising and Selling Magazine*, 5, 5-6.
- Stehle, W. (1990). Personalauswahl mittels biographischer Fragebogen. In H. Schuler & W. Stehle (Hrsg.), *Biographische Fragebogen als Methode der Personalauswahl* (2. Aufl., S. 17-57). Stuttgart: Hogrefe/Verlag für Angewandte Psychologie.
- Steiner, D.D. & Gilliland, S.W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, 81, 134-141.
- Stokes, G.S. (1994). Introduction and history. In G.S. Stokes, M.D. Mumford & W.A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. xv-xix). Palo Alto, CA: Consulting Psychologists Press.
- Wagner, R. (1949). The employment interview: A critical review. *Personnel Psychology*, 2, 17-46.
- Wanous, J.P. (1992). *Organizational entry: Recruitment, selection, orientation, and socialization*. Reading, MA: Addison-Wesley.

Fortsetzung Literatur

Fortsetzung Literatur

- Westhoff, K. & Kluck, M.-L. (1994). *Psychologische Gutachten schreiben und beurteilen* (2. Aufl.). Berlin: Springer.
- Weuster, A. (1994). *Personalauswahl und Personalbeurteilung mit Arbeitszeugnissen*. Göttingen: Hogrefe/Verlag für Angewandte Psychologie.
- Wiesner, W.H. & Cronshaw, S.F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, 61, 275-290.