

**„Biographie orientierte Personalauswahl
im Kontext angewandter Eignungsdiagnostik“**

Dissertation
zur Erlangung des Doktorgrades
der Philosophischen Fakultät
der Christian-Albrechts-Universität
zu Kiel
vorgelegt von
Rainer Petersen

Kiel
2002

Erstgutachter: Prof. Dr. Günter Köhnken
Zweitgutachter: Prof. Dr. Thomas Bliesener

Tag der mündlichen Prüfung: 23.10.2002
Durch die zweite Prodekanin Prof. Dr. Fleischmann
zum Druck genehmigt am: 28.11. 2002

Vorwort

An die Relevanz der vorliegenden Arbeit für die praktische Eignungsdiagnostik habe ich kaum einmal gezweifelt, daran, dass ich die Arbeit wirklich erfolgreich werde beenden dürfen jedoch häufiger. Insofern gebührt der erste und ganz besondere Dank meiner Frau Isabel, die nicht nur ihre eigenen beruflichen Ziele für mich zurücksteckte, sondern mir immer wieder den Rücken stärkte, mich motivierte und niemals den Glauben an den Abschluss verlor.

Dank gilt aber insbesondere auch Herrn Prof. Dr. Köhnken, der es mir ermöglichte, trotz meiner beruflichen Einbindung in Hamburg, die vorliegende Arbeit zu verfassen. Prof. Köhnken hat als Betreuer und Erstgutachter stets hilfreich, unterstützend und fördernd gewirkt. Herzlicher Dank gilt ferner Herrn Prof. Dr. Bliesener für seine Unterstützung als Zweitgutachter.

Herrn Dipl.-Psych. Martin Senkbeil sei für seine versierte Beratung in statistischen Fragen und die freundschaftliche Unterstützung gedankt. Nicht zuletzt möchte ich all jenen meinen aufrichtigen Dank aussprechen, die mich bei der Durchführung der Untersuchung und der Datenerhebung halfen. Hier sind insbesondere die Mitarbeiter der Polizei Hamburg zu erwähnen, die das Zustandekommen der Arbeit hilfreich unterstützten und mit großem Engagement begleiteten.

Abschließend ein Dank an jene, die es nicht mehr haben erleben dürfen, sich aber ganz besonders gefreut hätten und ohne deren Mitwirken nichts von dem jemals erschienen wäre: Erna Petersen, verstorben 2000 und Artur Petersen, verstorben 1981.

Inhaltsverzeichnis

1	Einleitung	6
2	Theoretischer und empirischer Hintergrund	10
2.1	Psychologische Personalauswahl	10
2.1.1	Grundlagen psychologischer Personalauswahl	10
2.1.2	Eignungsdiagnostische Verfahren zur Personalauswahl	21
2.1.2.1	Bewerbungsunterlagen	23
2.1.2.1.1	Vorauswahl und Bewerbungsunterlagen	23
2.1.2.1.2	Vorauswahl und Schulnoten	25
2.1.2.1.3	Schulnoten im Auswahlverfahren für den Polizeivollzugsdienst	26
2.1.2.2	Interview	30
2.1.2.3	Psychologische Testverfahren	33
2.1.2.4	Assessment-Center	36
2.1.2.5	Arbeitsproben	38
2.1.2.6	Computergestützte Verfahren	39
2.1.2.7	Biographische Daten	40
2.1.2.8	Weitere Verfahren	41
2.1.3	Personalauswahl am Beispiel der Einstellung in den Polizeivollzugsdienst	42
2.1.3.1	Allgemeine Grundlagen für die Einstellung	42
2.1.3.2	Einstellung in den Polizeivollzugsdienst	43
2.1.3.3	Vorauswahl bei Einstellung in den Polizeivollzugsdienst	44
2.2	Biographische Eignungsdiagnostik	46
2.2.1	Grundlagen biographischer Eignungsdiagnostik	46
2.2.1.1	Allgemeine Grundlagen biographischer Berufseignungsdiagnostik	47
2.2.1.2	Theoretischer Hintergrund biographischer Eignungsdiagnostik	52
2.2.2	Biographische Daten in der Eignungsdiagnostik	58
2.2.2.1	Konstruktion Biographischer Fragebögen	58
2.2.2.2	Anwendung Biographischer Fragebögen	63
2.3	Soziale Validität biographischer Daten	72
2.3.1	Zum Begriff der Validität	72
2.3.2	Zum Begriff der Sozialen Validität	77
3	Ableitung der Fragestellung	92
3.1	Konkretisierung der allgemeinen Fragestellung	92
3.2	Formulierung der Untersuchungsfragen	100
4	Untersuchungsplan	101
5	Durchführung der Untersuchung	105
5.1	Beschreibung der Stichproben	106
5.1.1	Stichprobe 1: Untersuchung zur Fragebogenkonstruktion	106
5.1.2	Stichprobe 2: Validierung des BfP	107

5.1.3	Stichprobe 3: Bestimmung der Retest-Reliabilität und Akzeptanzuntersuchung	107
5.2	Ablauf der Untersuchung	108
5.2.1	Erster Abschnitt: Fragebogenkonstruktion	108
5.2.1.1	Konzeption einer Fragebogenvorversion	109
5.2.1.2	Auswertung und Analyse der Fragebogenvorversion	112
5.2.1.3	Fragebogenendversion (BfP)	113
5.2.2	Zweiter Abschnitt: Validierung der Fragebogenendversion (BfP)	115
5.2.2.1	Durchführung zur Validierung der Fragebogenendversion	115
5.2.2.2	Validierungs- und Kreuzvalidierungsstichprobe	118
5.2.2.3	Validierungs- und Kreuzvalidierung des Fragebogens	118
5.2.3	Dritter Abschnitt: Akzeptanzuntersuchung und Retest-Reliabilität	119
5.2.3.1	Durchführung zur Bestimmung der Retest-Reliabilität	120
5.2.3.2	Durchführung zur Bestimmung der Sozialen Validität	120
6	Ergebnisdarstellung	122
6.1	Validierung des Biographischen Fragebogens (BfP)	122
6.1.1	Netto- und Grobgewichtung der Antwortalternativen in den Extremgruppen	122
6.1.2	Validierung des BfP in der Stichprobe	122
6.1.3	Validierung des BfP in der Kreuzstichprobe	124
6.2	Bestimmung der Reliabilität des BfP	125
6.3	Bestimmung der sozialen Validität des BfP	128
6.4	Beantwortung der Hypothesen	129
7	Interpretation und Diskussion	131
7.1	Zusammenfassung	131
7.2	Interpretation der Ergebnisse	137
7.3	Ausblick	149
8	Literatur	153
9	Anhang	165
9.1	Anhang A: Durchführung	165
9.2	Anhang B: Ergebnisse	173

1 EINLEITUNG

Produktivität und Effizienz eines Unternehmens, ebenso wie die Leistungsfähigkeit, Zufriedenheit und Belastung des einzelnen Arbeitnehmers hängen entscheidend davon ab, inwieweit die Leistungsmöglichkeiten des Individuums mit den Anforderungen des Arbeitsplatzes übereinstimmen (Rösler, 1992). Dennoch ist die Qualität von Millionen jährlich zu treffender Entscheidungen in der Personalauswahl so unterschiedlich wie die zur Anwendung kommenden Verfahren (Schuler, 1996). Dabei ist die Relevanz berufseignungsdiagnostischer Entscheidungen nach Hossiep (1995) von ungebrochener Tragweite. Gerade berufseignungsdiagnostische Entscheidungen legen den Handlungsspielraum für immer längere Zeiträume fest und stellen aus diesem Grund eine besonders wichtige Phase menschlichen Handelns dar, so dass getroffene Entscheidungen, bspw. über die Annahme oder Ablehnung eines Bewerbers für eine bestimmte Position, für die betroffenen Personen und Organisationen stets von besonderer Bedeutung sind. Wir können nach Schmidt und Hunter (1998) heute auf eine 85-jährige Forschungstradition zurückblicken, die diese Qualität zu sichern hilft und von der wir profitieren, wenn wir die Validität unterschiedlicher eignungsdiagnostischer Instrumente bestimmen. In diesem Kontext ist die prognostische Validität von zentraler Bedeutung für das psychologische Eignungsurteil. Die zur Beantwortung stehende Frage ist hierbei, welche von den zur Verfügung stehenden Bewerbern für eine vakante Position die besten Arbeitsergebnisse bzw. Ausbildungsergebnisse erwarten lassen. Die eingestellten Personen werden gemäß den zur Verfügung stehenden freien Arbeitsplätzen verteilt, darüber hinaus zur Verfügung stehende Interessenten zurückgewiesen.

Sucht ein Unternehmen neue Mitarbeiter¹ oder Auszubildende, so orientiert es sich primär an der zu erwartenden Leistung im unternehmerischen Sinne. Aber ein Unternehmen wäre schlecht beraten, wenn es dabei die Zufriedenheit und Gesundheit seiner Mitarbeiter außer Acht ließe. Potenzielle Zufriedenheit im zukünftigen Beruf ist ein wichtiges Kriterium, das schon im Auswahlprozess berücksichtigt werden muss. Ein unter- oder langfristig überforderter Mitarbeiter dürfte dauerhaft nicht zufrieden mit seiner Tätigkeit sein können, andererseits dürfte die Arbeitsleistung des Mitarbeiters nicht effektiv einsetzbar sein.

¹ In der vorliegenden Arbeit wurde zur Vereinfachung die männliche Form verwendet. Unabhängig hiervon sind selbstverständlich Personen beiderlei Geschlechts gemeint.

Trotz dieses hohen Anspruchs und ungeachtet einer langjährigen Forschungstradition sowie eines vorliegenden vielfältigen und qualitativ hochwertigen eignungsdiagnostischen Instrumentariums zur Personalauswahl, ist das folgende Beispiel nicht realitätsfern:

„ Im oberen Management einer Firma ist eine wichtige Position zu besetzen. Um den besten Kandidaten zu finden, wendet die Firma psychologische Testverfahren an. Das läuft ungefähr folgendermaßen:

Der erste Kandidat ist der Ausbildung nach Mathematiker. Ihm wird kurz erklärt, warum die Firma psychologische Tests anwendet, dann wird ihm die Testfrage gestellt: „Wie viel ist zwei mal zwei?“ Der Kandidat antwortet natürlich wie aus der Pistole geschossen: „4“. Der Personalchef bedankt sich bei ihm und schließt mit dem Satz: „Sie hören von uns.“

Der zweite Kandidat ist ein Diplom-Ingenieur. Der Ablauf ist derselbe wie zuvor beschrieben. Auf die Frage „Wie viel ist zwei mal zwei?“ zückt er seinen Rechenschieber und antwortet: „3,9 ungerade“. Im übrigen siehe oben.

Der dritte Kandidat ist ein Diplom-Kaufmann: Seine Antwort auf die Testfrage lautet: „Im Einkauf würde ich sagen drei, im Verkauf würde ich sagen fünf“. Ansonsten wie gehabt.

Und wer bekam die Position? – Der Schwiegersohn des Generaldirektors.“

Anders (1992; zitiert nach Wiendieck 1994)

Auch wenn es derartige Beispiele in der einen oder anderen Form immer geben wird, so sieht zumindest der grundsätzliche Anspruch an wissenschaftlich begründete Personalauswahl heute anders aus, denn falsche Personalentscheidungen können für ein Unternehmen erhebliche Kosten erzeugen und seine Wirtschaftlichkeit gefährden. Nicht ohne Grund wird von der „Ressource“ Mensch gesprochen und bezeichnet der moderne Personalverantwortliche sich selbst gern als „Human Resource Manager“. Nur mit einem wissenschaftlich abgesicherten professionellen Personalauswahlprozess können fehlerhafte Personalentscheidungen minimiert und damit Kosten gespart werden und kann es auch im Interesse der Mitarbeiter gelingen, die richtig qualifizierten Arbeitnehmer an die richtigen Stellen zu bringen.

Häufig kommen für die Personalauswahl geeignete Instrumente jedoch nicht zum Einsatz, da durch die Vorauswahl der Bewerbungsunterlagen potenziell geeignete Kandidaten gar nicht zum Auswahlverfahren zugelassen werden. Die wichtigste Informationsquelle bei Bewerbern um einen Ausbildungsplatz ist das Zeugnis, d.h. Schulnoten bestimmen über Annahme oder Ablehnung einer Bewerbung. Da in den letzten Jahren jedoch wiederholt Zweifel an der Nützlichkeit dieses Prädiktors für die Erfassung zukünftiger Leistungen in Ausbildung und Beruf auftraten, stellt sich die Frage nach geeigneten Alternativen. Dies ist insbesondere von

Bedeutung, da für die Inhalte der sonstigen Bewerbungsunterlagen die prognostische Validität mit $r = .18$ (Reilly & Chao, 1982) unzureichend ist.

Im Rahmen der eignungsdiagnostischen Forschung konnte für biographische Fragebögen in verschiedenen Studien eine hohe prognostische Validität belegt werden. Die Höhe des Validitätskoeffizienten variiert jedoch in unterschiedlichen Untersuchungen und in Abhängigkeit vom gewählten Kriterium zwischen $r = .33$ und $r = .52$ (Bliesener, 1992; Ghiselli, 1955; Schmitt, Gooding, Noe & Kirsch, 1984; Schmidt & Hunter, 1998). Damit liegen für verschiedene Berufsfelder Hinweise vor, dass dieses eignungsdiagnostische Instrument in Fragebogenform eine nützliche Alternative zur Vorauswahl mit Schulnoten darstellen könnte, auch wenn bislang noch keine Theorie zu biographischen Daten vorliegt.

Aufgrund der spezifischen Validierungen vorliegender biographischer Fragebögen, die entweder nicht für den Kontext der Personalauswahl entwickelt wurden oder nur für die Anwendung innerhalb ganz bestimmter Populationen gedacht sind, besteht für die Auswahl zur Einstellung in den Polizeivollzugsdienst allerdings die Notwendigkeit, einen neuen biographischen Fragebogen zu entwickeln. Mit der Neukonstruktion und Validierung verbindet sich aber auch die Möglichkeit, den häufig in diesem eignungsdiagnostischen Bereich vernachlässigten Aspekt der sozialen Validität bzw. Akzeptanz in Sinne von Schuler (1990) zu berücksichtigen.

Somit leitet sich für die Untersuchung die Frage ab, in wie weit mittels eines für die Einstellung in den Polizeivollzugsdienst entwickelten biographischen Verfahrens, die für andere Berufsfelder dokumentierte und wiederholt replizierte hohe Vorhersagegüte in vergleichbarer Höhe erschlossen werden kann. Bisher liegen keine Untersuchungen vor, welche die Eignung biographischer Fragebögen für die Personalvorauswahl im Polizeivollzugsdienst belegen.

Ein biographischer Fragebogen kann als nützlich gelten, wenn durch ihn prognostisch valide Eignungsvorhersagen möglich sind, d.h. wenn zwischen den zum Zeitpunkt t_1 erhobenen Prädiktoren (biographischen Daten im Fragebogen) und den zu einem Zeitpunkt t_2 erhobenen Kriterium (Testergebnis im Auswahlverfahren) ein signifikanter Zusammenhang nachweisbar ist. Daneben sollte jedoch auch die Frage der Akzeptanz eines eignungsdiagnostischen Auswahlinstrumentes für den jeweiligen Zweck und in der jeweiligen Population reflektiert

werden. Denn die Auswahlssituation stellt häufig den Erstkontakt zwischen Bewerber und Organisation dar. Aufgrund fehlender anderer Erfahrungen können Bewerber in diesem Stadium den Eindruck gewinnen, dass die Auswahlssituation und das hier von der Organisation gezeigte Verhalten repräsentativ für die Organisation seien. D.h., die Auswahlssituation vermittelt ein Bild der Organisation, welches, insbesondere durch abgelehnte Bewerber, die keine Gelegenheit finden, diesen selektiven Eindruck zu kompensieren, an andere potenzielle Bewerber weitergetragen wird. Die Frage, ob Personen, die das Auswahlverfahren bereits absolviert haben, anderen eine Bewerbung bei der Organisation empfehlen, hängt damit durchaus auch davon ab, welches im Auswahlverfahren wahrgenommene Image der Organisation zugeschrieben und weitergetragen wird. Daneben sollte die Akzeptanz aber ebenfalls originäres Interesse des Eignungsdiagnostikers selbst sein. Verfahren mit einer hohen Akzeptanz dürften die Kooperationsbereitschaft und die Motivation der Teilnehmer und somit die Validität der Eignungsdiagnose positiv beeinflussen (Nevo, 1986).

Für die soziale Akzeptanz eines Auswahlinstrumentes ist grundlegend, dass die Teilnehmer am Auswahlverfahren, die Items des biographischen Fragebogens als für den Zweck geeignet anerkennen. Eine wichtige Voraussetzung hierfür ist die Partizipation der Betroffenen schon zurzeit der Entwicklung und vor Einsatz als Selektionsinstrument. Derzeit liegen allerdings für den Bereich der biographischen Eignungsuntersuchungen kaum Belege für die Akzeptanz dieser Verfahren vor.

2 Theoretischer und empirischer Hintergrund

2.1 Psychologische Personalauswahl

2.1.1 Grundlagen psychologischer Personalauswahl

Ziel einer jeden eignungsdiagnostischen Strategie ist ein insgesamt möglichst günstiger Beitrag aller Personen zu dem angestrebten Organisationserfolg (Hossiep, 1995). Doch was sind eignungsdiagnostische Strategien oder treffender, was haben wir unter Berufseignungsdiagnostik zu verstehen? Hossiep (1995) gibt einen umfangreichen Überblick (S. 4-7) zu der sich mit dem Zeitgeist verändernden Definition der Berufseignungsdiagnostik. Beispielhaft seien an dieser Stelle zwei Definitionen genannt: Nach Schuler und Funke (1991) ist Eignungsdiagnostik „die Wissenschaft und Praxis der Zuordnung von Personen und Tätigkeiten. Die wissenschaftliche Seite ist die der Entwicklung von Theorien über den Zusammenhang von Fähigkeiten und Leistungen sowie Methoden, diese Merkmale zu messen und zu prognostizieren. Aufgabe der Eignungsdiagnostik in der Praxis ist es, sich dieser Methoden als Bestimmungs- und Entscheidungshilfen in einer Weise zu bedienen, die das Umfeld der wirtschaftlichen und sozialen Bedingungen berücksichtigt. Bei angemessener Anwendung liefert die berufsbezogene Eignungsdiagnostik nicht nur Entscheidungshilfen für die Organisation, die Mitarbeiter sucht und zu deren Weiterentwicklung sie beitragen möchte, sondern auch für den einzelnen Menschen, der sich für einen Beruf, einen Betrieb oder eine Umschulung zu entscheiden hat. Beide Seiten – Mensch und Organisation – profitieren von einem fähigkeits- und interessenentsprechenden beruflichen Einsatz, dessen Nutzen sich nicht nur in ökonomischen Größen bemessen lässt, sondern auch im Beitrag zur Zufriedenheit und Gesundheit, der dadurch zustande kommt, dass Über- und Unterforderung vermieden werden“ (S.5). Und unter dem Konzept der Eignung verstehen Schuler und Funke (1993) „Das Ausmaß von Übereinstimmung von Anforderungen des Arbeitsplatzes und der weiteren Arbeitsumgebung mit den Leistungsvoraussetzungen der Person. Dabei wird angenommen, dass der Grad dieser Übereinstimmung die Erfolgswahrscheinlichkeit für eine Berufstätigkeit mitbestimmt. In der Eignungsdiagnostik ist neben der persondiagnostischen Aktivität also als zweiter Bezugspunkt Situationsdiagnostik (Arbeits- und Anforderungsanalyse) gefordert“ (S.237).

Was letztendlich als beruflicher Erfolg angesehen wird, kann nach Auffassung der genannten Autoren vielfältig sein: Leistung, Sinnerleben, gesellschaftlicher Status, Zufriedenheit, psychische und physische Gesundheit, das Gefühl gefordert zu sein oder ein angemessenes Auskommen zu haben, die Überzeugung das zu tun, was den eigenen Fähigkeiten und Interessen entspricht oder der Eindruck sich persönlich weiterzuentwickeln. Die Abhängigkeit beruflichen Erfolges von einer Vielzahl Faktoren ist unbestritten. Auf der Basis familiärer und sozialer Hintergründe beeinflussen u.a. Einstellungen und Erwartungen, Ausbildungsmöglichkeiten, gesellschaftliche Rahmenbedingungen und die sich verändernde Situation auf dem Arbeitsmarkt die Chancen in beruflichen Feldern tätig bzw. erfolgreich tätig werden zu können. Eignungsrelevante Personenmerkmale können nach Hossiep (1995) nur in dem Ausmaß zur individuellen Berufserfolgsprognose beitragen, wie die Berufseignungsdiagnostik diese Personenmerkmale erfassen kann und diese zwischen dem Zeitpunkt der Prognose und der Zeitspanne auf die sich diese wiederum bezieht, unverändert bleiben oder die Veränderungen regelhaft und damit vorhersagbar vonstatten gehen. Die Berufseignungsdiagnostik ist keine Prozessdiagnostik, die einen Veränderungsprozess begleitet, sondern Statusdiagnostik in dem Sinne einer Stützung auf jeweils vorgefundene Merkmale oder Merkmalskombinationen.

Die Vorhersage beruflichen Erfolges wird zusätzlich dadurch erschwert, da alle für den Erfolg einer Tätigkeit relevanten Merkmale zum Zeitpunkt einer Entscheidung über Annahme oder Ablehnung eines Bewerbers kaum vollständig erfasst werden können oder gar bekannt sein dürften. Letzteres gilt insbesondere für Berufe, die sich hinsichtlich ihres Anforderungsprofils laufend verändern wie dies bspw. für den Polizeivollzugsdienst oder den Finanzdienstleistungssektor gelten dürfte.

Zu den Merkmalen, die als relativ stabil gelten und sich zumindest auch partiell relativ gut durch psychologische Auswahlinstrumente erfassen lassen, gehören nach Schuler und Funke (1993) in erster Linie Intelligenz, allgemeine Leistungsfähigkeit, soziale Kompetenz und Selbstvertrauen. So nennen bspw. für ausgezeichnete Manager-Leistungen Sarges und Weinert (1991) als die wichtigsten Persönlichkeitsmerkmale: überdurchschnittliche intellektuelle Fähigkeiten, hohe soziale Kompetenz und einen prägnanten Leistungs- und Gestaltungswillen.

Die Frage nach Feststellung der Eignung eines Bewerbers für einen Ausbildungsplatz oder die von ihm angestrebte berufliche Position hat, wie eingangs schon erwähnt, eine lange Tradition.

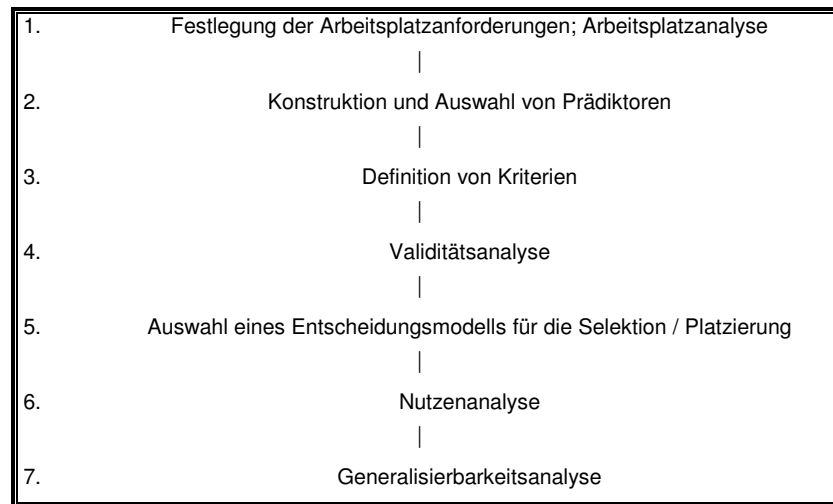
Da sich schon eine Vielzahl von Autoren mit den historischen Wurzeln der Berufseignungsdiagnostik beschäftigten (vgl.a. DuBois, 1970; Schuler & Moser, 1995), soll an dieser Stelle nur exemplarisch auf die geschichtlichen Hingründe eingegangen werden. Schon sehr früh wurden Verfahren zur Auswahl von Mitarbeitern entwickelt. Nach DuBois (1970) fanden in China vor mehr als 3000 Jahren die ersten Testbatterien zur Auswahl von Staatsbediensteten Anwendung. Wissenschaftlich fundierte Eignungsdiagnostik allerdings wurde erst mit Beginn des 20. Jahrhunderts entwickelt und implementiert. Zurzeit der Jahrhundertwende wurden nach Owens (1976) auch erste biographische Fragebögen zur Personalauswahl in der Versicherungsindustrie verwendet. Der Schwerpunkt lag bei diesen erstmals nach psychometrischen Prinzipien konstruierten Verfahren in der Anwendung im Produktions- und Dienstleistungsbereich (Schuler, 1996). Die Forderung nach systematischen Eignungsuntersuchungen ist nach Hossiep (1995) in aller Konsequenz erst im Verlauf des vergangenen Jahrhunderts erhoben worden. Hierbei ging es darum, jeden Einzelnen gemäß seiner Eignung bei der arbeitsteiligen Verrichtung von Tätigkeiten einzusetzen, und zwar gemäß dem Leistungsprinzip. In den 20er Jahren wurden die ersten Persönlichkeitstests eingesetzt (Anastasi, 1985) und nach ihrer ursprünglichen Verwendung als Schulreifetests fanden bald auch Intelligenztests zur Auswahl von Mitarbeitern Verbreitung. Verwiesen sei an dieser Stelle auf den berühmten Intelligenztest Army Alpha, zur Selektion amerikanischer Rekruten (vgl. a. Greif, 1995).

Wie unter Kapitel 2.1.2 näher beschrieben, ist die Verwendung standardisierter Testverfahren in der Berufseignungsdiagnostik heute weitgehend verbreitet und hat sich in den Augen vieler Praktiker wie Theoretiker bewährt. Neben Weiterentwicklungen und Anpassungen in den Bereichen der Intelligenz- und Persönlichkeitstestverfahren, finden zwischenzeitlich spezielle Funktions- und Eignungstests sowie selbst klinische Testverfahren Anwendung. Letztere wurden für diesen Kontext weder entwickelt noch entsprechend normiert.

Bevor jedoch eignungsdiagnostische Auswahlverfahren zur wissenschaftlich begründeten beruflichen Selektion oder Platzierung angewandt werden können, bedarf es einiger Vorarbeit.

Eine einfache übersichtliche Darstellung der einzelnen Stadien, die bei der Etablierung eines wissenschaftlichen eignungsdiagnostischen Personalausleseverfahrens durchlaufen werden sollten, findet sich bei Rösler (1992). Sie ist in Tabelle 1.2-1 schematisch wiedergegeben.

Tab. 1.2-1: Schematische Übersicht einzelner Stadien wissenschaftlich begründeter Personalauswahl nach Rösler (1992)



Der Arbeitsplatz ist hier definiert durch die Gesamtheit aller Aufgaben, die von dem jeweiligen Stelleninhaber bewältigt werden müssen. Zur Beschreibung dieser Aufgaben dient die Arbeitsplatzanalyse. Ziel dieser ist nach Kannheiser und Frieling (1992) u.a. die Erstellung einer empirischen Basis für die Konstruktion und die Auswahl von Prädiktoren und Kriterien sowie von Job-Typologien, anhand derer eine Validitätsgeneralisierung vorgenommen werden kann. Verfahren zur Arbeitsplatzanalyse sind eignungsdiagnostische Instrumente, mit denen überdauernde Eigenschaften eines Arbeitsplatzes erfasst werden sollen.

Zu den bekanntesten standardisierten Verfahren zählen bspw. der Fragebogen zur Arbeitsplatzanalyse (FAA) von Frieling und Hoyos (1978) und das Arbeitswissenschaftliche Erhebungsverfahren zur Tätigkeitsanalyse (AET) von Rohmert und Landau (1979). Als Beispiel für einen in Deutschland verbreiteten Fragebogen soll an dieser Stelle der FAA näher beschrieben werden. Die Intention des Fragebogens besteht in der Erfassung verhaltenswirksamer Aspekte eines Arbeitsplatzes. Er besteht aus vier Hauptteilen, die sich weiter in Abschnitte und Unterabschnitte untergliedern. Hierzu zählen „Informationsgewinnung und Informationsverarbeitung“; d.h. es soll die Frage beleuchtet werden, wo und wie der Beschäftigte die Informationen, die er bei der Ausübung seiner Tätigkeit verwendet, gewinnt. Unter der Rubrik „Arbeitsausführung“ werden jene Verhaltensweisen analysiert und aufgezeichnet, die beobachtbar sind sowie registriert, welche Werkzeuge und Geräte verwendet werden. „Arbeitsrelevante Beziehungen“ werden erfasst, um festzuhalten, welche Beziehungen zu anderen Personen bei der Ausübung der Tätigkeit erforderlich sind. Der vierte Hauptteil beschreibt „Umgebungseinflüsse und zusätzliche

Arbeitsbedingungen“ und damit das physische und soziale Milieu, in dem die Arbeit durchgeführt wird.

Insbesondere im nicht gewerblichen Bereich oder bei der Erfassung von Anforderungen leitender Angestellter bzw. Manager kommen häufig nicht standardisierte Verfahren zur Arbeitsplatzanalyse zur Anwendung. Beispiele wären hier Expertensitzungen aufgrund impliziter Annahmen über erfolgreiches Führungsverhalten (Kaufmann, 1984). Rosenfeld und Thornton (1976) haben zum aufgabenanalytischen Vorgehen eine Checkliste zur Durchführung von Interviews für Managertätigkeiten erarbeitet. Weitere Methoden wären Selbstbeobachtungen (Tagebuchmethode) durch den jeweiligen Stelleninhaber, Beobachtungen durch externe Beobachter oder die in diesem Kontext klassisch zu nennende Methode der kritischen Ereignisse nach Flanagan (1949). Hierbei werden in einem ersten Schritt jene Tätigkeiten (kritischen Ereignisse) beschrieben, die zu einem besonders erfolgreichen oder besonders wenig erfolgreichen Arbeitsverhalten geführt haben. Neben der Skalierung werden im zweiten Schritt die zur Bewältigung der kritischen Ereignisse notwendigen Verhaltensweisen benannt. In einem dritten Schritt werden dann die konkreten Verhaltensbeschreibungen als Anforderungen abgeleitet.

Personalauswahl ist ein prädiktiver Prozess, d.h. die gegenwärtigen Merkmale eines Bewerbers werden zur Vorhersage der Merkmale der späteren Arbeitskraft genutzt. Erfassungsmethoden berufsrelevanter psychologischer Voraussetzungen werden unter Punkt 2.1.3 näher beschrieben. Diese sogenannten Prädiktoren dienen als Entscheidungsgrundlage in der Personalauswahl, denn mit den vor oder während der Einstellung erhobenen Personenmerkmalen soll das spätere Verhalten, insbesondere der Erfolg eines Teilnehmers am Auswahlverfahren, in der Ausbildung oder am Arbeitsplatz, vorhergesagt werden. Nach Maukisch (1978) sind Prädiktoren diagnostische Verfahren, die Persönlichkeitsmerkmale beliebiger Art in irgendeiner Weise operationalisieren und für die Vorhersage eines Kriteriums eingesetzt werden. Hossiep (1995) erweitert diese Definition dahingehend, dass nicht nur Persönlichkeitsmerkmale, sondern auch andere Variablen zur Prognose des Berufserfolges herangezogen werden können. Wichtige Voraussetzung hierfür ist die Stabilität der Merkmale. Ohne sie wäre eine Vorhersage nicht möglich. Rapide Veränderungen ergeben sich bei Persönlichkeitsmerkmalen vorrangig in der Kindheit und im Alter. Merkmale, die bei der Personalauswahl von Interesse sind, in der Regel in der Altersgruppe zwischen 20 und 50 Jahren, bleiben jedoch im Laufe ihres Arbeitslebens relativ

stabil (Schmidt und Hunter, 1998). Beispielhaft verweisen die Autoren auf die allgemeine kognitive Fähigkeit, die sogenannte Intelligenz, als ein wesentliches und stabiles Merkmal eines Bewerbers. Langzeitstudien zeigen bei standardisierten Intelligenztests Retest-Reliabilitäten um $r = .81$.

Sollen die Konsequenzen von Personalentscheidungen als zutreffend bzw. „falsch“ oder „richtig“ klassifiziert werden können, so muss ein Maßstab anlegbar sein, an dem die Qualität der Entscheidungen bewertet werden kann. Bei dieser Übersetzung der Zielvorstellung eines Personalauswahlverfahrens in Verhaltensdimensionen und damit der Definition von Kriterien müssen die gleichen methodischen Anforderungen wie bei den Prädiktoren gelten, nämlich Objektivität, Reliabilität und Diskriminationsfähigkeit. Die Validität von Kriterien kann nach Hossiep (1995, S. 50) jedoch nicht empirisch geprüft werden. Er zitiert Hoyos (1956) der bemerkt: „Psychologische Prüfverfahren sind Konstruktionen, die an der Wirklichkeit geprüft werden, während die Kriterien die Wirklichkeit selbst sind. Bei Kriterien muss demnach gefragt werden, ob sie die wichtigsten und bedeutsamsten Aspekte einer Tätigkeit repräsentieren, d.h. für eine Tätigkeit relevant sind“. Nach Maukisch (1978) verstehen wir unter Kriterien Verhaltensweisen und Verhaltensergebnisse, die einer sozialen oder individuellen Bewertung unterliegen und geeignet sind, den Grad und die Art der Wertverwirklichung, der durch psychologische Intervention erreicht wird, zu erfassen. Kriterien lassen sich nach verschiedenen Gesichtspunkten klassifizieren. So wird beim Zeitpunkt der Messung unterschieden, ob das Kriterium ein Verhalten unmittelbar nach der Einstellung erfasst oder erst nach einem längeren Zeitraum des Verbleibs in der Organisation. Bei der Spezifität der Messung wird getrennt, ob das Kriterium ein konkretes Verhalten am Arbeitsplatz abbildet oder eher das Gesamtverhalten des Arbeitnehmers operationalisiert. Rösler (1992) unterscheidet weiter zwischen „harten“ und „weichen“ Kriterien. Als harte Kriterien werden dabei die der direkten Beobachtung zugänglichen Verhaltensweisen wie Fehlzeiten, Kündigungen, Unfälle etc. bezeichnet. Weiche Kriterien sind bspw. Beurteilungen des Verhaltens einzelner Arbeitnehmer durch Vorgesetzte oder Kollegen.

Dabei sollte nicht außer acht bleiben, dass auch sogenannte harte Kriterien subjektiven Wertungen unterliegen, denn bspw. wäre bei den Fehlzeiten auch zu bewerten, ob es sich um vermeidbares vs. nicht vermeidbares oder entschuldigtes vs. unentschuldigtes Fehlen handelt. Mittels einer Validitätsanalyse wird im nächsten Schritt einer wissenschaftlich begründeten Personalauswahl erfasst, in welchem Umfang ein Kriterium durch einen Prädiktor vorhersagbar ist. Nur wenn der Nachweis über einen Zusammenhang zwischen Prädiktor und

Kriterium vorliegt, ist es sinnvoll aufgrund der Prädiktorinformation zwischen geeigneten und weniger geeigneten Bewerbern um einen Ausbildungsplatz oder eine Stelle in einer Organisation zu entscheiden. Die entsprechende Kennzahl wird abgebildet durch die Korrelation zwischen den beiden Merkmalen, dem Validitätskoeffizienten. Zu unterscheiden ist hier zwischen der prädiktiven und konkurrenten Validität. Im ersten Fall wird der Prädiktor zum Zeitpunkt der Auswahl erhoben und das Kriterium erst zu einem späteren Zeitpunkt bspw. nach Abschluss der Ausbildung oder mehrjähriger Bewährung im Beruf. Bei der konkurrenten Validität werden dagegen Prädiktor und Kriterium zum gleichen Zeitpunkt an einer Stichprobe von bereits ausgewählten Personen erhoben. Wird der Prädiktor nicht schon als Selektionsinstrument genutzt, d.h. alle Bewerber werden eingestellt, kann der Zusammenhang zwischen Prädiktor und Kriterium durch die prädiktive Validität unverfälscht wiedergegeben werden. Dagegen kann der konkurrente Validitätskoeffizient zu einer verzerrten Annahme über den Zusammenhang führen. Bei einer bereits eingestellten Gruppe können sowohl Prädiktor wie auch Kriterium in der Varianz eingeschränkt sein (selegierte Stichprobe). Bereits am Arbeitsplatz gemachte Erfahrungen der Mitarbeiter können die erhobenen Prädiktoren beeinflussen und sich im Laufe der Arbeitszeit als ungeeignet herausgestellte Arbeitnehmer sind möglicherweise schon entlassen worden.

Widersprüchliche Ergebnisse unterschiedlicher Validitätsstudien unterstützten bis in die siebziger Jahre hinein die Annahme, dass sich die Validität eines Testverfahrens in verschiedenen Situationen unterscheidet. Validitätsschätzungen für dasselbe Verfahren und für denselben Beruf unterschieden sich erheblich in verschiedenen Untersuchungen. Diese sogenannte Theorie der situationalen Spezifität blieb aufrecht, bis gezeigt werden konnte, dass die meisten Unterschiede zwischen den Studien auf statistische Artefakte und Messfehler zurückzuführen waren. Häufig waren zu kleine Stichproben ursächlich und nicht tatsächliche Differenzen zwischen den Arbeitsplätzen (Schmidt und Hunter, 1977; Schmidt, Hunter, Pearlman & Shane, 1979). Nach Schuler (1996) liegt die Höhe eines durchschnittlichen Korrelationskoeffizienten in der Berufseignungsdiagnostik im Bereich von $r = .30$. Ein Wert im Bereich von $r = .50$ ist für die Prognose beruflichen Erfolges schon als sehr gut zu bezeichnen. Höhere Werte lassen sich nach Meinung des Autors nur erzielen, wenn mit einem multiplen Verfahren auf der Basis gründlicher Arbeitsanalysen vielfältige Anteile der Tätigkeitsanforderungen erfasst wurden.

Zu welchem Informationsgewinn eine Korrelation zwischen zwei dichotomen Merkmalen im Bereich von $r = .30$ führt, lässt sich nach Moser (1995) anhand des „binomial effect size display“ zeigen. Bei einer Korrelation von $r = .00$ wäre eine zufällige Zellenbesetzung in allen vier Zellen von 50 % zu erwarten. Wie in der Tabelle 2.2-1 dargestellt, kann bei einer Korrelation von $r = .30$ und vorliegendem A das Merkmal B in fast doppelter Häufigkeit erwartet werden als wenn A nicht vorläge.

Tab. 2.2-1: Binomial effect size display nach Moser (1995)

Binomial effect size display		B	
		liegt vor	liegt nicht vor
A liegt vor	65	35	
A liegt nicht vor	35	65	

Diese Relation wächst in Abhängigkeit von der zunehmenden Höhe der Korrelation auf das Dreifache (75 : 25) bei $r = .50$ und Vierfache (80 : 20) bei $r = .60$.

Der nächste Schritt in der Ablaufbeschreibung wissenschaftlich begründeter Personalauswahl, ist nach Rösler (1992) die Frage, nach welcher Strategie eine Entscheidung über geeignete bzw. nicht geeignete Bewerber zu treffen ist. Dabei ist immer zu berücksichtigen, dass Entscheidungen grundsätzlich nicht ohne das Risiko einer falschen Annahme bzw. Ablehnung getroffen werden können, da nicht von einer vollständigen Gültigkeit der diagnostischen Instrumente ausgegangen werden kann. Cronbach und Gleser (1965) verweisen auf 64 verschiedene Arten diagnostischer Entscheidungen, von denen viele in der Praxis jedoch nur eine untergeordnete Bedeutung besitzen. Fragen, ob das Unternehmen alle Personen nach einem standardisierten Verfahren testet (institutionelle Entscheidung), ob Annahmehquoten für eine das Angebot übersteigende Nachfrage festgelegt werden müssen oder nach welcher Klassifikation die ausgewählten Probanden welchen Aufgabenfeldern zugewiesen werden, müssen beantwortet werden.

Amelang und Zielinski (1994) unterscheiden im diagnostischen Prozess zwischen einer nicht sequentiellen und einer sequentiellen Vorgehensweise in der Personalauswahl. Bei der sogenannten nichtsequentiellen Batterie wird allen Probanden das gesamte Testprogramm vorgegeben und diejenigen ausgewählt, die den - optimal gewichteten - höchsten Summenwert erreicht haben. Ähnlich dem sogenannten „single screen“, nur dass hier alle Entscheidungen auf der Grundlage eines Testergebnisses fußen. Dagegen sind bei einer sequentiellen Vorgehensweise drei Grundmuster möglich. In der „pre-reject-Strategie“ werden alle Probanden, die zu einem festen Zeitpunkt des Verfahrens einen bestimmten Wert

nicht erreichen konnten, vom weiteren Verfahren ausgeschlossen. Die restlichen Teilnehmer absolvieren weitere Testteile. Die Entscheidung wird auf der Grundlage des Gesamtergebnisses beider Verfahrensstufen getroffen. Bei der „pre-accept-Strategie“ werden die Teilnehmer, die nach einem bestimmten Teil des Verfahrens einen gesetzten Wert überschritten haben, akzeptiert. Mit den verbleibenden Probanden werden dann weitere Verfahrensschritte durchgeführt. Eine vollständige sequentielle Strategie besteht aus einer Kombination beider Vorgehensweisen. Nach einem gesetzten Testabschnitt erfolgt die Aufteilung der Teilnehmer in drei Gruppen. Eine wird akzeptiert und eingestellt, eine andere definitiv abgewiesen und die dritte weiteren Auswahlstufen unterzogen. Der Einsatz unterschiedlich zusammengesetzter Testbatterien ist aber in der Praxis nicht immer ohne Folgen.

In diesem Kontext ist das „bandwidth-fidelity“-Dilemma (Schuler & Höft, 2001) erwähnenswert, das durch Cronbach (1990) publik wurde. Kommen eine Vielzahl von Verfahren zum Einsatz, um einen möglichst breiten Überblick über die Fähigkeiten und Eigenschaften der Bewerber zu gewinnen, geht dies häufig zu Lasten der Detailtiefe, da spezifische Einzelinformationen nicht mehr erfasst werden können. „Band with“ bezeichnet in diesem Zusammenhang die Komplexität und Bandbreite der Verfahren, „fidelity“ die Genauigkeit. Die Autoren verweisen zur Lösung des Problems auf die Symmetriehypothese von Cronbach und Gleser (1965), nach der, wenn ein eng umgrenztes Kriterium vorhergesagt werden soll, auch ein spezifischer Prädiktor einzusetzen ist. Ist dagegen ein allgemeines Globalmaß das Kriterium, so muss auch der Prädiktor ein vergleichbares Abstraktionsniveau aufweisen.

Die zentrale Frage bei der Selektion zwischen geeigneten und ungeeigneten Bewerbern in einem Personalausleseprozess zielt auf die Vermeidung von Entscheidungsfehlern. Fehler treten immer dann auf, wenn die Auswahl aufgrund von Prädiktorinformationen nicht mit der tatsächlich vorhergesagten Prädiktion übereinstimmt. Für den Fall von zwei möglichen Klassen (geeignet vs. nicht geeignet) lassen sich vier mögliche Kombinationen von Übereinstimmung und fehlender Übereinstimmung in Anlehnung an Kallus und Jahnke (1992) beschreiben, die in Tabelle 3.2-1 illustriert werden.

Somit lassen sich generell zwei Arten von Zuordnungsfehlern unterscheiden. Zum einen der Fehler erster Art: falsch Positive, d.h. Teilnehmer an einem Auswahlverfahren werden als

geeignet bezeichnet, obwohl sie ungeeignet sind. Zum anderen der Fehler zweiter Art: falsch Negative, d.h. die eigentlich geeigneten Teilnehmer werden als ungeeignet klassifiziert.

Tab. 3.2-1: Arten richtiger und falscher Zuordnung nach Kallus & Jahnke (1992)

		Auswahl:		Zuordnung:	
		geeignet		ungeeignet	
tatsächlich: geeignet	richtige Zuordnung	richtig als geeignet ausgewählt	Risiko: $1 - \alpha$	falsche Auswahl	fälschlich als bezeichnete Geeignete
					Risiko: α
tatsächlich: ungeeignet	falsche Auswahl	fälschlich als geeignet ausgewählte Ungeeignete	Risiko: β	richtige Auswahl	richtig als ungeeignet ausgewählte Ungeeignete
					Risiko: $1 - \beta$

Die Effizienz diagnostischer Entscheidungen lässt sich im Verhältnis der Anteile von richtig klassifizierten Positiven, falsch klassifizierten Positiven, falsch klassifizierten Negativen und richtig klassifizierten Negativen darstellen (vgl. auch bspw. Amelang & Zielinski, 1994).

Darüber hinaus ist für die Beurteilung der Effizienz eines Auswahlverfahrens insbesondere das Zusammenspiel von Grundquote, Selektionsrate und Prädiktorvalidität bedeutsam. Möchte man den Nutzen bestimmen, den ein eignungsdiagnostisches Verfahren für die Organisation besitzt, so ist der Validitätskoeffizient allein kein ausreichendes Maß. Die Auswahl von Bewerbern um einen Ausbildungsplatz anhand eines Prädiktors ist nur dann sinnvoll, wenn man unter den später Eingestellten mehr für die Ausbildung geeignete als ungeeignete finden kann und zudem das Verhältnis Geeigneter vs. Ungeeigneter besser ausfällt, als bei Nichtverwendung dieses Prädiktors. Taylor und Russel (1939) haben dieses Zusammenspiel tabellarisch anschaulich dargestellt. Sie wiesen erstmalig auf das Zusammenspiel von Grundquote, d.h. der Anteil der Geeigneten unter den Bewerbern und Selektionsquote, d.h. der Anteil der Ausgewählten unter den Bewerbern hin. Aus ihren veröffentlichten Tafeln lässt sich bei gegebener Validität und Grundquote ablesen, wie viel Geeignete man bei einer bestimmten Selektionsrate tatsächlich unter den Ausgewählten finden wird. In Tabelle 4.2-1 wird an drei verschiedenen Beispielen von Schuler (1996) anschaulich das Zusammenwirken der unterschiedlichen Parameter illustriert.

Deutlich wird aufgrund der Beispiele, dass die Verwendung eines validen Einstellungsverfahrens um so wichtiger wird, je geringer die Anzahl der Geeigneten unter den Bewerbern und je geringer die Selektionsquote ist. Ist ein Unternehmen gezwungen, jeden Bewerber einzustellen, dann hat ein Auswahlverfahren keinen praktischen Wert. Kann sich andererseits die Organisation leisten, nur die besten 1% der Bewerber einzustellen, dann ist der praktische Wert eines solchen Verfahrens durch die Selektion pro eingestellter Person relativ groß. Nach

Schmidt und Hunter (1998) liegt die Selektionsrate normalerweise etwa zwischen $r = .30$ und $.70$, einem Bereich also, der für eine Personalauswahl erheblichen praktischen Nutzen verspricht.

Tab. 4.2-1: Zusammenwirkungen von Grundquote, Selektionsrate und Prädiktorvalidität nach Schuler (1996)

Freie Stellen im Unternehmen = 50						
Anzahl der geeigneten Bewerber = 90 von 100						
Validität des Auswahlverfahrens				r = .00	r = .35	r = .65
Prozentsatz der	Geeigneten	unter	den	90 %	95 %	98 %
Eingestellten						
Freie Stellen im Unternehmen = 50						
Anzahl der geeigneten Bewerber = 50 von 100						
Validität des Auswahlverfahrens				r = .00	r = .35	r = .65
Prozentsatz der	Geeigneten	unter	den	50 %	61 %	73 %
Eingestellten						
Freie Stellen im Unternehmen = 10						
Anzahl der geeigneten Bewerber = 50 von 100						
Validität des Auswahlverfahrens				r = .00	r = .35	r = .65
Prozentsatz der	Geeigneten	unter	den	50 %	74 %	92 %
Eingestellten						

Cronbach und Gleser (1965) erweiterten den dargestellten Ansatz um den monetären Aspekt, denn der finanzielle Wert einer Verbesserung lässt sich auch in Geldeinheiten ausdrücken. Hier müssen zu den oben notwendigen Informationen weitere Faktoren wie bspw. die Kosten des Auswahlverfahrens berücksichtigt werden. Barthel und Schuler (1989) haben an einem realen Beispiel, der Auswahl von Außendienstmitarbeitern in einer Versicherungsgesellschaft, den Nutzen des Einsatzes eines biographischen Fragebogens berechnet. Immerhin ergab sich in diesem Fall aufgrund der Kalkulation nach einem Modell von Boudreau (1983) ein Nutzensbetrag von DM 473.265. Cascio (1991) zeigte, welche Kosten durch Kündigungen auf ein Unternehmen zu kommen und belegte eindrucksvoll die enormen Beträge, die falsche Personalentscheidungen erzeugen können. Als Grundlage seiner fiktiven Berechnung, wie sie in Tabelle 5.2-1 dargestellt wird, wurde ein 200-Betten-Krankenhaus mit 1200 Angestellten und einer monatlichen Kündigung von 2% herangezogen.

Gerade praktisch-diagnostisch tätige Psychologen verweisen nach Hossiep (1995) immer wieder auf die „Lächerlichkeit“, aus Kostengründen auf wissenschaftlich fundierte Auswahlmethoden zu verzichten. Denn es können, wie bspw. Marschner (1980) betont, die Kosten psychologisch fundierter Auswahlverfahren bereits bei der Senkung etwa der Durchfallrate von Auszubildenden in der Lehrabschlussprüfung um wenige Prozent eingespart werden. Zudem können heute, angesichts des technologischen Fortschritts, in allen Phasen der Datenerhebung und -verarbeitung bei der beruflichen Eignungsdiagnostik computergestützte Verfahren Anwendung finden, die zur weiteren Kostensenkung beitragen können.

Tab. 5.2-1: Fiktive Kostenrechnung eines Unternehmens nach Cascio (1991)

1. Kosten der Trennung	29679 \$
- Abschlussgespräch	- 6869 \$
- Administrative Kosten	- 3462 \$
- Trennungsentschädigung	- 135936 \$
- Arbeitslosigkeitssteuern	- 150528 \$
Kosten durch Neueinstellung	1519949 \$
- Suche nach neuen Mitarbeitern	- 82385 \$
- Administrative Kosten	- 17309 \$
- Auswahlinterviews	- 10385 \$
- Durchführung von Testverfahren	- 34560 \$
- Abteilungsbesprechungen	- 3266 \$
- Umzugs- und Reisekosten	- 1339200 \$
- Einführung des neuen Mitarbeiters	- 62924 \$
- Medizinische Untersuchung	- 25920 \$
Kosten durch Trainingsmaßnahmen	321800 \$
- Informationsmaterialien	- 2880 \$
- Trainingsprogramme	- 137480 \$
- Einarbeitung des Mitarbeiters	- 181440 \$
Gesamtkosten	2138543 \$
Kosten pro Mitarbeiter	7426 \$

Gerade angesichts steigender Personalkosten und nach wie vor hohem Bedarf an geeigneten Bewerberinnen und Bewerbern ist eine wissenschaftlich fundierte Personalauswahl heute wichtiger denn je. Die Verwaltungsmodernisierung der Öffentlichen Haushalte ist nicht zuletzt ein Beleg dafür, dass auch hier erkannt wurde, was unter betriebswirtschaftlichen Gesichtspunkten durch die freie Wirtschaft schon längst praktiziert wird. Der Aufwand der notwendig ist, um geeignete Interessenten für eine vakante Position auswählen zu können, ist auch unter wirtschaftlichen Aspekten in Relation zu dem Nutzen zu sehen, den er der Organisation verspricht. Wenn ausreichend Bewerber für eine einfach erlernbare Tätigkeit zur Verfügung stehen, ist ein aufwändig durchzuführendes Auswahlverfahren sicherlich nicht zu rechtfertigen. Angesichts der zunehmenden Komplexität vieler Berufe, der davon ableitbaren umfangreicheren Anforderungen und nicht zuletzt der Erkenntnis, dass Mitarbeiter, einmal umfangreich und teuer ausgebildet, ein Berufsleben lang weitere Investitionen für die Organisation rechtfertigen können müssen – Stichworte sind hier Human Ressource und strategische Personalentwicklung – ist heute und vor allem zukünftig eine professionelle Personalauswahl nicht nur im Interesse von Organisation und Bewerber, sondern fast schon betriebswirtschaftlich zwingend notwendig.

2.1.2 Eignungsdiagnostische Verfahren zur Personalauswahl

Die Darstellung eignungsdiagnostischer Verfahren in der Personalauswahl erfolgt an dieser Stelle nur im Kontext als kurzer Überblick. Detaillierte Übersichten über verbreitete Verfahrensweisen zur Auswahl externer und interner Bewerber und Verweise auf

weiterführende Literatur findet der interessierte Leser bspw. bei Sarges (1995a), Schuler (1996, 2001) oder Weinert (1998). Für einen umfangreichen Überblick neuerer wirtschaftspsychologischer Testverfahren sei auf Sarges & Wottawa (2001) verwiesen. Hier findet sich insbesondere eine übersichtliche Zuordnung getesteter Merkmale zu unterschiedlichsten Testverfahren und die Darstellung verschiedener Einsatzbereiche. Exemplarisch werden an dieser Stelle einige der gebräuchlichsten Auswahlverfahren der Berufseignungsdiagnostik vorgestellt.

Für eignungsdiagnostische Entscheidungen kommen praktisch alle über Mitarbeiter oder Bewerber verfügbare Datenquellen in Frage. Nach Hossiep (1995) werden personelle Auswahlentscheidungen in der Wirtschaft jedoch nur in wenigen Fällen mit Hilfe fundierter eignungsdiagnostischer Methoden getroffen. Er schätzt den Anteil professioneller psychologischer Beteiligung bei Auswahlentscheidungen rein quantitativ auf nur gut 10 %.

Eine „Hitliste“ der 20 meist verwendeten Testverfahren in der diagnostischen Praxis findet sich u.a. bei Schorr (1991), daran dürfte sich auch heute wenig geändert haben. Wie aus Tabelle 6.2-1 zu entnehmen, stehen an erster Stelle bei der Personalauswahl, von insgesamt 46,8% der befragten Arbeits- und Organisationspsychologen in der Berufspraxis angewandt, die Intelligenztests. Weiterhin werden Persönlichkeitstests, klinische und projektive Testverfahren angewandt und in der Übersicht finden sich auf Platz drei „spez. Funktionsüberprüfungs- und Eignungstests“ sowie auf Platz fünf „allgemeine Leistungstests“:

Tab. 6.2-1: Verwendung von Testverfahren durch Arbeits- und Organisationspsychologen nach Schorr (1991)

Verfahren	Anwendung durch A & O Psychologen
- Intelligenztest	46,8 %
- Persönlichkeitstest	31,2 %
- Spezielle Funktions- und Eignungstests	24,7 %
- Klinische Testverfahren	23,4 %
- Allgemeine Leistungstests	14,3 %
- Projektive Verfahren	5,2 %

Allerdings ist aus gleicher Quelle zu erfahren, dass 14,3 % der Wirtschaftspsychologen die Verwendung von Intelligenztests ablehnen. Dies ist insbesondere bemerkenswert, da nur 11,7 % sich gegen die Anwendung von Klinischen Testverfahren aussprechen, die in der Regel nicht für den Gebrauch in der Arbeits- und Organisationspsychologie konzipiert sind und für andere Anwendungsfelder normiert wurden.

Eine Übersicht, welche der wichtigsten Testverfahren sich im eignungsdiagnostischen Kontext als nützlich erwiesen haben, geben Schmidt & Hunter (1998) in Tabelle 7.2-1 .

Tab. 7.2-1: Prädiktive Validität von eignungsdiagnostischen Prädiktoren nach Schmidt & Hunter (1998)

Eignungsdiagnostische Instrumente	Validität
- Intelligenztests	r = . 51
- Arbeitsproben	r = . 54
- Interview (strukturiert)	r = . 51
- Interview (unstrukturiert)	r = . 38
- Biographische Daten	r = . 35
- Assessment Center	r = . 36
- Interessen	r = . 10
- Test zum Berufswissen	r = . 48
- Alter	r = -. 01

Zur generellen Problematik hinsichtlich der Aussagekraft und Bewertung von vorliegenden Validitätsergebnissen in der Berufseignungsdiagnostik sei der interessierte Leser bspw. auf Hossiep (1995) verwiesen. In Abbildung 7 werden Resultate veröffentlichter Metaanalysen aus 20 Jahren Forschung wiedergegeben, in welchen die prädiktive Validität eignungsdiagnostischer Instrumente untersucht wurde.

2.1.2.1 Bewerbungsunterlagen

Im unter Punkt 2.1.3 beschriebenen Auswahlverfahren für den Polizeivollzugsdienst spielt die Analyse von Bewerbungsunterlagen bei der Vorauswahl von Bewerbern eine besondere Rolle, so dass an dieser Stellen darauf ausführlicher eingegangen wird.

2.1.2.1.1 Vorauswahl und Bewerbungsunterlagen

In unterschiedlichen Berufsgruppen geben 80 – 90 % der deutschen Unternehmen an, die Bewerbungsunterlagen zur externen Personalauswahl heranzuziehen (Schuler, Frier und Kaufmann, 1993). Die erste und wichtigste Funktion bei der Sichtung von Bewerbungsunterlagen liegt, nach Hollmann und Reitzig (1995), in der Überprüfung von erfüllten grob definierten Voraussetzungen. Dazu zählen bspw. Alter, Zertifikate, Zeugnisse, gesundheitliche Voraussetzungen o.ä.. Die Vorauswahl trägt unter Einbeziehung auch formaler Aspekte zu einer ersten Reduzierung der Bewerbungen bei. Die Autoren beschreiben die Prüfung der Bewerbungsunterlagen häufig als eine Art „Schlechtesten-Auslese“; so dass nur ein Teil der Bewerber die nächste Auswahlstufe erreicht. Damit scheidet aus Bewerberpersicht die Stellensuche oft auf der Grundlage ihrer Unterlagen. Insoweit ist das oberflächliche Screening der Bewerbungsunterlagen Gegenstand eines für viele Bewerber wenig nachvollziehbaren Vorauswahlprozesses. Es dient in erster Linie dazu, ein große

Anzahl von Bewerbungen auf jene Zahl zu reduzieren, die in einer nächsten Phase zu bewältigen sind. Die hierfür von den Praktikern herangezogenen Hypothesen sind jedoch selten wissenschaftlich überprüft (Rastetter, 1999).

Neben dem Anschreiben wird weiter dem Lebenslauf besondere Bedeutung beigemessen. Enthaltene Fakten werden analysiert und dienen als Grundlage weiterer Interpretationen. Ziel ist es, auf ein Bild vom Bewerber zu schließen. Wie Schuler (1996) anmerkt, werden die im Lebenslauf enthaltenden Daten und Fakten einerseits als Belege verarbeitet und gelten andererseits als Indikatoren künftiger Bereitschaft und Verhaltensmuster. Hollmann und Reitzig (1995) verweisen bspw. auf Indikatoren im Lebenslauf (Auslandsaufenthalt, Vielfalt der bisherigen Stellen, außerberufliche Aktivitäten) die Analogieschlüsse auf Faktoren wie Selbstsicherheit, Selbstständigkeit oder Durchsetzungs- und Entscheidungsvermögen erlauben sollen. Statistisch sind solche Zusammenhänge jedoch nicht abgesichert.

Arbeitszeugnisse weisen dagegen offensichtlich einige Schwächen auf, die ihre Verwendbarkeit in der Personalauswahl einschränken. Reilly und Chao (1982) ermittelten für derartige Zeugnisse eine durchschnittliche Validität von $r = .14$. Insbesondere die Wohlwollenspflicht des Arbeitgebers, nach der keine ausgesprochen negativen Formulierungen Verwendung finden dürfen, sollte hierzu ihren Beitrag leisten. Zudem ist nicht auszuschließen, dass gute Zeugnisse auch dazu dienen, sich von einem unbequemen Arbeitnehmer zu trennen. Allein die genannte Wohlwollenspflicht lässt jedoch eine valide Beurteilung kaum zu. So entschied bspw. das Hamburger Landesarbeitsgericht 2002², dass ein Unternehmen dem neuen Arbeitgeber seines ehemaligen Mitarbeiters keine über das Zeugnis hinausgehenden Informationen zukommen lassen darf. Im vorliegenden Fall hatte die alte Firma mitgeteilt, dass sie nicht in der Lage sei, zusätzliche Informationen zu geben. Sie könne aber bestätigen, dass das vorliegende Zeugnis im Verlauf eines arbeitsgerichtlichen Prozesses erlangt worden sei. Da das zukünftige Unternehmen nach dieser Auskunft auf eine Einstellung verzichtete, wurde der ehemalige Arbeitgeber zur Zahlung eines entsprechenden Schadensersatzes verurteilt.

Eine fehlende Standardisierung lässt zudem, im Gegensatz zu Schulzeugnissen, eine Vergleichbarkeit fast nicht möglich erscheinen. Dagegen konnten Moser und Rhyssen (2001) für Referenzen eine prognostische Validität von $r = .20$ feststellen. Dabei zeigte ein Vergleich zwischen Ergebnissen der durch die Arbeitnehmer vorgelegten Referenzen und einem mit ihnen durchgeführten strukturierten Interview, dass die Prognose für Positionen ohne Führungsaufgaben in der Studie valider war als umgekehrt.

² 2 Sa 144/83

2.1.2.1.2 Vorauswahl und Schulnoten

Als wichtigster Einzelprädiktor künftiger Erfolge werden häufig die Schulnoten herangezogen. Sie gelten allgemein als valider Prädiktor des späteren Ausbildungserfolges. Bei berufserfahrenen Bewerbern werden dagegen Schulnoten und Zeugnisse kaum mehr berücksichtigt. Schon nach Althoff (1986) sind Schulzeugnisse eines der beliebtesten Kriterien für Personalentscheidungen. Wenn man heute die Vorauswahlkriterien von Unternehmen und öffentlichen Arbeitgeber reflektiert, dann stehen die Schulnoten immer noch an der Spitze, wenn es zur Entscheidung über „Einladung oder Ablehnung“ kommt. Bei diesen Analogieschlüssen (Rastetter, 1999) wird in der Praxis nicht nur von guten Noten in einem Fach auf die Kenntnisse in eben diesem Fach geschlossen, sondern weiter auf Intelligenz, Lernfähigkeit, Anpassungsfähigkeit, Belastbarkeit u.ä. und daraus wiederum auf den späteren Berufserfolg gefolgert. Nach Behrens und Merkel (1989) bedeuten bspw. gute Noten im Sport Einsatzfreude, in Singen und Werken Anpassungsfähigkeit und in Fremdsprachen Willenseinsatz. Bei Berthel (1995) gelten gute Noten als ein generelles Indiz für Engagement und Anpassungsbereitschaft. Für eine statistisch signifikante Beziehung zwischen Schulnoten und diesen Konstrukten fehlen allerdings die Nachweise. Dabei ist die Voraussetzung für eine sinnvolle Einschätzung von Schulnoten, die Kenntnis über die Notenverteilung in der Klasse oder der Schule, in der Regel nicht erfüllt.

Baron-Boldt, Funke und Schuler (1989) berichten für Schulnoten von einer prädiktiven Validität von $r = .41$ für den Ausbildungs- und $r = .46$ für den Studienerfolg. Validester Einzelprädiktor war in beiden Fällen die Mathematiknote. Über die Verlässlichkeit von Schulnoten liegen jedoch unterschiedliche Untersuchungsergebnisse vor. So fand Althoff schon 1986, dass außer der Mathematiknote, die Schulnoten von Abiturienten und Realschülern weder mit Ergebnissen eines durchgeführten Intelligenztestes noch mit denen des Ausbildungserfolges signifikant korrelierten. Er verglich u.a. die Ergebnisse von 120 Abiturienten und 96 Realschülern, die an einer Eignungsuntersuchung für den gehobenen bzw. mittleren Dienst der Kommunalverwaltung teilnahmen und die im Vorfeld nicht nach Zeugnisnoten vorausgewählt worden waren. Zwar konnte er feststellen, dass Absolventen von allgemeinbildenden Gymnasien hochsignifikant bessere Ergebnisse erzielten als Schüler anderer Schulformen, fand jedoch insgesamt, außer bei der Mathematiknote, keine signifikanten Zusammenhänge zwischen Schulnoten und den Ergebnissen der psychologischen Eignungsuntersuchung. Er kommt aufgrund seiner Ergebnisse zu dem

Schluss, dass insbesondere bei der Vorauswahl der Rückgriff auf ein Losverfahren gerechter sei als der auf die Schulzeugnisse. Auch Hollmann und Reitzig (1995) kommen zu einer ähnlichen Bewertung: Die Studien zur prädiktiven Validität von Zeugnisnoten weisen zwar gelegentlich erstaunlich hohe Korrelationen auf, das Gesamtbild insgesamt sei jedoch zu widersprüchlich, als dass ihre Verwendung empfohlen werden kann. In wie weit die hohe Varianz vorliegender Ergebnisse auf unterschiedliche Bedingungen zurückzuführen ist oder auf Stichprobenfehler und andere Artefakte kann mittels metaanalytischer Methoden festgestellt werden. Zur Validität von Schulnoten liegen zwischenzeitlich mehrere in- und ausländischen Metaanalysen vor, die Schuler (1998) zusammenfassend wiedergibt. In der Untersuchung von Baron-Boldt et al. (1989) wurden 15 Einzelstudien mit 27 unabhängigen Stichproben und 2555 Teilnehmern zusammengefasst. Die artefaktkorrigierte durchschnittliche Validität von Schulnoten bezüglich des Ausbildungserfolges lag bei $r = .41$. Die Gesamtnote sagt den Ausbildungserfolg besser voraus als jede Einzelnote, wobei validester Einzelprädiktor auch hier die Mathematiknote war ($r = .30$).

2.1.2.1.3 Schulnoten im Auswahlverfahren für den Polizeivollzugsdienst

Schulnoten stellen in unterschiedlichen Kombinationen in vielen Bundesländern bei der Auswahl von Bewerbern für den Polizeivollzugsdienst ein wichtiges Auswahlkriterium dar. Teilweise ist sogar das Vorliegen eines bestimmten Notendurchschnittes für die Annahme von Bewerbungen grundsätzlich erforderlich. Aufgrund dessen wurden von Petersen (2002) vorliegende Schulzeugnisse und die Ergebnisse im Auswahlverfahren für den Polizeivollzugsdienst miteinander verglichen. In dieser Untersuchung wurden die Zeugnisnoten von insgesamt 689 Teilnehmern am Auswahlverfahren für den gehobenen und 156 für den mittleren Polizeivollzugsdienst zu ihren erreichten Ergebnissen im Einstellungsverfahren in Beziehung gesetzt. Im Vorfeld fand keine Vorauswahl aufgrund der Zeugnisnoten oder anderer Kriterien statt, so dass eine Vorselektion vermieden werden konnte. An dieser Stellen wird aus ökonomischen Gründen nur auf die Daten der Teilnehmer eingegangen, die aufgrund der gegebenen sequentiellen Auswahlstrategie an allen Stufen des Auswahlverfahrens teilgenommen haben. Die Stichproben setzten sich wie folgt zusammen:

Im Verfahren für den mittleren Polizeivollzugsdienst konnten die Gesamtergebnisse von 87 Pbn. ausgewertet werden. Es nahmen 24,1 % ($n = 21$) Teilnehmer aus Hamburg am Verfahren teil, die restlichen 75,9 % ($n = 66$) Bewerber kamen aus fast allen anderen

deutschen Bundesländern. Das Alter der Bewerber lag zwischen dem 15. und dem 32. Lebensjahr, dabei hatten über 90 % der Bewerber das 28. Lebensjahr noch nicht vollendet, insgesamt 81,6 % waren jünger oder gleich 25 Jahre. Der Altersmittelwert lag bei 21,08 Jahren.

74 Bewerber bewarben sich um Einstellung in die Schutzpolizei, 13 in die Wasserschutzpolizei. Es nahmen insgesamt 51 Männer (58,6 %) und 36 Frauen (41,4 %) an der Erhebung teil. 9 Teilnehmer (10,3 %) waren nicht deutscher Herkunft. Die Schulbildung der Teilnehmer wird in Tabelle 8.2-1 wiedergegeben.

Tab. 8.2-1: Schulbildung der Teilnehmer im mittleren Dienst nach Petersen (2002)

Schulbildung	Anzahl	%
- Hauptschulabschluss	1	1,1
- Realschulabschluss	38	43,7
- Fachhochschulreife	18	20,7
- Hochschulreife	21	24,1
- andere	9	10,3
n =	87	99,9

* in der Regel einem Realschulabschluss gleichwertige Schulbildung

Im Verfahren für den gehobenen Polizeivollzugsdienst konnten die Gesamtergebnisse von 193 Pbn. ausgewertet werden. Es nahmen 20,7 % (n = 40) Teilnehmer aus Hamburg am Verfahren teil, die restlichen 79,3 % (n = 153) Bewerber kamen aus den übrigen deutschen Bundesländern. Das Alter der Bewerber lag zwischen dem 17. und dem 34. Lebensjahr, dabei hatten über 90 % der Bewerber das 28. Lebensjahr noch nicht vollendet, insgesamt 80,8 % waren jünger oder gleich 24 Jahre. Der Altersmittelwert lag bei 21,14 Jahren.

131 Bewerber bewarben sich um Einstellung in die Schutz-, 43 in die Kriminal- und 19 in die Wasserschutzpolizei. Es nahmen insgesamt 108 Männer (56,0 %) und 85 Frauen (44,0 %) an der Erhebung teil. 5 Teilnehmer (2,6 %) waren nicht deutscher Herkunft. Die Schulbildung der Teilnehmer verteilte sich wie in Tabelle 9.2-1 ersichtlich.

Tab. 9.2-1: Schulbildung der Teilnehmer im gehobenen Dienst nach Petersen (2002)

Schulbildung	Anzahl	%
- Fachhochschulreife	90	46,6
- Hochschulreife	78	40,4
- andere	25	13,0
n =	193	100,0

* abgeschlossenes Fach- oder Hochschulstudium

Bei den Bewerbern für den mittleren Polizeivollzugsdienst zeigte sich, dass weder die einzelnen Zeugnisnoten in den erhobenen Fächern Deutsch, Mathematik, Englisch, zweite Fremdsprache, Biologie, Physik, Chemie, Geschichte, Politik, Wirtschaft oder Sport signifikant mit dem erreichten Gesamtergebnis im Auswahlverfahren korrelierten, noch hinsichtlich des berechneten Notendurchschnittes der Einzelfächer.

Das Ergebnis bei den Bewerbern für den gehobenen Polizeivollzugsdienst ergab folgendes Bild: Zwar korrelierte die Deutschnote nicht signifikant mit den Testteilen, die auch die Deutschleistung abbilden sollen (Diktat und Bildbeschreibung), jedoch schwach ($r = .21$) mit dem Gesamtergebnis. Die von Althoff (1986) und anderen Autoren als validester Einzelprädiktor beschriebene Mathematiknote korrelierte weder mit den Ergebnissen der Intelligenz- und Leistungstests oder anderer Testteile noch mit dem Endergebnis signifikant. Keine der anderen oben beschriebenen Einzelnoten oder der errechnete Notendurchschnitt korrelierte signifikant mit dem von den Teilnehmern im Auswahlverfahren erreichten Endergebnis. Auf der Einzelprädiktorebene konnten signifikante Korrelationen einzig zwischen der zweiten Fremdsprache und dem Ergebnis im Diktat ($r = .44$) sowie der Sportnote und dem Ergebnis in der Sportprüfung ($r = .32$) nachgewiesen werden.

Die erschlossene Höhe des Validitätskoeffizienten hängt allerdings wesentlich von folgenden Faktoren ab: vom Grad dessen, was an Gemeinsamkeit durch den Test und das Kriterium erfasst und oft als „Zulänglichkeit“ des Tests bezeichnet wird, der Reliabilität des Tests und von der Reliabilität des Kriteriums. Insbesondere Reliabilität von Prädiktor und Kriterium sind zu hinterfragen. Von den einzelnen Komponenten des Testverfahrens, aus deren Gesamtergebnis sich das Kriterium zusammensetzt, kann zumindest für die standardisierten Testteile (CFT und IST 2000) eine ausreichend hohe Reliabilität belegt werden. Dagegen wurde die Reliabilität der weiteren Verfahrensteile (Diktat, Berichtsfertigung, Sporttest, Vorstellungsgespräch) nicht erschlossen und kann, insbesondere für das standardisiert durchgeführte Vorstellungsgespräch, allenfalls angenommen werden.

Angesichts der teilweise sehr unterschiedlichen schulischen Rahmenbedingungen in den verschiedenen Bundesländern und der darüber hinausgehenden Vielfalt von Schularten und –systemen kann eine ausreichende Reliabilität des Prädiktors erhofft, jedoch kaum noch angenommen werden. Aber auch die Objektivität von Schulnoten, das heißt die Frage, inwieweit unterschiedliche Bewertungen von Lehrern verschiedener Schularten und –systeme bei gleicher Leistung eines Schülers zu vergleichbaren Ergebnissen führt, muss kritisch hinterfragt werden. Beide Einschätzungen hinterfragen berechtigterweise das Ergebnis der vorliegenden Untersuchung, die um die Berücksichtigung dieser methodischen Voraussetzungen ergänzt durchgeführt werden sollte.

Gleichwohl ließen sich zur Erklärung dieser Ergebnisse weitere stützende Hypothesen heranziehen. Schon Ingenkamp kommt 1969 zu dem Schluss, dass in den Schulnoten die

Anpassung der Schüler an die in der jeweiligen Schule geforderten Verhaltensweisen stärker zum Tragen kommen als das eigentliche Leistungspotential. Schuler (1998) stellt fest, dass in Realschulzeugnissen ab den 80er Jahren ein deutlicher Abwärtstrend der Validität von Schulabschlüssen erkennbar ist. Die mittlere Korrelation lag nur noch bei $r = .26$. Zwar konnte dieser Trend für Abiturzeugnisse aus den 60er – 80er Jahren nicht festgestellt werden, inwieweit diese Aussage auch heute noch gilt, lässt sich jedoch angesichts vorliegender Ergebnisse nicht ohne Zweifel beantworten. Letztere berücksichtigend, bleibt der Bedarf weiterer Untersuchungen gegeben. Dabei ist auch die Frage besonderer Stichprobeneffekte bei Bewerbern für den Polizeivollzugsdienst zu reflektieren.

Hinterfragt werden sollten in diesem Zusammenhang auch die Auswirkungen, die sich durch eine verstärkte Verbreitung sog. Testknacker ergeben haben. Insbesondere seit den 90er Jahren bereiten sich viele Teilnehmer an Auswahlverfahren gezielt mit derartigen Übungsanweisungen (bspw. von Hesse und Schrader, 1992) vor. Wenn auch Untersuchungen belegen (Wottawa & Amelang, 1985), dass durch einfache Übungen die Unterschiede zwischen den Probanden größer werden, so ist nach den Autoren jedoch davon auszugehen, dass mit der Anwendung von Übungsheften es eher zu einer Akzentuierung der Ungleichheiten kommt. Hossiep (1995) zitiert in diesem Zusammenhang Fay (1992), die belegen konnte, dass in Durchstreichverfahren zur Messung der Konzentrationsleistung, die grundsätzlich starken Trainingseffekten unterliegen, zwar die Ergebnisse durch die Testtrainings stark verändert wurden, es hinsichtlich der Rangreihe jedoch keine Unterschiede gab. Wenn auch die verschiedenen Überlegungen, derzeit nur Anregungen für Untersuchungen, nicht aber konkrete Erklärungen bieten können, ist zwischenzeitlich zumindest zu hinterfragen, inwieweit bei der Bewertung der Bewerbungsunterlagen den Schulzeugnissen der bisher gegebene Stellenwert eingeräumt werden darf, um Entscheidungen hinsichtlich der Annahme oder Ablehnung eines Bewerbers zum Testverfahren zu treffen. Dies ist insbesondere aufgrund der in Kapitel 2.1.3 dargestellten Bedeutung der Schulnoten bei der Bewerbung für den Polizeivollzugsdienst wichtig. Andere Überlegung für eine ökonomische Vorauswahl der Bewerbungen aufgrund der eingereichten Unterlagen sind somit von Bedeutung. Reilly & Chao (1982) geben für die gesamten Bewerbungsunterlagen eine prognostische Validität von $r = .18$ an. Diese dürfte für Bewerbungsunterlagen als Vorauswahlinstrument nicht ausreichend sein.

Trotz dieser Ergebnisse werden nach Mohr (1990) nur durchschnittlich zehn Minuten für die Durchsicht von Bewerbungsunterlagen veranschlagt. Die innerhalb dieser kurzen Zeitspanne zu treffenden Entscheidungen über Annahme oder Ablehnung einer Bewerbung birgt ein hohes Fehlerpotential. Während der Fehler erster Art, d.h. zur Einstellung vorgesehene Bewerber, die nicht für die Tätigkeit geeignet sind, in der Regel durch nachfolgende validere Stufen im Auswahlverfahren kompensiert werden kann, hat der Fehler zweiter Art in diesem Kontext weitreichendere Folgen. Der Verzicht auf Bewerber, die für die Tätigkeit geeignet wären, aber aufgrund der fehlenden Validität der Bewerbungsunterlagen nicht zum Auswahlverfahren zugelassen worden, sind für die Organisation dann besonders kritisch, wenn die Selektionsquote niedrig gesetzt wird, d.h. aufgrund einer nicht ausreichend validen Vorauswahl zu wenig potenziell geeignete Bewerber eingeladen werden. Kompensierend müssen zusätzliche Auswahlverfahren durchgeführt werden, um ausreichend geeignete Bewerber zu finden, angesichts der sich damit verbindenden Sach- und Personalkosten keine ökonomische Arbeitsweise für eine Organisation.

2.1.2.2 Interview

Auf Interviews oder Vorstellungsgespräche bzw. Einstellungsgespräche wird in Auswahlverfahren aufgrund ihrer breiten Informationsquelle so gut wie nie verzichtet. Westhoff (2000) unterstreicht die allgemeine Bedeutung des Interviews für die psychologische Praxis als die mit weitem Abstand am häufigsten eingesetzte Methode zur Gewinnung von Informationen über menschliches Verhalten. Neben der verbreiteten Anwendung des Interviews auch im klinischen oder rechtspsychologischen Kontext, hat es als Auswahlinstrument in der Organisationspsychologie einen festen Platz gefunden. Nach Auswertung der Bewerbungsunterlagen sind Interviews insbesondere in der Personalauswahl die am häufigsten verwendete Methode (Schuler et al., 1993). Seine vielfältige Funktion wird von Weinert (1998, S.292) hervorgehoben. Es soll spezifische Informationen über die Person des Bewerbers liefern, die Organisation hat die Möglichkeit sich selbst darzustellen und Fragen zu beantworten und schließlich hat es eine Auswahlfunktion, die zu einer Entscheidung über Annahme oder Ablehnung des Bewerbers sowie einer möglichen Fortführung des Selektionsverfahrens führen kann.

Interviews werden als Vorstellungsgespräche sowohl strukturiert als auch unstrukturiert, d.h. ohne festgelegten Fragekatalog und festes System zur Beurteilung der Antworten der Bewerber, geführt. Nach Westhoff (2000) zeichnet sich aber in den letzten 50 Jahren eine

zunehmende Tendenz zur Strukturierung der Interviews ab. Häufig in einem Interview gestellte Fragen beziehen sich vor allem auf die bisherige Berufserfahrung und Berufsausbildung, auf Aspekte des Lebenslaufes und deren subjektive Verarbeitung oder auch auf persönliche Bereiche wie den familiären Hintergrund. Traditionelles Ziel des Interviews ist die Gewinnung von Informationen über den Entwicklungs- und Werdegang eines Bewerbers, dessen persönliche Eigenschaften, seine Fähigkeiten, Interessen und (Berufs-) Vorstellungen. Aus diesen soll die vermutete Bewährung in der Ausbildung oder im späteren Beruf abgeleitet werden.

Die Anwendung des Interviews gründet sich auf unterschiedliche theoretische Vorstellungen (Westhoff, 2000). Der Autor verweist auf Latham, Saari, Pursell und McDaniel (1980), die davon ausgehen, dass sich Menschen auch zukünftig entsprechend ihren überdauernden Vorstellungen und Zielen verhalten werden. Dagegen gehen Janz (1982) und auch Pulakos und Schmitt (1995) von dem diagnostischen Erfahrungsgrundsatz aus, dass der beste Prädiktor zukünftigen Verhaltens das entsprechende Verhalten in der Vergangenheit ist. Während hinsichtlich der Konstruktvalidität des Interviews noch einige Fragen offen bleiben müssen (Sarges, 1995b), lässt sich bezüglich der prädiktiven Validität annehmen, dass sich diese durch zunehmende Strukturierung verbessert. Für die prognostische Validität des strukturierten Interviews wird von Schmidt und Hunter (1998) ein Validitätskoeffizient von $r = .51$, gegenüber $r = .38$ für das nicht strukturierte Interview angegeben. Während andere Autoren zum strukturierten Interview ähnlich hohe Validitätskoeffizienten angeben (bspw. Wiesner & Cronshaw (1988) mit $r = .40$) liegen sie für nicht strukturierte Interviews teilweise erheblich niedriger (bspw. Hunter & Hunter, 1984 mit $r = .08 - .14$; Wiesner & Cronshaw, 1988 mit $r = .13$). Eine höhere Strukturiertheit des Interviews scheint wesentlich zur Verlässlichkeit des durch das Interview gewonnenen Urteils beizutragen.

Die geringe Vorhersagekraft unstrukturierter Interviews findet nach Hossiep (1995) einen zentralen Grund in der Informationsverarbeitung durch den Interviewer. So findet vermutlich bereits eine frühe Urteilsbildung statt, die für die weitere Informationsverarbeitung und Führung des Gespräches zentrales Gewicht hat. Zudem werden negative Informationen häufig überbewertet und emotionale Urteilskomponenten (wie bspw. Ähnlichkeit mit dem Interviewer oder andere Sympathieparameter) spielen eine das Urteil beeinflussende Rolle. Ähnlich sehen es auch Schuler und Funke (1993), die ebenso annehmen, dass die meisten Informationen, die Interviewer bei ihren Entscheidungen berücksichtigen, irrelevant sind. Frühe Eindrücke im Gespräch haben für die weitere Informationsverarbeitung dominierendes

Gewicht und steuern das Verhalten der Interaktionspartner. So belegte bspw. Sierverding (2000) in ihrer Untersuchung u.a., dass im standardisierten Interview solche Bewerber von den Beobachtern als erfolgreicher eingeschätzt werden, die eine angebotene Möglichkeit zur Selbstdarstellung nutzen und länger redeten als andere.

Nach Westhoff (2000) erwies es sich als grundsätzlich nützlich, wenn statt des traditionellen Interviews, das Persönlichkeitsmerkmale in ihrer Ausprägung zu erfassen sucht, sich die Interviews auf eine Anforderungsanalyse der Tätigkeiten gründen, für welche die Personen auszuwählen sind. So kombinieren neuere Interviewansätze o.g. theoretische Annahmen mit diesem Aspekt. Als ein beispielhafter Vertreter beim Versuch, die Höhe der Validität positiv zu beeinflussen, kann hier das Multimodale Interview von Schuler (1992a) genannt werden.

Es besteht, wie in Tabelle 10.2-1 dargestellt, aus einer feststehenden Abfolge von insgesamt sieben Komponenten und wurde mit dem Ziel konstruiert, möglichst viele der in der Interviewforschung festgestellten Defizite zu überwinden.

Tab.10.2-1: Aufbau des multimodalen Interviews nach Schuler (1992a.)

<p>1. Gesprächsbeginn</p> <ul style="list-style-type: none"> - Kurze informelle Unterhaltung - Bemühen um angenehme und offene Atmosphäre - Skizzierung des Verfahrensablaufes - Keine Beurteilung <p>2. Selbstvorstellung des Bewerbers</p> <ul style="list-style-type: none"> - Bewerber sprechen einige Minuten über ihren persönlichen und beruflichen Hintergrund - Beurteilung von sechs Dimensionen auf einer dreistufigen Skala <p>3. Freies Gespräch</p> <ul style="list-style-type: none"> - Interviewer stellt offene Fragen in Anknüpfung an die Selbstvorstellung und Bewerbungsunterlagen - Summarische Eindrucksbildung <p>4. Biographiebezogene Fragen</p> <ul style="list-style-type: none"> - Biographische oder Erfahrungsfragen werden aus Anforderungsanalysen abgeleitet - Beurteilung anhand einer dreistufigen verhaltensverankerten Skala <p>5. Realistische Tätigkeitsinformation</p> <ul style="list-style-type: none"> - Ausgewogene Information seitens des Interviewers über Arbeitsplatz und Unternehmen - Überleitung zu situativen Fragen <p>6. Situative Fragen</p> <ul style="list-style-type: none"> - Auf critical incident Basis konstruierte situative Fragen werden gestellt - Beurteilung der Antworten auf fünfstufiger verhaltenverankerter Skala <p>7. Gesprächsabschluss</p> <ul style="list-style-type: none"> - Fragen des Bewerbers - Zusammenfassung - Weitere Vereinbarungen
--

Als ein Beispiel für situative Fragen, geben die Autoren folgendes Szenario vor:

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

Beispielantwort 0 Punkte:

Ich sage dem Mitarbeiter, dass ich ihm gerne mehr gegeben hätte, dass aber die Geschäftsleitung keine weiteren Rahmen offen lässt.

Beispielantwort 2 Punkte:

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

Beispielantwort 4 Punkte:

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

Faktoren, die sich nach Schuler und Funke (1993) generell als förderlich für die Höhe der Validität des Vorstellungsgespräches ausgewirkt haben, sind in Tabelle 11.2-1 dargestellt.

Tab.11.2-1: Faktoren, welche die Höhe der Validität im Interview positiv beeinflussen nach Schuler & Funke (1993)

- | |
|---|
| <ul style="list-style-type: none"> - Anforderungsbezogene Gestaltung des Interviews - Beschränkung auf das Registrieren von Aspekten, Anforderungen, Merkmalen, die nicht anderweitig zuverlässiger gesammelt werden können. - Durchführung des Interviews in strukturierter bzw. (teil-)standardisierter Form - Auswahl von Interviewfragen möglichst nach empirischer Prüfung und nicht allein aufgrund subjektiver Evidenz - Versuchsweise Übernahme validierter Fragen - Trennung von Informationswahrnehmung und Entscheidung - Verwendung geprüfter und verankerter Skalen - Einsatz zusätzlicher Beurteiler - Gestaltung der Gewichtungs- und Entscheidungsprozedur nach psychometrischen Prinzipien - Vorbereitung der Interviewer durch Training |
|---|

Insbesondere die Strukturierung und der Anforderungsbezug haben sich in der Praxis als wirksame Moderatoren der Validität erwiesen.

2.1.2.3 Psychologische Testverfahren

In der Eignungsdiagnostik sind psychologische Testverfahren die am häufigsten verwendeten Instrumente. Nach Lienert (1989, S.7) ist ein Test ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung. Damit wird von Testverfahren verlangt, dass sie wissenschaftlich begründet sind, unter standardisierten Bedingungen durchgeführt werden, eine

Positionsbestimmung des untersuchten Individuums innerhalb einer Gruppe von Individuen erlauben und empirisch abgrenzbare Eigenschaften prüfen. Testverfahren sind prinzipiell für alle Berufsgruppen einsetzbar und daher Bestandteil vielfältiger Auswahlverfahren. Die i.d.R. Papier-Bleistift-Verfahren verfügen oft über eine hohe Validität, verursachen geringe Kosten und sind relativ leicht anwend- und auswertbar. Ein Überblick über Zahl und Vielfalt anwendbarer Testverfahren findet sich bei Brickenkamp (1975, 1983).

In der Berufseignungsdiagnostik werden häufig Testverfahren zur Erfassung der allgemeinen kognitiven Fähigkeiten (sog. Intelligenztests) und ihrer Komponenten wie Gedächtnis, Aufmerksamkeit, Belastungsfähigkeit oder räumliches Vorstellungsvermögen angewendet. Nach Brambring (1983) wären Beispiele für die in Privatunternehmen und Behörden am häufigsten angewandten Eignungstests der Intelligenz-Struktur-Test von Amthauer et al. (1970, 2000), das Leistungs-Prüf-System von Horn (1983) oder der Aufmerksamkeits-Belastungstest d2 von Brickenkamp (1972). Daneben werden von verschiedenen größeren Unternehmen und Organisationen auch eigene Testverfahren entwickelt und bedarfsangepasst eingesetzt. Bekannt dürfte insbesondere der Test für medizinische Studiengänge (TMS) sein, der zielgruppenspezifisch entwickelt wurde (Bartussek, Raatz & Stapf, 1986). Ein Beispiel für den Polizeivollzugsdienst ist der PatPol (Brenkelmann, 1983), der als Personal-Auslese-Test u.a. von der Polizei in Nordrhein-Westfalen eingesetzt wurde.

Der Nachweis der Validität von psychologischen Leistungstests zur Vorhersage von Berufserfolgen hat eine längere wissenschaftliche Tradition als der Nachweis für jede andere Methode (Hunter, 1986; Schmidt & Hunter, 1981). Entsprechend liegen heute eine große Zahl Untersuchungen der Validität kognitiver Fähigkeitstests vor. Schmidt und Hunter haben 1998 die Messung der Intelligenz, mit einer prädiktiven Validität von $r = .51$, als das wichtigste eignungsdiagnostische Instrument für Einstellungsentscheidungen bezeichnet. Nur Arbeitsproben sind valider, aber wesentlich aufwendiger und können nur bei Personen mit Berufskennnissen angewandt werden. So kommt Schuler (1996) auch zur Einschätzung, dass es praktisch keinen Beruf gibt, für den Maße intellektueller Fähigkeiten nicht zur Leistungsprognose beitragen können. Dabei lassen sich Ausbildungsleistungen besser vorhersagen als Kriterien der Berufsleistung. Hunter und Hunter (1984) kamen zu dem Ergebnis, dass die Validität der Leistungsprognose umso höher ausfällt, je höher die Komplexität des betroffenen Berufes ist. Andere Verfahren können häufig als Ergänzungen

zu Intelligenztests gesehen und im Zusammenhang mit der Erschließung der inkrementellen Validität bewertet werden.

In Tabelle 12.2-1 stellen Schmidt und Hunter (1998) die prädiktive Validität von eignungsdiagnostischen Prädiktoren und, bei Kombination mit einem zweiten Prädiktor, die inkrementelle Validität gegenüber Intelligenztests dar.

Tab. 12.2-1: Eignungsdiagnostische Prädiktoren kombiniert mit einem zweiten Prädiktor nach Schmidt & Hunter (1998)

Eignungsdiagnostische Instrumente	Validität	Inkrementelle Validität
- Intelligenztests	.51	---
- Arbeitsproben	.54	.12
- Interview (strukturiert)	.51	.14
- Interview (unstrukturiert)	.38	.07
- Biographische Daten	.35	.01
- Assessment Center	.36	.01
- Interessen	.10	.01
- Test zum Berufswissen	.48	.07
- Alter	-.01	.00

Die hier zu entnehmende Feststellung, dass bspw. biographische Messinstrumente zwar eine hohe prädiktive Validität besitzen, jedoch im Sinne der inkrementellen Validität nur zu einen durchschnittlichen Validitätszuwachs von $r = .01$ beitragen, begründet sich nach Ansicht der Autoren dadurch, dass biographische Daten in beträchtlichen Umfang ($r = .50$) mit allgemeinen kognitiven Fähigkeiten korrelieren. Ihrer Ansicht nach lässt dieses vermuten, dass biographische Daten ein indirektes Maß für mentale Fähigkeiten darstellen, zusätzlich zu dem, was sie zu anderen Merkmalen noch messen.

Neben Testverfahren zur Erfassung der kognitiven Fähigkeiten finden in der betrieblichen Auswahlpraxis häufig auch Persönlichkeitsfragebögen Anwendung. Bei Hossiep, Paschen und Mühlhaus (2000) findet sich ein Überblick über die derzeit in der Eignungsdiagnostik verwendeten Fragebogenverfahren. Hier zu nennen wären bspw. der 16-Persönlichkeitsfaktoren-Test (16 PF) von Schneewind, Schröder und Cattell (1994), das Neo-Fünf-Faktoren-Inventar (NEO-FFI) von Borkenau und Ostendorf (1993). Insbesondere der 16 PF darf heute fast als ein Klassiker unter den psychometrischen Persönlichkeitstests betrachtet werden, denn das Verfahren blickt auf etwa 50 Jahre praktische Anwendung zurück und wird besonders im anglo-amerikanischen Raum intensiv beforscht. Kritisch zu sehen bleibt allerdings, dass sich in der Vergangenheit wiederholt die 16 Faktoren nicht replizieren ließen.

Für Persönlichkeitstests werden generell geringere Validitätswerte berichtet als für Verfahren zur Erfassung kognitiver Fähigkeiten. Schmitt et al. (1984) berichten eine durchschnittliche Validität von $r = .15$, Reilly und Chao (1982) von $r = .18$. Zudem schien eine Einschätzung von Schuler (1986) lange Zeit Gültigkeit zu haben. Danach leide die empirische Brauchbarkeit von Persönlichkeitsfragebögen zur Prognose des Berufserfolges daran, dass die Frage in den Persönlichkeitstests so formuliert sind, dass entsprechend den vermuteten Erwartungen, die Antworten in einer Auslesesituation mitgeprägt sind von Wunschdenken, Erinnerungstäuschungen und bewussten Verfälschungen. Hossiep et al. (2000) verweisen jedoch auf neuere metaanalytische Ergebnisse, welche die Nützlichkeit von Persönlichkeitsverfahren auch für die Personalauswahl belegen. Sprechen 1965 Guion und Gottier Persönlichkeitsverfahren noch jede Nützlichkeit ab und sehen sogar die Gefahr negativer Validität, konzidieren Schmitt et al. (1984) immerhin nur noch die geringere Prädiktionskraft im Vergleich zu anderen Verfahren. Barrick und Mount untersuchten 1991 das Fünf-Faktoren-Modell der Persönlichkeit und konnten mit ihrer Metaanalyse differentielle Zusammenhänge zwischen einzelnen Traits und Maßen beruflichen Erfolges feststellen. Insbesondere das Persönlichkeitsmerkmal Gewissenhaftigkeit korrelierte in allen untersuchten Berufsgruppen positiv mit verschiedenen Erfolgskriterien. Salgado (1997) konnte bestätigen, dass die fünf Persönlichkeitsdimensionen auch länderübergreifend als valide für die Vorhersage des Berufserfolges angesehen werden können. Die Dimensionen Gewissenhaftigkeit und Emotionale Stabilität erwiesen sich dabei in unterschiedlichen europäischen Ländern und über verschiedene Berufsgruppen hinweg als valide Prädiktoren für unterschiedliche Maße beruflichen Erfolges.

2.1.2.4 Assessment-Center

Jeserich (1991) beschreibt das Assessment-Center als ein systematisches Verfahren zur qualifizierten Feststellung von Verhaltensleistungen bzw. Verhaltensdefiziten für mehrere Teilnehmer in bezug auf vorher festgelegte Anforderungen. Die in diesem Zusammenhang konstruierten unterschiedlichen Übungen sollen den Berufsalltag der zukünftigen Stelle simulieren. Als mögliche Zielsetzungen werden bspw. von Schuler (1992b) eine Vielzahl möglicher Anwendungsfelder genannt:

- Laufbahnplanung
- Potentialbeurteilung und -beratung
- Trainingsbedarfsanalyse

- Teamentwicklung
- Arbeitsplatzgestaltung
- Forschung.

Schätzungen von Thornton und Byham (1982) gehen davon aus, dass ca. 95% aller Assessment-Center für die Personal-Auswahl und Potentialentwicklung eingesetzt werden. Charakteristisch für das AC ist, dass mehrere Personen (i.d.R. 6 – 12) gleichzeitig als Beurteilte teilnehmen und die Einschätzung durch mehrere unabhängige Beobachter (i.d.R. im Verhältnis 2 : 1) vorgenommen werden.

Kleinmann (1997) unterstreicht die Aussage von Kompa (1989), wonach es das Assessment-Center in der Anwendungspraxis nicht gebe, sondern sich Assessment-Center-Verfahren nicht nur hinsichtlich der Vielfalt der Zielsetzungen, Beobachter- und Teilnehmerzusammensetzung oder Rotationstechniken unterscheiden, sondern insbesondere auch aufgrund der unterschiedlicher Übungen. Typischerweise jedoch finden sich häufig Aufgaben wie Rollenspiele, Gruppendiskussionen mit und ohne Rollenvorgabe, Postkörbe, Interviews, Vorträge und Präsentationen, Wirtschaftsspiele oder Organisations-, Analyse- und Entscheidungsaufgaben in einem AC. Eine Übersicht findet sich bei Jeserich (1991), Kleinmann (1997), Sarges (1996) und Schuler und Stehle (1992).

Das Abschneiden eines einzelnen Kandidaten bei den unterschiedlichen Aufgaben soll zum einen darüber Aufschluss geben, wie er zukünftig die gleichen Aufgaben bewältigen wird, zum anderen den Schluss darüber zulassen, wie er sich generell in seiner zukünftigen Position bewähren wird, deren Aufgabenfelder und vor allem Einzeltätigkeiten zum Teil noch unbekannt sind. Ob dies ein AC zu leisten in der Lage ist, wird auch hier durch die unterschiedlichen Zugänge zur Erfassung der Validität überprüft. Prognostisch valide Schlüsse wären dann möglich, wenn die Ergebnisse im AC in empirisch nachweisbarer Beziehung zum späteren Berufserfolg stehen. Dies wird durch eine Vielzahl Studien belegt, bspw. Schmidt & Hunter (1998), die einen Validitätskoeffizienten von $r = .36$ angeben. Kann die inhaltliche Validität häufig als gegeben angesehen werden, da die Aufgaben in einem Assessment-Center üblicherweise als für die Berufsausübung relevant angesehen werden, so sieht es dagegen mit dem Nachweise der Konstruktvalidität, d.h., ob mit den Übungen auch tatsächlich das erfasst wird, was das AC zu messen beansprucht, ungünstiger aus. „Bis dato muss als ungeklärt gelten, was den Urteilen der Beobachter zugrunde liegt – diejenigen Eigenschaften, die man eigentlich damit zu erfassen behauptet, sind es allenfalls zum Teil“

(Schuler, 1992, S. 13). Für eine valide Konstrukterfassung ist nach dem Multitrait-Multimethod-Ansatz von Campbell und Fiske (1959, zit. n. Höft & Funke, 2001) zu fordern, dass Messungen eines Konstruktes über verschiedene Messmethoden hinweg einen hohen Zusammenhang aufweisen (konvergente Validität), während Messungen verschiedener Konstrukte auch bei konstantgehaltener Methode einen geringen Zusammenhang aufweisen (diskriminante Validität). Lammers (1994, zit. n. Kleinmann 1997) sieht einen Grund der eingeschränkten Konstruktvalidität in den Beobachtern, da ihre messfehlerbehaftete Einschätzung insbesondere durch die Rotation zur Reduzierung des Korrelationskoeffizienten führe. Höft und Funke (2001, S.164f) stützen diese Annahme und führen zudem verfahrensspezifische Mängel, Mängel bei der Erfassung anforderungsrelevanter Verhaltensweisen und anforderungsbedingte Mängel als Quelle möglicher weiterer Ursachen an. Interessant ist in diesem Zusammenhang der Verweis auf die Metaanalyse von Scholz und Schuler (1993), die für Intelligenz, soziale Kompetenz, Dominanz und Selbstvertrauen die größten Zusammenhänge mit den AC-Gesamturteil ergab. Nach Klimoski und Brickner (1987) könnten allgemeine kognitive Fähigkeiten hinter den eigentlich zu erfassenden Anforderungen stehen.

2.1.2.5 Arbeitsproben

Unter Arbeitsproben werden standardisierte Aufgaben verstanden, die inhaltlich valide sind und erkennbar eine äquivalente Stichprobe des erfolgsrelevanten beruflichen Verhaltens darstellen. Die Arbeitsprobe wird auf der Basis von arbeitsplatzanalytischen Daten gezielt zur Prognose der Leistung an einem konkreten Arbeitsplatz entwickelt (Hossiep, 1995). Arbeitsproben sind somit praktische Simulationen eines Teilbereiches oder des gesamten Bereiches einer Tätigkeit. Es wird versucht, das erfolgsrelevante Verhalten am Arbeitsplatz durch inhaltliche Nachbildung der Anforderung unmittelbar zu prognostizieren. Anstelle eines Ergebnisses im Testverfahren wird von einer Verhaltensstichprobe auf ähnliches künftiges Verhalten geschlossen. Häufig werden Arbeitsproben eingesetzt, wenn handwerklich-industrielle Kenntnisse erfasst oder technische Fachkräfte eingestellt werden sollen. Die Abgrenzung von Arbeitsproben und psychologischen Testverfahren ist nicht immer ganz einfach und wird in der Praxis auch sehr unterschiedlich gehandhabt. Der deutlichste Unterschied zwischen Test und standardisierten Arbeitsproben liegt darin, dass Testverfahren abstrakte Konstrukte zu erfassen suchen, die direkt auf ein Verhalten schließen sollen.

Arbeitsproben dagegen erfassen erfolgsrelevantes Verhalten über direkte Beobachtung (Lang von Wins & von Rosenstiel, 1998).

Trotz der hohen prädiktiven Validität (Schuler et al., 1993: $r = .30$; Schmitt, Gooding, Noe & Kirsch, 1984: $r = .35$; Schmidt & Hunter, 1998: $r = .54$) haben Arbeitsproben einen entschiedenen Nachteil. Sie können nur dann zum Einsatz kommen, wenn Bewerber die Tätigkeit bereits kennen und sind somit aus diesem Grunde für größere und externe Bewerbergruppen kaum anwendbar. Als nachteilig dürfte sich auch der relativ große konstruktive Aufwand darstellen. Ein offenkundiger Vorteil der Arbeitsproben liegt dagegen in der hohen Augenscheinvalidität und Akzeptanz seitens der Bewerber. Zudem sind in den Arbeitsproben selbst relevante Informationen über die Anforderung der Tätigkeit enthalten.

2.1.2.6 Computergestützte Verfahren

Insbesondere auch die immer weiter Verbreitung findende Anwendung von Personalcomputern hat dazu beigetragen, dass computergestützte Testverfahren in den letzten Jahren in der Eignungsdiagnostik an Bedeutung gewonnen haben. Hornke berichtet schon 1991, dass Computer zum Berufsleben dazu gehören und deshalb sollten sie auch in einer modernen Diagnostik verwandt werden, die über Lebensbewältigungen Prognosen machen will. Der Einsatz von Computern ist heute praktisch in allen Phasen einer Testentwicklung, -durchführung und -auswertung sowie der Evaluation möglich. Häufig genannter Vorteil ist der Standardisierungsgrad bei Durchführung und Auswertung durch Einsatz der PC-Variante.

Hinsichtlich der Übertragbarkeit der Testgütekriterien bestehender Papier-Bleistift-Verfahren auf Computerversionen liegen allerdings eine Reihe widersprüchlicher Ergebnisse vor. Neubauer und Urban (1991) unterstreichen die Notwendigkeit die testtheoretischen Eigenschaften in jedem Einzelfall zu überprüfen. So konnten die Autoren bei einem Vergleich der Computer- und der Standardversion von Ravens APM (1958) signifikante Unterschiede hinsichtlich der Vergleichbarkeit der Normwerte feststellen. Nach Mead und Drasgow (1993, zit. n. Schuler & Höft, 2001) tritt die Äquivalenzproblematik besonders bei kognitiven Fähigkeitstests auf, die unter Zeitdruck bearbeitet werden müssen. Die Autoren vermuten, dass die unterschiedlichen motorischen Anforderungen der Verfahrensformen bei den Speed-Varianten besonders stark ins Gewicht fallen und für die Unterschiede verantwortlich zeichnen. Eine Gegenüberstellung von Vor- und Nachteilen von computergestützten

Verfahren gegenüber Papier und Bleistift Tests anhand unterschiedlicher Testgütekriterien findet sich bei Kubinger (1993).

Grundsätzlich lassen sich in der Computerdiagnostik nach Schuler und Höft (2001) grob drei Bereiche unterscheiden. Für viele Papier- und Bleistift-Verfahren liegen heute äquivalente Durchführungen für den Computer vor. Es handelt sich hierbei oft um eine einfache Übertragung der Version auf elektronische Datenverarbeitung, ohne weitergehende Vorteile der PC gestützten Bearbeitung zu nutzen. Diese Gruppe stellt mit Abstand den heute größten Anwendungsbereich dar. Die nächste Gruppe umfasst das adaptive Testen, d.h. Strategien der Testvorgabe, bei denen sich die Auswahl der Testitems am Leistungsniveau der jeweiligen Teilnehmer orientiert. Aufgrund des hohen Konstruktionsaufwandes sowie des Rückgriffes solcher Verfahren auf neuere testtheoretische Ansätze, die sich stark von der klassischen Testtheorie unterscheiden, finden adaptive Verfahren in der Eignungsdiagnostik bisher kaum Anwendung. Weiter verbreitet ist dagegen die dritte Gruppe der komplexen Simulationen. Gemeint sind Computerszenarien, in denen komplexe Systeme simuliert werden und die durch Intervention der Teilnehmer gesteuert werden müssen. Insbesondere bei der Durchführung komplexer Assessment Center gehören diese Verfahren heute fast zum Standard.

2.1.2.7 Biographische Daten

Die biographischen Daten der Bewerber werden häufig in standardisierten Fragebögen erhoben und umfassen den Gesamtbereich demographischer Gegebenheiten. Die Fragebögen dienen der Erfassung gemachter Lebenserfahrungen wie bspw. Erfahrungen in der Familie, in der Schule, mit Hobbys und anderen Beschäftigungen. So können Fragen zu bisherigen Ämtern in schulischen Organisationen oder Vereinigungen, zu ausgeübten Sportarten oder zu Erziehungsmethoden der Eltern in der Erhebung enthalten sein. Einfachere Formen begnügen sich mit standardisierten Fragen zu Daten des Lebenslaufs, ähnlich wie sie in vielen Bewerbungsbögen vorgegeben oder auch in Einstellungsinterviews erfragt werden. Bei einzelnen Formen werden dagegen auch subjektive Stellungnahmen, Emotionen, Einstellungen oder Interessen etc. erfasst.

Grundsätzliches Ziel ist es, Fragen in das Instrument einzuschließen, die während der Entwicklung des Fragebogens mit der späteren beruflichen Leistung korrelieren. Instrumente zur Erfassung biographischer Daten werden also in ersten Linie empirisch entwickelt.

Biographische Messinstrumente besitzen eine hohe prognostische Validität. Funke, Krauß, Schuler und Stapf (1987) berichten einen Validitätskoeffizienten von $r = .47$, Hunter und Hunter (1984) von $r = .37$, Schmidt & Hunter (1998) von $r = .35$. Damit kann die prognostische Validität von biographischen Daten mit der von Leistungstestverfahren verglichen werden und übertrifft die vieler andere Personalauswahlverfahren. Eine ausführliche Darstellung findet sich in Kapitel 2.2.

2.1.2.8 Weitere Verfahren

Im Zusammenhang mit alternativen gebräuchlichen Auswahlinstrumenten sind insbesondere graphologische Gutachten erwähnenswert. Trotz aller Kritik an graphologischen Methoden, insbesondere in der Eignungsdiagnostik, verweisen Schuler, Frier und Kaufmann (1993) darauf, dass in der Bundesrepublik, mehr noch als im europäischen Ausland, graphologische Methoden in nicht unerheblichem Ausmaß zur Personalauswahl herangezogen werden. Entgegen der Beliebtheit dieses Diagnostikinstrumentes ist der Beitrag an Vorhersagegüte eher gering. Hossiep (1995) verweist auf eine metaanalytische Studie von Neter und Ben-Shakhar aus dem Jahre 1989, nach deren Ergebnis ihr Beitrag zur Eignungsdiagnostik gegen Null tendiert. Eine Erklärung für die Verbreitung der Methode dürfte in ihrer Augenscheinvalidität zu finden sein. Nach Heinze (1995) erscheint die graphologische Systematik dem Laien plausibel durch den Zusammenhang zwischen dem spontanen Laieneindruck der Schriftmerkmale und gleichzeitiger Entsprechung der Eindrucksmerkmale mit graphologischen Deutehypothesen.

Ein weiteres an dieser Stelle zu nennendes Verfahren ist das sog. Peer-Rating. Hier erfolgt die Beurteilung eines einzelnen in einer Gruppe durch ein oder meist mehrere Gruppenmitglieder. Besondere Bedeutung bekommt das Peer-Rating nach Hossiep (1995) für Tätigkeiten, in denen Teameignung oder Führungsqualitäten eine besondere Rolle spielen.

Auf weitere Methoden z.B. Messung der Integrität und Gewissenhaftigkeit, Messung von Berufskennnissen, Verhaltenskonsistenz-Methode, Erfassung von Interessen und Fachkennnissen, Auswertung der Probezeit etc., soll an dieser Stelle nicht weiter eingegangen werden. Gleiches gilt für Verfahren der internen Personalauswahl wie Probezeit, Vorschlag durch Vorgesetzte, Leistungsbeurteilungen o.ä.. Der interessierte Leser wird an auf die weiterführende einschlägige Fachliteratur (bspw.: Hossiep, 1995; Sarges, 1995; Schuler, 1996, 2001; Weinert, 1998) verwiesen.

2.1.3 Personalauswahl am Beispiel der Einstellung in den Polizeivollzugsdienst

2.1.3.1 Allgemeine Grundlagen für die Einstellung

Im Unterschied zur Durchführung von Auswahlverfahren in der freien Wirtschaft werden im Öffentlichen Dienst die Rahmenbedingungen für die Personalauswahl in Gesetzen und Verordnungsblätter festgelegt und geregelt. Gemäß der Verordnung über die Laufbahn der Polizeivollzugsbeamten (HmbLVOPol) von 1992 gliedert sich die Einheitslaufbahn der Polizeivollzugsbeamten in die Laufbahnabschnitte I bis III. Danach erfasst:

- der Laufbahnabschnitt I den mittleren Dienst (Hauptwachtmeister – Hauptmeister)
- der Laufbahnabschnitt II den gehobenen Dienst (Kommissar – Erster Hauptkommissar)
und
- der Laufbahnabschnitt III den höheren Dienst (Rat – Leitender Direktor).

Nach den Vorschriften dieser Verordnung steht jedem Polizeivollzugsbeamten entsprechend seiner Eignung, Befähigung und fachlichen Leistung der Aufstieg in alle Ämter des Polizeivollzugsdienstes offen. Dieser umfasst die Dienstzweige der Schutz-, Kriminal- und Wasserschutzpolizei.

Die Grundlagen der Einstellung in den Polizeivollzugsdienst regeln die §§ 7 HmbLVOPol für den mittleren Dienst und 17 für den gehobenen Dienst. Auch eine direkte Einstellung in den höheren Dienst ist theoretisch nach § 24 HmbLVOPol möglich, wobei dieser Direkteinstieg aber derzeit nicht praktiziert wird. Dagegen ist sowohl für die Beamten des mittleren Dienstes durch § 15 HmbLVOPol und des gehobenen Dienstes durch § 22 HmbLVOPol der Aufstieg in die jeweils nächst höhere Laufbahn geregelt und möglich.

Weiter werden in der Laufbahnverordnung die formellen Voraussetzungen und Rahmenbedingungen hinsichtlich Alter, Schulbildung, Teilnahme an einer Einstellungsprüfung, Vorliegen einer gesundheitlichen Polizeivollzugstauglichkeit, Ausnahmen u.ä. für eine Einstellung in den Polizeivollzugsdienst geregelt.

2.1.3.2 Einstellung in den Polizeivollzugsdienst

Auf Grundlage der o.g. gesetzlichen Rahmenbedingungen findet die Einstellung von Bewerbern für den mittleren und gehobenen Dienst ihre Umsetzung in den sogenannten „Richtlinien des Polizeivollzugsdienstes“. Hier werden die Regularien für das Einstellungsverfahren (wie z.B. Fristen und Inhalt der Bewerbung, Ausschluss vom Verfahren, Testverfahren zur Eignungsfeststellung, Wiederbewerbungsmöglichkeiten) verfügt. Eine Präzisierung und Konkretisierung des Einstellungsverfahrens, Detailfragen zur Vorauswahl, Eignungsfeststellung sowie Bewertung der Testergebnisse, zu erbringende Mindestanforderungen u.ä. findet sich dann in den sogenannten „Ergänzenden Regelungen für das Einstellungsverfahren“. In beiden Fällen sind gemäß Personalvertretungsgesetz die Personalräte zu beteiligen.

Das Personalvertretungsgesetz (HmbPersVG) regelt die vertrauensvolle Zusammenarbeit zwischen den einzelnen Dienststellen und den Personalvertretungen (Personalräten). Zurzeit der Untersuchung wurden die Mitarbeiter der Dienstzweige Schutz-, Kriminal- und Wasserschutzpolizei von jeweils eigenen Personalräten vertreten. Seit Sommer 2001 hat die Polizei einen Gesamtpersonalrat, der sowohl die Interessen aller Vollzugsmitarbeiter wie auch die der Verwaltung vertritt. Die §§ 86 – 96 HmbPersVG regeln insbesondere die Beteiligung und das Mitbestimmungsrecht des Personalrats. Nach allgemeiner Rechtsauffassung hat der Personalrat u.a. bei Fragen der Einstellung bzw. allgemeinen Umfragen (Fragebögen) in der Polizei das Recht der Mitbestimmung. Ähnliches gilt für Personalfragebögen, die grundsätzlich gem. § 94 Abs. 1 Betriebsverfassungsgesetz (BetrVG) der Zustimmung durch den Betriebsrat unterliegen. Zwar erstreckt sich die Zustimmungspflicht nicht auf die eigentliche Einführung, wohl aber auf die inhaltliche Ausgestaltung der Fragebögen. Im Zweifelsfall wäre nach § 76 BetrVG durch eine paritätisch besetzte Einigungsstelle ein Einvernehmen zu erzielen

Im Folgenden wird der derzeitige Ablauf in den Einstellungsverfahren für den mittleren und gehobenen Polizeivollzugsdienst zusammengefasst beschrieben. Aufgrund des bestehenden Testgeheimnisses kann an dieser Stelle jedoch nicht auf alle Einzelheiten eingegangen werden.

Nach Durchsicht der schriftlich vorliegenden Bewerbungen und nach einer Vorauswahl (s.a. Punkt 2.1.3.3) ist es das Ziel des Auswahlverfahrens, die geistigen, körperlichen und persönlichen Voraussetzungen der Bewerber für den Polizeivollzugsdienst zu überprüfen.

Zu den Anwendung findenden eignungsdiagnostischen Instrumenten zählen ein Diktat, die Abfassung einer Bildbeschreibung und ein kulturfairer kognitiver Leistungstest. Letzterer wurde eingeführt, um Bewerbern ausländischer Herkunft gleiche Chancen gegenüber Bewerbern deutscher Herkunft einzuräumen. Bewerber des gehobenen Dienstes haben in einem weiteren Abschnitt in einer Klausur ihre Problemanalyse-, Urteils- und sprachliche Ausdrucksfähigkeit nachzuweisen.

Die Überprüfung der körperliche Eignung erfolgt in Form eines speziell angepassten Sporttestes. In fünf Übungen werden Schnelligkeit und Koordination, Kraft, Schnellkraft, Beweglichkeit und Koordination sowie Ausdauer erfasst.

Die Prüfung der persönlichen Eignung unterscheidet sich erneut für die Bewerber der beiden Laufbahnen. Bewerber für den mittleren Dienst nehmen an einem strukturierten und halbstandardisierten Interview teil. In einem vorgegebenen Zeitrahmen von ca. 30 Minuten pro Bewerber werden Aspekte wie bspw. Denk- und Urteilsvermögen, sprachliche Ausdrucksfähigkeit, Überzeugungskraft und emotionale Stabilität bewertet. Bewerber des gehobenen Dienstes führen ein ca. 30minütiges unstrukturiertes Vorstellungsgespräch und nehmen anschließend in einem Rundgespräch in Gruppen von 5-6 Personen zu einem aktuellen Thema Stellung.

Teilnehmer, die eine gewisse Mindestpunktzahl in den einzelnen Bereichen nicht unterschritten haben, werden abschließend hinsichtlich ihrer erreichten Gesamtpunktzahl in eine Rangreihe gebracht. Im Rahmen des Bedarfes werden die Bewerber aufgrund der so festgelegten Reihenfolge für die Einstellung vorgesehen. Auf weitere für eine Einstellung in diesem besonderen Kontext notwendige Voraussetzungen wie bspw. die Überprüfung der Vollzugstauglichkeit durch den Personalärztlichen Dienst oder Überprüfung der charakterlichen Eignung durch örtliche Polizeireviere, das Landeskriminalamt und das Bundeszentralregister wird an dieser Stelle nicht näher eingegangen.

2.1.3.3 Vorauswahl bei Einstellung in den Polizeivollzugsdienst

Auch die bisher praktizierte Vorauswahl der Bewerbungen orientiert sich an rechtlichen Vorgaben. Sie ist ebenfalls in den Richtlinien festgelegt und wird unterteilt in sogenannte

Muss-Kriterien, bei deren Nichterfüllung eine Ablehnung obligatorisch ist, sowie Kann-Kriterien, bei denen mindestens zwei oder mehr Kriterien erfüllt sein müssen, bevor es zu einer Ablehnung kommt.

Beispiele für Muss-Kriterien sind negative Hinweise im Zeugnis früherer Arbeitgeber zum Arbeits- oder Leistungsverhalten, unwahre Angaben (insbesondere zu vorhandenen Vorstrafen), zweimaliger Abbruch eines Studiums oder mehr als zwei schlechtere Schulnoten als „ausreichend“ in den Fächern Deutsch, Sport, Sozialkunde, Politik, Geschichte oder Mathematik. Beispielhafte Kann-Kriterien wären ein fehlendes Anschreiben, mangelnde Deutschkenntnisse im Bewerbungsschreiben, zwei schlechtere Schulnoten als „befriedigend“ in den o.g. Fächern oder eine schlechtere Note als „ausreichend“ in sonstigen Fächern, zweimal die Lehre abgebrochen, einmal das Studium abgebrochen.

Die Bewertung der Vorauswahlkriterien obliegt den Mitarbeitern der Einstellungsstelle der Polizei, die über keine weiteren eignungsdiagnostischen Kenntnisse verfügen. Die Auswahl der Kriterien beruht auf Erfahrungswerten der Mitarbeiter. Es sind empirische Annahmen und Eindrucksbildungen über Zusammenhänge zwischen Vorauswahlkriterium und dem erwarteten Abschneiden eines Bewerbers im Testverfahren bzw. der möglichen Eignung zur Berufsausübung ohne teststatistische Absicherung.

Die Funktion des Vorauswahlverfahren ist es in erster Linie, Bewerbungen hinsichtlich ökonomischer Gesichtspunkte zu reduzieren, eine statistische Absicherung der genannten Kriterien liegt wie erwähnt nicht vor.

Angesichts vorliegender Erfahrungen mit der bisherigen unzureichenden Vorauswahlpraxis, der Hinweise hinsichtlich einer möglicherweise nicht mehr ausreichenden prädiktiven Validität von Schulnoten und auch unter Beachtung der insgesamt sehr geringen Validität sonstiger Inhalte der Bewerbungsunterlagen (s.a. Punkt 2.1.2) sind Alternativen für eine verlässliche und ökonomische Erfolgsprognose und damit eine effektive Vorauswahl wünschenswert. Auch aufgrund des in den kommenden Jahren prognostizierten erhöhten Bedarfes an qualifiziertem Personal erscheint es notwendig, verlässliche Instrumente nutzen zu können. Warum sollte bspw. ein Bewerber, der einmal ein Studium abbrach und die Note „ausreichend“ in Politik im Zeugnis hat, nicht seine Eignung im Testverfahren überprüfen lassen dürfen. Weshalb sollten weiterhin Bewerber mit einer schlechten Deutschnote schon in der Vorauswahl eine Ablehnung erfahren, wenn die Zusammenhänge zwischen einzelnen Schulnoten unterschiedlicher Schulen in verschiedenen Bundesländern und dem Ergebnis

im Auswahlverfahren nicht empirisch signifikant nachweisbar ist. Insofern ist eine statistisch signifikante Absicherung der empirischen Annahmen der gewählten Datengeber wünschenswert und dient notwendigerweise auch der Rechtfertigung und Akzeptanz des gewählten Verfahrens. Nur wenn es gelingt, einen entsprechend signifikanten Zusammenhang zwischen Prädiktor und Kriterium, d.h. Vorauswahlkriterium und Ergebnis im Auswahlverfahren, herzustellen, ist die Anwendung von Vorauswahlfiltern nicht nur sinnvoll, sondern auch unter Kosten-Nutzen-Aspekten zwingend. Zudem lässt sich nur auf diesem Wege auch inhaltlich gegenüber Bewerbern die Vorauswahl vertreten.

2.2 Biographische Eignungsdiagnostik

2.2.1 Grundlagen biographischer Eignungsdiagnostik

Noch 1986 berichtete Schuler, dass in Deutschland erst seit wenigen Jahren versucht wird, wirkliche Alternativen zur Diagnose der Berufseignung mittels psychologischer Testverfahren zu finden. Zwei Erfolge waren in diesem Kontext besonders herauszuheben: Assessment-Center und Biographische Fragebogenverfahren. Wie schon ausführlich dargestellt, verfügen beide Verfahren über ausreichend hohe prädiktive Validität. Das Assessment-Center (vgl. a. Kapitel 2.1.2.4) konnte sich zwischenzeitlich als eine allgemein weit verbreitete Alternative im eignungsdiagnostischen Instrumentarium implementieren. Seine Verbreitung ging einher mit zunehmender Akzeptanz als eignungsdiagnostisches Auswahlinstrument bzw. adäquate Methode zur Potenzialfeststellung in der Personalentwicklung, wie auch hinsichtlich der grundlegenden Akzeptanz bei Teilnehmern und Personalverantwortlichen in Unternehmen. In der heutigen Zeit hinsichtlich seiner praktischen Anwendung wieder ein wenig in Vergessenheit geraten, ist dagegen die zweite Alternative. Die systematische Sammlung biographischer Daten in Fragebogenverfahren diente und dient im Rahmen der Berufseignungsdiagnostik vor allem der Selektion externer Bewerber. Biographische Fragebogen wurden zur Personalauswahl in unterschiedlichen Bereichen eingesetzt, in erster Linie gehören dazu Außendienstmitarbeiter, insbesondere im Versicherungsgewerbe, aber auch bei Auszubildenden, Militärangehörigen und Führungskräften. Doch schon in den frühen 80er Jahren zeigte sich, dass die Vielfalt der praktischen Anwendungen nicht im gleichen Maß mit vertiefender Forschung und Weiterentwicklung einherging.

2.2.1.1 Allgemeine Grundlagen biographischer Berufseignungsdiagnostik

Vor einer vertiefenden Auseinandersetzung mit dem Thema biographischer Daten bzw. biographisch orientierter Berufseignungsdiagnostik, ist zu klären, was genau unter dem Begriff zu verstehen ist. Nach Jäger (1992) umfassen biographische Daten Informationen, die über die Lebensgeschichte eines Individuums in Ausschnitten oder insgesamt Auskunft geben. Hierunter fallen jegliche Selbstbeschreibungen einer Person, zu der sowohl nachprüfbar objektive, aber auch subjektive Information über diese Person zu früheren, gegenwärtigen und zukünftigen Gegebenheiten zählen. Von biographischer Diagnostik kann nach Keßler (1982) gesprochen werden, wenn mit Hilfe biographischer Daten Diagnostik betrieben wird. Aus beiden Definitionen lässt sich ableiten, dass es Ziel der biographischen Berufseignungsdiagnostik ist, zu erkennen, welche lebensgeschichtlichen subjektiven oder objektiven Daten eines Bewerbers dem Eignungsdiagnostiker darüber Auskunft geben können, ob die Person für die von ihr angestrebte berufliche Position geeignet bzw. nicht geeignet ist. Die Bereiche, in denen dies grundsätzlich möglich ist, sind vielfältig. In der klinischen Diagnostik interessiert bspw. die Frage, welche Ereignisse in der Vergangenheit eines Patienten zu möglichen Belastungen in der heutigen Zeit führten. Die Pädagogische Diagnostik hinterfragt, welche früheren Lern-, und Sozialisationsprozesse heutigen Erfolg oder Misserfolg erklären könnten und in der beruflichen Eignungsdiagnostik steht das Bemühen im Vordergrund, mit Hilfe biographischer Daten ein verlässliche Prognose über die zukünftige Leistung in der Ausbildung oder im Beruf zu machen. Auch wenn der Wert und die Bedeutung biographischer Daten in den verschiedenen psychologischen Anwendungsfeldern unterschiedlich ausgeprägt ist, steht im Vordergrund dieser Arbeit ausschließlich die Verwendung biographischer Daten in der beruflichen Eignungsdiagnostik. Hier werden biographische Daten im weitesten Sinne dazu genutzt, um aus dem bisherigem Verhalten eines Bewerbers um einen Ausbildungsplatz oder eine vakante Stelle in einer Organisation, eine Prognose hinsichtlich zukünftigen Verhaltens abzugeben.

Aus dem von Hossiep (1995) dargestellten Überblick möglicher Determinanten beruflichen Erfolges (Tabelle 1.2-2) wird deutlich, dass unterschiedliche Daten als Prädiktoren beschrieben werden, die dem Bereich biographischer Daten zugeordnet werden können.

Hierbei läge es auf der Hand anzunehmen, dass es sich in erster Linie um verifizierbare biographische Daten handelt, die besonders aussagekräftig und zur gewünschten Prognose beitragen. Gemeint sind dabei Informationen aus einer Bewerbung, die „harten“ Charakter

haben, wie bspw. Zeugnisnoten oder konkrete Arbeitserfahrungen. Jedoch sind manche biographische Informationen kaum verifizierbar und haben trotzdem eine hohe prädiktive Validität. So konnte z.B. das biographische Item „Bauten Sie jemals ein flugtüchtiges Modellflugzeug?“ im Zweiten Weltkrieg den Erfolg eines Flugtrainings fast ebenso gut vorhersagen, wie die gesamte Air Force Battery (vgl. Henry 1965, zit. n. Keßler, 1982).

Tab. 1.2-2: Mögliche Determinanten beruflichen Erfolges nach Hossiep (1995)

Mögliche Determinanten beruflichen Erfolges	
-	Persönliches Wertesystem / Werthaltung
-	Kultureller Hintergrund
-	Sozioökonomische Sozialisation
-	Erfahrung
-	Ausbildung
-	Erziehung
-	Interesse / Motivation
-	Physische Eigenschaften
-	Alter und Geschlecht
-	Charaktereigenschaften
-	Kognitive / motorische Fähigkeiten
-	Persönliches Lebensschicksal
-	Sonstige Merkmale der Person

Wie in Kapitel 2.1.2 dargestellt, ist die Analyse der Lebensläufe häufig Gegenstand einer ersten Auswahl von Bewerbern, um eine ausgeschriebene Position in einem Betrieb oder einer Organisation zu besetzen. Annahmen, dass die Biographie eines Menschen keineswegs nur durch zufällige äußere Einflüsse bestimmt wird, rechtfertigten diese Vorgehensweise. Aus den Angaben in Schul- oder Ausbildungszeugnissen, Lebenslauf, Referenzen früherer Arbeitgeber und speziellen Fragebögen werden Informationen über die Vergangenheit eines Bewerbers gewonnen. Die sich hier reflektierenden Ereignisse, Verhaltensweisen und Leistungen werden zur eignungsdiagnostischen Prognose herangezogen. Dieses Vorgehen beruht wesentlich auf der Annahme, dass menschliches Verhalten über einen längeren Zeitraum stabil bleibt.

Auf die Bedeutung von Zeugnisnoten und ihre prädiktive Validität wurde schon in Kapitel 2.1.2.1 verwiesen. Hinweise mit biographischen Charakter gemäß o.g. Definition finden sich aber besonders in den Lebensläufen der Bewerber, die fast jeder Bewerbung beiliegen und in der Regel vom Unternehmen in freier oder tabellarischer Form eingefordert werden. Hier sind es häufig Informationen über gegebene berufliche Qualifikationen, die den Lebensläufen durch Personalverantwortliche entnommen werden. Denn aus der Sicht vieler Praktiker spielt die berufliche Erfahrung eine besondere Rolle bei der Auswahl von neuen Mitarbeitern. Schuler und Marcus (2001) kommen jedoch bei ihrer Bewertung zu dem Ergebnis, dass die intuitive Bedeutungseinschätzung der Praktiker sich kaum in metaanalytischen Studien belegen lässt. Die aus dem Lebenslauf ersichtlichen Elemente der Berufserfahrung seien eher

mäßig valide. Problematisch bleibt bei der Bewertung biographischer Prädiktoren aus Bewerbungsunterlagen zudem, dass die Inhalte i.d.R. nicht standardisiert sind oder nur teilweise als halbstandardisiert gelten können (bspw. Arbeitszeugnisse früherer Arbeitgeber) und aufgrund der im o.g. Kapitel dargestellten Hintergründe nur eine geringe Aussagekraft besitzen. Wohl auch wegen fehlender Vergleichbarkeit und Standards fordern einige Unternehmen die Bewerber grundsätzlich zu Angaben in einem vorgegebenen standardisierten Personalfragebogen auf. Die Auswahl der in diesen zusammengefassten Fragen ist nach Schuler und Marcus (2001) jedoch häufig beliebig und den jeweiligen Vorlieben der Verantwortlichen im Unternehmen angepasst. Historisch betrachtet, liegt in diesem Verfahren aber der Ursprung des biographischen Fragebogens. Auf der Grundlage der in derartigen Kontexten erfassten Items wurde schon früh zu Beginn des letzten Jahrhunderts begonnen, einzelne Informationen hinsichtlich ihrer prädiktiven Validität zu überprüfen.

Doch was genau ist unter einem biographischen Fragebogen zu verstehen, wie ist dieser von anderen Fragebogen wie Persönlichkeitsfragebogen oder Personalfragebogen abzugrenzen?

Für Owens (1976) sind biographische Fragebogen Verfahren zur Selbstbeschreibung, die sich aus Items zusammensetzen, welche dem Beantwortenden erlauben sich selber in lebensgeschichtlichen / demographischen, erfahrungs- oder einstellungsbezogenen Variablen zu beschreiben. Von diesen Variablen wird wiederum vermutet, dass sie in Verbindung zu persönlichen Strukturen, Einstellungen oder Erfolgen im sozialen, erzieherischen oder beruflichen Bereichen stehen (S.612 ff, Übers. d. Verf.). In der Regel werden dem Beantwortenden zu einer Frage unterschiedliche Antwortalternativen vorgegeben, aus denen er die für ihn passende aussuchen muss.

Biographische Fragebogen unterscheiden sich von klassischen Personalfragebogen, da letztere fast ausschließlich demographische Variablen erfassen (Dunette und Kirchner, 1960). Attribute, die ein biographisches Item als solches qualifizieren und von anderen unterscheidet hat Mael (1991) systematisiert. Als Kernaussage seiner Arbeit ist zu entnehmen, dass biographische Items historischen Charakter haben sollten, d.h. sich auf konkrete Ereignisse in der Vergangenheit beziehen. Hypothetische Fragen sowie Fragen nach Verhaltensabsichten wie bspw. „Was würden Sie tun wenn ...?“ gehören nicht in einen biographischen Fragebogen.

Die o.g. Definition von Owens und die in vielen biographischen Fragebogen gegebene Möglichkeit Summenscores zu bilden, macht dennoch eine Unterscheidung zu

Persönlichkeitsfragebogen schwierig. Dies wird auch von Jäger (1992) betont. Er nimmt an, dass mit biographischen Fragebogen eher inventarisiert wird, während Persönlichkeitsfragebogen eher zur Testung herangezogen werden. Hollmann (1991) reflektiert diese Annahme kritisch und verweist auf Pawlik (1976) und Schmidtke (1992), wonach eine Inventarisierung zur Beantwortung von klinischen Fragestellungen im Vordergrund steht, dagegen bei biographischen Fragebogen, die zur Personalsektion eingesetzt werden, nicht. Knoblauch (1986) stellt in Tabelle 2.2-2 zur besseren Abgrenzung exemplarisch biographische Items anderen Fragetypen gegenüber.

Tab. 2.2-2: Biographische Frage im Vergleich mit anderen Fragetypen nach Knoblauch (1986)

Orientierung an	Item
Biographie	Wie viele Bücher über Antibiotika haben Sie in den letzten zwei Jahren gelesen?
Persönlichkeit	Würden Sie lieber allein neue Antibiotika erforschen oder lieber beim Vertrieb dieser Präparate mitwirken?
Interesse	Interessieren Sie sich für die Bedeutung und Wirkungsweise von Antibiotika?
Meinung	Werden Antibiotika häufig unnötigerweise verschrieben?
Wissen	Welches Antibiotikum hat Sir Alexander Fleming entdeckt?

Der wesentliche Unterschied zu Persönlichkeitstests dürfte jedoch im Sinne von Mael (1991), im zeitlichen Bezug der Fragestellungen zu finden sein. Fragen in einem biographischen Fragebogen stellen überwiegend einen Bezug zur Vergangenheit her. Knoblauch (1986) weist zudem darauf hin, dass biographische Fragebögen, im Gegensatz zu Leistungs- und Persönlichkeitstests, parallel die Prognose der Ausbildungs- wie auch die Berufsbewährung gewährleisten können.

Einem weiteren Unterscheidungsaspekt kommt auch Bedeutung zu bei der Bewertung der Fälschungssicherheit biographischer Fragebögen. Werden fehlende Offenheit oder die Bereitschaft zu unwahren Angaben bei Fragen zum eigenen Erleben und Empfinden schon bei der Konstruktion eines Persönlichkeitsfragebogen durch eine eigene Skala, eine sogenannte „Offenheitsskala“, berücksichtigt (vgl. hierzu a. Mummendey, 1987), ist dies in biographischen Fragebögen nicht vorgesehen. Bevor jedoch die Hintergründe hierfür erörtert werden können, sollte an dieser Stelle, angesichts der teilweise schon mehrere Jahre zurückliegenden Bezugspunkte in der Biographie eines Bewerbers, die Begrifflichkeit „Fälschungssicherheit“ näher definiert werden. Gerade hier ist zu erwarten, dass Antworten durch Erinnerungsfehler beeinflusst werden. Damit können teilweise durch den Beantwortenden zwar falsche Angaben im zeitlichen Rückblick auf seine Biographie erfolgen, diese sind jedoch nicht zwangsläufig vorsätzlich gemacht worden. Es sollte also die bewusste Fälschung biographischer Fragebögen von irrtümlich gemachten Angaben abgegrenzt werden. Treffender dürfte daher in diesem Zusammenhang der Begriff der

Glaubwürdigkeit der von den Beantwortenden gemachten Angaben sein, wie er in der forensischen Psychologie Anwendung findet. Glaubwürdige und damit nicht gefälschte Angaben sind von den bewusst gemachten falschen Antworten zu trennen. Insbesondere in der forensischen Aussagepsychologie erfährt Glaubwürdigkeit eine konkrete und spezifische Beschreibung. Es steht hier der Inhalt einer Aussage im Vordergrund, die vergleichbar mit der Bewertung der Antworten in einem biographischen Fragebogen gesehen werden kann. Nach Köhnken (1990) und Wegener, Köhnken und Steller (1988) wird Glaubwürdigkeit als Wahrheitsvorsatz innerhalb einer Aussage verstanden und damit als Abwesenheit von intendierten Täuschungen. So schreibt Köhnken (1990, S.4): "Glaubwürdigkeit liegt vor, wenn ein Kommunikator an einen Rezipienten eine Information vermittelt, von der er glaubt, dass sie zutreffend ist, wenn also der Kommunikator keine Täuschungsabsicht hat. Entscheidendes Bestimmungsstück ist die Intentionalität." Täuschungen liegen nach Köhnken (1990) dann vor, wenn zumindest zwei Personen beteiligt sind, wobei eine als Kommunikator absichtlich eine falsche Information vermittelt und diese an eine zweite als Rezipienten richtet.

Bei Fragestellungen, die einen länger zurückliegenden zeitlichen Kontext beleuchten sollen, ist damit zu rechnen, dass Erinnerungsverzerrungen sowie falsche Einschätzungen aus heutiger Sicht zu nicht wahrheitsgerechten Antworten führen, von denen der Beantwortende jedoch ehrlich glaubt, dass es sich wie dargestellt verhalten habe. Damit gemeint sind in biographischen Fragebogen Aspekte einer Biographie, deren Phasen teilweise schon viele Jahre zurück liegen, wie bspw. Fragen nach dem Arbeitsverhalten in der Schule oder Ausbildung, den Mitgliedschaften in Organisationen, in denen der Bewerber schon in seiner Jugend früh verantwortliche Positionen hat übernehmen können etc.. Neben der motivationalen Komponente, der bewussten Fälschung der Angaben im Fragebogen, sind also unbeabsichtigte Fehler und kognitive Ursachen von einer nicht wahrheitsgemäßen Beantwortung, wie z.B. Gedächtnisstörungen, Wahrnehmungsverzerrungen o.ä. abzugrenzen.

Auch wenn die Differenzierung der Begrifflichkeit im o.g. Kontext noch keinen Eingang in die Bewertung der Fälschungssicherheit biographischer Fragebögen gefunden haben mag, verweisen Baehr und Williams (1967) darauf, dass die generelle Gefahr der Fälschung bei biographischen Fragebogen wesentlich geringer sei als bei Persönlichkeitsfragebögen, da neben Meinungen und Einstellungen hauptsächlich Tatsachendaten erhoben werden, die schwieriger zu fälschen sind. Die geringe Neigung die Daten zu fälschen, überträgt sich auch auf die anderen Items eines Fragebogens. Zu vermuten ist hier die Annahme, dass ein Teil der

erhobenen Daten wie Schulabschluss, Qualifizierungsnachweise oder Fortbildungszertifikate in der Regel durch vorgelegte Belege vom Bewerber zu verifizieren sind. Hollmann (1991) zitiert zudem unterschiedliche Autoren (bspw. Keating, Patterson & Stone, 1959; Kavanagh & York, 1972, Blum & Stumpf, 1983), welche die Fälschungssicherheit empirisch belegen konnten, kann aber auch keine weitergehende Erklärung aufzeigen. Neben der genannten teilweisen Überprüfbarkeit einzelner Items, sieht er sie auch durch die hohe Retest-Reliabilität biographischer Fragebogen begründet. Die Ergebnisse vorliegender Untersuchungen sprechen dafür, dass sich biographische Fragebogen im Rahmen der Berufseignungsdiagnostik zumindest fälschungstabil nutzen lassen.

2.2.1.2 Theoretischer Hintergrund biographischer Eignungsdiagnostik

Das generelle Interesse der letzten Jahren hinsichtlich der Entwicklung von Theorien über Ursachen und Zusammenhänge der beruflichen Leistung gilt auch für die biographisch orientierte Eignungsdiagnostik. Allgemeines Ziel ist es, die psychologischen Prozesse, die der Leistung zugrunde liegen und diese bestimmen, besser zu verstehen. Schmidt und Hunter (1998) verweisen auf Schmidt und Rothstein (1994), die postulieren, dass einige Verfahren Konstruktkombinationen messen und die tatsächlichen Konstrukte unbekannt sind. Ein Beispiel hierfür wären Vorstellungsgespräche, die vermutlich eine Kombination aus früheren Erfahrungen, mentalen Fähigkeiten und einer Reihe von Persönlichkeitsmerkmalen erfassen. So ist nach den Autoren ein wesentlicher Grund, warum allgemeine kognitive Fähigkeiten einen guten Prädiktor darstellen, im vorhandenen Berufswissen zu finden. Intelligenter Menschen eignen sich schneller und mehr Berufswissen an als andere.

Es existiert trotz verschiedener Bemühungen bislang immer noch keine Theorie biographischer Daten. Die theoretische Verknüpfung von gewonnenen Einzeldaten zu einem Konstrukt Persönlichkeit fehlt. Seit den 40er Jahren wird Kritik an der klassischen Form der Anwendung biographischer Daten geübt. Im Vordergrund steht der Vorwurf, dass es sich bei der biographischen Diagnostik um einen ausschließlich empirisch orientierten Ansatz handelt, der in keiner Beziehung zu den wichtigsten Ergebnissen der Verhaltens- oder Persönlichkeitsforschung steht. Oder wie Savage (1986, S.76) es zusammenfasst: „Wir mögen wissen, dass ein Item Individuen unterscheiden kann in Bezug auf irgendwelche Kriterien; aber solange wir nicht verstehen, warum es das tut, und wie sich das zu anderen Items verhält, gibt es keine verlässliche Basis für die Behauptung, es sei irgend etwas Wissenschaftliches

oder Rationales an diesem Prozess. Er ist reduziert auf a priori-Annahmen über statistische Methoden.“

Trotz der in unterschiedlichen Untersuchungen wiederholt belegten prädiktiven Validität dieses Auswahlinstrumentes, fehlt es immer noch an ausreichend Studien, welche die Konstruktvalidität biographischer Fragebögen zu erfassen suchen und belegen. Dies mag mit ein Grund sein, warum der Einsatz solcher Fragebögen in erster Linie auf das Feld der eignungsdiagnostischen Auswahl beschränkt ist. Hier sind die Fragebögen häufig nur für eine stark eingeschränkte Population und auch nur für ausgesuchte Kriterien validiert. Eine Generalisierung der erzielten Ergebnisse auf andere Anwendungsfelder ist aus diesem Grund häufig nicht möglich, von den Anwendern aber oft auch nicht beabsichtigt. Nach Hollmann und Reitzig (1995) wird ein ausgeliehener Fragebogen, der speziell für bestimmte zu besetzende Stellen konzipiert wurde, zwangsläufig zu einer unzusammenhängenden Sammlung von Indikatoren ohne Konstrukt. So werden zwar biographische Fragebögen insbesondere hinsichtlich ihrer prädiktiven Validität im Zusammenhang mit anderen eignungsdiagnostischen Auswahlinstrumenten genannt. Oft handelt es sich dabei jedoch nur um mehr oder weniger umfangreiche und heterogene Itemsammlungen ohne jede Auswertungshilfe oder Normwerte. Die eigentliche Prozedur der Testkonstruktion und die Ermittlung eines Auswerteschlüssels muss für jeden Anwendungsfall neu durchlaufen werden (Hollmann, 1991).

Die Validität ist die wichtigste Bedingung, die ein eignungsdiagnostisches Verfahren erfüllen muss (siehe auch Kapitel 2.3). Mittels eines biographischen Fragebogens sollte es möglich sein, richtige Schlüsse aus den erhobenen Messwerten zu ziehen. Nur durch diesen Interpretationsschritt ist es möglich, aus demselben biographischen Item eines Fragebogens auf den späteren Berufserfolg eines Probanden zu schließen. Dass biographische Fragebögen über eine ausreichend prädiktive Validität verfügen, wurde schon wiederholt dargestellt. Auf die hier erfasste kriterienbezogene Validität, aber auch auf die Inhaltsvalidität wird unter Kapitel 2.3 eingegangen. An dieser Stelle sollen ausschließlich Hinweise zur Konstruktvalidität von biographischen Fragebogenverfahren skizziert werden. Hier steht die Frage der psychologischen Bedeutung eines Testresultates, in diesem Fall das Antwortverhalten eines Probanden im Fragebogen, im Vordergrund. Es wird konkret danach gefragt, was der Fragebogen erfasst, was er misst. Hierbei wird ein bestehender Zusammenhang zwischen dem Antwortverhalten im Fragebogen und einem nicht direkt beobachtbaren, sondern zu erschließenden psychologischem Konstrukt nachgewiesen

(Lienert, 1989). Das hierfür zu wählende Vorgehen entspricht nach Cronbach (1970) der wissenschaftlichen Theorieentwicklung und Hypothesenprüfung. So ist einerseits der Schluss vom erreichten Testergebnis eines Antwortenden im Fragebogen auf ein persönlichkeitspsychologisches Konstrukt möglich, andererseits lassen sich bestimmte Antwortverhalten in der Analyse von Fragebögen durch psychologische Konstrukte erklären.

Cronbach (1970) beschreibt in Tabelle 3.2-2 Bedingungen im Vorgehen, um die Bedeutung eines Testergebnisses herauszustellen. Grundsätzlich sind dabei für jedes eignungsdiagnostische Verfahren eine größere Anzahl von Studien notwendig.

Tab.3.2-2: Drei Bedingungen, im Vorgehen zur Konstruktvalidierung eines Testverfahrens nach Cronbach (1970)

1.	Auf der Grundlage bisheriger Beobachtung und psychologischen Wissens eine Vermutung, welche Konstrukte ein Testergebnis erklären kann
2.	Auf der Grundlage von Überlegungen und Schlussfolgerungen die Ableitung prüfbarer Hypothesen aus einer Theorie
3.	Abschließend die Durchführung empirischer Studien, um vorliegende Hypothesen zu testen

In Abhängigkeit von den Hypothesen dürfte die jeweilige Vorgehensweise unterschiedlich aussehen. Funke und Schuler (1986) stellen in Anlehnung an Lienert (1969), Cronbach (1970), Anastasi (1982), Michel und Conrad (1982) beispielhaft die verschiedenen Techniken dar, von denen mehrere zur Anwendung kommen und bei denen der Methode der Faktorenanalyse meist eine besondere Bedeutung zu kommt. Im folgenden soll ansatzweise auf unterschiedliche Zugangsmöglichkeiten zur Bestimmung der Konstruktvalidität eingegangen werden.

Bei der Korrelation des Testergebnisses mit einem Außenkriterium können erste Schlüsse gezogen werden, welche psychologischen Variablen in einem Testergebnis enthalten sind. So könnte der Testwert eines biographischen Fragebogens neben der Korrelation mit dem Kriterium Verkaufsleistung eines Verkäufers auch mit dem Kriterium für Kontaktfähigkeit hinsichtlich eines statistischen Zusammenhanges überprüft werden. Auf diese Weise wären möglicherweise erste Hinweise auf einen psychologischen Einflussfaktor für die Verkaufsleistung identifizierbar.

Eine weitere korrelationsstatistische Technik bietet die Erfassung eines Zusammenhanges mit anderen Testverfahren. Voraussetzung hierfür ist jedoch, dass die Validitätsbereiche der verwendeten Testverfahren bekannt sind. Hohe Zusammenhänge sind mit Testverfahren zu

erwarten, die dem vermuteten Konstrukt des biographischen Fragebogens nahe stehen („convergent validity“ n. Campbell & Fiske, 1959). Andererseits sollten die Zusammenhänge mit konstruktfernen Testverfahren niedrig bis gar nicht gegeben sein („discriminat validity“). Funke und Schuler (1986) verweisen in diesem Zusammenhang auf die Untersuchung von Eberhardt und Muchinski (1982), die versuchten, sechs berufliche Hauptinteressenrichtungen durch biographische Daten vorherzusagen, was für den Bereich wissenschaftlicher Interessen am besten gelang. Aufgrund der oben dargestellten Annahme, dass das Antwortverhalten in biographischen Fragebogen im Zusammenhang mit kognitiven Fähigkeiten zu sehen ist, wäre demnach auch ein hoher statistischer Zusammenhang mit kognitiven Fähigkeitstests zu erwarten. Andererseits könnte man die Hypothese aufstellen, dass bei dem Antwortverhalten eines erfolgreichen Verkäufers nur ein geringer Zusammenhang mit dem psychologischen Konstrukt „Introversion“ zu erwarten wäre.

Eine weitere Möglichkeit bietet die Faktorenanalyse. Mittels dieses Verfahrens lassen sich eine oft unübersehbare Anzahl von Interkorrelationen auf eine überschaubare Zahl von Faktoren reduzieren. Die Interpretation dieser Faktoren bietet dann wiederum Gelegenheit, wichtige Zusammenhänge aufzudecken. So ließe sich bspw. ein biographischer Fragebogen zusammen mit einem zeitgleich erhobenen Persönlichkeitsfragebogen faktorenanalisieren.

Liegen bereits Gruppen vor, von denen Unterscheidungen bekannt und definiert sind, können die vorliegenden Gruppen hinsichtlich ihrer Unterschiede im Antwortverhalten eines biographischen Fragebogens untersucht werden. Albright und Glennon (1961) konnten feststellen, dass sich forschungsinteressierte und managementinteressierte Wissenschaftler in einem biographischen Fragebogen bezüglich Arbeits- und Sozialverhalten signifikant unterschieden.

Weitere Zugangsmöglichkeiten zur Erschließung der Konstruktvalidität wie experimentelle Variationen, Testwiederholung oder Analyse der Items siehe Funke und Schuler (1986).

Mögliche theoretische Erklärungen für die hinter biographischen Daten stehenden Konstrukte finden sich nach Jäger (1992) in Lern- und Sozialisationstheorien. Diese können nach seiner Meinung herangezogen werden, um biographische Daten theoretisch zu fundieren. So erklärt Cassens (1966), dass Verhaltensweisen, die von einer Person in der Vergangenheit gezeigt wurden und sich als erfolgreich erwiesen haben, sowie Merkmale und Ereignisse aus dem zurückliegenden Leben als die besten Prädiktoren zukünftigen Verhaltens gelten. Elemente

der Lebens- und der Lerngeschichte könnten für die Validität biographischer Daten mitverantwortlich sein. Zwar unterstützt auch Owens (1976, S. 625) diese Annahme mit den Worten „...the best predictor of what a man will do in the future is what he has done in the past“; doch darf bei dieser augenscheinlich plausiblen Aussage nicht außer acht gelassen werden, dass gerade im biographischen Fragebogen größtenteils nicht nur konkretes Verhalten hinterfragt wird, sondern auch Einstellungen und Beweggründe. Ein derartiger Induktionsschluss beim Einsatz biographischer Fragen, aber auch anderer eignungsdiagnostischer Instrumente wie z.B. Schulzeugnisse, ist aber nicht zwingend, denn oftmals werden Bewerber für ihnen vollständig neue Tätigkeiten ausgewählt, bei denen gleichartiges früheres Verhalten nicht unbedingt gefordert sein muss (Funke & Schuler 1986).

1967 traten Baehr und Williams mit ihrer Untersuchung der bestehenden Kritik entgegen, die Anwendung biographischer Fragebogen in der Eignungsdiagnostik sei reine Empirie ohne theoretische Absicherung der Verfahrensweise oder Ergebnisse. Sie entwickelten einen 150 biographische Fragen umfassenden Fragebogen. Anschließend wurde er Vertretern unterschiedlicher Berufsgruppen wie bspw. Chemikern, Führungskräften, Führungsnachwuchs, Außendienstmitarbeitern, Verkaufsleitern, Vorarbeitern, Verwaltungsbeamten und Lehrern vorgegeben. Die durchgeführte Faktorenanalyse mit der Hauptachsenmethode und schließlich obliquen Rotation führte zu 15 Faktoren wie bspw. Leistungen im Schul- oder Hochschulbereich, Ausbildungszeitraum, Leistungsbereitschaft, Kontaktfähigkeit und -bereitschaft, finanzielle Verantwortlichkeit, frühe familiäre Verantwortung, realistische und konstruktive Einstellungen in der Kindheit, familiärer Hintergrund und stabile Arbeitssituation. Durch eine anschließend gerechnete Varianzanalyse konnten die Autoren belegen, dass die gefundenen Faktoren zwischen den Berufsgruppen diskriminieren und in einem zweiten Schritt ihrer Untersuchung, dass anhand der Faktoren auch der Berufserfolg bei Führungskräften prognostizierbar ist.

Funke und Schuler leisteten 1986 mit ihrer Untersuchung einen weiteren Beitrag zur Konstruktaufklärung biographischer Fragebogen. Sie verwendeten die Experimentalform eines Fragebogens, der Fragen zu den Bereichen Schule, Berufsverhalten und -motivation, Freizeitverhalten und -interessen, Familie, soziale Aktivitäten sowie Selbstbild und Attributionsstil enthielt. Die Datenerhebung erfolgte an zwei unterschiedlichen Stichproben: Zum einen handelte es sich um 65 Bewerber für eine Ausbildung zum Industriekaufmann, zum anderen um 91 Auszubildende zur chemisch-technischen bzw. pharmazeutisch-

technischen Assistentin. Als Außenkriterium lagen für die erste Stichprobe u.a. Ergebnisse des Freiburger Persönlichkeitsinventars (FPI) von Fahrenberg, Selg und Hampel (1978), des Intelligenz-Struktur-Tests-70 (Amthauer, 1970), des Aufmerksamkeits-Belastungs-Tests d2 (Brickenkamp, 1981) sowie des Berufs-Interessen-Tests (BIT) von Irlle (1955) vor. Für die zweite Stichprobe wurden die Ergebnisse der Ausbildungsleistung, der Angaben über die Arbeitszufriedenheit aus dem Arbeitsbeschreibungsbogen von Neuberger und Allerbeck (1978), aus Unterformen des IST 70 sowie aus dem Berufseignungstest (BET) von Schmale und Schmidtke (1966) als Kriterium herangezogen.

Die anschließende Faktorenanalyse, die zur Interpretation und Benennung der gefundenen Faktoren durchgeführt wurde, ergab acht Faktoren, die insgesamt 41,8 % der Varianz aufklärten. Diese Faktoren ließen sich vier Bereichen (Tabelle 4.2-2) zuordnen.

Im Gesamtergebnis kamen die Autoren u.a. zu dem Schluss, dass es eine befriedigende Übereinstimmung von Berufsinteressenfaktoren aus dem biographischen Fragebogen mit BIT-Skalen sowie einen engen Zusammenhang zwischen Berufsinteressenfaktoren mit persönlichen Fähigkeiten (IST) sowie Persönlichkeitseigenschaften (FPI) gibt.

Weitere Möglichkeiten, wie bereits bei der Konstruktion von biographischen Fragebögen berufsrelevante Konstrukte berücksichtigt werden können, zeigt bspw. Hollmann (1990, S. 71 – 78) auf.

Wenn sich auch insbesondere die Forschergruppe um William S. Owens in den 70er Jahren im angloamerikanischen Raum um neue Ansätze bemühte, die einen Weg weg vom Vorwurf des Empirismus aufzeigen sollten (s.a. Kapitel 2.2.2.1), geht Keßler noch 1982 davon aus, dass biographische Daten allgemein von der Forschung vernachlässigt wurden. Denn auch wenn die Biographieforschung im deutschsprachigen Raum eine lange Tradition besitzt (Thomae und Jüttemann, 1987), sind, im Gegensatz zum angloamerikanischen Raum, die Bestrebungen zur Entwicklung biographischer Fragebogen in Deutschland noch relativ jung (Owens, 1976; Stehle, 1986). In den letzten Jahren hat es zudem keine neuen Veröffentlichungen gegeben, die einen neuen Beitrag zur Klärung der Konstruktvalidität liefern können. Angesichts der häufigen Anwendung in der Praxis und der prädiktiven Validität biographischer Fragebögen ist und bleibt hier ein Forschungsdefizit, das dringender Klärung bedarf.

Tab.4.2-2: Faktorenstruktur einzelner Inhaltsbereiche eines biographischen Fragebogens nach Funke und Schuler (1986)

Bereich	Faktor Stichprobe 1	Faktor Stichprobe 2
Interessen		
- kaufmännisch-verwaltend	1	1
- naturwissenschaftlich	2	4
- schriftstellerisch	3	6
- technisch-handwerklich	4	3
- häuslich und Natur	5	2
- Bankwesen	6	5
Selbstbild		
- Extraversion	1	4
- Neurotizismus	2	6
- Leistungsmotivation	3	2
- Gewissenhaftigkeit	4	1
- Selbstsicherheit	5	5
- Rigidität	6	3
Attribution schlechter Leistungen		
- intern-variabel	1	1
- extern-variabel	2	2
- extern-stabil	3	3
Allgemeine Biographie		
- Zurückgezogenheit	1	1
- Extraaktivitäten	5	2
- Berufliche Entscheidungen	3	3
- Schulische Leistung	4	4
- Introversion	7	5
- Soziale Herkunft	2	7

2.2.2 Biographische Daten in der Eignungsdiagnostik

Trotz eines bestehenden Defizits in der theoretischen Erklärung, warum die Faktoren, die unseren Lebenslauf markieren, einen vorhersagenden Beitrag in Auswahlverfahren leisten können, fanden und finden nach wie vor Fragebögen unterschiedlichster Form und Konstruktion in der Praxis Anwendung. Fragen hinsichtlich der Art, nach der diese Fragebögen konstruiert und wie sie in der eignungsdiagnostischen Praxis verwendet werden, sind im folgenden Kapitel dargestellt.

2.2.2.1 Konstruktion Biographischer Fragebögen

Bei der Konstruktion biographischer Fragebögen lassen sich historisch zwei klassische Vorgehensweisen unterscheiden, der empirische Ansatz oder die sog. Summen-Score-Modelle und der rationale oder auch faktorenanalytische Ansatz.

Aus der Zeit vor dem ersten Weltkrieg datieren die Ursprünge des empirischen Ansatzes. Nachdem ein sogenannter Arbeitsfragebogen mit 100 – 200 Items nach inhaltlichen Gesichtspunkten zusammengestellt wurde, wird in einem zweiten Schritt jedes Item daraufhin untersucht, ob es in statistisch bedeutsamer Weise Zusammenhänge zu einem

Erfolgskriterium aufweist. In der Regel werden zwei Extremgruppen gewählt, die sich in der Ausprägung ihres beruflichen Erfolges deutlich unterscheiden. Ziel ist es, die Items zu suchen, die beide Gruppen auch in der Beantwortung signifikant charakterisieren. Ein empirisch nachweisbarer Unterschied führt dazu, dass das betreffende Item in den endgültigen Testwert eingeht. Die verbleibenden Items werden aus der Skala entfernt, bzw. in der Auswertung nicht berücksichtigt. Diese sogenannte Kontrastgruppenmethode ist dabei vollkommen unabhängig vom Inhalt des Items und keineswegs auf biographische Items beschränkt. Prinzipiell ließe sich auch ein Item wie die signifikant unterschiedliche Körpergröße von erfolgreichen vs. nicht erfolgreichen Stelleninhabern entsprechend berücksichtigen. Ein prominentes Beispiel dieser Vorgehensweise außerhalb des biographischen Kontextes ist der MMPI (Minnesota Multiphasic Personality Inventory) von Hathaway und McKinley (1951). Hier stand anfänglich eine Liste von 1000 Items zur Verfügung, die sich auf allgemeine Gesundheit, familiäre und eheliche Beziehungen, sexuelle und religiöse Einstellungen sowie emotionale Zustände bezogen. Die Items wurden unterschiedlichen, von psychiatrischen Gutachtern klassifizierten Probandengruppen (Schizophrene, Hysteriker, Hypochonder usw.) zur Bearbeitung vorgegeben. Jene Items, welche die Patienten signifikant von den Kontrollgruppen diskriminierten wurden schließlich in Skalen zusammengefasst.

Das bei Einsatz von Kontrastgruppen häufigste Verfahren ist der Ansatz des „Weighted application blank approach“ von England (1961, 1971). Hier wird jedes Item aufgrund seiner Kriteriumsvalidität ausgesucht und entsprechend gewichtet. Dies entspricht, das Skalenniveau bei dem Vergleich nicht beachtend, einem regressionsanalytischen Ansatz. In der Itemanalyse wird für jeden Prädiktor (bei England jede Antwortalternative) anhand der Prognosegüte ein empirisches Gewicht ermittelt. Die empirische Itemgewichtung führt in der dafür eingesetzten Stichprobe zu einer fast maximalen Kriteriumsvalidität. Diese Vorgehensweise führt allerdings dazu, dass der Itemschlüssel die Besonderheiten der Stichprobe berücksichtigt und kaum auf andere Kontexte übertragbar ist.

Aufgrund seines historischen Einflusses und der Bedeutung für die vorliegende Arbeit soll an dieser Stelle auf den Ansatz von England ausführlicher eingegangen werden.

Entsprechend den Vorgaben von England (1961, 1971) wird im ersten Schritt für jede Antwortalternative der Prozentsatz der jeweiligen Kontrastgruppe (erfolgreiche Mitarbeiter vs. wenig erfolgreiche Mitarbeiter) errechnet, der die vorgegebene Alternative gewählt hat.

An einem Beispiel-Item eines biographischen Fragebogens „Welche Einstellung beschreibt Ihre bisherige Arbeitsweise am ehesten?“ wird dieser Schritt in Tabelle 5.2-2 beschrieben.

Tab.5.2-2: Prozentuale Antworthäufigkeit in Kontrastgruppen nach England (1961, 1971)

Antwortalternativen	Erfolgreiche Mitarbeiter	Weniger erfolgreiche Mitarbeiter
a) Einen gemachten Plan halte ich konsequent ein	12,9 %	20,2 %
b) Einen gemachten Plan versuche ich möglichst einzuhalten	39,2 %	33,2 %
c) Ich plane ein Ziel, den Weg dorthin gehe ich spontan	8,1 %	14,4 %
d) Ich plane nicht, sondern entscheide stets aus der Situation heraus	12,2 %	18,3 %
e) Ob ich plane oder nicht hängt immer vom Ziel ab.	27,7 %	13,9 %

Wie in Tabelle 6.2-2 dargestellt, wird im zweiten Schritt aus den beiden Prozentwerten der Kontrastgruppen schließlich die Differenz berechnet.

Tab.6.2-2: Prozentdifferenz der Kontrastgruppen nach England (1961, 1971)

Item	Erfolgreiche Mitarbeiter	Weniger erfolgreiche Mitarbeiter	Prozentdifferenz
a	12,9 %	20,2 %	- 7,3 %
b	39,2 %	33,2 %	6,0 %
c	8,1 %	14,4 %	- 6,3 %
d	12,2 %	18,3 %	- 6,1 %
e	27,7 %	13,9 %	13,8 %

Die errechneten Differenzen werden dann in einem dritten Schritt nach den von Strong (1926) tabellierten Nettogewichten zugeordnet (Tabellen 7.2-2 und 8.2-2), die eine Aussage über die Richtung und das Ausmaß treffen, in dem die Antwortalternative zwischen den beiden Gruppen differenziert.

Tab.7.2-2: Tabellierte Nettogewichte der Prozentdifferenzen nach Strong (1926)

Strong`s Tables of Net Weights for Differences in Per Cent					
To be used when both per cent are between 8 and 92		To be used when one per cent is between 3 and 7, or 93 and 97		To be used when one per cent is between 0 and 2, or 98 and 100	
Diff. in Per Cents	Net Weight	Diff. in Per Cents	Net Weight	Diff. in Per Cents	Net Weight
69	27	69	27	69	28
68	26	68	26	68	27
67	25	67	25	67	26
66	24	66	24	66	25
65	23	64 – 65	23	65	24
64	22	63	22	63 – 64	23
62 – 63	21	62	21	62	22
61	20	60 – 61	20	60 – 61	21
60	19	58 – 59	19	59	20
58 – 59	18	57	18	57 – 58	19
56 – 57	17	55 – 56	17	55 – 56	18
54 – 55	16	53 – 54	16	53 – 54	17
52 – 53	15	50 – 52	15	51 – 52	16
50 – 51	14	48 – 49	14	49 – 50	15
48 – 49	13	45 – 47	13	46 – 48	14
45 – 47	12	42 – 44	12	43 – 45	13
42 – 44	11	39 – 41	11	40 – 42	12
39 – 41	10	35 – 38	10	36 – 39	11
36 – 38	9	31 – 34	9	32 – 35	10
33 – 35	8	27 – 30	8	28 – 31	9
29 – 32	7	23 – 26	7	24 – 27	8
24 – 28	6	19 – 22	6	19 – 23	7
21 – 23	5	15 – 18	5	15 – 18	6
16 – 20	4	11 – 14	4	11 – 14	5
12 – 15	3	7 – 10	3	7 – 10	4
8 – 11	2	4 – 6	2	4 – 6	3
3 – 7	1	2 – 3	1	2 – 3	2
0 – 2	0	0 – 1	0	1	1
				0	0

Tab.8.2-2: Nettogewichte der Prozentdifferenzen im Beispiel nach Strong (1926)

Strong's Tables of Net Weights for Differences in Per Cent	
a	- 1
b	1
c	- 1
d	- 1
e	- 3

Dabei empfiehlt England (Tabellen 9.2-2 und 10.2-2) auf die Information über das Ausmaß zu verzichten, um den Itemschlüssel zu vereinfachen und sich nicht zu stark an Zufallseinflüsse bei der gegebenen Stichprobe anzupassen. Daher werden nur bedeutsame Nettogewichte berücksichtigt und in einheitliche Grobgewichte verwandelt.

Tab. 9.2-2: Tabellierte Grobgewichte nach England (1961, 1971)

Table of Assigned Weights Derived from Net Weights	
- 4 or less	0
3, 2, 1, 0 or -1, -2, -3	1
4 or more	2

Tab. 10.2-2: Grobgewichte im Beispiel nach England (1961, 1971)

Item	Net Weight (nach Strong)	Assigned weights (nach England)
a	- 1	1
b	1	1
c	- 1	1
d	- 1	1
e	- 3	1

Items mit unzureichender Trennschärfe werden aus dem Fragebogen entfernt oder in der Auswertung nicht berücksichtigt. Der eigentliche Prädiktorwert eines biographischen Fragebogens wird dann ermittelt, indem die von den Teilnehmern gewählten Antwortalternativen, versehen mit dem jeweiligen Punktwert, aufsummiert werden.

Ein methodisches Problem, das an dieser Stelle jedoch nicht unerwähnt bleiben darf, ist die sog. immanente kompensatorische Skalenbildung (Knoblauch, 1986). So können Bewerber mit gleicher Punktezahl Defizite in ganz unterschiedlich wichtigen Anforderungen aufweisen, ohne dass die Möglichkeit besteht, diese bei der Beurteilung des Summenscores zu bemerken oder zu berücksichtigen.

Neben dieser empiristisch orientierten Vorgehensweise haben, insbesondere auch aufgrund der fehlenden inhaltlichen Interpretierbarkeit der durch diese Form gewonnenen Skalen, andere Konstruktionsmethoden Verbreitung gefunden. Mittels des sog. rationalen Ansatzes wird auf der Basis umfangreicher und sorgfältiger Arbeitsanalysen versucht, Gruppen von

biographischen Items zu quantifizieren, die eine interpretierbare Menge von Konstrukten messen. Dabei müssen die Items eines biographischen Fragebogens mit den Fähigkeiten, dem Wissen und Können der Arbeitsbeschreibungen übereinstimmen. Typisch für diese Vorgehensweise ist die Nutzung der Faktorenanalyse, mit der die Scores der Items für jede Dimension errechnet werden. Diese Werte werden dann zu bestimmten Kriterien in Beziehung gesetzt (Brush & Owens, 1979; Owens & Schoenfeldt, 1979). Statt der individuellen Items werden hier Faktor-Scores als Prädiktoren verwendet.

Klassischerweise ist an erster Stelle die theoriegeleitete Definition von Konstrukten, die bspw. mit beruflichem Erfolg im Zusammenhang stehen, zu erwähnen. Bei dieser deduktiven Vorgehensweise werden vor der empirischen Datenerhebung zur Messung der Konstrukte, biographische Items formuliert sowie Itemschlüssel und Skalen festgelegt. Nach Schuler und Moser (2001) liegt das vordringliche Ziel damit nicht in der Berechnung der Itemgewichte, sondern in der Itemgenerierung. Vordergründiges Ziel ist dabei die Maximierung der Inhaltsvalidität und die Interpretierbarkeit der Skalen.

Ein alternativer rationaler Ansatz findet sich mit der induktiven Methode. Mit Hilfe einer Faktorenanalyse wird dabei die interne Struktur eines Fragebogens analysiert. Ziel ist es, homogene Konstrukte, d.h. Faktoren zu identifizieren, die möglichst einen großen Teil der im Fragebogen enthaltenen Informationen auf eine möglichst geringe Anzahl von Faktoren reduzieren. Durch die Zusammenfassung der Items, welche die gewonnenen Faktoren gut repräsentieren, erfolgt die anschließende Skalierung.

Ein weiterer Ansatz - das sogenannte Subgrouping – entstand durch Arbeiten der Gruppe um Owens (1976; Owens & Schoenfeldt, 1979). In diesem theoretisch entwicklungspsychologischen Ansatz geht es um die Identifizierung stabiler Muster von Lebensereignissen, die sich zu einer generalisierbaren Beschreibung typischer Lebensereignisse eignen. Die objektivierbaren Ereignisse in einer Lebensgeschichte werden dann als Persönlichkeitstypen interpretiert. Nach Schuler und Marcus (2001) stellt die Subgrouping Methode den bislang ambitioniertesten Versuch dar, den biographischen Ansatz als eigenständiges theoretisches Konzept in der Eignungsdiagnostik zu etablieren. Mit Hilfe einer Clusteranalyse werden dabei nicht die Items nach Ähnlichkeit zusammengefasst, sondern Personen, die sich durch eine ähnliche Lebensgeschichte auszeichnen. Beispiele

verschiedener männlicher und weiblicher biographischer Subgruppen finden sich bei Owens und Schoenfeldt (1979).

Wenn auch die rationale Vorgehensweise bei der Konstruktion biographischer Fragebögen hinsichtlich der Interpretierbarkeit der Skalen und insbesondere bezogen auf die Konstruktvalidität dem empiristischen Vorgehen deutlich überlegen ist, hat sie doch für die Praxis einen entscheidenden Nachteil. Vergleichsstudien zwischen dem empirischen und dem rationalen Ansatz belegen die Überlegenheit des empirischen Vorgehens hinsichtlich der prädiktiven Validität (Mitchell & Klimoski, 1982; Hornick, James & Jones, 1977). Allerdings ist einschränkend anzumerken, dass bis heute nur wenige vergleichende Studien vorliegen (Weinert, 1998). Auf der anderen Seite verglichen Hough und Paullin (1994) in ihrer Metaanalyse systematisch den empirischen, deduktiven und induktiven Ansatz hinsichtlich der Kriteriumsvalidität. Dabei fanden die Autoren eine minimal höhere Validität des deduktiven Ansatzes gegenüber dem empirischen und des empirischen gegenüber dem induktiven.

2.2.2.2 Anwendung Biographischer Fragebögen

Wie schon im Kapitel 2.2.1 dargestellt, wird bei der Verwendung biographischer Fragebögen in der Berufs- und Eignungsdiagnostik immer wieder die überwiegend zufriedenstellende prädiktive Validität herausgehoben. Neben der relativ einfachen und ökonomischen Anwendungsmöglichkeit von Fragebogenverfahren generell liegt hier sicherlich eine zentrale Ursache für Verbreitung und Beliebtheit des Auswahlinstrumentes. Asher (1972) geht soweit zu behaupten, dass biographische Daten in diesem Kontext, verglichen mit anderen Merkmalen, wie bspw. Intelligenz, Persönlichkeit, Interessen und Eignungen, die besseren Prädiktoren seien. Diese Annahme wird allerdings, wie in Kapitel 2.1.2 dargestellt, von den Ergebnissen Schmidt & Hunter's (1998) u.a. nicht gestützt. Wenn gleich die biographischen Fragebögen unbestritten zu den Auswahlinstrumenten mit einer ausreichend hohen prädiktiven Validität zählen, bleibt auch heute noch unreflektiert, was genau diese Vorhersagegüte determiniert.

Tabelle 9.2-2 bietet eine Übersicht von Item Beispielen, wie sie in der biographischen Forschung häufig verwendet werden (Weinert, 1998). Vergleichbare Items finden sich heute in vielen zur Personalauswahl eingesetzten biographischen Fragebogen bzw. Personalbogen.

Tab.9: Beispiele biographischer Items, die in der Forschung häufig verwendet werden nach Weinert (1998)

<p>1. Persönliche Angaben</p> <ul style="list-style-type: none"> - Alter - Verheiratet / ledig - Wie viele Jahre verheiratet - Anzahl der zu versorgenden Kinder <p>2. Schulausbildung</p> <ul style="list-style-type: none"> - Art der Schulausbildung - Abschlüsse - Schulbildung des Partners - Beliebte Fächer - Noten <p>3. Sozioökonomischer Status</p> <ul style="list-style-type: none"> - Regelmäßige Ausgaben - Schulden - Wohneigentum - Erwartetes Einkommen <p>4. Interessen</p> <ul style="list-style-type: none"> - Hobbies - Sportaktivitäten - Bevorzugte Arbeiten - Wichtigste Freizeitaktivitäten <p>5. Allgemeine Hintergrundinformationen</p> <ul style="list-style-type: none"> - Beruf - Berufe von nahen Verwandten - Militärdienst / Rang - Elterliche Familie <p>6. Berufserfahrung</p> <ul style="list-style-type: none"> - Frühere Anstellungen - Anzahl früherer Berufe - Dauerstellung - Kündigungsgrund <p>7. Persönliche Charakteristika</p> <ul style="list-style-type: none"> - Einverständnis, sich an andere Orte versetzen zu lassen - Selbstvertrauen - Zentrale Grundbedürfnisse, die erfüllt werden müssen - Arbeiten und Anstellungen, die bevorzugt werden
--

Biographische Daten werden in der psychologischen Diagnostik, wie schon erwähnt, seit vielen Jahren verwendet. In den meisten Fällen handelt es sich dabei um Daten, die durch Interviews in Form einer sogenannten Anamnese erhoben werden. Häufig geschieht dies durch direkte Befragung der betroffenen Person oder alternativ ihrer Verwandten und Bekannten. In der Regel erfolgt die Verarbeitung dieser Daten in einer eher unsystematischen Art und Weise, d.h. durch Eindrucksbildung. Der Beurteilende sammelt die unterschiedlichen Ergebnisse zu einem subjektiven Gesamteindruck und macht von diesem Gesamtbild seine Entscheidung für oder gegen eine Intervention oder über die eingeschätzte zukünftige Leistung abhängig.

In einer systematisierteren Form fanden biographische Fragebogen im deutschsprachigen Raum jahrelang überwiegend im klinischen Bereich Verwendung und Verbreitung. Zu den bekannteren Fragebogenverfahren dieser Art gehören sicherlich das Mannheimer Biographische Inventar (MBI) von Jäger,R., Berbig,E., Gossler,H., Hagen,J., Lieblich,W. und Schafheutle, R. (1973) und das Biographische Inventar zur Diagnose von

Verhaltensstörungen (BIV) von Jäger, R. et al. (1976). So wurde bspw. das MBI, analog der bestehenden Annahme in der biographischen Eignungsdiagnostik, davon ausgehend konzipiert, dass für die Vorhersage zukünftigen Verhaltens das bisherige Verhalten der beste Prädiktor sei.

Mit Hilfe dieses Fragebogens sollte durch eine Quantifizierung anamnestischer Daten die Bestimmung der von dem Probanden erlebten Umweltsituation ermöglicht werden. 11 Skalen erfassten Inhaltbereiche wie z.B. Familiensituation, Lebensgewohnheiten, Interessen, soziale Aktivitäten, Kontaktverhalten oder Arbeitsverhalten. Daneben aber auch Skalen, die nach der eingangs geschilderten Definition unhinterfragt nicht in biographische Fragebogen gehören, wie bspw. „Kreatives Verhalten“, „Durchsetzungsmotivation“ oder „Leistungsmotivation“.

Auch im Rahmen der „Life-Event-Forschung“ wurde versucht, Zusammenhänge zwischen vorhandenen durch Stress bedingten Krankheitssymptomen und biographischen Daten zu ermitteln. Die „Social Readjustment Rating Scale“ von Holmes und Rahe (1967) gilt in diesem Zusammenhang als bekanntestes Beispiel. In dieser Skala sind kritische Lebensereignisse wie bspw. Tod des Ehepartners, Scheidung, Heirat, gesellschaftliche Veränderungen usw. entsprechenden Belastungsgewichten (life change units = LCU) zugeordnet. Anhand dieser Gewichte können individuelle Belastungen errechnet und das Risiko von physiologischen Stressauswirkungen kalkuliert werden.

Die Autoren stellten fest, dass kranke Menschen in dem Jahr vor ihrer Erkrankung viel höhere LCU-Werte aufwiesen als gesunde Menschen. Ein besonders vielsagender Grenzwert war 300 LCU's. Das heißt: wenn sich die Belastungen eines Menschen innerhalb eines Jahres zu mehr als 300 Lebensveränderungseinheiten summierte (Tabelle 10.2-2), entwickelte diese Person mit großer Wahrscheinlichkeit ein schweres Gesundheitsproblem, in vielen Fällen eine Virus- oder Bakterieninfektion.

Tab. 10.2-2: Auswirkungen der LCU-Werte auf die Erkrankungswahrscheinlichkeit nach Holmes und Rahe (1967)

<p>Ein Wert von 150 oder darunter steht für ein relativ niedriges Ausmaß an Veränderung im Leben und für eine geringe Gefahr von durch Stress herbeigeführten gesundheitlichen Störungen.</p> <p>Ein Wert zwischen 150 und 300 bedeutet, dass in den nächsten beiden Jahren mit 50%iger Wahrscheinlichkeit eine gesundheitliche Störung eintreten wird.</p> <p>Ein Wert von über 300 Punkten lässt nach dem statistischen Vorhersagemodell von Holmes und Rahe die Wahrscheinlichkeit auf über 80 % ansteigen.</p>

Jedoch blieben Kritikpunkte an den Untersuchungen und damit einhergehenden Schlussfolgerungen nicht aus. So bestand die Eichstichprobe nur aus weißen Amerikanern,

erscheint die additive Aufrechnung der einzelnen Events fragwürdig und der Fragebogen enthält eine Reihe von Items, die gleichsam direkte oder soziale Folgen einer auftretenden Erkrankung beinhalten. Zudem ist aus heutiger Sicht die Grundannahme der life event Forschung, kritische Lebensereignisse wirkten sich bei allen Individuen nahezu gleich aus, fragwürdig. Es wurde bereits früh vermutet, dass sich Krankheitsrisiken nicht allein aus der Art und Weise bestimmter Lebensveränderungen vorhersagen lassen, sondern dass sie sehr wahrscheinlich ebenso stark durch körperliche und vor allen Dingen psychologische Faktoren der betroffenen Personen beeinflusst werden (Comer, 1995).

Eine Übersicht biographischer Fragebogen für den klinischen Kontext findet sich u.a. bei Keßler (1982). Jäger weist jedoch schon 1976 darauf hin, dass mit diesen Verfahren Selektionen im nicht-klinischen Kontext auf keinen Fall vorgesehen sind.

Anders sieht es für biographische Fragebogen im eignungsdiagnostischen Kontext aus, die explizit für die Selektion im nicht-klinischen Kontext entwickelt wurden. Stehle (1983) gibt in seiner Arbeit eine Übersicht über Historie und Vielfalt der Anwendungsfelder biographischer Fragebogen. Die auch heute noch gängige Praxis hat einen langen geschichtlichen Hintergrund. Owens (1976) beschrieb die erste Verwendung biographischer Daten in der Eignungsdiagnostik für das Jahr 1894: Zur Auswahl von Versicherungsverkäufern wurde vorgeschlagen, biographische Daten zu erheben. Die Bewerber wurden angehalten, einen Fragebogen mit einer Liste vorgegebener Fragen u.a. nach derzeitigen und vergangenen Wohnorten, Familienstand, Empfehlungen und Erfahrungen im Verkauf von Versicherungen zu beantworten. Dieser Fragebogen gilt heute als der erste biographische Fragebogen, der im eignungsdiagnostischen Kontext Anwendung fand. Nach dieser anfänglich sehr eingeschränkten Verwendung biographischer Fragebogen, fanden eine Vielzahl neuer biographischer Ausleseverfahren in den vergangenen Jahrzehnten, insbesondere im angelsächsischen Raum, Verbreitung (vgl. u.a. Stehle, 1983, 1986). In den 20er Jahren fanden Fragebogen häufig Anwendung, um bei Verkäufern den Verkaufserfolgsindex zu prognostizieren. In diesem Kontext wurden insbesondere Fragebogen zur Auswahl von Warenhaus- und Versicherungsverkäufern entwickelt. So konnte bspw. Goldsmith (1922) zeigen, dass eine Gewichtung von Items in einem Personalfragebogen eine bessere Erfolgsprognose erlaubte, als die intuitive Beurteilung. Er gab drei Gruppen von jeweils 50 guten, mittleren und schlechten Verkäufern den Fragebogen zur Beantwortung vor und fand entsprechende Unterscheidungen zwischen den Gruppen.

Als Anwendungsfeld dominierte die Auswahl im Versicherungsbereich bis in die Mitte des 20. Jahrhunderts. Zu dieser Zeit erfolgte eine fast universelle Ausweitung auf weitere Bereiche, zunächst auf das Militär und die Industrie, einschließlich der Auswahl von Führungspotenzial. Im industriellen Bereich wurden mittels biographischer Fragebogen gewerbliche und kaufmännische Arbeitnehmer ausgewählt, sie fanden aber auch zunehmend für die Auswahl von Führungskräften, erfolgreiche Wissenschaftlern und im militärischen Bereich Anwendung. Geprägt waren diese ersten biographisch orientierten eignungsdiagnostischen Instrumente durch ein weitgehend empiristisches Vorgehen in der Konstruktion. Dabei wurden einzelne Items anhand ihrer empirisch ermittelten Validität zur Prognose eines Außenkriteriums ausgewählt und gewichtet. Aber auch das Grundaxiom biographischer Fragebogen, dass nämlich der beste Prädiktor zukünftigen Verhaltens, das Verhalten in der Vergangenheit sei, wurde teilweise unterlaufen. Auch Selbsteinschätzungen, die eher für Persönlichkeitsfragebogen typisch sind, fanden in biographischen Fragebogen Aufnahme. Entscheidend war allein, dass empirische Zusammenhänge mit dem Kriterium erkennbar waren. Aus heutiger Sicht dürften insbesondere die 60er und 70er Jahre als eine Sternstunde in der jüngeren Forschungsgeschichte gelten.

In dieser Zeit überprüften Baehr und Williams (1968) mit ihrem im Kapitel 2.2.1.2 beschriebenen 1967 selbst entwickelten „Personal History Index“ den Verkaufserfolg von Verkäufern im Lebensmittelbereich. In der Kreuzvalidierung konnten sie eine Korrelation zwischen vorhergesagter und tatsächlicher Verkaufleistung von $r = .50$ beschreiben.

Rawls (1968) untersuchte mit einem biographischen Fragebogen die Eignung von Hochschulabsolventen für Managementpositionen. Von 150 Führungskräften eines Unternehmens wurden, nach Kriterien wie Gehaltstufe und hierarchische Einordnung im Unternehmen, die 30 erfolgreichsten und die 30 am wenigsten erfolgreichen Führungskräfte ausgewählt. Neben zwei Persönlichkeitstests bekamen die 60 Führungskräfte einen biographischen Fragebogen mit 179 Fragen vorgelegt. Anschließend wurde eine Gruppe von Studenten, entsprechend ihrer Ähnlichkeiten im Persönlichkeitsprofil, ausgewählt. Sie füllten den gleichen Fragebogen aus. Hierbei stellten sich 87 der Items als trennscharf in beiden Gruppe heraus, so dass Rawls zu dem Schluss gelangte, dass der verwandte biographische Fragebogen in der Lage sei, potenziell erfolgreiche Führungskräfte unter den Studenten zu identifizieren.

Erfolgreiche obere Führungskräfte mit einer vergleichbaren Stichprobe unterer Führungskräfte wurden von Dill, Hilton und Reimann (1960) miteinander verglichen. Beide Stichproben stimmten hinsichtlich der Variablen Ausbildung und Alter sowie Unternehmenszugehörigkeit überein. Dabei konnte durch die Auswertung biographischer Fragebogen festgestellt werden, dass erfolgreiche Führungskräfte ein höheres Maß an Leistungsmotivation besitzen, bereits beim Berufseintritt konkrete Karriereziele verfolgen, von der Verstärkung durch den jeweiligen Vorgesetzten unabhängig sind, sich aktiv an der Definition der eigenen Aufgaben beteiligen, stark nach Macht, Autonomie und Status streben, sich stärker mit dem Unternehmen identifizieren und emotional von Untergebenen wesentlich unabhängiger sind. Wie Stehle (1995) erwähnt, kommen Rawls und Rawls (1974) in einem Übersichtsartikel über Entwicklungen im Bereich der Führungskräfteauswahl zu dem Schluss, dass biographische Fragebogen generell sehr gute Prädiktoren zur Vorhersage des Managementenerfolges darstellen.

Stehle (1983; s.a. Richter & Stehle, 1986) untersuchte die Möglichkeit, mittels eines biographische Fragebogens künftige Außendienstmitarbeiter aus Bewerbern mit größtenteils berufsfremden Vorerfahrungen auszuwählen. Für ein großes deutsches Versicherungsunternehmen übersetzte er den von Baehr und Williams (1967) entwickelten Fragebogen. Er erfasste die konkurrente Validität des Fragebogens durch einen Extremgruppenvergleich, d.h. die Korrelation mit einem zukünftig beobachtbaren Kriterium wurde ersetzt durch die Korrelation mit einem zeitlich koexistenten Kriterium. Als erfolgreich wurden die Außendienstmitarbeiter klassifiziert, die seit zwei Jahren bezüglich ihrer Verkaufsleistung die höchsten Rangplätze erzielten. Entsprechend als weniger erfolgreich jene, welche die letzten Rangplätze einnahmen. Zur Auswertung des biographischen Fragebogens entschied sich Stehle für die Methode des „Weighted application blank approach“ nach England (1971) (s.a. Pkt.2.3.2). Nach Kreuzvalidierung konnte die erschlossene Validität von $r = .45$ bestätigt werden.

Aus dem Zeitraum von 1950 bis in die 80er Jahre liegen eine Vielzahl unterschiedlicher Untersuchungen überwiegend aus dem angloamerikanischen Raum vor, die auf eine ausreichend hohe Kriteriumsvalidität schließen lassen. Allerdings variiert die Höhe der prädiktiven Validität biographischer Daten in unterschiedlichen Untersuchungen. In der Metaanalyse von Schmidt und Hunter wurde der durchschnittliche Validitätskoeffizient, bezogen auf alle in der Analyse berücksichtigten Kriterien wiedergegeben. Verschiedene

Autoren verweisen auf die Abhängigkeit der Höhe der Validität im biographischen Fragebogen vom gewählten Kriterium. So kommt Ghiselli (1955) bei seiner systematischen Zusammenstellung von verschiedenen eignungsdiagnostischen Verfahren zu einer mittleren Validität biographischer Verfahren von $r = .44$, wenn das Kriterium der Aus- und Weiterbildungserfolg war und $r = .41$ bei dem Kriterium berufliche Bewährung. Die Metaanalyse von Schmitt, Gooding, Noe und Kirsch (1984) ergab eine Validität von $r = .52$, wenn als Kriterium die Gehaltshöhe gewählt wurde, dagegen nur $r = .23$ wenn die Ausbildungsleistung herangezogen wurde. 1993 kommt Bliesener (zit. n. Schuler, 1996) zu einer ähnlichen Einschätzung. Er bezog 162 Ergebnisse von Einzelstudien mit insgesamt 90000 Probanden in seine Analyse ein und errechnete eine Validität von $r = .39$. Schon 1992 hatte Bliesener 75 Primärstudien in einer Metaanalyse berücksichtigt. Dabei schwankten die ungewichteten beobachteten Validitätskoeffizienten zwischen den Extremwerten $r = -.23$ und $r = .84$. Wenn die einzelnen Validitätskoeffizienten um die Stichprobengröße korrigiert wurden, lag der mittlere Koeffizient bei $r = .33$. Insgesamt kommt der Autor zu dem Ergebnis, dass biographische Verfahren auch unter Berücksichtigung methodischer Artefakte eine gewisse Validität für die Prognose berufserfolgsrelevanter Kriterien aufweisen. Savage verglich 1986 13 englische Studien, bei denen in den Jahren 1982 – 1985 bei fast 1000 Personen Erfahrungen mit biographischen Daten zur Bestimmung der Übereinstimmungsvalidität bei vorhandenen Mitarbeitern gewonnen werden konnten. Die von ihm erfassten Validitätskoeffizienten lagen für Mitarbeiter aus den Berufsbereichen Management, Bankwesen, Öffentliche Verwaltung, Verkauf und der Industrie in einem Bereich von $r = .39 - .87$. Dabei lagen 10 der 13 Studien in einem Bereich von $r = .50 - .79$. Bei der Bewertung der veröffentlichten Ergebnisse sollte jedoch nicht unberücksichtigt bleiben, dass auch hier das sog. Schubladenproblem eine Rolle spielen dürfte, d.h. in der Regel werden eher signifikante Ergebnisse veröffentlicht, dagegen landen die nicht-signifikanten in den Schubladen der Untersucher.

Wie aus den vorliegenden Studien zu entnehmen, kann festgehalten werden, dass die prädiktive Validität biographischer Daten durch eine ausreichende Zahl von Untersuchungen und unterschiedlichen Metaanalysen belegt werden konnte. Leider bleibt jedoch die Forderung nach einer theoriegeleiteten Itemauswahl häufig unerfüllt. Vielfach werden Items biographischer Fragebogen nach rein empirischen Gesichtspunkten zusammengefasst. Dies ist zudem zweckmäßiger, da theoretisch abgeleitete Items nach Bliesener (1992) eher auch eine Kreuzvalidierung überstehen und damit nicht ausschließlich durch Eigenschaften der

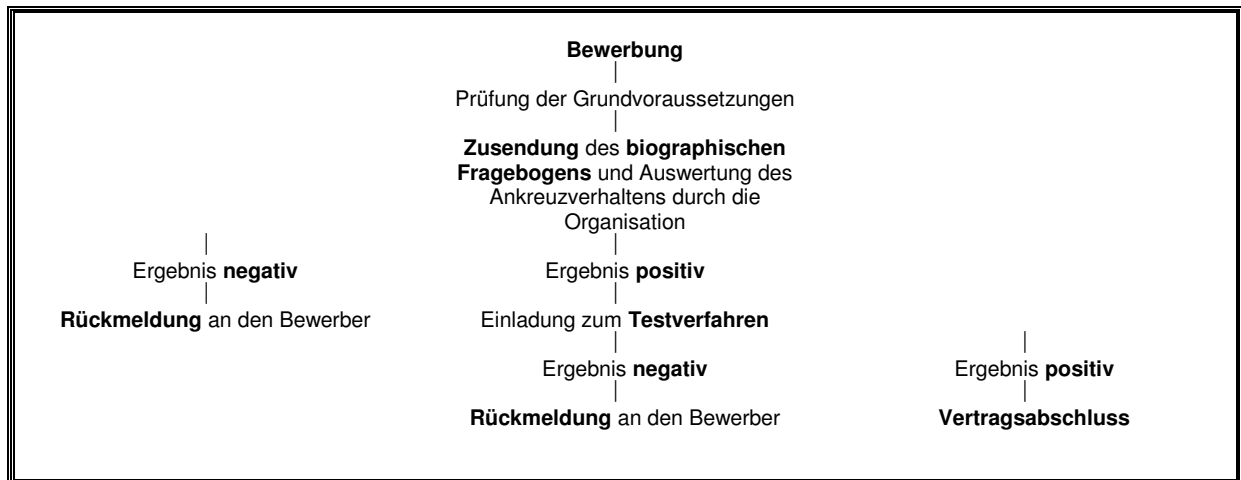
jeweiligen Validierungsstichprobe charakterisiert sind. Dennoch liegt der Schluss nahe, dass aufgrund steigender Nachfrage nach qualifiziert ausgebildeten Arbeitnehmern sowie gut ausgebildeten Schülern und damit verbunden, ständig steigenden Kosten für Weiter-, Fort- und Ausbildung, schon die Personalvorauswahl ein entscheidender Faktor bei Kosten-Nutzen Überlegungen ist. Die Personalendauswahl kann nur so gut sein, wie es die Personalvorauswahl zulässt. Hier könnten erfasste biographische Datenquellen eine Alternative zur bisher praktizierten Vorauswahl darstellen.

In Kapitel 2.1.2 wurde bereits beschrieben, dass die Personalendauswahl in bundesdeutschen Unternehmen und bei öffentlichen Arbeitgebern zwar auf relativ hohem Niveau durchgeführt wird, der Aufwand für die Vorauswahl dagegen in der Regel relativ gering ist (Seibt und Kleinmann (1991)). Die Autoren befragten 200 Unternehmen hinsichtlich ihrer Praktiken bei der Personalvorauswahl. Auffallend war die absolute Dominanz der Bewerbungsunterlagen. Jedoch wurden nur durchschnittlich zehn Minuten für die Durchsicht dieser Datengeber veranschlagt. Die innerhalb dieser kurzen Zeitspanne zu treffenden Entscheidung über Annahme oder Ablehnung einer Bewerbung birgt fast zwangsläufig ein hohes Fehlerpotenzial. Die Gefahr wertvolle Arbeitskräfte für ein Unternehmen durch vorschnelle Urteile zu verlieren ist demnach erheblich. Demnach fehlen nach wie vor geeignete Instrumente, welche die Vorauswahl ökonomisch und verlässlich gestalten und dabei möglichst auf die Akzeptanz der Bewerber stoßen und somit ein Verfahren, das ökonomisch und kostengünstig einsetzbar ist und eine möglichst hohe Trefferquote verspricht.

Neben der Verwendung biographischer Fragebogen als zusätzliches Instrument im traditionellen Vorstellungsgespräch, sind sie als traditionelles Instrument einer effektiven Vorauswahl einsetzbar (Stehle, 1995). In Anlehnung an die Darstellung des Autors, ist in Tabelle 11.2-2 die klassische Vorgehensweise zur Vorauswahl bei großen Bewerberzahlen beschrieben.

Aufgrund der belegten hohen prädiktiven Validität biographischer Daten und ihrer besonderen Eignung als Auswahlinstrument externer Bewerber (Schuler, 1996), ihrem dagegen geringen Beitrag zu inkrementellen Validität (Schmidt & Hunter 1998), bieten sich biographische Fragebögen besonders zur Vorauswahl an. Bei steigender Nachfrage nach qualifizierten Bewerbern und den damit einhergehenden Kosten für Auswahlverfahren ist angesichts betriebswirtschaftlicher Aspekte eine effektive Vorauswahl um so wichtiger.

Tab. 11.2-2: Klassische Vorgehensweise bei der Vorauswahl auf der Grundlage eines biographischen Fragebogens nach Stehle (1995)



Durch die zu vermutende nicht mehr ausreichende Filterfunktion von Schulnoten müssen eine große Anzahl Bewerber zur Feststellung der Eignung zum Auswahlverfahren eingeladen werden. Dabei entstehen nicht nur hohe Kosten für die Organisation, die u.a. Personal für die Durchführung des Verfahrens sowie ggf. Übernachtungsplätze für auswärtige Teilnehmer bereitstellen muss. Auch für die Bewerber bedeutet die Teilnahme finanzieller und organisatorischer Aufwand (Reisekosten, Freistellung von der Schule oder dem bisherigen Arbeitgeber). Zudem werden durch die Einladung zum Auswahlverfahren häufig Hoffnungen auf den Traumberuf geweckt, die dann aufgrund des Ausschlusses wegen nicht ausreichender Leistungen schon im Auswahlverfahren zerschlagen werden.

Im Gegensatz zu einer Vorauswahl auf Basis der Bewerbungsunterlagen gründet sich die biographische Diagnostik nicht auf Induktions- oder Analogieschlüsse, sondern auf den empirisch ermittelten Zusammenhang zwischen Indikatoren und Kriterien (Sarges, 1995). Trotz der berechtigten Kritik aufgrund des immer noch fehlenden theoretischen Hintergrundes, besitzen biographische Messinstrumente wie dargestellt eine hohe prädiktive Validität. Da die jedoch im Sinne der inkrementellen Validität nur zu einen durchschnittlichen Validitätszuwachs von $r = .01$ beitragen, wäre ihr Einsatz als Auswahlinstrument im Zusammenhang mit den fast obligatorisch angewandten – sowohl prädiktiv als auch konstruktvaliden - Intelligenztestverfahren wenig nützlich. Schmidt und Hunter (1998) begründen diese Erkenntnis dadurch, dass biographische Daten in beträchtlichen Umfang (ca. $r = .50$) mit allgemeinen kognitiven Fähigkeiten korrelieren. Ihrer Ansicht nach lässt dieses vermuten, dass biographische Daten ein indirektes Maß für mentale Fähigkeiten darstellen, zusätzlich zu dem, was sie zu anderen Merkmalen noch messen.

Ein neben der prädiktiven Validität weiterer wichtiger Aspekt ist die Akzeptanz eines Auswahlinstrumentes. Auch wenn dieser sogenannten Sozialen Validität in den vergangenen Jahren wenig Beachtung geschenkt wurde, gewinnt sie gerade auch angesichts der zunehmenden Konkurrenz um qualifizierte Schulabgänger und Teilnehmer am Auswahlverfahren zunehmend an Bedeutung und sollte in konzeptionelle Vorüberlegungen eingebunden werden.

2.3 Soziale Validität biographischer Daten

2.3.1 Zum Begriff der Validität

Die Validität oder Gültigkeit eines Tests gibt den Grad der Genauigkeit an, mit der ein Test das Persönlichkeitsmerkmal oder die Verhaltensweise, das bzw. die er messen oder vorhersagen soll, tatsächlich misst oder vorhersagt. Vollkommen valide wäre ein Test, wenn die Ergebnisse einen unmittelbaren und fehlerfreien Rückschluss auf den Ausprägungsgrad des zu erfassenden Persönlichkeits- oder Verhaltensmerkmals zulassen, wenn also der individuelle Testpunktwert eines Probanden diesen auf der Merkmalskala eindeutig lokalisiert (Lienert & Raatz, 1994, S. 10).

Die Frage, welchen Anforderungen ein Test genügen muss, um aufgrund eines Testergebnisses auf die tatsächliche Ausprägung des getesteten Merkmals schließen zu können, ist Gegenstand der Testtheorie (Bortz, 1984, S. 134). Grundlegend für die klassische Testtheorie sind folgende vier Axiome:

1. Das Testergebnis setzt sich additiv aus dem „wahren Wert“ (true score) und dem Messfehler zusammen.
2. Bei wiederholten Testanwendungen kommt es zu einem Fehlerausgleich, d.h. der Erwartungswert des Messfehlers ist Null.
3. Die Höhe des Messfehlers ist unabhängig vom Ausprägungsgrad des getesteten Merkmals.
4. Die Messfehler verschiedener Testanwendungen (bei verschiedenen Personen oder Testwiederholungen bei einer Person) sind voneinander unabhängig.

Aus diesen vier Axiomen ergeben sich Strategien zur Ermittlung der Gütekriterien eines Tests.

Neben den Hauptgütekriterien Objektivität und Reliabilität stellt die Validität das dritte, jedoch wichtigste Gütekriterium dar, welches ein guter Test erfüllen sollte. Unter Reliabilität wird nach Lienert (1989) der Grad an Genauigkeit verstanden, mit dem ein Test ein

bestimmtes Persönlichkeits- oder Verhaltensmerkmal misst, unabhängig davon, ob der Test dieses Merkmal auch zu messen beansprucht. Die Objektivität eines Tests ist dann gegeben, wenn verschiedene Untersucher bei derselben Versuchsperson zu gleichen Ergebnissen gelangen. Objektivität und Reliabilität haben primär die Rolle, günstige Voraussetzungen für das Erreichen einer hohen Validität zu schaffen (Amelang & Zielinski, 1994, S. 113).

Bei gegebener hoher Validität erlauben die Ergebnisse eines Tests oder Fragebogens die Generalisierung aus dem numerischen Relativ der Testskalen auf das empirische Relativ des interessierenden Zielmerkmals oder einen Schluss aus dem beobachteten Verhalten in der Testsituation auf Verhalten außerhalb der Testsituation (Michel & Conrad, 1982). In der Literatur werden jedoch unterschiedliche Zugänge zur Erschließung der Validität unterschieden. Dabei ist laut der American Psychological Association (APA 1974, zit. nach Jäger, 1986) insbesondere zwischen drei Hauptarten, nämlich der Kriteriums-, der Inhaltlichen- sowie der Konstruktvalidität, zu differenzieren. Weitere Validitätsbegriffe stellen Spezifikationen dieser drei Validitätsarten dar, die zueinander in Beziehung stehen.

Die Validität eines Tests wird jedoch nicht direkt gemessen, sondern aus Messungen erschlossen (Jäger, 1986). An dieser Stelle soll nicht vertiefend auf die verschiedenen methodischen Zugänge der Validitätsbestimmung eingegangen werden, eine ausführliche Darstellung findet sich bspw. bei Lienert und Raatz (1994).

Für das generelle Verständnis erscheint es aber sinnvoll, beispielhaft in diesem Zusammenhang die drei o.g. Hauptarten zu beschreiben.

Bei der Bestimmung der Inhaltlichen Validität ist der Test bzw. sind seine Elemente so beschaffen, dass sie das zu erfassende Persönlichkeitsmerkmal oder die in Frage stehende Verhaltensweise repräsentieren. Der Test selbst stellt das optimale Kriterium für das Persönlichkeitsmerkmal oder die Verhaltensweise dar. Inhaltliche Validität wird einem Test in der Regel durch ein Rating von Experten zugebilligt (Lienert & Raatz, S. 10 f.). Bei Jäger (1986) ist die so genannte Kontentvalidität definiert als Repräsentativität der Itemstichprobe eines Tests für eine klar definierte Grundgesamtheit (Population) von Verhaltensweisen. Ein Beispiel wäre die Beherrschung der Grundrechenarten.

Dagegen wird aufgrund theoretischer Erwägungen und anhand sich daran anschließender empirischer Untersuchungen entschieden, ob ein Test ein bestimmtes Konstrukt erfasst. Die so erschlossene Konstruktvalidität zielt direkt auf die psychologische Analyse der einem Test zugrundeliegenden Eigenschaften und Fähigkeiten ab. Damit auf Beschreibungsmerkmale,

die nicht in eindeutiger Weise operational erfassbar sind, sondern theoretischen Charakter haben, wobei jedoch eine empirische Basis gegeben ist.

Ein Test etwa, der den individuellen Ausprägungsgrad von Angst messen soll, hätte dann eine hinreichende Konstruktvalidität, wenn nachgewiesen wurde, dass die vom Test erfassten Merkmale in genügender Übereinstimmung mit dem theoretischen Konstrukt „Angst“ stehen. (Lienert & Raatz, 1994, S. 11). Zum Thema Konstruktvalidität siehe auch Kapitel 2.2.1.2.

Während sich im Fall der Inhaltlichen Validität und der Konstruktvalidität im allgemeinen keine Maßzahl für den Grad der Validität eines Testes ermitteln lässt, ist dies bei der kriterienbezogenen Validität durchaus möglich. Denn um von erreichten Ergebnissen in einem Test oder Fragebogen (Prädiktor) auf ein bestimmtes Zielmerkmal (Kriterium) schließen zu können, bedarf es des Korrelationsschlusses zur statistischen Absicherung. Hierzu werden einer Stichprobe von Probanden der Test bzw. Fragebogen zur Bearbeitung vorgelegt. Anschließend wird überprüft, inwieweit die Enge des linearen Zusammenhanges zwischen Prädiktor und Kriterium gegeben ist, d.h. die Ergebnisse im Testverfahren mit dem Zielmerkmal übereinstimmen.

Die so erschlossene Höhe des Validitätskoeffizienten hängt, wie schon in Kapitel 2.1 dargestellt, im wesentlichen von drei Faktoren ab:

- a) vom Grad dessen, was an Gemeinsamkeit durch den Test und das Kriterium erfasst und oft als „Zulänglichkeit“ des Tests bezeichnet wird,
- b) von der Reliabilität des Tests und
- c) von der Reliabilität des Kriteriums.

Die kriterienbezogene Validität eines Tests ist umso größer, je größer die Gemeinsamkeit des von Test und Kriteriums erfassten Merkmalanteils ist. Andererseits ist diese umso geringer, je geringer die Reliabilitätskoeffizienten ausfallen (Lienert & Raatz, 1994).

Als Unterformen der kriteriumsbezogenen Validität lassen sich die konkurrente sowie die prädiktive Validität voneinander unterscheiden.

Bei der konkurrenten Validität werden Test- und Kriteriumswerte nahezu gleichzeitig erhoben, bei der prädiktiven Validität dagegen zu einem Zeitpunkt t_1 die Testergebnisse und zu einem späteren Zeitpunkt t_2 die Punktwerte des Kriteriums. Allerdings verweisen Amelang und Zielinski (1994) nicht zu unrecht auf die nicht immer mögliche trennscharfe Unterscheidung zwischen diesen beiden Arten zur Bestimmung der Validität. Ebenso wie man den schulischen Erfolg auf Intelligenz zurückführen kann, ist es möglich, Intelligenz mit der schulischen Leistung zu erklären.

Grundsätzlich kommt gerade auch im eignungsdiagnostischen Kontext der prädiktiven Validität der höhere Stellenwert zu, da nur zeitlich Früheres verantwortlich sein kann für

zeitlich Späteres. Insofern spricht gerade in der Berufseignungsdiagnostik vieles für die Nützlichkeit eines Auswahlinstrumentes, wenn dessen ermitteltes Testergebnis den Ausbildungserfolg zu einem späteren Zeitpunkt vorherzusagen vermag.

Zusammenfassend kann Jäger (1986, S 284) erklärend zitiert werden: „Konstruktvalidierung dient der Theorieentwicklung, die ihrerseits das Ziel einer wissenschaftlichen Erklärung und Prognose von Verhalten verfolgt und an diesem als Kriteriumsverhalten zu validieren ist. Kontentvalidierung kann Bausteine zur Konstruktentwicklung beitragen. Kriteriumsvalidierung ist essentieller Bestandteil der Konstruktvalidierung. Das Ensemble der Kriteriumsvaliditäten liefert Informationen, die zur Verankerung von Konstrukten in der Realität des Verhaltens erforderlich sind und damit ihrer empirischen Spezifikation dienen“. Insbesondere die o.g. Abhängigkeiten des Korrelationskoeffizienten machen deutlich, dass die mögliche Höhe der Validität eines zur Eignungsdiagnostik herangezogenen Auswahlinstrumentes nur so hoch sein kann, wie es die gegebene Reliabilität von Prädiktor und Kriterium zulassen.

Bei der Berücksichtigung der Genauigkeit eines Testverfahrens (Reliabilität) erscheint es sinnvoll, auch den Begriff der damit einhergehenden Bandbreite des eignungsdiagnostischen Instrumentes zu reflektieren. Das Konzept der Bandbreite eines Prädiktors wurde zuerst von Cronbach (1949; zit. nach Hossiep, 1995) auf psychologische Testverfahren angewandt. Die damit gemeinte Flexibilität eines Auswahlinstrumentes lässt sich am Beispiel des Interviews veranschaulichen. Das Einstellungsinterview ist ein Verfahren mit einer hohen Bandbreite, d.h. es können vielfältige Bereiche an Informationen damit gewonnen werden. Dagegen ist die Messpräzision des Interviews eher gering. Auf der anderen Seite wäre ein speziell für die Auswahl eines konkreten hochtechnisierten Berufes konstruiertes apparatives Auswahlinstrument hinsichtlich seiner Messpräzision vermutlich jedem anderen Verfahren überlegen, hätte jedoch nur eine geringe Bandbreite. Bei der Prognosestellung geht ein großer und inhaltlich vielfältiger Bereich an Informationen (große Bandbreite) in der Regel mit einer Einschränkung oder dem Verlust an Fidelität (geringe Messgenauigkeit) einher. Da jedoch unter der Reliabilität der Grad der Genauigkeit verstanden wird, mit dem ein Test ein bestimmtes Persönlichkeits- oder Verhaltensmerkmal misst, werden eignungsdiagnostische Auswahlinstrumente, die als Prädiktoren herangezogen werden, immer dann eine niedrigere Reliabilität erwarten lassen, wenn sie möglichst breite Aussagemöglichkeiten über die Eignung eines Bewerbers anstreben. Zur generellen Problematik bei der Auswahl und

Bewertung von Prädiktoren und Kriterien sei der interessierte Leser bspw. auf Hossiep (1995) verwiesen. Allgemein kann jedoch angenommen werden, dass der wahre Zusammenhang zwischen Prädiktor und Kriterium immer unterschätzt wird, da bei beiden mit Messfehlern gerechnet werden muss, so dass die Prognostizität von wissenschaftlich fundierten Auswahlverfahren nach Kleinevoss und Sonnenberg (1987) generell eher unter- als überschätzt wird.

Zentraler Punkt für die Bewertung der Verwendung findenden Instrumente ist, wie oben dargestellt, die mit Hilfe des Korrelationskoeffizienten abgebildete Enge des Zusammenhanges zwischen Prädiktor und Kriterium. Die Validierung eignungsdiagnostischer Untersuchungsverfahren besteht in der Analyse der hypothetisch abgeleiteten und durch die Prädiktoren operationalisierten Personenmerkmale. Durch die kriterienbezogene Validierung wird mittels eines empirischen Verfahrens aufgezeigt, in welchem Ausmaß Rückschlüsse auf andere Verhaltensweisen möglich sind. In der eignungsdiagnostischen Praxis geht es bei der Maximierung der prädiktiven Validität in der Regel darum, aus einer Vielzahl möglicher diagnostischer Instrumente diejenigen auszuwählen, die mit einem möglichst kleinen Aufwand und geringen Kosten eine größtmögliche Enge des Zusammenhanges zwischen gewählten Prädiktoren und Kriterien aufzeigen.

Dabei wird der Wert der prognostischen Validität eines Selektionsinstrumentes daran gemessen, wie hoch der prozentuale Anteil der Varianz des Kriteriums ist, der durch die Varianz des Prädiktors aufgeklärt werden kann. Dieser im Determinationskoeffizienten seinen Ausdruck findende Prozentwert wird bestimmt durch das Quadrat von Prädiktor und Kriterium. Korreliert bspw. ein zur Auswahl eingesetztes Instrument zu $r = .50$ mit dem gewählten Kriterium, so ermöglicht dieses Verfahren die Aufklärung von 25 % der Kriteriumsvarianz. Bei der Interpretation und beabsichtigten Erhöhung von Validitätskoeffizienten ist unter Nützlichkeitsaspekten grundsätzlich zu berücksichtigen, welchen Effekt eine mögliche Optimierung mit sich bringt. Zur Nützlichkeit von Auswahlverfahren siehe auch Kapitel 2.2.1.

Gemäß der von Lienert (1989) eingeführten allgemeinen Richtlinien über die erforderliche Höhe von Validitätskoeffizienten, muss ein Verfahren in einem solchen Umfang valide sein, dass seine Anwendung eine bessere Vorhersage ermöglicht als seine Unterlassung. Prinzipiell könne man daher mit jedem Validitätskoeffizienten zufrieden sein, der sich statistisch auf dem 1% Niveau absichern ließe. Allerdings sind niedrige Validitätskoeffizienten unter $r = .30$

für die praktische Anwendung als Einzeltest relativ bedeutungslos, jedoch im Rahmen einer Testbatterie unter Umständen gut einsetzbar. Nach Jäger (1986) macht sich die praktische Verwendbarkeit eines Auswahlinstrumentes jedoch nicht ausschließlich an der Höhe des Korrelationskoeffizienten fest. Für ihn gelten als Bewertungsmaßstab zum einen die Signifikanz des Abstandes seiner Validität von der eines Würfelbechers ($r = .00$) und die Abstände zu den alternativ zur Verfügung stehenden Instrumenten. So ließe sich Jägers Betrachtung mit der Feststellung zusammenfassen, dass jedes Testverfahren, welches eine statistisch abgesicherte Validität nachweisen kann (und somit bessere Vorhersagen erlaubt als ein Würfelbecher) immer dann über eine ausreichend hohe Validität verfügt, wenn kein alternatives Testverfahren zur Erfassung der beabsichtigten Merkmalsausprägung zur Verfügung steht.

Aus den in Kapitel 2.2 dargestellten Untersuchungen aus dem Zeitraum von 1950 bis in die 80er Jahre ist gemäß dieser Definition eine ausreichend hohe Kriteriumsvalidität biographischer Daten für die Berufseignungsdiagnostik anzunehmen. Die hier zitierten unterschiedlichen Metaanalysen belegen durchschnittliche Validitätskoeffizienten zwischen $r = .33 - .87$ (Bliesener, 1993, 1996; Savage, 1986; Schmitt, Gooding, Noe & Kirsch, 1984; Schmidt & Hunter, 1998).

2.3.2 Zum Begriff der Sozialen Validität

Die Nützlichkeit eines Auswahlverfahrens, wie sie in den vorherigen Kapiteln für biographische Daten belegt werden konnte, reicht heute allein nicht aus. Darüber hinaus ist fast genau so wichtig, dass Probanden, die in einem Verfahren ausgewählt bzw. nicht ausgewählt werden, das Instrument, mit dem diese Selektion stattfand, zumindest als dafür geeignet akzeptieren. Die Auswahl-situation stellt häufig den Erstkontakt zwischen Bewerber und Organisation dar. Aufgrund fehlender anderer Erfahrungen können Bewerber in diesem Stadium den Eindruck gewinnen, dass die Auswahl-situation und das von der Organisation gezeigte Verhalten repräsentativ für die Organisation seien. D.h., die Auswahl-situation vermittelt ein Bild der Organisation, welches, insbesondere durch abgelehnte Bewerber, die keine Gelegenheit finden, diesen selektiven Eindruck zu kompensieren, an andere potentielle Bewerber weitergetragen wird. Die Frage, ob Personen, die das Auswahlverfahren bereits absolviert haben, anderen eine Bewerbung bei der Organisation empfehlen, hängt damit durchaus auch davon ab, welches im Auswahlverfahren wahrgenommene Image der Organisation zugeschrieben und weitergetragen wird. Daneben sollte die Akzeptanz aber auch

originäres Interesse des Eignungsdiagnostikers selbst sein. Verfahren mit einer hohen Akzeptanz dürften die Kooperationsbereitschaft und die Motivation der Teilnehmer und somit die Validität der Eignungsdiagnose positiv beeinflussen (Nevo, 1986).

In diesem Zusammenhang muss auch wieder die bisher fehlende Konstruktvalidierung biographischer Fragebogen gesehen werden (Kapitel 2.2.1), die nach Schuler und Stehle (1983) nicht dazu beiträgt, die Akzeptanz unter Bewerbern zu erhöhen. Letztendlich wird mit der Frage nach der Akzeptanz ein Dilemma der Eignungsdiagnostik deutlich: Verfahren, die hinsichtlich ihrer prognostischen Validität eine hohe Vorhersagegüte bieten, finden durch die Teilnehmer häufig nur eine geringe Wertschätzung. So werden Testverfahren in der Regel schlechter beurteilt als zur Auswahl herangezogene Schulnoten und das, obwohl psychologische Leistungstests bessere Prognosen hinsichtlich bspw. des Ausbildungserfolges erlauben (Hossiep et al, 2000). Dagegen erfreuen sich nicht standardisierte Interviews einer hohen Beliebtheit, auch wenn sie insgesamt keine hohe prognostische Validität besitzen. Kersting (1998) verweist aber zu Recht darauf, dass das Konzept der sozialen Akzeptanz nicht mit zu hohen Erwartungen überfrachtet werden darf. „Eine hundertprozentige Akzeptanz von Auswahlverfahren wird es solange nicht geben, solange Leistungsmessungen anstrengend und stressig sind, solange mit Hilfe dieser Verfahren Personen zurückgewiesen werden, solange Bewerber mit zum Teil unrealistischen Erwartungen in die Auswahlverfahren gehen und solange sich gegensätzliche Interessen von Bewerbern und Organisationen im Auswahlverfahren manifestieren“ (S.73).

Das Konzept der sozialen Validität wurde von Schuler und Stehle (1983) eingeführt und von Schuler (1990) weiterentwickelt. Ein weiterer Ansatz zur Erfassung der sozialen Akzeptanz von Auswahlverfahren wurde 1993 von Gilliland vorgelegt. In diesem Modell wird eine Unterscheidung zwischen der Wahrnehmung der Auswahl-situation und der Auswahlentscheidung getroffen. Dabei bestehen für beide Komponenten unterschiedliche Regeln, deren Verletzung die soziale Akzeptanz eines Auswahlverfahrens negativ beeinflussen kann. Da das Modell im deutschsprachigen Raum bisher wenig Beachtung fand, steht in dieser Arbeit das Konzept von Schuler im Vordergrund.

Allerdings ist die von Schuler ursprünglich gewählte Bezeichnung „Konzept der sozialen Validität“ nicht ganz unmissverständlich. Die Akzeptanz eines Verfahrens ist nicht in empirischer Abhängigkeit von der Validität eines eignungsdiagnostischen Instrumentes zu sehen. Wenn auch die gegebene prädiktive Validität oder Konstruktvalidität eines

Auswahlinstrumentes dazu beitragen kann, dass ein Auswahlverfahren eher von Bewerbern akzeptiert wird, ist dies doch nicht zwangsläufig die Folge. Insofern ist die von Schuler (1990) vorgeschlagene Alternativbezeichnung der sozialen Akzeptanz treffender.

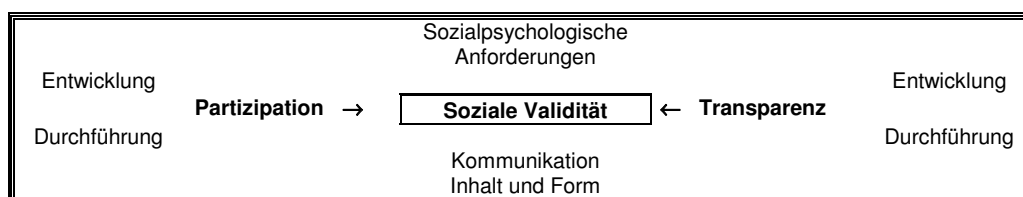
Letztendlich ist das Erleben von Probanden in Auswahl-situationen sowie deren Bewertung verschiedener eignungsdiagnostischer Verfahren noch weitgehend unerforscht.

Verstanden und zusammengefasst werden unter dem Begriff der sozialen Validität alle Maßnahmen, die diagnostische Situationen zu sozial akzeptablen Situationen machen. Zu nennen sind insbesondere vier wesentliche Komponenten:

1. Mitteilung relevanter Informationen über wichtige Aufgaben und Organisationsmerkmale
2. Partizipation an der Entwicklung und Anwendung der diagnostischen Verfahren
3. Transparenz hinsichtlich des Auswahlverfahrens
4. Angemessene Kommunikation der Ergebnisse

Die Komponenten lassen sich graphisch wie folgt darstellen (Tabelle 1.2-3):

Tab.1.2-3: Komponenten der sozialen Validität nach Schuler und Stehle (1983)



Unter sozialpsychologischen Anforderungen verstehen die Autoren die diagnostische Verwertung von Informationen über das Sozialgefüge einer Organisation. Gemeint sind damit in erster Linie das Organisationsklima und der gelebte Führungsstil. Partizipation bedeutet die Einbeziehung von Organisationsmitgliedern, wenn es um die Entwicklung und Durchführung von Auswahlverfahren geht. Dabei ist nach Aussage der Autoren die Mitwirkung während des gesamten Prozesses der Entwicklung, angefangen bei der Arbeitsplatzanalyse bis hin zur Entscheidung über Annahme oder Ablehnung eines Bewerbers, zu gewährleisten. Weiter sollte die diagnostische Situation ausreichend transparent sein und damit für den Teilnehmer einsichtig und augenscheinlich sowie die Beurteilungsmaßstäbe nachvollziehbar. Schließlich sollte die Kommunikation so gestaltet werden, dass ein wechselseitiger Informationsaustausch gefördert wird. Dabei ist aber durch die Wahrung einer gewissen Distanz Rücksicht auf die Persönlichkeitssphäre des Kandidaten zu nehmen.

Schuler (1990, 1998) hat die wesentlichen erlebnisrelevanten Aspekte einer Auswahl-situation aus Sicht der Teilnehmer an einem Auswahlverfahren konkretisiert und, wie in Tabelle 2.2-3 dargestellt, übersichtlich zusammengefasst.

Tab. 2.2-3: Erlebnisrelevante Aspekte einer Auswahl-situation nach Schuler (1990, 1998)

<p>1. Information</p> <ul style="list-style-type: none"> - über die Aufgabenbereiche der Tätigkeit - über die erfolgskritischen Anforderungen - über die wichtigsten Organisationsmerkmale und –ziele - über die Organisationskultur und –stil - über Möglichkeiten persönlicher und beruflicher Entwicklung und weitere Aspekte, die sich als bedeutsam für Leistung und Befinden erwiesen haben und Selbstselektion erleichtern <p>2. Partizipation / Kontrolle</p> <ul style="list-style-type: none"> - im engeren Sinn als Beteiligung an der Gestaltung der Auswahl-situation oder –instrumente oder an der Entscheidung (in direkter oder repräsentativer Form mittels Arbeitnehmervertretung) - im weiteren Sinne als Möglichkeit, Kontrolle über die Situation auszuüben und über das eigene Verhalten oder über das Verhalten oder die Entscheidung relevanter anderer, oder verstanden als Freiheit von der Machtausübung anderer <p>3. Transparenz</p> <ul style="list-style-type: none"> - der Auswahl-situation inklusive der handelnden Personen, ihrer Rollen, Intentionen und Kompetenzen sowie der Verhaltenserwartungen an die Bewerber - der Bedeutung und des Aufgabenbezugs der diagnostischen Instrumente - des Bewertungsprozesses und der Bewertungsregeln, d.h. der Beurteilungskriterien, Standards, Prinzipien des diagnostischen Schlusses und der Aggregation von Daten sowie der Transformation der Daten in Urteile oder der Urteile in Entscheidungen - des diagnostischen Prozesses in einer Form, die Selbstbeurteilung begünstigt (und in deren Konsequenz Selbstselektion erleichtert wird, wie bspw. durch Arbeitsproben oder via sozialem Vergleich im Assessment Center) <p>4. Urteilskommunikation / Feedback</p> <ul style="list-style-type: none"> - diagnostische Information durch die Verfahren und die Beurteiler - inhaltlich: offen, wahrhaftig, bezogen auf Erfolgswahrscheinlichkeiten und Entwicklungsmöglichkeiten - formal: verständlich (semantisch und pragmatisch), rücksichtsvoll, unterstützend; Selbsteinsicht, Integration in das Selbstkonzept und informierte Entscheidung der Kandidaten erleichternd
--

In der nachfolgenden Tabelle 3.2-3 werden von Schuler (1998) beispielhaft idealtypische Fragen vorgestellt, mit denen die ausgewählten Aspekte der sozialen Validität bei Teilnehmern am Auswahlverfahren zu erfassen sind.

Tab.3.2-3: Ausgewählte Fragen zu den vier Aspekten der sozialen Validität nach Schuler (1998)

<p>1. Information</p> <ul style="list-style-type: none"> - Ich wurde über meine künftigen Aufgaben informiert - Mir ist klar, welche künftigen Anforderungen an mich gestellt werden - Ich fühle mich über das Unternehmen gut informiert - Die Aufgabe hilft mir, meine beruflichen Pläne zu präzisieren <p>2. Partizipation / Kontrolle</p> <ul style="list-style-type: none"> - Ich konnte meine Stärken zum Einsatz bringen - Ich konnte die Situation zu meinen Gunsten beeinflussen - Ich konnte durch eigene Anstrengung etwas erreichen - Ich fühlte mich ernstgenommen und respektiert <p>3. Transparenz</p> <ul style="list-style-type: none"> - Die Aufgabenstellung war verständlich - Mir ist klar, welche Fähigkeiten untersucht wurden - Mir ist klar, welches Verhalten von mir erwartet wurde - Ich kann mir vorstellen, dass hierdurch berufliche Eignung festgestellt wird <p>4. Urteilskommunikation / Feedback</p> <ul style="list-style-type: none"> - Ich wurde offen und klar informiert, wie ich abgeschnitten habe - Ich wurde fair und rücksichtsvoll über mein Abschneiden informiert - Durch die Aufgabe weiß ich jetzt mehr über meine Fähigkeiten als zuvor - Die Aufgabe hat mir geholfen, mir über meine berufliche Eignung klar zu werden

Unter diesen Gesichtspunkten ist es zur Erhöhung der sozialen Validität biographischer Fragebogen insbesondere wichtig, auch Betroffene, d.h. die Bewerber und Interessenvertreter (Betriebsräte) bei der Einführung des Auswahlinstrumentes zu beteiligen. Die Darstellung eines gewissen Berufsbezuges bietet die Grundlage für eine hohe Augenscheinvalidität. Durch das Weglassen sozial nicht erwünschter Fragen entsteht bei den Bewerbern kein Gefühl einer Verletzung der Privatsphäre.

Bei den Konstrukten sollte nach Hossiep et al. (2000) darauf geachtet werden, dass die Bedeutung der einzelnen Dimensionen gut vermittelt wird und den Kandidaten deren Relevanz für das Berufsleben plausibel erscheint. So sind bspw. Skalen mit Kunstwörtern zu vermeiden, deren Interpretation unter Umständen schwierig ist. Weitere Aspekte sind nicht im Verfahren selbst begründet, sondern durch den Eignungsdiagnostiker beeinflussbar. So ist es förderlich, Teilnehmern die Funktionsweise psychologischer Testverfahren und deren Bewertung zu erklären. Hierzu zählt bspw. das Prinzip, dass die Beschreibung einer Person stets relativ zum Mittelwert der jeweiligen Vergleichsgruppe erfolgt (Wottawa, 1980). Wichtig ist aber auch eine angemessene Rückmeldung an die Teilnehmer. Ähnlich wie es im Assessment Center zum Standard geworden ist, dass die Beobachteten eine differenzierte Rückmeldung unabhängig von ihrem erreichten Erfolg erhalten, sollte eine ausführliche Erklärung der erreichten Ergebnisse und die Hinweise auf persönliche Weiterentwicklungsmöglichkeiten oder Verbesserungsmöglichkeiten grundsätzlich Bestandteil eines Auswahlverfahrens sein. Da dies nach Durchführung psychologischer Testverfahren leider nicht die Regel ist, wäre dies ein Ansatz, der ohne großen Aufwand einen Beitrag zur Erhöhung der sozialen Validität liefern könnte. Scheitert ein Bewerber aufgrund seiner Angaben in einem biographischen Fragebogen, kann ihm differenziert zurückgemeldet werden, warum andere Mitbewerber bevorzugt ausgewählt wurden und welche Hintergründe sich mit der Auswahl verbinden. Sind es in einem solchen Fragebogen Items wie Berufserfahrung oder Übernahme von Verantwortung, die zu einem hohen Summenscore beitragen, kann die Empfehlung durchaus lauten, dass sich jugendliche Bewerber erst einmal in einem anderen als dem gewünschten Beruf bewähren oder in anderen Kontexten lernen, Verantwortung zu übernehmen. Hossiep et al. (2000) verweisen auf Benziman (1986), wonach gute Erfahrungen damit gemacht werden konnten, wenn auch die Teilnehmer die Gelegenheit zum Feedback über das durch sie wahrgenommene Verfahren bekamen.

Bisherige Untersuchungsergebnisse lassen vermuten, dass das Erleben der Bewerber nicht in erster Linie durch deren subjektive Belastung im Auswahlverfahren beeinflusst wird, sondern eher durch Situationsparameter im o.g. Sinne, d.h. vorhandene vs. fehlende Informationsmöglichkeiten, wahrgenommene Aspekte der Kontrollierbarkeit und die Augenscheinvalidität eines Verfahrens (Fruhner, Schuler, Funke & Moser, 1991). Nach Graudenz (1987) wünschen bspw. zwei Drittel der Probanden mehr Vorinformationen zur Eignungsuntersuchung. Angesichts der zunehmenden Konkurrenz auf dem Arbeitsmarkt um qualifizierte Arbeitnehmer und gute Schulabgänger sind Kennwerte zu Erleben, Bewertung und Akzeptanz eignungsdiagnostischer Untersuchungen nicht nur im Interesse von Teilnehmern eines Auswahlverfahrens zu betrachten, sondern insbesondere im ureigenen Interesse einer jeden werbenden Organisation.

Gleichwohl kritisiert Kersting (1998) die bisherige Praxis bei der Erforschung der sozialen Akzeptanz von Personalauswahlverfahren. So zeigen sich bei einigen Arbeiten Defizite auf der konzeptionellen Ebene, auf der Ebene der eingesetzten Messinstrumente zur Erfassung der sozialen Akzeptanz und /oder auf der Stichprobenebene. So werden bspw. diagnostische Verfahren mit unvergleichbaren Zielsetzungen, wie Interviews einerseits und Leistungstests andererseits miteinander verglichen. Es fehlten vor allem binnenanalytische Vergleiche von eignungsdiagnostischen Instrumenten innerhalb einer Verfahrensgruppe. Weiter verweist Kersting auf die teilweise willkürliche Kategorisierung einzelner Verfahren. So etwa wird bei Fruhner und Schuler (1991) unter der Oberkategorie „Tests“ zugleich nach der Akzeptanz von Persönlichkeitsfragebogen und nach der von Leistungstests gefragt. Hinsichtlich der gewählten Stichproben bleibt zu bemängeln, dass nur solche Personen befragt werden, welche die zu beurteilenden Verfahren und den Ernstfall von Personalauswahlentscheidungen aufgrund eigener Informationen aus eigener Erfahrung kennen. Die zu beklagende Praxis sind jedoch Studenten, die nach der Akzeptanz von Auswahlverfahren befragt werden, ohne Kenntnis des eignungsdiagnostischen Ernstfalles oder der zur Frage stehender Verfahren aus eigener Erfahrung.

Mit der in den letzten Jahren rückläufigen Anzahl geeigneter Bewerber und des teilweise zunehmenden Anspruches an neue Mitarbeiter, hat das Interesse am Personalmarketing deutlich zugenommen (Moser, Stehle & Schuler, 1992).

Qualifizierte Mitarbeiter erhalten heute häufig mehrere Stellenangebote, da sie sich in der Regel parallel für verschiedene Berufe und in unterschiedlichen Organisationen bewerben. So

darf sich eine Organisation heutzutage nicht ausschließlich auf die Gewinnung neuer Mitarbeiter und Auszubildender konzentrieren, sondern muss sich gegenüber möglichen Nachwuchskräften, insbesondere jenen, die schon im Auswahlverfahren erfolgreich waren, möglichst positiv darstellen können. Denn wie Moser und Zempel (2001) beschreiben, wird der Bedarf an Personalmarketing durch das Verhältnis von Angebot und Nachfrage bestimmt. Je geringer die qualifizierte Nachfrage an gegebenen Arbeitsplätzen ist, desto größer ist der Bedarf an Personalmarketing.

Ein im Sinne der sozialen Akzeptanz wie auch hinsichtlich des Marketings wichtiger Aspekt, der für Bewerber von Bedeutung sein kann, ist die Transparenz und das Wissen um die zur Anwendung kommenden Auswahlverfahren. Dabei gehören grundsätzlich Verfahren, die für den Teilnehmer über eine gewisse Augenscheinvalidität verfügen, nach wie vor zu den beliebtesten Instrumenten aus Sicht der Bewerber. So fanden Fruhner et al. (1991), kaum überraschend, auf Platz eins das Einstellungsinterview. In Tabelle 4.2-3 wird eine Reihenfolge wiedergegeben, die befragte Hochschulabsolventen auf die Frage nannten, mit welchem Verfahren sie ausgewählt werden möchten.

Tab. 4.2-3: Reihenfolge beliebter Auswahlinstrumente bei Hochschulabsolventen nach Fruhner, Schuler, Funke & Moser, 1991

1.	Vorstellungsgespräch
2.	Arbeitsprobe
3.	Praktikumleistung
4.	Zeugnisnoten
5.	Psychologische Testverfahren
6.	Lebenslauf
7.	Schriftproben
8.	Losverfahren

Die relative schlechte Platzierung psychologischer Testverfahren wird durch die Autoren dahingehend erhellt, da sich die Ablehnung vielfach auf Persönlichkeitstests und weniger auf kognitive oder Intelligenztests bezieht. Insbesondere jene Beurteiler, die noch keine Erfahrungen mit einem genannten Auswahlinstrument machten, beurteilten dieses als weniger positiv. Insgesamt schätzen nach Jäger (1995) bei 15 – 20 Jahre alten Bewerbern nur 45 % die Bedeutung von Eignungsdiagnostik als hoch bis sehr hoch ein. 28 % messen der Eignungsüberprüfung einen niedrigen bis sehr niedrigen Wert bei.

Obwohl die berufliche Eignungsdiagnostik zwischenzeitlich weit verbreitet etabliert und ihre Errungenschaften und Erkenntnisse anerkannt sind, gab es auch immer wieder Kritik und dies nicht nur von Teilnehmern im Auswahlverfahren, sondern auch von Psychologen selbst (vgl. a. Hesse & Schrader, 1991; Sarges, 1995a; Schuler & Stehle, 1983, Weinert, 1998). Es wurde

der Eignungsdiagnostik vorgeworfen, sie diene in erster Linie den Organisationen, orientiere sich an fragwürdigen Leistungsnormen und missachte die Bedürfnisse der Probanden. Ethische Fragen, wie jene nach der Stellung und Rechtfertigung von Persönlichkeitstests dürften bei biographischen Fragebogen ähnlich zu bewerten sein. Bedeutung erfährt insbesondere die Frage nach dem Recht einer Organisation, die neue Arbeitnehmer sucht, in die Privatsphäre eines Menschen einzudringen. In diesem Zusammenhang sind auch Möglichkeiten eines Missbrauchs von Daten und deren Weitergabe an Dritte zu erwähnen. Kritik betrifft aber auch die Validität des Testmaterials. So wurden Meinungen geäußert, dass die angewandten Testmethoden kein wahres bzw. valides Bild von Fähigkeiten und Qualifikationen geben, dass der abgefragte Testinhalt mit den Anforderungen in der beruflichen Situation wenig gemein habe und dass Testverfahren nur sehr kleine Facetten des Gesamtspektrums menschlicher Fähigkeiten abbilden können. Weiter wird den Testverfahren teilweise abgesprochen, dass sie in der Lage sind, die wesentlichen Variablen zu erfassen, die im Berufsleben von Bedeutung seien (Talent, Initiative, Stehvermögen etc.) und vor allem durch sie Situationsvariablen nicht berücksichtigt werden. Letzteres ließe damit wenig Aussagefähigkeit darüber zu, wie sich eine Person unter anderen als den Konstellationen der Prüfungssituation verhalten würde. Gerade aufgrund der wiederkehrenden Kritik an der eignungsdiagnostischen Praxis und dem Instrumentarium lässt sich die Notwendigkeit der theoretischen Fundierung, der empirischen Überprüfung und der Erfassung der sozialen Akzeptanz ableiten.

Zudem weisen Schuler und Stehle (1985) darauf hin, dass das Verhältnis zwischen Testteilnehmer und Diagnostiker in der Praxis nicht durch die üblichen Normen der gesellschaftlichen Interaktion gekennzeichnet ist, sondern es hierbei eher um eine Form der Unterordnung gehe. Fruhner et al. (1991) kritisieren den Ablauf diagnostischer Situationen vielfach als undemokratisch und verweisen auf das Fehlen von Transparenz, Kontrollmöglichkeiten sowie gleichgewichtiger Partizipation, d.h. fehlender sozialer Validität.

Trotz der teilweise mangelnden Sorgfalt in der bisherigen Praxis bei der Erfassung der sozialen Akzeptanz von eignungsdiagnostischen Instrumenten (Kersting, 1998) liegen heute verschiedene Studien vor.

Wie Bewerber die in vielen Unternehmen üblichen Ehrlichkeits- bzw. Integritätstests bewerten, untersuchten 1987 Ryan und Sackett. Unter drei unterschiedlichen Bedingungen: ehrlich zu antworten, sich möglichst gut darzustellen und wie ein potenziell interessierter

Bewerber zu antworten, wurden 148 Studenten mit Arbeitserfahrung in drei Gruppen aufgeteilt und als Stichprobe herangezogen. Im Ergebnis zeigte sich, dass die Bewertungen der potenziell interessiert antwortenden Gruppe mit den Ergebnissen der ehrlich antwortenden Gruppe nahezu vergleichbar waren. Starke negative Bewertungen von Integritätstests wurden nicht gefunden. Die meisten Befragten fanden die Verwendung solcher Testverfahren angemessen. Auch hatte die Anwendung keinen Einfluss auf das Erleben und Bewerten derjenigen Unternehmen, die solche Tests als Auswahlinstrument einsetzen.

Das Erleben zweier unterschiedlicher diagnostischer Situationen in der Personalauswahl, die Teilnahme an einem psychologischen Test sowie einem Vorstellungsgespräch, untersuchten Fruhner, Schuler, Funke und Moser (1991). Von den 909 teilnehmenden Studenten hatten etwa die Hälfte Vorerfahrungen mit Testverfahren, die andere mit Interviews. Als generelles Ergebnis zeigte sich, dass Einstellungsinterviews hinsichtlich aller erfassten Aspekte der sozialen Akzeptanz besser beurteilt wurden als psychologische Testverfahren. Zusätzlich bewirkte die Vorerfahrung, dass erfahrene Bewerber das jeweilige Auswahlinstrument positiver, transparenter und weniger belastend erlebten als unerfahrene Bewerber.

Köchling und Körner untersuchten 1996 die Frage, wie Ausbildungsplatzbewerber eines Großunternehmens der Industrie ihre Eignungsuntersuchung mit Hilfe psychologischer Testverfahren und eines Vorstellungsgesprächs erleben und bewerten. Die Probanden durchliefen unterschiedliche Intelligenz- und Leistungstests sowie ein abschließendes Interview. Um den Befindlichkeitsverlauf der Bewerber zu erheben, wurde den Teilnehmern viermal über den Tag verteilt die Befindlichkeits-Skala (Bf-S; v. Zerssen, 1975) vorgelegt. Weiter wurden die Probanden gebeten, anhand konstruierter Fragebögen einzuschätzen, wie unterschiedliche Aspekte der Auswahl-situation von ihnen erlebt wurden. Wichtig waren den Autoren insbesondere Fragen nach den Eindrücken der Teilnehmer bezüglich der dargebotenen Information, der Partizipation und Kontrollierbarkeit sowie der Transparenz. Auf der Grundlage der Ergebnisse des ersten Untersuchungstages wurden verschiedene Maßnahmen zur Verbesserung von Ablauf und Erscheinungsbild der Untersuchung aus Sicht der Bewerber durchgeführt. Ziel war ein positiveres Situationserleben und einen positiveren Eindruck der Kandidaten zu erreichen. Nach Neugestaltung fühlten sich die Bewerber im Auswahlverfahren signifikant besser über Merkmale und Ziele des Unternehmens informiert. Insgesamt konnte auch das subjektive Befinden der Bewerber während des gesamten Auswahlverfahrens, insbesondere das nach Durchführung der psychologischen Eignungstests,

deutlich verbessert werden. Weiterhin wurde die Zufriedenheit der Kandidaten mit den zu Beruf und Unternehmen erhaltenen Informationen, die von ihnen erlebte Mitwirkungsmöglichkeit bei der Gestaltung der Test- und Gesprächssituationen sowie ihr Vertrauen in die Tauglichkeit des Vorstellungsgesprächs als Auswahlinstrument bedeutsam erhöht. Der in der Situation enthaltene Leistungsdruck wurde kaum noch bemängelt. Auch hier wurde wieder das Vorstellungsgespräch von den Bewerbern positiver erlebt als die Testverfahren. Leider äußern sich die Autoren nicht zu der Frage, inwieweit die Verbesserungen dazu beigetragen haben, dass die vom Unternehmen angebotenen Plätze für die ausgewählten Bewerber auch wirklich von diesen angenommen wurden. Damit wäre zumindest ein Indiz für die längerfristige Stabilität des Eindrucks im Auswahlprozess und die mittelbare positive Wirkung auf die Entscheidung hinsichtlich der Annahme bzw. Ablehnung eines Ausbildungsplatzes durch geeignete Kandidaten zu sehen.

Informationen darüber, ob weniger beliebte Testverfahren sogar zu einem Verzicht auf eine Bewerbung führen können, sind nur wenige vorhanden. Poortinga, Coetsier, Meuris, Miller, Samsonowitz, Seisedos und Schlegel (1982); zit. n. Moser und Zempel (2001) schätzen den Anteil der Bewerber, die lieber auf eine Bewerbung verzichten als sich einem psychologischen Test auszusetzen auf nur ca. 2 %. Fruhner et al. (1991) befragten mehrere hundert Studierende und stellten fest, dass zwar 84,3 % freiwillig an einem Vorstellungsgespräch teilnehmen würden, jedoch nur 45 % an einem psychologischen Test. Ob dies allerdings bedeutet, dass die Bewerber dann auf eine Vorstellung verzichten oder die Anwendung nur ein negatives Licht auf das Unternehmen wirft, bleibt offen.

Ziel der Arbeit von Kersting (1998) war es, die Moderation des Urteils über die soziale Akzeptanz insbesondere durch Personenvariablen zu untersuchen. Dazu verglich er Intelligenztests einerseits und computergestützte Problemlöseszenarien andererseits. Die Untersuchungsgruppe setzte sich aus 103 überwiegend männlichen Angehörigen des Polizeivollzugsdienstes zusammen. Zur Akzeptanzuntersuchung wurde ein extra für diesen Zweck konstruierter Akzeptanz-Fragebogen entwickelt, mit dem die Dimensionen Transparenz, Partizipation und positives Erleben erfasst wurden. Der Autor kommt zu dem Ergebnis, dass im Gegensatz zu den Gerüchten über eine pauschale Ablehnung von Intelligenztests, die Teilnehmer Verfahren durchaus differenziert beurteilen. Das positive Erleben der Problemlöseszenarien kann als ein Indiz dafür gewertet werden, dass den Teilnehmern diese Aufgaben „mehr Spaß“ machen als Intelligenzaufgaben. Intelligenztests

übertreffen dagegen die Problemlöseszenarien bezüglich der Kontrollierbarkeit im Sinne einer höheren wahrgenommenen Qualität der Messung. Weiter wurde den Problemlöseszenarien im direkten Vergleich eine größere Realitätsnähe (face-validity) zugesprochen. Hinsichtlich der Abhängigkeit in der Einschätzung von Personenfaktoren, hat vor allem die eigene Leistung einen moderierenden Effekt. Auf selbstwertbedrohende Informationen, wie Misserfolg im Diagnoseverfahren, reagieren die Teilnehmer mit einer selbstwertdienlichen Abwertung des Verfahrens. In der Untersuchung zeigte sich ein Haupteffekt der Intelligenz auf die Beurteilung der Intelligenztests unter dem Gesichtspunkt Kontrollierbarkeit / Messqualität: Personen mit einer niedrigeren Intelligenzprägung sprachen bspw. den Tests eine geringere Kontrollierbarkeit zu als Personen, die über eine hoch ausgeprägte allgemeine Intelligenz verfügten. Ältere Menschen schätzten daneben die Messqualität von Intelligenztests skeptischer ein als Jüngere. Zudem zeigt sich, dass die Akzeptanz von Auswahlverfahren nicht unidimensional, sondern mehrdimensional erfasst werden sollte. Ein Verfahren hat nicht nur einen Akzeptanzwert, sondern mehrere, die zusätzlich noch in Abhängigkeit von Person- und Situationsmerkmalen variieren.

Dass Organisationen nach bereits erfolgter Bewerbung Gelegenheit haben, sich bei möglichen Nachwuchskräften im Vergleich zu anderen in Frage kommenden Arbeitgebern möglichst positiv darzustellen, hängt auch von der Gestaltung der jeweiligen Auswahlbedingungen ab. Wie verschiedene Untersuchungen belegen, spielen die im Rahmen der Personalauswahl gewonnenen Eindrücke bei der späteren Wahl des Arbeitgebers durchaus eine Rolle. So konnten Schmitt und Coyle schon 1972 zeigen, dass die Absicht von Bewerbern, ein Stellenangebot zu akzeptieren sowie ihre Einstellung bezüglich der um sie werbenden Organisation, eng mit der subjektiven Wahrnehmung des Interviewprozesses zusammenhängen. Wichtig waren hier insbesondere die wahrgenommene Persönlichkeit des Interviewers, die Art und Weise der Gesprächsführung und die Angemessenheit und Güte der Information über die zu besetzende Stelle. Dieses Ergebnis unterstützt die Erkenntnis von anderen Autoren wie Schuler (1989) sowie Schuler und Moser (1992), dass das Interview neben der Personalauswahl auch der Darstellung von Organisation und Arbeitsplatz, zur Beantwortung von Fragen des Bewerbers und der Öffentlichkeitsarbeit dient.

Wie Köchling und Körner (1996) zutreffend darstellen, ist durch derartig gewonnene Eindrücke das Image einer Organisation für Nachwuchskräfte direkt betroffen. Bewerber tragen Informationen über die Personalauswahlpraxis, das Verhalten der

Personalverantwortlichen und damit ihren Eindruck des Betriebsklimas nach außen und verfehlen damit kaum die Wirkung auf andere potentielle Nachwuchskräfte. So ist es für Unternehmen wichtig und sinnvoll, die Durchführung der Personalauswahl auch unter dem Aspekt der Imagepflege zu betrachten. Hierfür spricht insbesondere, dass ebenso bei der Imagebildung der Halo-Effekt (Forsyth, 1987) zum Tragen kommt, d.h. dass ein besonderer positiver oder negativer Einzeleindruck als typisch für das gesamte Klima bzw. den generellen Umgang mit den Mitarbeitern wahrgenommen wird. Fast noch wichtiger ist diese Erkenntnis in ihrer Bedeutung für die Teilnehmer am Testverfahren, die als nicht geeignet wieder nach Hause geschickt werden. Sie haben keine Gelegenheit, ein vermeintlich gewonnenes Bild über die Organisation durch weitergehende spätere Erfahrungen zu relativieren und neue Eindrücke zu integrieren. Gleichwohl tragen sie aber ihre Eindrücke weiter und halten so möglicherweise für die Organisation interessante Nachwuchskräfte von einer Bewerbung ab.

Doch wie werden biographische Fragebogen als eignungsdiagnostische Instrumente durch die Teilnehmer bewertet? In den wenigen vergleichenden Untersuchungen zur Akzeptabilität unterschiedlicher Auswahlinstrumente schnitten biographische Fragebogen gegenüber anderen Testverfahren in den USA besser ab, in Frankreich gleich (Steiner & Gilliland, 1996) und in Deutschland etwas schlechter (Schuler & Marcus, 2001).

An der Spitze stehen fast immer Interviews und Arbeitsproben, dabei landeten die biographischen Fragebogen zumindest in der Regel im Mittelfeld, die Graphologie dagegen deutlich am Ende.

Gezielt untersucht wurde die Akzeptanz biographischer Fragebogen von Stehle (1983) an 26 Außendienstmitarbeitern und sieben Betriebsratsmitgliedern in der Versicherungsindustrie im Rahmen seiner in Kapitel 2.2 dargestellten Validierungsstudie. Beiden in die Akzeptanzuntersuchung eingebundenen Gruppen wurden die Fragen des Verwendung findenden biographischen Fragebogens mit zwei zusätzlich versehenen Skalen vorgegeben. Jeder Teilnehmer hatte die Aufgabe zu bewerten, ob die vorgegebenen Fragen in einem Einstellungsinterview sowie einem Personalfragebogen als uneingeschränkt zulässig, bedingt zulässig oder unzulässig betrachtet werden können. Zur Beurteilung wurde den Teilnehmern der Untersuchung die deutsche Übersetzung des von Baehr und Williams (1967) entwickelten biographischen Fragebogens vorgelegt. Zusammengefasst lässt sich anmerken, dass Fragen, die sich auf bisherige Berufserfahrungen beziehen als absolut problemlos angesehen wurden.

Einzigste Ausnahme war hier eine Frage, die Auskunft hinsichtlich einer möglichen finanziellen Verschlechterung aufgrund eines Stellenwechsel erbat. Diese Frage wurde im Fragebogen wie auch im Interview als unzulässig bewertet. Deutlich geringer war die Bereitschaft, Inhalte eines Fragebogens zu akzeptieren, wenn es sich um die Schul- und Ausbildungszeit handelte. Hier war die Tendenz, Fragen als unzulässig zu bewerten, deutlich höher. Durchgängig abgelehnt wurden Fragen, die inhaltlich den Bereich Kindheit und Verhalten der Eltern repräsentierten. Eine Frage wie bspw. „Wie oft waren Sie im Verhältnis zu anderen Kindern während ihrer Kindheit krank?“ wurde durchgängig als unzulässig betrachtet. Andererseits fand eine Frage nach der Anzahl der Krankheitstage im letzten Jahr sowohl im biographischen Fragebogen wie auch in einem Interview Zustimmung. Von den 151 zur Beurteilung vorgelegten Fragen wurden 36 im Interview und 90 in einem biographischen Fragebogen für unzulässig betrachtet. Die Teilnehmer zeigten demnach eine ausgeprägtere Bereitschaft Informationen über die eigene Person im Interview zu geben als im Fragebogen.

Bei der Arbeit von Stehle handelt es sich um die dem Autor einzig bekannte, in der ein zur Anwendung kommender biographischer Fragebogen hinsichtlich seiner Akzeptanz bei Teilnehmern eines Auswahlverfahrens hinterfragt wird. Mehrere Studien zur sozialen Validität von eignungsdiagnostischen Verfahren richten sich auf die soziale Validität von Assessment-Centern, deren soziale Validität im Vergleich zu anderen eignungsdiagnostischen Verfahren nach Holling und Leippold (1991) als relativ hoch anzunehmen ist. Schuler und Stehle (1983) betonen die hohe soziale Validität, die ein Assessment Center u.a. dadurch hat, dass für die Beteiligten über das gesamte Verfahren – einschließlich des Ergebnis-Feedbacks – eine hohe Transparenz gegeben ist. Wiendieck (1994) geht sogar soweit zu behaupten, dass zwar der prognostische Wert eines AC's (bspw. Thorntorn & Byham, 1982: $r = .37$) angesichts der Aufwendigkeit des Verfahrens keineswegs als zufriedenstellend zu bezeichnen sei. Der eigentliche Gewinn des Verfahrens vielmehr in der hohen Akzeptanz liege, die sich weitgehend auf den Glauben an seine Validität stütze.

Zum gängigen Verfahren bei der Durchführung eines Assessment Center gehört heute, die Teilnehmer im Vorfeld darüber zu informieren, welche Dimensionen mit den dazugehörigen Verhaltensoperationalisierungen in welchen Übungen handlungsrelevant sind. Das Wissen um die relevanten Dimensionen soll die Kandidaten motivieren, sich bezüglich dieser Dimensionen besonders gut zu präsentieren (Kleinmann, 1997). Wie der Autor belegen konnte, wurden die Teilnehmer, die über die Anforderungsdimensionen und bewerteten

Verhaltensweisen informiert wurden, im Assessment Center besser bewertet als Teilnehmer, die nicht informiert wurden. Man kann annehmen, dass die Teilnehmer durch die vorab geleistete Aufklärung ein stärkeres Gefühl der Kontrolle erfahren, da sie wissen, was genau Beobachter beobachten und welches Verhalten jeweils pro Übung relevant ist.

Auch die Frage nach der Akzeptanz von Beurteilungsverfahren wurde beleuchtet. Marcus und Schuler (2001) beurteilten die Akzeptabilität als zwar nicht ausreichende, aber unbedingt notwendige Bedingung, dass Leistungsbeurteilungen ihre Funktion erfüllen können. In diesem Kontext werden die Reaktionen von Beurteilern und Beurteilten als das Ausmaß beschrieben, in dem Prozess und Ergebnisse der Beurteilung als gerecht, d.h. fair und der Wirklichkeit entsprechend, empfunden werden. Unzureichend akzeptierte Beurteilungssysteme können negative Reaktionen provozieren, die von Gleichgültigkeit bis zur offenen Reaktanz reichen. Die Autoren verweisen auf zusammenfassende Ergebnisse von Dickinson (1993), dem zufolge darf mit guter Akzeptanz der Leistungsbeurteilungen dann gerechnet werden, wenn folgende Bedingungen erfüllt sind: kurze Beurteilungsintervalle, Existenz eines formalen Beurteilungssystems, hohe Sach- und Gesprächsführungskompetenz der Beurteiler, Gelegenheit für die Beurteilten, Einwände vorzubringen, als relevant erachtete Beurteilungsdimensionen und auf empfundene Schwächen des Systems wird mit Veränderung reagiert. Eine Schlüsselrolle hinsichtlich der Zufriedenheit mit dem System spielt auch hier die Partizipation der Betroffenen während der Entwicklung und Einführung.

Wie unter Kapitel 2.1.3 dargestellt, orientiert sich die Einführung neuer eignungsdiagnostischer Instrumente neben gesetzlichen Rahmenbedingungen am Personalvertretungsgesetz und dem Betriebsverfassungsgesetz. Hier wird die vertrauensvolle Zusammenarbeit zwischen den einzelnen Betrieben und den Personalvertretungen bzw. Personalräten geregelt. Die generelle Notwendigkeit der Akzeptanz von Fragebogenverfahren wird auch durch die hier geregelte Zustimmungspflicht der Personalräte unterstrichen.

Schuler und Marcus (2001) geben in Tabelle 5.2-3 einen Überblick für Fragebereiche aus Personalfragebögen, die rechtlich zulässig bzw. unzulässig sind.

Neben diesen rechtlichen Aspekten sind bei der Neukonstruktion eines biographischen Fragebogens aufgrund der im Kapitel dargestellten Hintergründe auch Möglichkeiten zu bedenken, die es erlauben, die Soziale Validität positiv zu beeinflussen.

Tab. 5.2-3: Beispiel für Fragebereiche, die in Personalfragebögen unzulässig sind nach Schuler & Marcus (2001)

Fragebereich	zulässiges Beispiel	unzulässiges Beispiel
Allgemeine identifizierende Merkmale	Name, Anschrift, Geburtsdatum	---
Rechtsstatus	Staatsangehörigkeit, Kranken- und Rentenversicherung	---
Familie	Familienstand, Geburtsdaten von Ehepartnern und Kindern	Heiratsabsichten, intime Beziehungen
Bildungsweg	Ausbildung, Zeugnisnoten, Lehrgänge	---
Bisherige Berufstätigkeit	Arbeitsplätze, Arbeitgeber, Kontinuität	---
Einkommen	letztes Einkommen	falls ohne Bezug zu Eignung
Vermögensverhältnisse	bei leitenden Angestellten in Vertrauensstellung	falls ohne Bezug zur Position
Vorstrafen	bei Einschlägigkeit, z.B. Unterschlagung bei Kassierern	bei mangelnder Einschlägigkeit
Religions- und Parteizugehörigkeit	in Tendenzbetrieben, z.B. konfessionelle Einrichtungen, politische Parteien	grundsätzlich unzulässig
Gewerkschaftszugehörigkeit	in Tendenzbetrieben, z.B. Arbeitgeberverbänden	grundsätzlich unzulässig
Schwangerschaft	in Sonderfällen, z.B. Infektionsgefahr, Tänzerinnen	nach europäischem Recht grundsätzlich unzulässig
Gesundheitszustand	bei Bedeutung für die Arbeitstätigkeit und Leistungsfähigkeit; gefährliche Infektionserkrankungen	allgemein gehaltene Fragen über frühere und derzeitige Krankheiten

Als ein wesentliches Fazit der Überlegungen von Schuler und Stehle (1983) sowie Schuler (1990, 1998) darf festgehalten werden, dass wenn ein eignungsdiagnostisches Instrument, so prognostisch valide es auch immer sein mag, nicht von den Teilnehmern akzeptiert wird, für deren Auswahl es konzipiert wurde, nicht zu erwarten ist, dass dies ohne Auswirkung auf das Verhalten potenziell Interessierter bleiben wird. Zum einen kann es sich direkt auf das Antwortverhalten auswirken, wenn Teilnehmer Fragen nicht oder nicht korrekt beantworten. Zum anderen vermittelt ein nicht akzeptiertes Instrument bei Bewerbern einen negativ besetzten Eindruck von der Organisation, der unter Umständen generalisiert wird und bei bestehender Alternative zum Verzicht auf die Stelle oder den Ausbildungsplatz führen kann. Angesichts bestehender Konkurrenzsituation auf dem Arbeitsmarkt im Wettbewerb um geeignete Stellenbewerber kann sich dies heute kaum eine Organisation mehr leisten. In der Folge sollten die Bemühungen einer Organisation bei der Planung eines Auswahlverfahrens auch die dargestellten Aspekte der Sozialen Validität wie Information, Partizipation, Transparenz und Feedback beachten.

3 Ableitung der Fragestellung

3.1 Konkretisierung der allgemeinen Fragestellung

In Kapitel 2 konnte der praktische Stellenwert der Anwendung biographischer Daten im Kontext der eignungsdiagnostischen Personalauswahl dokumentiert werden. Durch eine Vielzahl von Studien in unterschiedlichen Kontexten wurde die prädiktive Validität dieses Auswahlinstrumentes belegt. Wie dargestellt, besitzen biographische Daten im Vergleich zu Intelligenztests ($r = .51$), Arbeitsproben ($r = .54$), strukturiert geführten Interviews ($r = .51$), Assessment Center ($r = .36$), eine ausreichend hohe prädiktive Validität.

Allerdings variiert die Höhe der Kriteriumsvalidität biographischer Daten in unterschiedlichen Untersuchungen und in Abhängigkeit vom gewählten Kriterium. In der Metaanalyse von Schmidt und Hunter wurde der durchschnittliche Validitätskoeffizient mit $r = .35$ wiedergegeben. Ghiselli (1955) belegt eine mittlere Validität biographischer Verfahren von $r = .44$, wenn das Kriterium der Aus- und Weiterbildungserfolg war und $r = .41$ bei dem Kriterium berufliche Bewährung. Schmitt, Gooding, Noe und Kirsch (1984) errechneten eine Validität von $r = .52$, wenn als Kriterium die Gehaltshöhe gewählt wurde, dagegen nur $r = .23$ wenn die Ausbildungsleistung herangezogen wurde. Bei Bliesener (1992) lag der mittlere Koeffizient bei $r = .33$. Trotz der sich hier ausdrückenden Variabilität kann Schuler (1996, S. 47) die Nützlichkeit von eignungsdiagnostischen Instrumenten mit Validitätskoeffizienten in dieser Höhe aufzeigen (vgl. a. Kapitel 2).

Wie weiter in Kapitel 2.3 dargestellt, geht es in der eignungsdiagnostischen Praxis bei der Maximierung der prädiktiven Validität in der Regel darum, aus einer Vielzahl möglicher diagnostischer Instrumente diejenigen auszuwählen, die mit einem möglichst kleinen Aufwand und geringen Kosten eine größtmögliche Enge des Zusammenhanges zwischen gewählten Prädiktoren und Kriterien aufzeigen. Für die Erreichung dieser Zielsetzung scheinen biographische Fragebogen als Vorauswahlinstrument eine gute Voraussetzung mitzubringen.

Aufgrund des Aufwandes der Durchführung und der Notwendigkeit zur Gewährleistung der Gütekriterien wie bspw. Ökonomie des Auswahlverfahrens oder Standardisierung der Durchführungsbedingungen kämen Intelligenztests, Arbeitsproben, Interviews oder Assessment Center als Instrumente zur Vorauswahl nicht in Betracht.

Steigende Personalkosten und ein hoher Bedarf an geeigneten Bewerberinnen und Bewerbern stützen heute zusätzlich die Notwendigkeit einer wissenschaftlich fundierten Personalauswahl. Auch Öffentliche Haushalte orientieren sich zunehmend an betriebswirtschaftlichen Gesichtspunkten. Damit kommt zu dem Aufwand, der notwendig ist, um geeignete Interessenten für eine vakante Position auszuwählen, eine zusätzliche ökonomische Komponente hinzu. Angesichts der zunehmenden Komplexität vieler Berufe, der davon ableitbaren umfangreicheren Anforderungen und nicht zuletzt der Erkenntnis, dass Mitarbeiter, einmal umfangreich und teuer ausgebildet, ein Berufsleben lang weitere Investitionen für die Organisation rechtfertigen können müssen – Stichworte sind hier Human Resource Management und strategische Personalentwicklung – ist eine professionelle Personalauswahl nicht nur im Interesse von Organisation und Bewerber, sondern fast schon betriebswirtschaftlich zwingend notwendig.

Derzeit wird die Vorauswahl von Bewerbungsunterlagen in erster Linie durch den Einzelprädiktor Schulnoten bestimmt. Schon nach Althoff (1986) sind Schulzeugnisse eines der beliebtesten Kriterien für Personalentscheidungen. Bei Unternehmen und öffentlichen Arbeitgebern stehen die Schulnoten immer noch an der Spitze, wenn es zur Entscheidung über „Einladung oder Ablehnung“ kommt. Gute Noten gelten dabei u.a. als ein generelles Indiz für Engagement und Anpassungsbereitschaft (Berthel, 1995). Für eine statistisch signifikante Beziehung zwischen Schulnoten und diesen Konstrukten fehlen allerdings die Nachweise. Auch die Voraussetzung für eine sinnvolle Einschätzung von Schulnoten, die Kenntnis über die Notenverteilung in der Klasse oder der Schule, sind in der Regel nicht erfüllt. In Kapitel 2.1.2 wurde aufgezeigt, dass derzeit auch für die Einstellung in den Polizeivollzugsdienst, bei der Vorauswahl von Bewerbern, die Schulnoten eine entscheidende Einflussgröße darstellen.

Wie dargestellt, sind allerdings in den letzten Jahren jedoch wiederholt Zweifel an der Nützlichkeit dieses Prädiktors für die Erfassung zukünftiger Leistungen in Ausbildung und Beruf aufgetreten. Hintergründe hierfür sind möglicherweise in der Vielfalt der Schulsysteme in den verschiedenen Bundesländern und den sehr unterschiedlichen Anforderungen an einzelnen Schulen zu suchen (vgl. a. Kapitel 2.1.2.1). Damit einhergehend sind direkte Auswirkungen auf die Gütekriterien (insbesondere Objektivität und Reliabilität) des Prädiktors Schulnoten zu vermuten und in weiteren Studien zu untersuchen.

Wenn auch die verschiedenen Überlegungen derzeit nur Anregungen für Untersuchungen, nicht aber konkrete Erklärungen bieten können, ist zwischenzeitlich zumindest zu hinterfragen, in wie weit bei der Bewertung der Bewerbungsunterlagen den Schulzeugnissen der bisher gegebene Stellenwert eingeräumt werden darf, um Entscheidungen hinsichtlich der Annahme oder Ablehnung eines Bewerbers zum Testverfahren zu treffen. Dies ist insbesondere aufgrund der in Kapitel 2.1.3.3 dargestellten Bedeutung der Schulnoten bei der Bewerbung für den Polizeivollzugsdienst wichtig. Andere Überlegungen für eine ökonomische Vorauswahl der Bewerbungen aufgrund der eingereichten Unterlagen sind somit von Bedeutung. Reilly & Chao (1982) geben für die gesamten Bewerbungsunterlagen eine prognostische Validität von $r = .18$ an. Diese dürfte für Bewerbungsunterlagen als Vorauswahlinstrument nicht ausreichend sein.

Angesichts vorliegenden Erfahrungen mit der bisherigen unzureichenden Vorauswahlpraxis, der Hinweise auf eine möglicherweise nicht mehr ausreichende prädiktive Validität von Schulnoten und auch unter Beachtung der insgesamt sehr geringen Validität sonstiger Inhalte der Bewerbungsunterlagen (s.a. Punkt 2.1.2) sind Alternativen für eine verlässliche und ökonomische Erfolgsprognose und damit eine effektive Vorauswahl wünschenswert. Auch aufgrund des in den kommenden Jahren prognostizierten erhöhten Bedarfes an qualifiziertem Personal erscheint es notwendig, verlässliche Instrumente nutzen zu können. Warum sollte bspw. ein Bewerber, der einmal ein Studium abbrach und die Note „ausreichend“ in Politik im Zeugnis hat, nicht seine Eignung im Testverfahren überprüfen lassen dürfen. Weshalb sollten weiterhin Bewerber mit einer schlechten Deutschnote schon in der Vorauswahl eine Ablehnung erfahren, wenn die Zusammenhänge zwischen einzelnen Schulnoten unterschiedlicher Schulen in verschiedenen Bundesländern und dem Ergebnis im Auswahlverfahren nicht empirisch signifikant nachweisbar sind. Insofern ist eine statistisch signifikante Absicherung der empirischen Annahmen der gewählten Datengeber wünschenswert und dient notwendigerweise auch der Rechtfertigung und Akzeptanz des gewählten Verfahrens. Nur wenn es gelingt, einen entsprechend signifikanten Zusammenhang zwischen Prädiktor und Kriterium, d.h. Vorauswahlkriterium und Ergebnis im Auswahlverfahren, herzustellen, ist die Anwendung von Vorauswahlfiltern nicht nur sinnvoll, sondern auch unter Kosten-Nutzen Aspekten zwingend. Zudem lässt sich nur auf diesem Wege auch inhaltlich gegenüber Bewerbern die Vorauswahl vertreten.

In der eignungsdiagnostischen Praxis scheitert aus Bewerbersicht die Stellensuche oft schon auf Grundlage der eingereichten Zeugnisse. Auch wenn in vielen Unternehmen, das Screening der Bewerbungsunterlagen in erster Linie dazu dient, eine große Anzahl von Bewerbungen auf jene Zahl zu reduzieren, die in einer nächsten Phase berücksichtigt werden können, sollten die von Personalverantwortlichen hierfür herangezogenen Hypothesen, wissenschaftlichen Ansprüchen genügen.

Die nur sehr geringe inkrementelle Validität biographischer Daten, im Vergleich zu den in vielen Auswahlverfahren standardmäßig durchgeführten Intelligenztests, von $r = .01$ (Schmidt & Hunter, 1998) lässt dieses Instrument als zusätzlich eingesetzten Datengeber kaum nützlich erscheinen. Deshalb bietet sich aus gleichen Gründen (hohe prädiktive und geringe inkrementelle Validität) an, mittels eines gezielt konstruierten biographischen Fragebogens eine Personalvorauswahl zu initiieren. Denn nach Seibt und Kleinmann (1991) kann die Effizienz der Personalendauswahl nur so gut sein wie dies die Personalvorauswahl zulässt.

Auch wenn sich biographische Fragebögen in unterschiedlichen Erhebungen als für den Zweck der Vorauswahl nützlich erwiesen haben, muss immer noch offen bleiben, was letztendlich genau durch die Anwendung biographischer Fragebogen in der Personalauswahl erfasst wird (s.a. Kapitel 2.2.1.2 zur Konstruktvalidität biographischer Daten). Trotz der wiederholt belegten prädiktiven Validität, fehlt es immer noch an Studien, welche die Konstruktvalidität biographischer Fragebogen zu erfassen suchen und belegen. Dies mag mit ein Grund sein, warum der Einsatz solcher Fragebogen in erster Linie auf das Feld der eignungsdiagnostischen Auswahl beschränkt ist. Hier sind die Fragebogen häufig nur für eine stark eingeschränkte Population und auch nur für ausgesuchte Kriterien validiert. Eine Generalisierung der erzielten Ergebnisse auf andere Anwendungsfelder ist aus diesem Grund häufig nicht möglich, von den Anwendern aber oft auch nicht beabsichtigt. Angesichts der nach wie vor bestehenden Defizite wären weitere Erkenntnisse als Beitrag zur Erschließung der Konstruktvalidität auch im Rahmen dieser Untersuchung wünschenswert. Aufgrund der konkreten Praxisanwendung, die sich mit dem Ergebnis der Fragestellung ergeben soll, und der sich daraus ergebenden zeitlichen Rahmenbedingungen ist eine Ausweitung der Fragestellung auf den Bereich der Konstruktvalidierung nicht realisierbar. Zudem wurde in den Verhandlungen mit den Personalvertretern zugesichert, notwendige Daten ausschließlich kurzfristig personengebunden zu erheben und für weitergehende Berechnungen nur das abschließende Gesamtergebnis zu Grunde zu legen. Die Feststellung, dass bspw.

biographische Messinstrumente zwar eine hohe prädiktive Validität besitzen, jedoch im Sinne der inkrementellen Validität nur zu einem durchschnittlichen Validitätszuwachs von $r = .01$ beitragen, begründet sich dadurch, dass biographische Daten in beträchtlichen Umfang ($r = .50$) mit allgemeinen kognitiven Fähigkeiten korrelieren (Schmidt & Hunter, 1998). Nach Ansicht der Autoren lässt dies vermuten, dass biographische Daten ein indirektes Maß für mentale Fähigkeiten darstellen. Spätere Untersuchungen sollten diese Annahme aufgreifen und Zusammenhänge zwischen den erfassten mentalen Fähigkeiten von Teilnehmern des Auswahlverfahrens und ihrem Ergebnis im neu konzipierten biographischen Fragebogen reflektieren.

Letztendlich ist jeder weitere Schritt der Erschließung der prädiktiven Validität biographischer Daten nach Jäger (1986) essentieller Bestandteil der Konstruktvalidierung und liefert möglicherweise weitere Informationen, die zur Verankerung von Konstrukten in der Realität erforderlich sind. In anschließenden Untersuchungen wären beispielweise lineare Zusammenhänge der Gewichte einzelner Faktoren eines biographischen Fragebogens mit den Subtests des Testverfahrens zu analysieren. Die dargestellten Untersuchungen und vorliegenden Ergebnisse rechtfertigten aber aufgrund des hohen praktischen Nutzens die Anwendung im Kontext der Personalauswahl. Ungeachtet davon bleibt die weitergehende Konstruktaufklärung in zukünftigen Untersuchungen wünschenswert und notwendig, um auch die Wissenschaftlichkeit eines nützlichen Verfahrens zu belegen.

Wir können nach Schmidt und Hunter (1998) heute auf eine 85-jährige Forschungstradition zurückblicken, von der wir profitieren, wenn wir die Validität unterschiedlicher eignungsdiagnostischer Instrumente bestimmen. In diesem Kontext ist die prognostische Validität von zentraler Bedeutung für das psychologischen Eignungsurteil.

Ausgehend von der in den vorangegangenen Untersuchungen für unterschiedliche Berufsfelder dokumentierten Nützlichkeit (vgl. a. Kapitel 2.2) biographischer Fragebogen sowie der Notwendigkeit, ein geeignetes Instrument für eine effektivere Vorauswahl von Bewerbern zu finden, leitet sich die Frage ab, inwieweit mittels eines für den speziellen Kontext des Polizeivollzugsdienstes entwickelten biographischen Verfahrens dieser Funktion auch hier erfolgreich entsprochen werden kann. Denn wie Jäger (1986) konzidiert, macht sich die praktische Verwendbarkeit eines Auswahlinstrumentes zwar nicht ausschließlich an der Höhe des Korrelationskoeffizienten fest, aber eine ausreichend hohe Validität ist immer dann anzunehmen, wenn eine statistisch abgesicherte Validität nachgewiesen werden kann und

wenn kein alternatives Testverfahren zur Erfassung der beabsichtigten Merkmalsausprägung zur Verfügung steht. Angesichts der geringen prädiktiven Validität sonstiger Bewerbungsunterlagen (siehe Kapitel 2.1.2.1) und der beschriebenen Schwierigkeiten bei der Anwendung von Schulnoten als Vorauswahlkriterium erscheint es sinnvoll zu überprüfen, inwieweit biographische Fragebogen auch als eignungsdiagnostisches Vorauswahlinstrument für Bewerber für den Polizeivollzugsdienst in Frage kommen können.

Bisher liegen keine Untersuchungen vor, welche die Eignung biographischer Fragebogen in diesem Kontext belegen. Vorliegende biographische Fragebogen wurden häufig nicht für die Personalauswahl entwickelt oder sind nur für eine stark eingeschränkte Population validiert. Wie schon erwähnt, ist eine Generalisierung der erzielten Ergebnisse auf andere Anwendungsfelder kaum möglich und von den Anwendern oft auch nicht beabsichtigt. Zudem liegen häufig nur Übersetzungen von Fragebogen aus anderen Ländern vor. Ziel dieser Untersuchung ist es, aufgrund des bestehenden Defizits deutschsprachiger Fragebogen für diesen Kontext einen Fragebogen zu entwickeln, der für die eignungsdiagnostische Auswahl von Bewerbern für den Polizeivollzugsdienst gedacht ist. Auch dieser Fragebogen wird ausschließlich für die eingeschränkte Population (Bewerber für den Polizeivollzugsdienst) und auch nur für ausgesuchte Kriterien validiert. Durch eine Neukonstruktion eröffnen sich aber auch weitere Möglichkeiten. Schon in der Konstruktionsphase können Aspekte der sozialen Validität bzw. Akzeptanz (Schuler, 1990) berücksichtigt werden.

Als nützlich kann ein biographischer Fragebogen gelten, wenn durch ihn prognostisch valide Eignungsvorhersagen möglich sind, d.h. wenn zwischen den zum Zeitpunkt t1 erhobenen Prädiktoren (biographischen Daten im Fragebogen) und den zu einem Zeitpunkt t2 erhobenen Kriterium (Testergebnis im Auswahlverfahren) ein signifikanter Zusammenhang nachweisbar ist. Im Rahmen der Erschließung der prädiktiven Validität ist in dieser Arbeit zu erfassen, in wie weit die Annahme einer hohen Vorhersagegüte auch für die Auswahl im Polizeivollzugsdienst übernommen werden kann. Neben dieser notwendigen Nützlichkeit eines Verfahrens wurde in Kapitel 2.3 jedoch auch die Frage nach der Akzeptanz von eignungsdiagnostischen Auswahlinstrumenten betont und herausgestellt.

Der nachfolgenden Darstellung von Köchling (1999) ist noch einmal zusammenfassend zu entnehmen, warum die Akzeptanz von Auswahlverfahren für Organisationen von großer Bedeutung ist:

1. Die Einstellung der Bewerber gegenüber dem Unternehmen wird verbessert und damit die Bereitschaft zum Eintritt in die Organisation erhöht.
2. Bewerber schließen von den in einem Auswahlverfahren gemachten Erfahrungen auf das Image der Gesamtorganisation.
3. Bewerber geben die in einem Auswahlverfahren gemachten Erfahrungen an ihr soziales Umfeld weiter und vervielfältigen so ihre Eindrücke.
4. Eindrücke wirken gezielt und effektiv, da es sich bei den eingeladenen Kandidaten um einen repräsentativen Ausschnitt aus der Bevölkerungszielgruppe des Unternehmens handelt.
5. Kostenintensive und imageschädliche Fluktuation wird vermieden.

Die Anmerkungen belegen, dass die Nützlichkeit eines Auswahlverfahrens heute allein nicht mehr ausreicht. Darüber hinaus ist es fast genau so wichtig, dass Probanden, die in einem Verfahren ausgewählt bzw. nicht ausgewählt werden, das Instrument, mit dem diese Selektion stattfand, zumindest als dafür geeignet akzeptieren.

Für die soziale Akzeptanz des genutzten Auswahlinstrumentes ist grundlegend, dass die Teilnehmer am Auswahlverfahren den biographischen Fragebogen als für den Zweck geeignet anerkennen. Eine wichtige Voraussetzung hierfür ist, wie in Kapitel 2.3 dargestellt, die Partizipation der Betroffenen schon zur Zeit der Entwicklung des Verfahrens sowie in der Phase vor Einsatz als Selektionsinstrument.

Nach Schuler (1990) geht der Begriff der sozialen Validität über die einfache Akzeptanz eines Auswahlinstrumentes hinaus. Wie dargestellt, versteht er unter diesem Begriff alle Maßnahmen, die diagnostische Situationen zu sozial akzeptablen Situationen machen. Zu nennen sind dabei insbesondere vier wesentliche Komponenten:

1. Mitteilung relevanter Informationen über wichtige Aufgaben und Organisationsmerkmale
2. Partizipation an der Entwicklung und Anwendung der diagnostischen Verfahren
3. Transparenz hinsichtlich des Auswahlverfahrens
4. Angemessene Kommunikation der Ergebnisse.

Partizipation bedeutet die Einbeziehung von Organisationsmitgliedern, wenn es um die Entwicklung und Durchführung von Auswahlverfahren geht. Dabei ist nach Aussage der Autoren die Mitwirkung während des gesamten Prozesses der Entwicklung, angefangen bei der Arbeitsplatzanalyse bis hin zur Entscheidung über Annahme oder Ablehnung eines

Bewerbers, zu gewährleisten. Weiter sollte die diagnostische Situation ausreichend transparent sein und damit für den Teilnehmer einsichtig und „augenscheinvalid“ sowie die Beurteilungsmaßstäbe nachvollziehbar.

Unter diesen Gesichtspunkten ist es zur Erhöhung der sozialen Validität biographischer Fragebogen insbesondere wichtig, Betroffene, d.h. die Bewerber und Interessenvertreter (Betriebsräte) bei der Entwicklung und Einführung des Auswahlinstrumentes zu beteiligen. Die Darstellung des Berufsbezuges im Fragebogen, die Begründung der Notwendigkeit eines alternativen Vorauswahlinstrumentes und der Überblick über gegebene Erfahrungen mit dem Prädiktor bieten dabei die Grundlage für einen umfassenden Informationsstand, eine hohe Augenscheinvalidität und Transparenz. Durch das Weglassen sozial nicht erwünschter Fragen entsteht bei den Teilnehmern weiter kein Gefühl einer verletzten Privatsphäre.

Wie Kersting (1998) allerdings zu Recht hervorhebt, sollten nur solche Personen befragt werden, welche die zu beurteilenden Verfahren und den Ernstfall von Personalauswahlentscheidungen aufgrund eigener Informationen und aus eigener Erfahrung kennen. Häufig wurden jedoch Studenten als Stichprobe herangezogen und nach der Akzeptanz von Auswahlverfahren befragt. Ihnen fehlen aber in der Regel Kenntnisse über den eignungsdiagnostischen Ernstfall oder die zur Frage stehenden Verfahren aus eigener Erfahrung. Aus diesem Grund sollen in dieser Untersuchungen nur Teilnehmer berücksichtigt werden, die beide Voraussetzungen mitbringen.

Für den Bereich der biographischen Eignungsuntersuchungen liegen als Beleg für die Akzeptanz dieser Verfahren kaum Untersuchungen vor. Wie in Kapitel 2.3 beschrieben, legte Stehle (1983) einer Stichprobe aus dem Versicherungsbereich die Übersetzung des von Baehr und Williams (1967) entwickelten biographischen Fragebogens vor. Von den 151 zur Beurteilung vorgelegten Fragen wurden 90 in einem biographischen Fragebogen für unzulässig betrachtet. Aufgrund der frühzeitigen Beteiligung schon in der Konstruktionsphase, sollte das Ziel dieser Untersuchung sein, eine deutlich höhere Akzeptanz zu erreichen.

Auch wenn eine differenziertere Erfassung der Akzeptanz sinnvoll erscheint, denn wie Kersting (1998) betont, ist die Akzeptanz von Auswahlverfahren nicht unidimensional,

sondern mehrdimensional (s.a. Kapitel 2.3), kann im Rahmen dieser Studie aus ökonomischen und praktischen Gründen nicht umfassend auf diesen Aspekt eingegangen werden. Die dem Schuler'schen Konzept (1990) entnommenen Aspekte der Transparenz und Partizipation sollten aber zumindest als wesentliche Faktoren Berücksichtigung finden.

Gerade weil den späteren Teilnehmern des Auswahlverfahrens schon bei der Entwicklung eines Fragebogenverfahrens ein Mitspracherecht eingeräumt werden sollte, sprach dies gegen die Möglichkeit der Übersetzung existierender biographischer Fragebogen. Zudem fehlen national wie international Fragebogen, die sich in vergleichbaren Kontexten, d.h. der Vorauswahl von Bewerbern für den Polizeivollzugsdienst, bewährt haben. Aus diesen Gründen wurde entschieden, als Basis für die Vorauswahl mit biographischen Daten einen eigenen Fragebogen zu entwickeln und diesen hinsichtlich seiner prädiktiven Validität und der Akzeptanz bei den Betroffenen zu untersuchen.

3.2 Formulierung der Untersuchungsfragen

Aus o.g. Überlegungen und zur Ergänzung der Aussagekraft bestehender Untersuchungen hinsichtlich der prädiktiven Validität von biographischen Fragebogen in anderen beruflichen Bereichen sowie der Notwendigkeit hinsichtlich der Akzeptanz eines neu konzipierten Fragebogens als eignungsdiagnostisches Auswahlinstrument, ergaben sich für die vorliegende Untersuchung folgende Fragestellungen:

- 1) Erfassung und Bewertung der prognostischen Validität eines biographischen Fragebogens:
Durch die Anwendung eines biographischen Fragebogens lassen sich die Gruppen erfolgreiche und nicht erfolgreiche Teilnehmer im Auswahlverfahren statistisch signifikant vorhersagen.
- 2) Erfassung der Stabilität (Retest-Reliabilität) eines biographischen Fragebogens:
Der entwickelte biographische Fragebogen ist als Grundlage für die Erfassung der Validität ausreichend reliabel (stabil).
- 3) Erfassung und Bewertung der sozialen Validität (Akzeptanzuntersuchung) eines biographischen Fragebogens:
Die Anwendung des zu entwickelnden biographischen Fragebogens als Vorauswahlinstrument wird durch Teilnehmer des Auswahlverfahrens akzeptiert.

4 Untersuchungsplan

Für die Untersuchung der in Kapitel 3 dargestellten Fragestellungen ergab sich folgender Plan für die Durchführung:

Zum Zeitpunkt der Untersuchung lag ein für die Untersuchungszwecke geeigneter biographischer Fragebogen nicht vor. Zudem ist eine Voraussetzung zur Gewährleistung der sozialen Validität gegeben, wenn den später von dem Selektionsinstrument betroffenen Teilnehmern des Auswahlverfahrens schon bei dessen Konstruktion ein Mitspracherecht eingeräumt werden kann. Der zur Vorauswahl im Rahmen der eignungsdiagnostischen Untersuchung deshalb zu konzipierende biographische Fragebogen ist zum Einsatz innerhalb einer bestimmten Population - Bewerbern um einen Ausbildungsplatz bei der Polizei - gedacht. Grundsätzlich ist eine Generalisierung der Ergebnisse nicht angestrebt, gleichwohl wurde die Gewährleistung allgemeiner testtheoretischer Rahmenbedingungen bei der Konstruktion des Fragebogens berücksichtigt.

Die Konzeption des biographischen Bewerberfragebogens für den Polizeivollzugsdienst (BfP), die Durchführung der Untersuchung und abschließende Validierung sowie Bestimmung der Retest-Reliabilität und Akzeptanzuntersuchung erfolgte nach allgemeinen testtheoretischen Grundsätzen, unter Berücksichtigung der Hauptkriterien psychologischer Diagnostik: Objektivität, Reliabilität und Validität.

Zur Konstruktion des Fragebogens und zur Beantwortung der Fragestellungen dieser Untersuchung war es Voraussetzung, möglichst mit der späteren Zielpopulation vergleichbare Gruppen zu bilden. Aus diesem Grund wurden für die Konstruktion des biographischen Fragebogens Bewerber herangezogen, die an einem aktuellen Auswahlverfahren für eine Einstellung in den Polizeivollzugsdienst teilnahmen (Stichprobe 1). Für den zweiten Teil der Untersuchung wurde auf der Grundlage der in der ersten Stichprobe erhobenen Daten ein biographischer Fragebogen faktorenanalytisch konstruiert. Für die anschließende Validierung des neu konzipierten Fragebogens (BfP) wurden Teilnehmer eines ein Jahr später durchgeführten Auswahlverfahrens herangezogen (Stichprobe 2). Diese zweite Gruppe war im Laufe der Untersuchung in erfolgreiche und nicht erfolgreiche Teilnehmer des Auswahlverfahrens zu unterteilen. Zur Bestimmung der Retest-Reliabilität und der sozialen Validität wurden wiederum Probanden aus vergleichbaren Populationen gewonnen (Stichprobe 3).

Da es wünschenswert war, gerade zur Erschließung der sozialen Validität kritische Anmerkungen zu erfassen, wurde für Stichprobe 3 nicht auf Teilnehmer am Auswahlverfahren zurückgegriffen, sondern auf bereits eingestellte Auszubildende, die das gleiche Auswahlverfahren ein Jahr zuvor durchlaufen hatten. Dieser Rückgriff auf ausschließlich zuvor erfolgreiche Teilnehmer sollte sozial erwünschte Antworten bei Einschätzungen der Fragen vermeiden und kritische Anmerkungen aufgrund eigener gemachter Erfahrungen ermöglichen. Um die Motivation bei der Bewertung der biographischen Items zu erhöhen, wurde den Probanden die Bedeutung des späteren Vorauswahlinstrumentes erklärt und die Hintergründe, bspw. die derzeit unzureichende Vorauswahl aufgrund von Schulnoten, dargelegt. Weiter wurden die Teilnehmer dieser Stichprobe dahingehend motiviert, indem ihnen deutlich gemacht wurde, dass das Vorauswahlverfahren dazu beitragen wird, ihre späteren Berufskollegen auszuwählen.

Die gleiche Stichprobe 3 wurde auch zur Bestimmung der Retest-Reliabilität herangezogen. Zum ersten Messzeitpunkt (t1) fand die Ersthebung der Daten (Beantwortung der vorläufigen Endversion des BfP) und zum zweiten Messzeitpunkt (t2) – acht Wochen später - die Zweiterhebung. An letztere schloss sich dann die Erhebung zur Erschließung der sozialen Validität an, um mögliche Erinnerungseffekte nicht zu verstärken. Aufgrund sich während der Untersuchung verändernder interner Rahmenbedingungen innerhalb der Personalvertretung war leider eine ursprünglich zusätzlich geplante und abgestimmte gleichzeitige Durchführung der Akzeptanzuntersuchung bei Personalratsvertretern nicht mehr möglich. Jedoch wurde der Personalrat bei der Auswahl der Items für die Vorversion (Stichprobe 1) beteiligt.

Die Teilnehmerzahl in den unterschiedlichen Untersuchungsphasen (Datenerhebung zur Fragebogenkonstruktion, Validierung und Erschließung der sozialen Validität sowie Bestimmung der Retest-Reliabilität) richtete sich nach den im Rahmen der Arbeit gegebenen Möglichkeiten und den zur Anwendung statistischer Verfahren notwendigen Voraussetzungen.

Die Durchführung der Untersuchung mit einer biographischen Itemsammlung, die als Vorversion des späteren Fragebogens zusammengestellt wurde, fand mit der Stichprobe 1 1998 statt. Ca. 1 Jahr später wurde der Stichprobe 2 die vorläufige Endversion (BfP) zur Validierung des Fragebogens in gleicher Form vorgegeben, wie sie zukünftig für die Vorauswahl interessierter Bewerber gedacht ist. Im Zeitraum dieser Erhebung wurde der

gleiche Fragebogen der Gruppe Auszubildender (Stichprobe 3) zur Bestimmung der Retest-Reliabilität an zwei unterschiedlichen Messzeitpunkten vorgegeben. Im Anschluss an den zweiten Messzeitpunkt wurde die Erhebung zur Bestimmung der sozialen Validität durchgeführt. Nach Abschluss der Datenerhebung in der zweiten Stichprobe im November 2000 wurde die Bestimmung der prädiktiven Validität auf der Grundlage des von England (1961, 1971) entwickelten Ansatzes (Weighted application blank approach) durchgeführt. An der Untersuchung zur Validierung nahmen 2/3 der Teilnehmer der Stichprobe 2 teil, 1/3 der Teilnehmer derselben Stichprobe wurden zur Bestimmung der Kreuzvalidierung randomisiert ausgewählt.

Als **abhängige Variablen** ergaben sich in dieser Untersuchung:

- ⇒ Prädiktive Validität der Netto- und Grobgewichte des BfP in der Stichprobe
- ⇒ Prädiktive Validität der Netto- und Grobgewichte des BfP in der Kreuzstichprobe
- ⇒ Stabilität des Ankreuzverhaltens (Retest-Reliabilität)
- ⇒ Akzeptanz einzelner Items des BfP (Soziale Validität)

Als **unabhängige Variablen** ergaben sich in dieser Untersuchung:

- ⇒ Ankreuzverhalten erfolgreicher vs. nicht erfolgreicher Testteilnehmer in der Stichprobe und Kreuzstichprobe
- ⇒ Itemgewichtung in den Extremgruppen
- ⇒ unterschiedliche Messzeitpunkte (t1 und t2)

In der folgenden Übersicht (Tabelle 1.4) wird der Untersuchungsablauf zusammengefasst dargestellt:

Tab. 1.4: Darstellung des Untersuchungsablaufes

	Stichprobe 1	Stichprobe 2			Stichprobe 3
Biographische Itemsammlung Vorversion	Beantwortung des Fragebogens	kein Treatment			kein Treatment
Biographischer Fragebogen Endversion (BfP)	kein Treatment	Beantwortung des Fragebogens			kein Treatment
Erfassung des Ergebnisses im Eignungstest	kein Treatment	Extremgruppe Erfolg	Extremgruppe kein Erfolg	neutrale Gruppe	kein Treatment
Biographischer Fragebogen Endversion (BfP) Netto- und Grobgewichte	kein Treatment	Gewichtung der Items durch Antwortalternativen der erfolgreichen Teilnehmer	kein Treatment	kein Treatment	kein Treatment
Biographischer Fragebogen Endversion (BfP) Validierung	kein Treatment	Stichprobe: 2/3 der Teilnehmer Bestimmung der prädiktiven Validität anhand des Kriteriums „Testergebnis“.	Kreuzstichprobe: 1/3 der c Bestimmung der prädiktiven Validität anhand des Kriteriums „Testergebnis“.	kein Treatment	
Biographischer Fragebogen Endversion (BfP) t1	kein Treatment	kein Treatment			Beantwortung des Fragebogens
Biographischer Fragebogen Endversion (BfP) t2	kein Treatment	kein Treatment			Beantwortung des Fragebogens
Biographischer Fragebogen Endversion Soziale Validität	kein Treatment	kein Treatment			Bewertung der Items des Fragebogens

5 Durchführung der Untersuchung

Für die Durchführung der Untersuchung war, wie aus Kapitel 3 und 4 abzuleiten, folgende Vorgehensweise geplant:

Im ersten Schritt wurde 1998 der Stichprobe 1 die Itemsammlung in Form einer Vorversion eines biographischen Fragebogens vorgegeben. Aufgrund der Ergebnisse der Analyse und faktorenanalytischen Auswertung wurde anschließend der endgültige biographische Bewerberfragebogen - Bewerberfragebogen für den Polizeivollzugsdienst (BfP) - entwickelt. Im folgenden Jahr (1999) bekam die Stichprobe 2 die so erstellte Endversion des BfP zur Validierung vorgegeben, wie sie auch zukünftig für die Auswahl interessierter Bewerber gedacht ist. Zum Abschluss der Datenerhebung in der Stichprobe 2 wurde die Validität des BfP nach dem Ansatz des „Weighted application blank approach“ von England (1961, 1971) erschlossen. Im ersten Schritt wurden hierzu Extremgruppen von den im Testverfahren erfolgreichen bzw. nicht erfolgreichen Teilnehmern gebildet. Dem schloss sich eine Gewichtung des Ankreuzverhaltens der jeweiligen Teilnehmer an. Im letzten Schritt erfolgte die Validierung des BfP. Hierzu wurden Daten von 2/3 der Teilnehmer randomisiert zur Validierung herangezogen, die Daten des verbliebenen Drittels zur Bestimmung der Kreuzvalidität genutzt. Im gleichen Erhebungszeitraum wurde eine identische Version des BfP einer Gruppe Auszubildender (Stichprobe 3) zur Bestimmung der Retest-Reliabilität zu zwei unterschiedlichen Messzeitpunkten vorgelegt. Im Anschluss an den zweiten Messzeitpunkt wurde mit derselben Gruppe die Erhebung zur Bestimmung der sozialen Validität (Akzeptanzuntersuchung) durchgeführt.

Eine Übersicht der Stichproben ist in Tabelle 1.5 dargestellt:

Abb. 1: Übersicht der Stichproben zum Ablauf der Untersuchung

Stichprobe 1	Stichprobe 2	Stichprobe 3
Konstruktion eines biographischen Fragebogens	Validierung des konstruierten biographischen Fragebogens (BfP)	Bestimmung der Retest-Reliabilität und Sozialen Validität des Fragebogens

5.1 Beschreibung der Stichproben

Die beiden Hauptstichproben 1 und 2 setzen sich aus Teilnehmern der Auswahlverfahren in den Jahren 1999 (Stichprobe 1) und 2000 (Stichprobe 2) zusammen. Die Vergleichbarkeit beider Stichproben mit der zur Auswahl vorgesehenen Population darf als gegeben angenommen werden (vgl. a. Stichprobenbeschreibung). Die Teilnahme war für die Bewerber freiwillig. Sie wurden jedoch ausdrücklich darauf hingewiesen, dass die Beantwortung des Fragebogens erwünscht sei, Bestandteil der Bewerbung ist und ihre Antworten u.U. Gegenstand des zum Einstellungsverfahren gehörenden Interviews sein können. So konnte insbesondere für die Validierung der Endversion eine Vergleichbarkeit mit der späteren Verwendung zur Vorauswahl von Bewerbern gewährleistet werden. Einzelne nicht vollständig ausgefüllte Fragebogen wurden von der Untersuchung ausgeschlossen.

5.1.1 Stichprobe 1: Untersuchung zur Fragebogenkonstruktion

Insgesamt nahmen im Jahr 1999 616 Bewerber an der Erhebung für die Vorversion des Fragebogens teil. Dabei handelte es sich um Teilnehmer, die sich um eine Einstellung zum August des Jahres 2000 beworben hatten.

28 % (n = 172) Teilnehmer kamen aus Hamburg, die restlichen 72 % (n = 444) aus allen anderen deutschen Bundesländern, mit Ausnahme des Saarlandes, aus dem keine Bewerbungen vorlagen. Die Altersstruktur der Bewerber lag zwischen dem 15. und dem 40. Lebensjahr, dabei hatten über 90 % der Bewerber das 30. Lebensjahr noch nicht vollendet. Insgesamt 80 % waren jünger oder gleich 25 Jahre, der Altersmittelwert lag bei 20,94 Jahren. Die beschriebene Altersverteilung ergibt sich aus der Tatsache, dass es sich um Bewerbungen um eine Ausbildungsplatz handelt, d.h. eher von jüngeren Teilnehmern am Verfahren auszugehen ist. 599 Teilnehmer bewarben sich um Einstellung in die Schutzpolizei, 47 um Einstellung in die Wasserschutzpolizei. Es nahmen insgesamt 398 Männer und 218 Frauen an der Erhebung teil. Die Schulbildung der Teilnehmer ist Tabelle 1.5-1 zu entnehmen:

Tab. 1.5-1: Schulbildung der Stichprobe 1

Schulbildung	Anzahl	%
Hauptschulabschluss	55	8,9
Realschulabschluss	458	74,4
Fachhochschulreife	49	8,0
Hochschulreife	32	5,2
andere	22	3,6
n = 616		

* in der Regel dem Realschulabschluss gleichgestellte Abschlüsse von Bewerbern ausländischer Herkunft oder aus den neuen Bundesländern (Polytechnische Oberschule)

5.1.2 Stichprobe 2: Validierung des BfP

An der zweiten Erhebung zur Validierung der Endversion des Fragebogens (BfP) nahmen im Jahr 2000 579 Personen teil. Auch hier handelte es sich wieder um Bewerber um eine Einstellung, diesmal für den Einstellungstermin August des Jahres 2001.

24,4 % (n = 141) der Teilnehmer kamen aus Hamburg, 75,6 % (n = 438) aus den anderen deutschen Bundesländern. Auch hier war einzige Ausnahme das Saarland, aus dem keine Bewerbungen vorlagen. Das Alter der Bewerber lag zwischen dem 15. und dem 40. Lebensjahr, dabei hatten über 96 % der Bewerber das 30. Lebensjahr noch nicht vollendet. Insgesamt 83,1 % waren jünger oder gleich 25 Jahre, der Altersmittelwert dieser Stichprobe lag bei 20,9 Jahren. 534 Personen bewarben sich um Einstellung in die Schutzpolizei, 45 um Einstellung in die Wasserschutzpolizei. Es nahmen insgesamt 364 Männer und 215 Frauen an der Erhebung teil, die, wie in Tabelle 2.5-1 dargestellt, folgende Schulbildung hatten:

Tab. 2.5-1: Schulbildung der Stichprobe 2

Schulbildung	Anzahl	%
Hauptschulabschluss	47	8,1
Realschulabschluss	432	82,7
Fachhochschulreife	64	11,1
Hochschulreife	35	6,0
andere	1	0,2
n = 579		

* in der Regel dem Realschulabschluss gleichgestellte Abschlüsse von Bewerbern ausländischer Herkunft oder aus den neuen Bundesländern (Polytechnische Oberschule)

5.1.3 Stichprobe 3: Bestimmung der Retest-Reliabilität und Akzeptanzuntersuchung

Die dritte Datenerhebung wurde zur Bestimmung der Retest-Reliabilität und der sozialen Validität (Akzeptanz) des BfP mit Auszubildenden durchgeführt, die ein Jahr zuvor das gleiche Auswahlverfahren erfolgreich durchlaufen hatten und eingestellt wurden.

Da es wünschenswert war, gerade für die Erschließung der sozialen Validität kritische Anmerkungen zu erfassen und nur Personen zu beteiligen, die über konkrete Erfahrungen mit dem Auswahlverfahren verfügen (s.a. Kapitel 2.3), wurde nicht auf Teilnehmer am Auswahlverfahren zurückgegriffen, sondern auf bereits eingestellte Auszubildende, die das gleiche Auswahlverfahren ein Jahr zuvor durchlaufen hatten. So konnten zudem sozial erwünschte Antworten bei Einschätzungen der biographischen Items vermieden werden. Die Motivation der Probanden wurde erhöht, indem die Bedeutung des späteren

Vorauswahlinstrumentes erklärt und die Hintergründe dargelegt wurden. Den Teilnehmern wurde deutlich gemacht, dass das Vorauswahlverfahren dazu beitragen wird, ihre späteren Berufskollegen auszuwählen.

In der ersten Phase nahmen im Zeitraum Mai – Juli 2000 insgesamt 54 Auszubildende an der Untersuchung zur Bestimmung der Retest-Reliabilität teil. Das Alter der Teilnehmer lag zwischen dem 18. und dem 31. Lebensjahr, dabei hatten über 94,4 % der Teilnehmer das 30. Lebensjahr noch nicht vollendet. Insgesamt waren 79,6 % jünger gleich 25 Jahre, der Altersmittelwert lag bei 23 Jahren. Es handelte sich hierbei um 47 Auszubildende der Schutzpolizei und 7 der Wasserschutzpolizei. Insgesamt nahmen 34 Männer und 20 Frauen an der Erhebung teil. Die Schulbildung wurde bei dieser Untersuchung nicht noch einmal erhoben, da davon auszugehen war, dass sie der Verteilung der Bewerber der letzten Jahre entsprach.

Für die zweite Phase zur Bestimmung der sozialen Validität wurde auf die gleiche Stichprobe zurückgegriffen. Im Anschluss an die Datenerhebung zum Messzeitpunkt t₂ bekamen die Teilnehmer einen weiteren Fragebogen zur Erfassung der Akzeptanz des BfP vorgegeben.

5.2 Ablauf der Untersuchung

Während der ersten Datenerhebung bekamen die Probanden (Stichprobe 1) die Vorversion des BfP erst am Tag ihrer Teilnahme am Auswahlverfahren zur Bearbeitung vorgelegt. Um für die Endversion des BfP eine vergleichbare Bearbeitungssituation sicherzustellen, wurde hier ein anderes Vorgehen gewählt. Da die späteren Bewerber den Fragebogen zu Hause ausfüllen würden, wurde der Fragebogen beim Versand den Bewerbungsunterlagen beigelegt und die Bewerber gebeten, den BfP gemeinsam mit ihren weiteren Bewerbungsunterlagen einzureichen. Die spätere Zuordnung für die Auswertung wurde ermöglicht, indem dem Fragebogen die gleiche Erfassungsnummer zugeordnet wurde, wie sie die einzelnen Bewerber im Einstellungsverfahren bekamen.

5.2.1 Erster Abschnitt: Fragebogenkonstruktion

Die Entwicklung der Fragebogenvorversion fand zwischen Ende 1998 und Frühjahr 1999 statt.

5.2.1.1 Konzeption einer Fragebogenvorversion

Bei der Konzeption der Vorversion waren die besonderen Bedingungen bei der Verwendung von Fragebogen im Öffentlichen Dienst zu antizipieren und zu berücksichtigen. Der Einsatz von Personalfragebogen bedarf grundsätzlich der Zustimmung des Personalrates, des Personalamtes sowie des Datenschutzbeauftragten. Aufgrund dieser Rahmenbedingungen fanden die Empfehlungen von Schuler (1996) und Schuler et al. (1993) besondere Berücksichtigung, welche die Autoren für die Konstruktion biographischer Fragebögen im Hinblick auf deren Akzeptanz geben. Danach sollten sensible Bereiche wie Fragen zu Familie, finanzieller Lage und Gesundheit der Probanden grundsätzlich nicht aufgenommen werden. Fragen mit Anforderungs- und Leistungsbezug werden nach Meinung der Autoren dagegen im allgemeinen gut akzeptiert. Dieser Empfehlung wurde auch bei der Konzeption des BfP entsprochen.

Eine weitere Orientierung zur konzeptionellen Vorgehensweise ergab sich aus früheren Untersuchungen (bspw. Hollmann, 1991; Jäger et al. 1973, 1976; Stehle, 1983; Strasser, 1975) sowie aus dem Vorschlag von Schuler (1993, 1996), der für die Konstruktion biographischer Fragebogen die in Tabelle 1.5-2 dargestellten Schritte empfiehlt:

Tab. 1.5-2: Schritte zur Konstruktion eines biographischen Fragebogens nach Schuler et al. (1993)

Schritte bei der Konstruktion eines biographischen Fragebogens	
1.	Wie soll der Fragebogen in der Potentialbeurteilung eingesetzt werden?
2.	Klärung der Erfolgskriterien und Formulierung der Zielkriterien für erfolgreiche Mitarbeiter
3.	Entscheidung über den Konstruktionsansatz: Summen-Score-Modell oder faktorenanalytisches Modell
4.	Ermittlungen der Anforderungen und ihres biographischen Kontexts
5.	Analyse von Fluktuationsgründen und biographischen Kontextfaktoren
6.	Fragebogenkonstruktion
	- Sammlung von Itemkonzepten
	- Itemformulierung
	- Itemauswahl
	- Akzeptanzüberprüfung durch Befragung einer relevanten Gruppe
	- Erhebung an einer möglichst großen Stichprobe
	- Statistische Bereinigung der Fragebogens
	- Kreuzvalidierung an einer zweiten Stichprobe
	- Entwicklung der Kriterien für „Erfolg“
	- Validierung und Bewährungskontrolle

Zur Reflektion der methodischen Vorgehensweise wurde auch die Entwicklung von Fragebogen in anderen Kontexten berücksichtigt. Beispiele fanden sich u.a. in der klinischen (Ehlers, Stangier, Dohn & Gieler, 1993; Melfsen, Florin & Walter, 1999; Pook, Tuschen-Caffier, Schnapper, Speiger, Krause & Florin, 1999) oder sozialpsychologischen Forschung (Guski, Wichmann, Rohrman & Funke, 1978). Zu grundlegenden Prinzipien der klassischen Testtheorie und zur Fragebogenkonstruktion, soweit in dieser Arbeit nicht weiter auf sie

eingegangen wird, sei auf Bortz und Döring (1995), Lienert (1989), Lienert und Raatz, (1994), Mummendey (1987) und Tränkle (1983) verwiesen.

Der erste Abschnitt der Konstruktion war im Sinne Schulers (1993) die Sammlung von Itemkonzepten, Itemformulierung und abschließender Itemauswahl. Auf der Grundlage einer inhaltlichen Orientierung an bisher bereits Verwendung findenden biographischen Fragebogen, wurden zur weiteren Itemsammlung qualitative Interviews mit erfahrenen Polizeibeamten durchgeführt. Dies war insbesondere wichtig, um den speziellen Kontext und die Akzeptanz zu berücksichtigen, für den die spätere Verwendung des Fragebogens gedacht war. Vor endgültiger Fertigstellung der Vorversion wurden die Items und Itemkonzepte noch einmal mit Polizeibeamten und Auszubildenden abschließend diskutiert und bei fehlender Akzeptanz oder Verständlichkeit entfernt bzw. angepasst.

Bei Entwicklung der Skalenstufen wurde auf die Verwendung einer mittleren Antwortkategorie verzichtet. Nach Mummendey (1987) ergeben sich bei einer mehrstufigen Antwortskala mit ungerader Kategorienzahl für den Probanden eher Schwierigkeiten als Vorteile. Vorteile wären in der möglichen Interpretation der Antworthäufigkeit einzelner Items, d.h. wie häufig das Item angekreuzt wurde und dem damit verbundenen Interesse oder Desinteresse an der Fragestellung zu sehen. Jedoch sollten Vor- und Endversion des Fragebogens für die spätere Vergleichbarkeit der Durchführungsbedingungen bzw. im Sinne der Durchführungsobjektivität die gleiche Form aufweisen. Für die Endversion war es zudem wegen der beiden zu bildenden Extremgruppen nützlich, dass die Probanden sich für ein Urteil entscheiden, d.h. keine mittlere Antwortkategorie vorgegeben wird.

Da die Auswertung der Endversion zur Vorauswahl im realen Bewerbungskontext dienen sollte, konnte den Teilnehmern schon im Vorverfahren keine Anonymität zugesichert werden. Tränkle (1983) verweist jedoch hierzu auf Untersuchungen von Kepes & True (1967) und Fuller (1974), die bei der Sichtung verschiedener Ergebnisse die Annahme unterstützen, dass insbesondere bei gegebener Zusicherung der Vertraulichkeit der Datenverwendung, der Einfluss von Nicht-Anonymität auf das Antwortverhalten eher befürchtet als real sei. Insofern wurde es im gesamten Verfahren als ausreichend erachtet, den Teilnehmern die Vertraulichkeit mündlich (Stichprobe 1) bzw. schriftlich (Stichprobe 2) zu versichern.

Bevor der erstellte Item - Pool der Stichprobe 1 vorgegeben und anschließend empirischen und statistischen Prozeduren unterzogen wurde, fand eine weitere Revision der gesammelten biographischen Fragen statt. Mummendey (1987, S.63) verweist auf Edwards (1957), der eine Reihe von allgemeinverbindlichen Regeln für die Formulierung von Feststellungen in Fragebogen vorlegte. An dieser Stelle der Fragebogenkonstruktion erfolgte deshalb ein letztes Mal eine Überprüfung auf gute Verständlichkeit, notwendige Kürze, Prägnanz der Formulierungen, Art der Fragerichtung und die Reihenfolge der Darbietung der Items. Die Vorversion wurde unterschiedlichen Personen (Psychologen der Polizei, Mitarbeitern des Polizeivollzugsdienstes, Teilnehmer der während der Konzeptionsphase laufenden Auswahlverfahren und Auszubildenden im Polizeivollzugsdienst) vorgelegt. Änderungsvorschläge und Bearbeitungspräferenzen wurden abschließend in die Vorversion eingearbeitet. Letzter Schritt vor Druck der Vorversion war eine Abstimmung unterschiedlicher Layout-Versionen. Diese wurden zufällig ausgewählten Teilnehmern eines laufenden Einstellungsverfahrens vorgelegt. Die von den Teilnehmern mit deutlicher Mehrheit bevorzugte Version ging letztendlich in Druck.

Tabelle 2.5-2 beinhaltet eine Übersicht über die Fragen der endgültigen Vorversion. Sie enthielt insgesamt 192 Items, die vor empirischer Prüfung nach inhaltlichen Kriterien vier unterschiedlichen Kategorien zugeordnet wurden.

Tab. 2.5-2: Übersicht der zugeordneten Items der Vorversion des BfP

1. Allgemeines = 58 Items	
a) demographische / lebensgeschichtliche Inhalte	Anzahl der Items: 9
b) Interessen / Einstellungen	Anzahl der Items: 49
2. Aktivitäten und Freizeit = 40 Items	
a) demographische / lebensgeschichtliche Inhalte	Anzahl der Items: 18
b) Interessen / Einstellungen	Anzahl der Items: 22
3. Familie und Jugend = 15 Items	
a) demographische / lebensgeschichtliche Inhalte	Anzahl der Items: 4
b) Interessen / Einstellungen	Anzahl der Items: 11
4. Schule / Beruf = 79 Items	
a) demographische / lebensgeschichtliche Inhalte	Anzahl der Items: 29
b) Interessen / Einstellungen	Anzahl der Items: 50

Um frühzeitigen Einwänden hinsichtlich der Akzeptanz des Fragebogens zu entgegnen, fanden schon vor der ersten Datenerhebung Gespräche mit Vertretern des Personalrates statt. Die beabsichtigte Untersuchung wurde vorgestellt und die geplante Verwendung des Fragebogens erläutert. Nach Meinung des Verfassers war es im Sinne einer angestrebten hohen Akzeptanz schon zu diesem frühen Zeitpunkt wichtig, gegenüber dem Personalrat die Transparenz des Verfahrens zu signalisieren, ihn zu beteiligen und damit die Möglichkeit eines Einwandes gegen einzelne Items oder die Gesamtintentionen des Verfahrens zu

eröffnen. Für die Vorgabe des Fragebogens als Vorversion wurden, unter der Bedingung, dass die Teilnahme freiwillig ist und ausschließlich für die Untersuchung notwendige Daten nur kurzfristig gespeichert werden, keine Einwände vorgebracht. Diese Vorbedingung wurde in der Untersuchung gewährleistet.

Die Datenerhebung für die Vorversion des BfP fand vom 13.05. – 19.10.1999 statt. Den Teilnehmern wurde mit der entsprechenden Instruktion, unter Wahrung stabiler Durchführungsbedingungen, die Vorversion am Morgen ihrer jeweiligen Teilnahme am laufenden Auswahlverfahren vorgelegt. In der Instruktion wurden die Teilnehmer insbesondere darum gebeten, spontan und zügig möglichst jede Frage zu beantworten. Weiter wurde der Hinweis aufgenommen, dass es keine richtigen oder falschen Antworten gibt und die Antworten u.U. Gegenstand des Vorstellungsgespräches im späteren Auswahlverfahrens sein können.

5.2.1.2 Auswertung und Analyse der Fragebogenvorversion

Ziel der Itemanalyse des Fragebogens war zu entscheiden, welche Items beibehalten und in die spätere Endversion des biographischen Fragebogens aufgenommen werden. Traditionelle Vorgehensweise ist dabei nach Mummendey (1987) die Prüfung des Schwierigkeitsgrades und der Trennschärfe jedes Items.

Der Schwierigkeitsindex einer Aufgabe ist gleich dem prozentualen Anteil P , der auf dieser Aufgabe entfallenden richtigen Antworten zur Analytestichprobe von der Größe N . Damit liegt der Schwierigkeitsindex bei schwierigen Aufgaben niedrig, bei leichten Aufgaben hoch. Eine sinnngemäße Übertragung auf Fragebogen, bei denen es weder richtige noch falsche Antworten gibt, berücksichtigt die Tendenz der Probanden zu zustimmenden vs. ablehnenden, bzw. bejahenden oder verneinenden Antworten (Lienert, 1989).

Hintergrund ist die Auslese und Eliminierung solcher Fragen und Feststellungen, die von fast allen Teilnehmern bzw. keinem im Sinne des zu messenden Konstruktes bejaht werden, d.h. zur Unterscheidung zwischen verschiedenen Individuen nicht brauchbar sind. Als optimal wird ein mittlerer Schwierigkeitsindex von $p = 50$ angesehen, konventionell werden Items mit einem Schwierigkeitsindex von kleiner $p = 20$ und größer $p = 80$ aus dem Fragebogen entfernt.

Der Trennschärfekoeffizient wird nach Lienert (1989) als Korrelationskoeffizienten zwischen der Aufgabenantwort - der richtigen oder falschen - und dem Rohwert definiert. Dieser Grad des Zusammenhanges zwischen der richtigen oder falschen Antwort und dem Gesamtscore gibt an, wie stark jedes einzelne Item den Endwert bereits vorhersagt. Übertragen auf den Fragebogen gibt der Trennschärfekoeffizient den Grad des Zusammenhanges zwischen der Bejahung und dem Gesamtscore an. Ein hoher Trennschärfekoeffizient besagt, dass sich an einer Frage „gute und schlechte“ Personen deutlich unterscheiden, da die „guten“ die Frage bspw. bejahen, die „schlechten“ sie dagegen verneinen oder auslassen. Grundsätzlich sind möglichst hohe Trennschärfen erstrebenswert. Werte zwischen $r = .3$ und $r = .5$ gelten als mittelmäßig, Werte größer $r = .5$ als hoch zu bezeichnen.

5.2.1.3 Fragebogenendversion (BfP)

Im ersten Schritt nach der Datenerhebung wurde eine Häufigkeitsanalyse über alle Items gerechnet. Items mit einem Schwierigkeitsindex kleiner $p = 20$ und größer $p = 80$ wurden entfernt. Bei dichotomen Items wurde zudem eine Standardabweichung von $SD > .35$, bei vierstufigen Items $SD > 1.0$ als wünschenswertes Kriterium zum Verbleib erachtet. Mit den nicht eliminierten Items wurde im zweiten Schritt eine Hauptkomponentenanalyse mit anschließender orthogonaler Varimaxrotation durchgeführt. In der PCA zeigten 30 extrahierte Faktoren einen Eigenwert > 1 und klärten 67,3 % der Gesamtvarianz auf. Dabei war der Varianzbeitrag der einzelnen Faktoren nicht substantiell. Die ersten 4 Faktoren ($\lambda_1 = 5,219$, $\lambda_2 = 4,927$, $\lambda_3 = 2,929$, $\lambda_4 = 2,735$) luden am höchsten, klärten aber zusammen nur 12,3 % der Varianz auf. Die Varianzaufklärungen der Faktoren 5 – 30 lagen im Bereich von 2,7 % und 1,6 %. Hough und Paullin (1994) konzidieren die generelle Heterogenität biographischer Fragebogen. Faktorenanalysen biographischer Itempools fanden i.d.R. eine große Zahl einzelner Komponenten (10 – 30 Faktoren), die jedoch nur einen geringen Prozentsatz der gesamten Varianz aufklären (zwischen 20 und 50 %). Angesichts dieser Ergebnisse muss zwar auch bei dem BfP eine nur mäßige Datenreduktion eingeräumt werden, jedoch hinsichtlich der über das übliche Maß hinausgehenden Varianzaufklärung von 67,3 % bleibt der Anteil nicht erklärter Varianz und damit Verlust an Informationen akzeptabel. Die nicht auf einem der 30 Faktoren ladenden Items wurden entfernt. Abschließend wurde eine Reliabilitätsanalyse (Modell Alpha) pro Faktor gerechnet. Als ausreichende Trennschärfe für den BfP wurde $r = .25$ angenommen. Für die vorläufige Endform ergaben sich 12 Skalen, die sich inhaltlich sinnvoll beschreiben ließen, mit insgesamt 43 Items.

Die statistischen Kennwerte sowie die Anzahl der Items der Skalen sind in Tabelle 3.5-2 zusammengefasst.

Tab.3.5-2: Statistische Kennwerte der Reliabilitätsanalyse

	Skala	Mittelwert	Standardabweichung	Trennschärfe	Cronbach's Alpha	Anzahl der Items
1	Vorhandenes Berufsbild Polizei	9,3	2,2	0,48	0,69	4
2	Sportliche Freizeitaktivitäten	24,8	3,9	0,35	0,67	8
3	Soziales und intellektuelles Engagement	13,5	2,2	0,43	0,67	5
4	Allgemeine Lebens- und Berufserfahrung	6,4	2,2	0,50	0,69	3
5	Spezielle Berufserfahrung	5,2	2,1	0,35	0,54	3
6	Gesellschaftliches Ansehen	11,1	1,7	0,34	0,55	4
7	Frühe finanzielle Selbstständigkeit	6,9	2,2	0,34	0,53	3
8	Passives Freizeitverhalten	7,7	1,3	0,36	0,54	3
9	Aktives Freizeitverhalten	6,1	1,2	0,34	0,52	3
10	Gemachte Erfahrungen für den Berufswunsch	6,4	2,3	0,33	0,51	3
11	Entspannendes Freizeitverhalten	4,8	1,2	0,49	0,66	2
12	Negative Erfahrungen im Beruf oder Leben	4,2	1,2	0,36	0,53	2

* Mittelwert der Trennschärfe aller Items einer Skala

Die interne Konsistenz nach Cronbach's Alpha des Fragebogens lag zwischen $\alpha = .51$ für die Skala „Gemachte Erfahrungen für den Berufswunsch“ und $\alpha = .69$ für die Skalen „Allgemeine Lebens- und Berufserfahrung“ sowie „Vorhandenes Berufsbild Polizei“. Neben seiner Abhängigkeit von der Interkorrelation der Items (Alpha ist umso höher, je höher die Item-Interkorrelation ist), hängt die Höhe des Alpha-Koeffizienten von der Anzahl der Items pro Skala ab. Je mehr Items pro Skala, desto höher tendenziell Cronbach's Alpha. Angesichts der teilweise geringen Itemzahl kann die durchschnittliche Höhe für den Zweck der Verwendung des Fragebogens als ausreichend erachtet werden. Durch Testverlängerung würde nach Lienert (1989, S. 244) schon eine Erhöhung der Itemzahl von 2 auf 3 (Skala 11) und 2 auf 4 (Skala 12) eine Verbesserung der Reliabilität auf $r = .70$ bringen.

Neben den 43, wurden weitere 13 Items, die entweder in der Hauptkomponentenanalyse auf keinem Faktor luden oder in der Reliabilitätsanalyse eliminiert wurden, nach inhaltlichen Gesichtspunkten und aufgrund ihrer Häufigkeitsverteilung als geeignet zur Diskrimination für die angestrebten Extremgruppen bewertet und zusätzlich in den Fragebogen aufgenommen.

Die Antwortmöglichkeiten der Items sind unterschiedlich codiert. 19 Items enthalten jeweils die vier Antwortstufen 1 = gar nicht, 2 = kaum, 3 = ziemlich, 4 = außerordentlich. Bei 14 Items konnten die vier Antwortstufen 1 = eher nie, 2 = eher selten, 3 = eher oft und 4 = eher immer gewählt werden. 10 Items beinhalteten Antwortstufen wie 1 = keine, 2 = 1 – 3 und 3 = 4 oder mehr. Die restlichen 13 Items sind nominalskaliert, wobei hier die Probanden zwischen vorgegebenen Kategorien zu wählen haben. Die Aufnahme in die Endversion des Fragebogens geschah aufgrund inhaltlicher Überlegungen, da sich in der Häufigkeitsverteilung eine deutliche Trennung zwischen unterschiedlichen Antwortalternativen abzeichnete. Die Darstellung des Fragebogens und die Häufigkeitsverteilungen der berücksichtigten Items finden sich im Anhang A.

Letztendlich fand für die endgültige Version des BfP eine Reduktion der 192 Items auf 56 statt. Der Fragebogen wurde aus Gründen der Vergleichbarkeit zur Vorversion wieder in ähnliche Bereiche unterteilt. Nach einer einleitenden Instruktion und vier anschließenden demographischen Angaben, die nicht in die Auswertung einfließen und nur zum „warm up“ gedacht sind, enthält der Teil „Allgemeines“ insgesamt 13 Items. Es folgen die Bereiche „Aktivitäten und Freizeit“ mit 23 und „Schule und Beruf“ mit 20 Items.

5.2.2 Zweiter Abschnitt: Validierung der Fragebogenendversion (BfP)

5.2.2.1 Durchführung zur Validierung der Fragebogenendversion

Die mit Teilnehmern durchgeführte Validierung des Fragebogens (Stichprobe 2) fand im Zeitraum vom 24.04.2000 bis 14.11.2000 statt. Gemeinsam mit den Bewerbungsunterlagen bekamen die Bewerber den Fragebogen ca. 12 – 16 Wochen vor Beginn des Auswahlverfahrens per Post übersandt. Sie wurden gebeten, den Fragebogen zu beantworten und diesen gemeinsam mit den weiteren Bewerbungsunterlagen zurückzusenden. Um eine bessere Vergleichbarkeit mit der Echtsituation zu gewährleisten, wurde den Teilnehmern in der Instruktion zum Fragebogen mitgeteilt, dass ihre Angaben Gegenstand des zum Auswahlverfahren gehörenden Vorstellungsgespräches sein können. Dies mag ein Grund mit sein, dass der Rücklauf der Fragebögen bei 97,4 % lag. Um die Zuordnung der Fragebogen zu den Ergebnissen des Einstellungstestverfahrens zu ermöglichen, wurden im Fragebogen die gleiche Erfassungsnummer der Bewerber aufgenommen, wie sie auch obligatorisch den weiteren Bewerbungsunterlagen zugeordnet war.

Für die später notwendige Bildung der Extremgruppen (England, 1961,1971; vgl.a. Kapitel 2.2.2.1) wurden im ersten Schritt die Ergebnisse der Teilnehmer am Einstellungsverfahren ausgewertet.

Die Grundlage für die Extremgruppenbildung waren die erreichten Ergebnisse der Bewerber der Stichprobe 2 in den einzelnen Testabschnitten des Einstellungsverfahrens. Die Testteile wurden in Kapitel 2.1.3.2 beschrieben. In der Tabelle 4.5-2 findet sich eine Übersicht, aus der die Ausscheidungsgründe der Teilnehmer am Einstellungsverfahren deutlich werden.

Tab. 4.5-2: Übersicht der Ausscheidungsgründe der Teilnehmer der Validierungsstichprobe

Ausscheidungsgrund	Teilnehmer	Prozent
1. Diktat	191	33,0 %
2. Bericht	45	7,8 %
3. Leistungstest	54	9,3 %
4. Sporttest	51	8,8 %
5. Vorstellungsgespräch	12	2,1 %
6. Sonstige Gründe	2	0,3 %
7. Nichterreichen der Mindestpunkte	81	14,0 %
erfolgreich im Gesamtverfahren	143	24,7 %
	n = 579	

Bei der Bewertung der Ausscheidungsgründe sind die Auswirkungen des sequentiellen Auswahlverfahrens zu berücksichtigen. Für die einzelnen Verfahrensteile (Diktat, Bericht etc.) sind Mindestanforderungen zu erfüllen, um weiter teilnehmen zu können. So schieden bspw. 33 % der Teilnehmer (n = 191) aufgrund ihres im Diktat erreichten Ergebnisses aus dem Verfahren aus. Sie erreichten nicht die Mindestanforderung von 4 Punkten auf einer neunstufigen Skala. In der Folge nahmen in den jeweils aufbauenden Verfahrensteilen immer nur jene Teilnehmer teil, welche die Mindestanforderungen im vorherigen erfüllen konnten. Letztendlich verblieben 143 Teilnehmer, die in allen Testteilen des Auswahlverfahrens erfolgreich waren.

Gemäß der in Kapitel 2.2.2.1 dargestellten Methode des „Weighted application blank approach“ (England, 1961, 1972) war zuerst eine Entscheidung für ein Kriterium zur Extremgruppenbildung zu treffen und anschließend die Teilnehmer in zwei Extremgruppen aufzuteilen. Gemäß der Fragestellung wurden in der Extremgruppe 1 die im Gesamtverfahren erfolgreichen Teilnehmer zusammengefasst. Extremgruppe 2 repräsentierte dagegen jene Teilnehmer, die vom Verfahren aufgrund unzureichender Einzelleistungen ausgeschlossen wurden. Nach inhaltlichen Überlegungen wurden die Ausscheidungsgründe Diktat, Bericht, kognitiver Leistungstest und Vorstellungsgespräch als Kriterium für die Extremgruppe 2 festgelegt. Der Sporttest wurde nicht aufgenommen. Die langjährige Erfahrung zeigte, dass

das Bestehen des Sporttestes abhängig von der speziellen Vorbereitung der Teilnehmer ist. Bei gegebener Wiederholungsmöglichkeit erzielt ein Großteil der Teilnehmer die festgelegten Anforderungen. Eine Wiederholungsmöglichkeit bei den anderen Testteilen ist zur Vermeidung von Lerneffekten jedoch nicht vorgesehen. Eine dritte neutrale Gruppe bildeten die sonstigen Teilnehmer. Sie setzte sich zusammen aus Teilnehmern, die aufgrund eines unzureichenden Ergebnisses im Sporttest, dem Nichterreichen der geforderten Mindestpunktzahl (Summe von 25 Punkte in den Verfahrensteilen Diktat, Bericht, Leistungstest und Sport) oder sonstiger Gründe (z.B. Ermittlungen, Verzicht) aus dem Verfahren ausgeschieden waren. In der Tabelle 5.5-2 findet sich eine übersichtliche Darstellung der Aufteilung in den drei Gruppen.

Tab. 5.5-2: Grundlage der Zuordnung in Extremgruppen und neutrale Gruppe

Extremgruppe 1		Extremgruppe 2	Neutrale Gruppe
Erfolgreicher Abschluss des gesamten Testverfahrens	des	Ausschluss vom Verfahren aufgrund nicht ausreichender Leistungen in einem der folgenden Verfahrensteile: - Diktat - Berichtsfertigung - kognitiver Leistungstest - Vorstellungsgespräch	Ausschluss vom Verfahren aufgrund nicht ausreichender Leistungen im Sporttest, dem Nichterreichen der vorgegebenen Mindestpunktzahl oder sonstiger Gründe wie bspw. Verzicht oder Ermittlungen

Aus dem gewählten Verteilungsschlüssel ergibt sich folgender Anteil von Bewerbern in den drei Gruppen: Extremgruppe 1 = 143 Teilnehmer, Extremgruppe 2 = 302 Teilnehmer und neutrale Gruppe = 134 Teilnehmer.

Von den 143 Teilnehmern der Extremgruppe 1 (erfolgreiche Teilnehmer) kamen 21,0 % (n = 30) aus Hamburg, die anderen 79,0 % (n = 113) Bewerber aus anderen Bundesländern. Das Alter der Bewerber lag zwischen dem 16. und dem 33. Lebensjahr, dabei hatten über 93,7 % der Bewerber das 30. Lebensjahr noch nicht vollendet. Insgesamt 79,7 % waren jünger oder gleich 25 Jahre. Der Altersmittelwert lag bei 22,5 Jahren.

Aus Tabelle 6.5-2 ist die Verteilung der Schulbildung der Teilnehmer ersichtlich. Es bewarben sich 123 Personen um Einstellung in die Schutzpolizei und 20 um Einstellung in die Wasserschutzpolizei. Es handelte sich um 97 Männer und 46 Frauen.

Tab. 6.5-2: Schulbildung der Extremgruppe 1 (erfolgreiche Teilnehmer)

Schulbildung	Anzahl	%
Hauptschulabschluss	3	2,1
Realschulabschluss	93	65,0
Fachhochschulreife	23	16,1
Hochschulreife	23	16,1
andere	1	0,7
	n = 143	

Die 302 Teilnehmer der Extremgruppe 2 (nicht erfolgreiche Teilnehmer) kamen zu 29,8 % (n = 90) aus Hamburg, die restlichen 70,2 % (n = 212) aus anderen Bundesländern. Das Alter lag zwischen dem 15. und dem 40. Lebensjahr, dabei hatten über 96,4 % der Bewerber das 30. Lebensjahr noch nicht vollendet, insgesamt 85,4 % waren jünger oder gleich 25 Jahre. Der Altersmittelwert lag bei 20,1 Jahren. 286 Bewerber bewarben sich um Einstellung in die Schutzpolizei und 16 um Einstellung in die Wasserschutzpolizei. 178 Teilnehmer waren männlich, 124 weiblich. Tabelle 7.5-2 zeigt die Schulbildung der Teilnehmer.

Tab. 7.5-2: Schulbildung der Extremgruppe 2 (nicht erfolgreiche Teilnehmer)

Schulbildung	Anzahl	%
Hauptschulabschluss	34	11,2
Realschulabschluss	237	78,5
Fachhochschulreife	24	7,9
Hochschulreife	7	2,3
andere	0	0
n = 302		

5.2.2.2 Validierungs- und Kreuzvalidierungsstichprobe

Für die Aufteilung von Validierungs- und Kreuzvalidierungsstichprobe wurde folgende Vorgehensweise gewählt: Aufteilung der Stichprobe 2 in jeweils eine Validierungs- und eine Kreuzvalidierungsstichprobe im Verhältnis 2 : 3. Damit wurden jeweils 2/3 der Teilnehmer zur Bestimmung der Validität herangezogen, 1/3 zur Kreuzvalidierung.

In der Tabelle 8.5-2 ist das Ergebnis der Zuordnung der Teilnehmer der Stichprobe 2 (n = 579) zur Validierungsstichprobe und Kreuzstichprobe dargestellt:

Tab. 8.5-2: Ziehung der Validierungsstichprobe und Kreuzstichprobe

Validierungsstichprobe (2/3)	Kreuzvalidierungsstichprobe (1/3)
$\frac{579 * 66,7}{100} = 386,2 = 386$	$\frac{579 * 33,3}{100} = 192,8 = 193$

Die Fragebogennummern aller Teilnehmer der Stichprobe 2 wurden in eine Urne gelegt. Die ersten 193 gezogenen Nummern (Ziehung ohne Zurücklegen) wurden der Kreuzvalidierungsstichprobe zugeordnet.

5.2.2.3 Validierung- und Kreuzvalidierung des Fragebogens

Wie unter Kapitel 2.2.2.1 ausführlich beschrieben, ist das bei Einsatz von Extremgruppen häufigste Verfahren der Ansatz des „Weighted application blank approach“ von England

(1961, 1971). Hier wird jedes Item aufgrund seiner Kriteriumsvalidität ausgesucht und entsprechend gewichtet. In der Itemanalyse wird für jede Antwortalternative, die einem Prädiktor entspricht, anhand der Prognosegüte ein empirisches Gewicht ermittelt.

Im Sinne der Erschließung der prädiktiven Validität wurden zum ersten Messzeitpunkt (t1) von den Teilnehmern das Antwortverhalten im Fragebogen und zum zweiten Messzeitpunkt (t2) ihr erreichtes Testergebnis im Auswahlverfahren erfasst. Im ersten Schritt wurde für jede Antwortalternative der Prozentsatz der jeweiligen Extremgruppe errechnet (s.a. 5.2.2.1), welche die vorgegebene Alternative wählte. Aus dem späteren Zweck des Fragebogens, dem Einsatz als Vorauswahlinstrument, ergab sich das zu wählende Kriterium. Es wurde durch das erreichte Testergebnis beider Extremgruppen operationalisiert. Um die Vergleichbarkeit mit der späteren Auswahl-situation zu gewährleisten, wurde den Teilnehmern zu t1 der Fragebogen mit den weiteren Bewerbungsunterlagen nach Hause gesandt. Das Kriterium zu t2 wurde bei Teilnahme am Auswahlverfahren ca. 12 - 16 Wochen nach der Bewerbung erhoben. Als Grundlage der Validierung wurde anschließend das Antwortverhalten der Teilnehmer gemäß den Vorgaben im „Weighted application blank approach“ (England, 1961, 1972) ausgewertet sowie die Prozentwertdifferenzen errechnet. In der Folge wurden die Prozentwertdifferenzen in Netto- und Grobgewichte umgerechnet. Items mit unzureichender Trennschärfe wurden aus dem Fragebogen entfernt bzw. in der Auswertung nicht berücksichtigt. Der eigentliche Prädiktorwert des biographischen Fragebogens wurde ermittelt, indem die von den Teilnehmern gewählten Antwortalternativen der trennscharfen Items, versehen mit dem jeweiligen Punktwert, aufsummiert werden. Die Gewichte wurden dann im nächsten Schritt mit dem Kriterium, d.h. dem jeweiligen Testergebnis der Teilnehmer im Auswahlverfahren, korreliert. Die gleiche Vorgehensweise wurde analog in der Kreuzvalidierungsstichprobe umgesetzt.

5.2.3 Dritter Abschnitt: Akzeptanzuntersuchung und Retest-Reliabilität

Die mit den Auszubildenden im Polizeivollzugsdienst (Stichprobe 3) durchgeführte dritte Erhebung zur Bestimmung der Retest-Reliabilität und sozialen Validität des Fragebogens (BfP) fand am 10.05.00 (t 1) und 06.07.00 (t 2 und soziale Validität) statt. Die Teilnehmer der Untersuchung hatten ein Jahr zuvor das gleiche Auswahlverfahren durchlaufen. Auf diese Weise konnte der häufig bestehenden Kritik an gewählten Stichproben begegnet werden. Es wurden nur Personen befragt, welche die zu beurteilenden Verfahren und den Ernstfall von

Personalauswahlentscheidungen aufgrund eigener Informationen und aus eigener Erfahrung kennen. Zum Zeitpunkt der Untersuchung befanden sich die Auszubildenden im ersten Jahr ihrer Ausbildung und sind daher mit potenziellen Bewerbern aufgrund ihrer relativ kurz zurückliegenden Erfahrungen im Auswahlverfahren relativ gut vergleichbar.

5.2.3.1 Durchführung zur Bestimmung der Retest-Reliabilität

Der Stichprobe 3 wurde zu den beiden Messzeitpunkten jeweils die Originalversion des BfP vorgegeben. Um die Zuordnung der Fragebogen zu beiden Messzeitpunkten (t_1 und t_2) zu ermöglichen, wurden die Teilnehmer gebeten, an vorgesehener Stelle ein Codewort zu verwenden. Aufgrund falscher Angaben bzw. fehlenden Möglichkeit der Zuordnung konnten insgesamt nur 47 der 54 Fragebögen ausgewertet werden. Somit konnte den Teilnehmern der Stichprobe 3 die Anonymität ihrer Angaben zugesichert werden. Um Erinnerungseffekte zu vermeiden, lag ein Zeitabstand von 8 Wochen zwischen den Datenerhebungen der ersten (t_1) und zweiten Vorgabe (t_2). Dieser Zeitraum erschien angesichts der unter Kapitel 5.2.1.3 dargestellten Komplexität des BfP` s mit 56 Items als ausreichend.

5.2.3.2 Durchführung zur Bestimmung der Sozialen Validität

Im Sinne des Konzeptes der sozialen Validität von Schuler et al. (1983) sowie Schuler (1990, 1998) war ein wichtiges Ziel, die Akzeptanz der Items zu erfassen. Damit einher ging die Intention, bei Erschließung der prognostischen Validität, den Fragebogen um die Items zu bereinigen, deren Akzeptanz nicht gegeben ist.

Neben den Teilnehmern der Stichprobe 3 waren zur Bestimmung der sozialen Validität auch Vertreter der Personalräte vorgesehen. Zwischen der Ersterhebung mit der Vorversion und damit verbundener Abstimmung in der Vorgehensweise mit den Personalratsvertretern und der Fertigstellung der Endversion, ergab sich durch Neuwahlen der Personalvertretung eine Situation, die eine Beteiligung des Personalrates verhinderte. Aus diesem Grund wurde versucht, weitere Auszubildende aus parallelen Ausbildungsgruppen zu gewinnen. Insgesamt konnten 11 zufällig ausgewählte Pbn geworben werden, so dass an der Bestimmung der sozialen Validität insgesamt 65 Personen beteiligt werden konnten.

Um bewerten zu können, in wie weit sich die Neukonzeption und damit einhergehend, die frühe Beteiligung von Betroffenen bei der Gestaltung des Fragebogens auswirken wird, wurde der Fragebogen zur Erschließung der sozialen Validität an der einzig bekannten Untersuchung in diesem Kontext, der von Stehle (1983), orientiert.

Wie in Tabelle 9.5-2 beispielhaft dargestellt, wurde den Teilnehmern der Untersuchung der BfP mit folgenden zusätzlichen Kategorien vorgegeben.

Tab. 9-5-2: Beispiel aus dem Fragebogen zur Erfassung der sozialen Validität

Die nachfolgenden Fragen würde ich in einem Bewerberfragebogen betrachten als:	
Wie wichtig ist Ihnen, sich im Sozialbereich außerhalb von Schule / Beruf zu engagieren?	zulässig unzulässig
Wie wichtig ist Ihnen, sich in der Politik außerhalb von Schule / Beruf zu engagieren?	zulässig unzulässig

Ein vergleichendes Kriterium für ausreichende Akzeptanz war die Höhe der Akzeptanz des Fragebogens von Stehle (1983). Da aber Vorgaben hinsichtlich der grundsätzlichen Höhe des Akzeptanzgrades fehlen bzw. kaum möglich sind (s.a. Kapitel 2.3), wurde in dieser Untersuchung nach inhaltlichen Überlegungen der Mindestgrad der Zustimmung abgewogen. Hierbei fanden sowohl die Anzahl der Teilnehmer an der Bestimmung der sozialen Validität sowie die Bedeutung der Bewertung der Akzeptanz einzelner Fragen ihre Berücksichtigung. Aus diesem Grund sollten nicht mehr als 15 % der Teilnehmer eine Frage als unzulässig für die Vorauswahl von Bewerbern bewerten. Um diesem hohen Anspruch hinsichtlich der erwünschten Akzeptanz gerecht werden zu können, wurde schon bei der Zusammenstellung der Fragen für die Vorversion darauf geachtet, keine Fragen zu wählen, die möglicherweise Akzeptanzprobleme bereiten könnten.

Beispielhaft sind die Fragebogen zur Bestimmung der Retest-Reliabilität und der sozialen Validität im Anhang A dargestellt.

6 Ergebnisdarstellung

6.1 Validierung des Biographischen Fragebogens (BfP)

6.1.1 Netto- und Grobgewichtung der Antwortalternativen in den Extremgruppen

Wie von England (1961, 1971) vorgegeben, wurden im ersten Schritt die Prozentwertdifferenzen zwischen den Extremgruppen „erfolgreiche Teilnehmer am Testverfahren“ (Extremgruppe 1 = 143 Teilnehmer) und „nicht erfolgreiche Teilnehmer am Testverfahren“ (Extremgruppe 2 = 302 Teilnehmer) bestimmt. Gemäß der von Strong (1926) und England (1961, 1971) vorgegebenen Gewichtungen wurden zum Vergleich die Netto- und Grobgewichte bestimmt (s. a. Kapitel 2.2 und 5).

Eine Übersicht der einzelnen trennscharfen Items, mit ihren Netto- bzw. Grobgewichten, ist in Tabelle 1.6-1 dargestellt. Vollständig sind die Ergebnisse der Gewichtung im Anhang B dargestellt.

Damit weisen insgesamt 16 der 56 Items des Fragebogens eine ausreichend hohe Trennschärfe für die Diskrimination zwischen den Gruppen erfolgreiche bzw. nicht erfolgreiche Testteilnehmer auf.

6.1.2 Validierung des BfP in der Stichprobe

Für die Validierung wurden die Items mit unzureichender Trennschärfe in der Auswertung nicht berücksichtigt. Der eigentliche Prädiktorwert des biographischen Fragebogens wurde ermittelt, indem die von den Teilnehmern gewählten Antwortalternativen der gewichteten 16 trennscharfen Items aufsummiert wurden.

Tab. 1.6-1: Items mit hohen Netto- und Grobgewichten im BfP

Extremgruppe N = 143: erfolgreiche TN N = 302: nicht erfolgreiche TN	erfolgreich		nicht erfolgreich		Prozent- differenz	Netto- gewicht	Grob- gewicht
	absolute Häufigkeiten	relative Häufigkeiten	absolute Häufigkeiten	relative Häufigkeiten			
4:Wie wichtig ist Ihnen ein großer Freundeskreis für Ihre Lebensplanung?							
kaum	39	41,1	49	4,4	16,7	4	2
8:Wie häufig verdienten Sie parallel zur Schule eigenes Geld?							
eher oft	50	52,6	47	23,5	29,1	7	2
11:Wie häufig waren Sie im Verkauf beschäftigt?							
eher oft	28	29,5	22	10,9	18,6	4	2
12:Wie häufig waren Sie in Verwaltung / Büro beschäftigt?							
eher immer	6	6,3	3	1,5	4,8	3	1
13:Wie häufig waren Sie in sonstigen Bereichen beschäftigt?							
eher oft	49	51,6	39	19,6	32,0	7	2
15:In wie viel Vereinen / Gruppen sind oder waren Sie Mitglied?							
3 oder mehr	39	41,1	46	22,9	18,2	4	2
17:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Wochenzeitungen?							
1 bis 3	68	71,6	97	48,3	23,3	5	2
19:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Büchern?							
4 oder mehr	37	38,9	21	10,4	28,5	7	2
22:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Hobbies?							
4 oder mehr	77	81,1	129	64,2	16,9	4	2
27:Wie wichtig ist Ihnen das Interessengebiet Wirtschaft?							
ziemlich	55	57,9	80	40,0	17,9	4	2
30:Wie wichtig ist Ihnen die Freizeitaktivität Lesen?							
außerordentlich	19	20,0	10	5,0	15,0	5	2
42:Hatten Sie parallel zur Schule / Ausbildung / Beruf Funktionen inne?							
über einen längeren Zeitraum	40	42,1	52	25,9	16,2	4	2
43:Wie wichtig war für Sie das Vorhandensein von freien Stellen bei dem Wunsch, den Polizeiberuf zu ergreifen?							
kaum	43	45,3	45	22,5	22,8	5	2
48:Wie wichtig war für Sie ein Praktikum bei der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?							
gar nicht	58	61,1	82	41,0	20,1	4	2
52:Welche Vorstellungen von Teamarbeit teilen Sie am ehesten?							
verspricht eher Erfolg als Einzelarbeit	66	69,5	101	50,2	19,3	4	2
56:Wie häufig haben Sie Erfahrungen mit Widerstand gegen neue Ideen in Schule / Ausbildung / Beruf gemacht?							
eher oft	35	36,8	37	18,5	18,3	4	2

Die Gewichte wurden dann im nächsten Schritt mit dem definierten Kriterium, d.h. dem jeweiligen Testendergebnis der Teilnehmer im Auswahlverfahren, korreliert. Die gleiche Vorgehensweise wurde analog in der Kreuzvalidierungsstichprobe umgesetzt.

In den Tabelle 2.6-1 und 3.6-1 sind die Validitätskoeffizienten zwischen den von den Teilnehmern in der Validierungsstichprobe (n = 386) erreichten Ergebnissen im Auswahlverfahren und der jeweiligen Nettogewichte und Grobgewichte des BfP dargestellt.

Da der im Rahmen des Auswahlverfahrens durchgeführte Sporttest auf das Endergebnis einen besonderen Einfluss haben könnte, wurden die Koeffizienten sowohl mit, wie auch ohne Ergebnisse der Sportbewertung berechnet.

Tab. 2.6-1: Validitätskoeffizient Testergebnis (ohne Sporttest) mit Netto- und Grobgewichten des BfP in der Validierungsstichprobe

Korrelationen: Validierungsstichprobe				
		Ergebnis (ohne Sport)	Summe Nettogewichte	Summe Grobgewichte
Ergebnis (ohne Sport)	Korrelation nach Pearson	1,000	,523**	,504**
	Signifikanz (2-seitig)	,	,000	,000
	N	386	386	386
Summe Nettogewichte	Korrelation nach Pearson	,523**	1,000	,969**
	Signifikanz (2-seitig)	,000	,	,000
	N	386	386	386
Summe Grobgewichte	Korrelation nach Pearson	,504**	,969**	1,000
	Signifikanz (2-seitig)	,000	,000	,
	N	386	386	386

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Tab. 3.6-1: Validitätskoeffizient Endergebnis mit Netto- und Grobgewichten des BfP in der Validierungsstichprobe

Korrelationen: Validierungsstichprobe				
		Endergebnis	Summe Nettogewichte	Summe Grobgewichte
Endergebnis	Korrelation nach Pearson	1,000	,522**	,502**
	Signifikanz (2-seitig)	,	,000	,000
	N	386	386	386
Summe Nettogewichte	Korrelation nach Pearson	,522**	1,000	,969**
	Signifikanz (2-seitig)	,000	,	,000
	N	386	386	386
Summe Grobgewichte	Korrelation nach Pearson	,502**	,969**	1,000
	Signifikanz (2-seitig)	,000	,000	,
	N	386	386	386

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Die errechneten Validitätskoeffizienten zwischen Testergebnis mit und ohne Sporttest unterscheiden sich nicht signifikant voneinander (z-Wert der Nettogewichte = 0,0190346; z-Wert der Grobgewichte = 0,037052). Der Sporttest hat keinen statistisch signifikanten Einfluss.

Damit konnte, bezogen auf das Gesamtergebnis in der Validierungsstichprobe, für den BfP eine prädiktive Validität von

für die Nettogewichte: $r = .522$ und

für die Grobgewichte: $r = .502$

erschlossen werden.

6.1.3 Validierung des BfP in der Kreuzstichprobe

In den Tabelle 4.6-1 und 5.6-1 sind die Validitätskoeffizienten zwischen den von den Teilnehmern in der Kreuzstichprobe ($n = 193$) erreichten Ergebnissen im Auswahlverfahren und der jeweiligen Nettogewichte und Grobgewichte des BfP dargestellt.

Tab. 4.6-1: Validitätskoeffizient Testergebnis (ohne Sporttest) mit Netto- und Grobgewichten des BfP in der Kreuzstichprobe

Korrelationen: Kreuzvalidierung				
		Ergebnis (ohne Sport)	Summe Nettogewichte	Summe Grogewichte
Ergebnis (ohne Sport)	Korrelation nach Pearson	1,000	,582**	,559**
	Signifikanz (2-seitig)	,	,000	,000
	N	193	193	193
Summe Nettogewichte	Korrelation nach Pearson	,582**	1,000	,976**
	Signifikanz (2-seitig)	,000	,	,000
	N	193	193	193
Summe Grobgewichte	Korrelation nach Pearson	,559**	,976**	1,000
	Signifikanz (2-seitig)	,000	,000	,
	N	193	193	193

** : Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Tab. 5.6-1: Validitätskoeffizient Endergebnis mit Netto- und Grobgewichten des BfP in der Kreuzstichprobe

Korrelationen: Kreuzvalidierung				
		Endergebnis	Summe Nettogewichte	Summe Grogewichte
Endergebnis	Korrelation nach Pearson	1,000	,579**	,558**
	Signifikanz (2-seitig)	,	,000	,000
	N	193	193	193
Summe Nettogewichte	Korrelation nach Pearson	,579**	1,000	,976**
	Signifikanz (2-seitig)	,000	,	,000
	N	193	193	193
Summe Grobgewichte	Korrelation nach Pearson	,558**	,976**	1,000
	Signifikanz (2-seitig)	,000	,000	,
	N	193	193	193

** : Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Auch hier unterscheiden sich die errechneten Validitätskoeffizienten zwischen Testergebnis mit und ohne Sporttest nicht signifikant voneinander (z-Wert der Nettogewichte = 0,04410198; z-Wert der Grobgewichte = 0,0141648). Der Sporttest hat keinen statistisch signifikanten Einfluss.

Damit konnte, bezogen auf das Gesamtergebnis in der Kreuzstichprobe, für den BfP eine prädiktive Validität von

für die Nettogewichte: $r = .579$ und

für die Grobgewichte: $r = .558$

erschlossen werden.

6.2 Bestimmung der Reliabilität des BfP

Wie in Kapitel 5.2.3.1 dargestellt, nahmen insgesamt 47 Teilnehmer der Stichprobe 3 an beiden Erhebungen (t1 und t2) teil.

Nach Mummendey (1987) können für Fragebogen grundsätzlich Reliabilitätskoeffizienten um $r = .80$ als gut gelten. In der vorliegenden Arbeit werden Retest-Reliabilitäten von $r \geq .70$ als adäquate Reliabilität angenommen.

Eine Übersicht einzelner Items mit adäquater Reliabilität findet sich in Tabelle 1.6-2. Vollständig sind die Ergebnisse zur Bestimmung der Reliabilität im Anhang B dargestellt.

Tab. 1.6-2: Items des BfP mit adäquater Retest-Reliabilität in Stichprobe 3

Biographische Frage (BfP)		r =
2	Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf im Sozialbereich zu engagieren?	.7312
4	Wie wichtig ist Ihnen ein großer Freundeskreis für Ihre Lebensplanung?	.7781
6	Welche Vorstellung von Verantwortung teilen Sie am ehesten?	.7063
8	Wie häufig verdienen Sie parallel zur Schule eigenes Geld?	.7848
9	Wie häufig verdienen Sie während der Ferien eigenes Geld?	.7028
12	Wie häufig waren Sie in Verwaltung / Büro beschäftigt?	.7256
15	In wie viel Vereinen / Gruppen sind oder waren Sie Mitglied?	.7512
19	Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Büchern?	.7234
20	Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Fernsehen?	.7741
23	Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Freundschaften?	.7090
24	Wie oft betreiben Sie Mannschaftssportarten?	.7601
28	Wie wichtig ist Ihnen das Interessengebiet Politik?	.7038
29	Wie wichtig ist Ihnen das Interessengebiet Geschichte?	.7681
30	Wie wichtig ist Ihnen die Freizeitaktivität Lesen?	.7268
32	Wie wichtig sind Ihnen Aktivitäten im Verein als Freizeitaktivität?	.9611
34	Wie häufig sehen Sie im Fernsehen Spielfilme / Serien?	.7457
36	Wie häufig sehen Sie im Fernsehen Sportsendungen?	.7672
39	Wie häufig sprachen Sie mit Anderen über Ihre berufliche Zukunft?	.7401
40	Mit wie viel Jahren haben Sie sich für Ihren jetzigen Berufswunsch entschieden?	.7132
41	Wie viel Jahre sind Sie schon berufstätig?	.8829
42	Hatten Sie parallel zur Schule / Ausbildung / Beruf Funktionen inne?	.7382
43	Wie wichtig war für Sie das Vorhandensein von freien Stellen bei dem Wunsch, den Polizeiberuf zu ergreifen?	.7852
44	Wie wichtig waren für Sie die Verdienstmöglichkeiten bei dem Wunsch, den Polizeiberuf zu ergreifen?	.7330
48	Wie wichtig war für Sie ein Praktikum bei der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?	.7985
54	Welche Zeiteinteilung bevorzugen Sie am ehesten?	.8609

Wie aus Tabelle 1.6-2 zu entnehmen, weisen insgesamt 25 der 56 Items eine adäquate Retest-Reliabilität auf. Die Koeffizienten liegen zwischen $r = .71$ bei der Frage 6: „Welche Vorstellung von Verantwortung teilen Sie am ehesten?“ und $r = .96$ bei der Frage 32: „Wie wichtig sind Ihnen Aktivitäten im Verein als Freizeitaktivität?“

31 Items befinden sich im Bereich einer marginalen Retest-Reliabilität und weisen Korrelationskoeffizienten zwischen $r = .51$ (Frage 5: „Wie wichtig ist ihnen, was andere Menschen über Sie denken?“) und $r = .69$ (Frage 35: „Wie häufig sehen Sie im Fernsehen Dokumentarfilme?“).

Einzig die Frage 22: „Wie viele Stunden verwenden Sie durchschnittlich für Ihre Hobbys?“ hatte mit $r = .44$ nur knapp ausreichende Retest-Reliabilität.

Da Korrelationen keine Maßzahlen sind und dem entsprechend ihre Mittelwerte nicht interpretierbar wären, wurde die durchschnittliche Retest-Reliabilität des BfP durch Transformation in Fishers Z-Werte bestimmt.

Der auf diese Weise bestimmte Korrelationskoeffizient betrug über alle Items $r = .69$. Er liegt nur knapp unter dem Bereich der als adäquat zu bezeichnenden Reliabilität und kann somit als insgesamt zufriedenstellend betrachtet werden.

In Tabellen 2.6-2 und 3.6-2 findet sich eine Übersicht der Reliabilitätskennwerte der 16 trennscharfen Items. 9 der 16 trennscharfen Fragen finden sich in der Gruppe „Items mit adäquater Reliabilität“ wieder. Die verbleibenden 7 Items weisen inadäquate Kennwerte auf.

Tab. 2.6-2: Trennscharfe Items mit adäquater Retest-Reliabilität

Trennscharfe Items mit adäquater Reliabilität		r =
2	Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf im Sozialbereich zu engagieren?	.7312
8	Wie häufig verdienten Sie parallel zur Schule eigenes Geld?	.7848
12	Wie häufig waren Sie in Verwaltung / Büro beschäftigt?	.7256
15	In wie viel Vereinen / Gruppen sind oder waren Sie Mitglied?	.7512
19	Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Büchern?	.7234
30	Wie wichtig ist Ihnen die Freizeitaktivität Lesen?	.7268
42	Hatten Sie parallel zur Schule / Ausbildung / Beruf Funktionen inne?	.7382
43	Wie wichtig war für Sie das Vorhandensein von freien Stellen bei dem Wunsch, den Polizeiberuf zu ergreifen?	.7852
48	Wie wichtig war für Sie ein Praktikum bei der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?	.7985

Tab. 3.6-2: Trennscharfe Items mit nicht adäquater Retest-Reliabilität

Trennscharfe Items mit nicht adäquater Reliabilität		r =
11	Wie häufig waren Sie im Verkauf beschäftigt?	.6133
13	Wie häufig waren Sie in sonstigen Bereichen beschäftigt?	.6518
17	Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Wochenzeitungen?	.5552
22	Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Hobbies?	.4394
27	Wie wichtig ist Ihnen das Interessengebiet Wirtschaft?	.6419
52	Welche Vorstellungen von Teamarbeit teilen Sie am ehesten?	.5385
56	Wie häufig haben Sie Erfahrungen mit Widerstand gegen neue Ideen in Schule / Ausbildung / Beruf gemacht?	.6426

Auch hier wurde wieder die durchschnittliche Retest-Reliabilität der trennscharfen Items nach Transformation in Fishers Z-Werte bestimmt. Der mittlere Korrelationskoeffizient für trennscharfe Items betrug ebenfalls $r = .69$.

6.3 Bestimmung der sozialen Validität des BfP

Vollständig sind die Ergebnisse zur Bestimmung der sozialen Validierung im Anhang B dargestellt. Es nahmen insgesamt 65 Teilnehmer an der Akzeptanzuntersuchung teil.

Aufgrund der in Kapitel 5 dargestellten inhaltlichen Überlegungen wurden die Items des BfP von den Teilnehmern der Stichprobe 3 in Anlehnung an die Untersuchung von Stehle (1983) als zulässig oder unzulässig für die Personalauswahl im Polizeivollzugsdienst bewertet. Da sich in der Literatur keine theoretischen Vorgaben finden lassen, die als Grundlage für eine ausreichend hohe Akzeptanz einer Frage herangezogen werden können bzw. aufgrund dessen in einem Fragebogen entsprechende Items zu eliminieren wären, wurde in eigener Überlegung das Ziel angestrebt, dass grundsätzlich nicht mehr als 15 % der Teilnehmer eine Frage als unzulässig bewerten sollten. Die Fragen mit absoluter Zustimmung (Akzeptanz) sind in Tabelle 1.6-3 dargestellt, die mit eher geringer Zustimmung in Tabelle 2.6-3.

Tab. 1.6-3: Items des BfP mit absoluter Zustimmung

Items (BfP) bewertet als		zulässig		unzulässig	
		n	%	n	%
1	Wie wichtig ist Ihnen, was andere Menschen über Sie denken?	65	100,0	--	--
21	Wie viel Stunden verwenden Sie durchschnittlich für Sport?	65	100,0	--	--
24	Wie oft betreiben Sie Mannschafts- oder Einzelsportarten?	65	100,0	--	--
29	Wie wichtig ist Ihnen das Interessengebiet Geschichte?	65	100,0	--	--
31	Wie wichtig sind Ihnen TV oder Radio als Freizeitaktivität?	65	100,0	--	--
52	Welche Vorstellung von Teamarbeit teilen Sie am ehesten?	65	100,0	--	--

Tab. 2.6-3: Items des BfP mit eher geringer Zustimmung

Items (BfP) bewertet als		zulässig		unzulässig	
		n	%	n	%
8	Wie häufig verdienen Sie parallel zur Schule eigenes Geld?	57	87,7	8	12,3
36	Wie häufig sehen Sie im Fernsehen Sportsendungen?	56	86,2	9	13,8
49	Wie wichtig waren für Sie Verwandte bei der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?	57	87,7	8	12,3
51	Welche Art der Bekanntgabe besonderer Leistungen bevorzugen Sie am ehesten?	55	84,6	10	15,4

Lediglich Frage 51 (Welche Art der Bekanntgabe besonderer Leistungen bevorzugen Sie am ehesten?) fand nicht die gewünschte Akzeptanz in der Stichprobe und unterschritt die geforderte Mindestzustimmung von 85% knapp. 15,4 % (n = 10) der Teilnehmer bewerteten die Frage in einem biographischen Bewerberfragebogen als für die Vorauswahl unzulässig. Auch wenn 55 (84,6%) Teilnehmer die Frage für zulässig erachteten, sollte die Frage im Sinne einer erwünschten hohen sozialen Validität aus dem Fragebogen entfernt werden. Da die Frage jedoch, wie aus Tabelle 1.6-1 ersichtlich, hinsichtlich ihrer Gewichtung keine Trennschärfe zwischen erfolgreichen und nicht erfolgreichen Teilnehmer im Testverfahren

besitzt, wurde das Item in der Berechnung des Summenscores nicht berücksichtigt. Das Item kann für zukünftige Anwendungen des Fragebogens ohne direkten Einfluss auf die prädiktive Validität entfernt werden. Eine korrigierende Berechnung des Validitätskoeffizienten ist somit nicht notwendig. Die Zustimmung bei den 16 trennscharfen Items ist in Tabelle 3.6-3 dargestellt.

Tab. 3.6-3: Akzeptanz trennscharfer Items des BfP

Trennscharfe Items wurden bewertet als		zulässig		unzulässig	
		n	%	n	%
2	Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf im Sozialbereich zu engagieren?	63	96,9	2	3,1
8	Wie häufig verdienten Sie parallel zur Schule eigenes Geld?	57	87,7	8	12,3
11	Wie häufig waren Sie im Verkauf beschäftigt?	60	94,3	5	7,7
12	Wie häufig waren Sie in Verwaltung / Büro beschäftigt?	61	93,8	4	6,2
13	Wie häufig waren Sie in sonstigen Bereichen beschäftigt?	60	94,3	5	7,7
15	In wie viel Vereinen / Gruppen sind oder waren Sie Mitglied?	59	90,7	6	9,2
17	Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Wochenzeitungen?	61	93,8	4	6,2
19	Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Büchern?	63	96,9	2	3,1
22	Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Hobbies?	63	96,9	2	3,1
27	Wie wichtig ist Ihnen das Interessengebiet Wirtschaft?	60	94,3	5	7,7
30	Wie wichtig ist Ihnen die Freizeitaktivität Lesen?	64	98,5	1	1,5
42	Hatten Sie parallel zur Schule / Ausbildung / Beruf Funktionen inne?	64	98,5	1	1,5
43	Wie wichtig war für Sie das Vorhandensein von freien Stellen bei dem Wunsch, den Polizeiberuf zu ergreifen?	59	90,7	6	9,2
48	Wie wichtig war für Sie ein Praktikum bei der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?	61	98,5	4	6,2
52	Welche Vorstellungen von Teamarbeit teilen Sie am ehesten?	65	100,0	--	--
56	Wie häufig haben Sie Erfahrungen mit Widerstand gegen neue Ideen in Schule / Ausbildung / Beruf gemacht?	59	90,7	6	9,2

Die durchschnittliche Zustimmung bei allen Items des Fragebogens lag bei 95,24 % , die durchschnittliche Akzeptanz der 16 trennscharfen Items bei 94,78 %.

6.4 Beantwortung der Hypothesen

Aufgrund der unter Punkt 6.1 – 6.3 dargestellten Ergebnisse lassen sich die unter Kapitel 3 aufgeworfenen Fragestellungen wie folgt beantworten:

1. Zur prognostischen Validität des konzipierten biographischen Fragebogens ist festzustellen, dass sich mit 16 von 56 Items des BfP die Gruppen erfolgreiche und nicht erfolgreiche Teilnehmer im Auswahlverfahren statistisch signifikant vorhersagen lassen. Hinsichtlich der Nettogewichte ließ sich eine prognostische Validität von $r = .52$ und hinsichtlich der Grobgewichte von $r = .50$ bestimmen. Beide für eine Personalauswahl außerordentlich zufriedenstellende Ergebnisse konnten auch in der Kreuzstichprobe bestätigt werden (Nettogewichte: $r = .58$, Grobgewichte: $r = .56$). Eine besondere Berücksichtigung des Sporttestes ist bei der Anwendung des Fragebogens nicht

notwendig. Der Sporttest hatte weder in der Hauptstichprobe noch in der Kreuzstichprobe einen signifikanten Einfluss auf das Ergebnis.

2. Der durch Fishers Z- Transformation bestimmte durchschnittliche Retest-Reliabilitätskoeffizient zeigt mit $r = .69$ eine ausreichend hohe Stabilität. Dies gilt sowohl für den gesamten biographischen Fragebogen wie auch für die 16 trennscharfen Items.
3. Im Rahmen der Erfassung der Akzeptanz des biographischen Fragebogens kann auf der Grundlage der Ergebnisse konzidiert werden, dass der BfP von den erfahrenen Teilnehmern eines Auswahlverfahrens weitestgehend als Vorauswahlinstrument akzeptiert wird. Die durchschnittliche Zustimmung bei allen Items lag bei 95,24 % , die durchschnittliche Akzeptanz der 16 trennscharfen Items bei 94,78 % und kann damit in beiden Fällen als sehr zufriedenstellend bezeichnet werden. Nur eine einzige Frage erfüllte nicht die Grundanforderung an die gewünschte Höhe der Akzeptanz, d.h. eine Zustimmung von mindestens 85 % der Teilnehmer. Da das Item sich als nicht trennscharf bei der Unterscheidung zwischen im Testverfahren erfolgreichen und nicht erfolgreichen Teilnehmern erwies, kann es ohne Verlust an Prädiktionskraft aus dem Fragebogen entfernt werden.

7 Interpretation und Diskussion

7.1 Zusammenfassung

Ziel der vorliegenden Arbeit war die Erschließung der prädiktiven Validität und Akzeptanz eines neu konzipierten biographischen Fragebogens (BfP = Bewerberfragebogen für den Polizeivollzugsdienst), der alternativ zu Schulnoten als Vorauswahlinstrument für Teilnehmer am Auswahlverfahren für die Einstellung in den Polizeivollzugsdienst Anwendung finden soll.

Der praktische Stellenwert biographischer Daten im Kontext der eignungsdiagnostischen Personalauswahl konnte in der Vergangenheit in unterschiedlichen beruflichen Bereichen dokumentiert werden. In einer Vielzahl von Studien wurde die prädiktive Validität dieses Auswahlinstrumentes belegt. Biographische Daten besitzen nach Schmidt und Hunter (1998) im Vergleich zu Intelligenztests ($r = .51$), Arbeitsproben ($r = .54$), strukturiert geführten Interviews ($r = .51$) oder Assessment Centern ($r = .36$) grundsätzlich eine gute prädiktive Validität. Die Höhe ihrer Kriteriumsvalidität variiert in unterschiedlichen Untersuchungen und in Abhängigkeit vom gewählten Kriterium zwischen $r = .33$ und $r = .52$ (Bliesener, 1992; Ghiselli, 1955; Schmitt, Gooding, Noe & Kirsch, 1984; Schmidt & Hunter, 1998).

Die derzeitige Vorauswahl im Einstellungsverfahren für den Polizeivollzugsdienst wird anhand der Bewerbungsunterlagen und hier in erster Linie durch den Einzelprädiktor Schulnoten bestimmt. Wie in Kapitel 2.1.2.1 dargestellt, sind in den letzten Jahren wiederholt Zweifel an der Nützlichkeit dieses Prädiktors für die Erfassung zukünftiger Leistungen in Ausbildung und Beruf aufgetreten (Althoff, 1986; Hollmann & Reitzig, 1995; Ingenkamp, 1969; Petersen, 2002; Rastetter, 1999; Schuler, 1998).

Die nur sehr geringe inkrementelle Validität biographischer Daten, im Vergleich zu den in vielen Auswahlverfahren standardmäßig durchgeführten Intelligenztests, von $r = .01$ (Schmidt & Hunter, 1998) lässt dieses eignungsdiagnostische Instrument als zusätzlich eingesetzten Datengeber kaum nützlich erscheinen. Deshalb erscheint es nicht sinnvoll, zusätzlich zum Intelligenztest einen biographischen Fragebogen für die Personalauswahl im Polizeivollzugsdienst zu verwenden.

Aus dieser Überlegung leitete sich die Frage ab, inwieweit ein gezielt konzipierter biographischer Fragebogen alternativ als eignungsdiagnostisches Vorauswahlinstrument für Bewerber für den Polizeivollzugsdienst in Frage kommen könnte. Denn nach Seibt und Kleinmann (1991) kann die Effizienz der Personalauswahl nur so gut sein wie dies die Personalvorauswahl zulässt. Bisher liegen keine Untersuchungen vor, die den Einsatz biographischer Fragebogen in diesem Kontext rechtfertigen. Vorliegende biographische Fragebogen wurden häufig nicht für die Personalauswahl entwickelt oder sind nur für eine stark eingeschränkte Population validiert. Eine Generalisierung der erzielten Ergebnisse auf andere Berufsfelder ist kaum möglich und von den Konstrukteuren oft auch nicht beabsichtigt.

Durch eine Neukonstruktion eröffnete sich aber auch die Möglichkeit, vielfach vernachlässigte Aspekte der sozialen Validität bzw. Akzeptanz eines Auswahlinstrumentes (Schuler, 1990) zu berücksichtigen. In Kapitel 2.3 wurde neben der notwendigen Nützlichkeit eines Verfahrens mit hoher Prädiktionskraft die Bedeutung der Akzeptanz von eignungsdiagnostischen Auswahlinstrumenten betont und herausgestellt (Kersting, 1998; Köchling, 1999; Schuler, 1990). Zurzeit liegen allerdings kaum Untersuchungen für den Bereich der biographisch orientierten Eignungsuntersuchung vor.

Um eine angestrebte hohe Akzeptanz zu erreichen, war es insbesondere wichtig, Betroffene, d.h. die Bewerber und Interessenvertreter (Betriebsräte), schon bei der Entwicklung und Einführung des Auswahlinstrumentes zu beteiligen.

Somit galt es im ersten Schritt der Untersuchung, als zukünftigen Prädiktor für die Vorauswahl, einen neuen biographischen Fragebogen zur Vorauswahl von Bewerbern für den Polizeivollzugsdienst zu konstruieren. Für die Konzeption dieser Vorversion waren sowohl die besonderen Bedingungen bei der Verwendung von Fragebogen im Öffentlichen Dienst als auch die Akzeptanzuntersuchung zu berücksichtigen. Bspw. wurden sensible Bereiche wie Fragen zu Familie, finanzieller Lage und Gesundheit der Probanden grundsätzlich nicht aufgenommen. Nach einer inhaltlichen Orientierung wurden aus bestehenden unterschiedlichen biographischen Fragebogen für den beabsichtigten Zweck geeignet erscheinende Items zusammengestellt. Zur ergänzenden Itemsammlung wurden qualitative Interviews mit erfahrenen Polizeibeamten durchgeführt. Letzteres war insbesondere wichtig, um den speziellen beruflichen Kontext und die Akzeptanz zu berücksichtigen, für den die spätere Verwendung des Fragebogens gedacht war. Bevor der erstellte Item - Pool der ersten

Stichprobe vorgegeben und anschließend empirischen und statistischen Prozeduren unterzogen werden konnte, fand eine weitere Revision der gesammelten biographischen Fragen statt. Allgemeinverbindliche Regeln für die Formulierung von Feststellungen wurden beachtet. So erfolgte ein letztes Mal eine Überprüfung auf gute Verständlichkeit, notwendige Kürze, Prägnanz der Formulierungen, Art der Fragerichtung und die Reihenfolge der Darbietung der Items. Vor endgültiger Fertigstellung der Vorversion wurden die Items noch einmal mit Polizeibeamten, Auszubildenden und Personalratsvertretern abschließend diskutiert und bei fehlender Akzeptanz oder Verständlichkeit entfernt bzw. angepasst.

Die endgültige Vorversion des Fragebogens enthielt 192 Items, die vor empirischer Prüfung nach inhaltlichen Kriterien vier Kategorien zugeordnet wurden. 58 Items erfassten allgemeine demographische und lebensgeschichtliche Inhalte, 40 sonstige Aktivitäten und Freizeitverhalten. 15 Fragen befassten sich mit dem Erlebten in Familie und Jugend sowie 79 mit gemachten Erfahrungen in Schule und Beruf.

An der Datenerhebung mit der Vorversion des Fragebogens nahmen im Jahr 1999 616 Bewerber teil. Dabei handelte es sich um Teilnehmer eines zum damaligen Zeitpunkt beginnenden Auswahlverfahrens für die Einstellung in den Polizeivollzugsdienst des kommenden Jahres. Die Probanden bekamen die Vorversion des Fragebogens am Tag ihrer Teilnahme am Auswahlverfahren zur Bearbeitung vorgelegt.

Ziel der anschließenden Itemanalyse war zu entscheiden, welche Fragen beibehalten und in die spätere Endversion des biographischen Fragebogens aufgenommen werden sollten. Im ersten Schritt wurde eine Häufigkeitsanalyse über alle Items gerechnet und Fragen mit einem Schwierigkeitsindex kleiner $p = 20$ und größer $p = 80$ entfernt. Mit den verbliebenen Items wurde im zweiten Schritt eine Hauptkomponentenanalyse mit anschließender orthogonaler Varimaxrotation durchgeführt. In der PCA zeigten 30 extrahierte Faktoren einen Eigenwert > 1 und klärten 67,3 % der Gesamtvarianz auf. Dabei war der Varianzbeitrag der einzelnen Faktoren nicht substantiell. Die ersten 4 Faktoren ($\lambda_1 = 5,219$, $\lambda_2 = 4,927$, $\lambda_3 = 2,929$, $\lambda_4 = 2,735$) luden am höchsten, klärten aber gemeinsam nur 12,3 % der Varianz auf. Die Varianzaufklärungen der Faktoren 5 – 30 lagen jeweils im Bereich von 2,7 % und 1,6 %. Es musste zwar beim vorliegenden Fragebogen eine nur mäßige Datenreduktion eingeräumt werden, angesichts der allgemein bei biographischen Fragebogen über das übliche Maß

hinausgehenden Varianzaufklärung von 67,3 %, bleibt der Anteil nicht erklärter Varianz und damit der Verlust an Informationen akzeptabel.

Abschließend wurde eine Reliabilitätsanalyse (Modell Alpha) pro Faktor gerechnet. Für die vorläufige Endform ergaben sich 12 Skalen mit insgesamt 43 Items, die sich inhaltlich sinnvoll beschreiben ließen. Die interne Konsistenz des Fragebogens nach Cronbach`s Alpha lag zwischen $\alpha = .51$ und $\alpha = .69$. Neben seiner Abhängigkeit von der Interkorrelation der Items (Alpha ist umso höher, je höher die Item-Interkorrelation ist), hängt die Höhe des Alpha-Koeffizienten von der Anzahl der Items pro Skala ab. Je mehr Items pro Skala, desto höher tendenziell Cronbach`s Alpha. Angesichts der teilweise geringen Itemzahl konnte die durchschnittliche Höhe für den Zweck der Verwendung des Fragebogens als ausreichend erachtet werden.

Aufgrund ihrer Häufigkeitsverteilung als geeignet zur Diskrimination für die angestrebten Extremgruppen bewertet, wurden zusätzlich weitere 13 Items nach inhaltlichen Gesichtspunkten in den Fragebogen aufgenommen.

Somit fand für die endgültige Version des biographischen Fragebogens, dem BfP (Bewerberfragebogen für den Polizeivollzugsdienst), eine Reduktion der 192 Items in der Vorversion auf letztendlich 56 Fragen statt. Der Fragebogen wurde aus Gründen der Vergleichbarkeit zur Vorversion wieder in ähnliche Bereiche unterteilt. Nach einer einleitenden Instruktion und vier anschließenden demographischen Angaben, die nicht in die Auswertung einfließen und nur zum „warm up“ gedacht waren, enthält der Teil „Allgemeines“ insgesamt 13 Items. Es folgen die Bereiche „Aktivitäten und Freizeit“ mit 23 und „Schule und Beruf“ mit 20 Items. Die Skalen des BfP sind in Tabelle 1.7-1 dargestellt.

Tab.1.7-1: Skalen und Anzahl der Items im BfP

	Skala	Anzahl der Items
1	Vorhandenes Berufsbild Polizei	4
2	Sportliche Freizeitaktivitäten	8
3	Soziales und intellektuelles Engagement	5
4	Allgemeine Lebens- und Berufserfahrung	3
5	Spezielle Berufserfahrung	3
6	Gesellschaftliches Ansehen	4
7	Frühe finanzielle Selbstständigkeit	3
8	Passives Freizeitverhalten	3
9	Aktives Freizeitverhalten	3
10	Gemachte Erfahrungen für den Berufswunsch	3
11	Entspannendes Freizeitverhalten	2
12	Negative Erfahrungen im Beruf oder Leben	2

Die mit 579 Probanden einer zweiten Stichprobe durchgeführte Untersuchung zur Erschließung der prognostischen Validität des BfP fand 2000 statt. Gemeinsam mit den Bewerbungsunterlagen bekamen die Bewerber des aktuell beginnenden Auswahlverfahrens für die Einstellung in den Polizeivollzugsdienst des Jahres 2001 den Fragebogen ca. 12 – 16 Wochen vor Beginn des Verfahrens per Post übersandt. Um eine bessere Vergleichbarkeit mit der Echtsituation zu gewährleisten, wurde den Teilnehmern in der Instruktion zum Fragebogen mitgeteilt, dass ihre Angaben Gegenstand des zum Auswahlverfahren gehörenden Vorstellungsgespräches sein können.

Gemäß der gewählten Methode des „Weighted application blank approach“ (England; 1961, 1972) war zur Beantwortung der Fragestellung zu diesem Zeitpunkt die Bildung von Extremgruppen und die Erhebung eines Kriteriums notwendig. Im ersten Schritt wurden deshalb die Ergebnisse der Teilnehmer am Auswahlverfahren ausgewertet. Die Grundlage für die Extremgruppenbildung waren die erreichten Ergebnisse der Bewerber der Stichprobe 2 in den einzelnen Testabschnitten des Verfahrens. In der Extremgruppe 1 wurden die im Gesamtverfahren erfolgreichen Teilnehmer zusammengefasst. Extremgruppe 2 repräsentierte jene Teilnehmer, die vom Verfahren aufgrund unzureichender Einzelleistungen ausgeschlossen wurden. Eine dritte neutrale Gruppe bildeten die sonstigen Teilnehmer. Sie setzte sich aus Teilnehmern zusammen, die aufgrund eines unzureichenden Ergebnisses im Sporttest, dem Nichterreichen der geforderten Mindestpunktzahl oder sonstiger Gründe aus dem Verfahren ausgeschieden waren. Im Sinne der Erschließung der prädiktiven Validität wurden somit zum ersten Messzeitpunkt (t1) das Antwortverhalten im Fragebogen (Prädiktor) und zum zweiten Messzeitpunkt (t2) das erreichte Testergebnis im Auswahlverfahren (Kriterium) erfasst.

Anschließend wurde für jede Antwortalternative der den Fragebogen beantwortenden Teilnehmer der Prozentsatz der jeweiligen Extremgruppe errechnet, welche die vorgegebene Alternative wählte. Entsprechend dem Ansatz des „Weighted application blank approach“ (England; 1961, 1972) wurden anschließend die Prozentwertdifferenzen gebildet und diese in Netto- und Grobgewichte umgerechnet. Items mit unzureichender Trennschärfe wurden aus dem Fragebogen entfernt bzw. in der Auswertung nicht berücksichtigt. Der eigentliche Prädiktorwert des biographischen Fragebogens wurde ermittelt, indem die von den Teilnehmern gewählten Antwortalternativen der trennscharfen Items, versehen mit dem jeweiligen Punktwert, aufsummiert wurden. Die Gewichte wurden dann im nächsten Schritt

mit dem Kriterium, d.h. dem jeweiligen Testendergebnis der Teilnehmer im Auswahlverfahren, korreliert.

Die gleiche Vorgehensweise wurde analog in einer Haupt- und Kreuzstichprobe umgesetzt. Hierzu wurden für die Untersuchung die teilnehmenden Probanden in Validierungs- und Kreuzvalidierungsstichprobe aufgeteilt. 2/3 der Teilnehmer wurden zur Bestimmung der Validität herangezogen, 1/3 zur Kreuzvalidierung.

In der dritten Erhebung dieser Arbeit wurde mit 65 Auszubildenden des Polizeivollzugsdienstes die Retest-Reliabilität und die soziale Validität (Akzeptanz) des BfP erschlossen. Die Untersuchung fand im Anschluss an die Erhebung in der zweiten Stichprobe statt. Die Teilnehmer hatten ein Jahr zuvor das gleiche Auswahlverfahren durchlaufen. Auf diese Weise konnte der bestehenden Kritik an häufig gewählten Stichproben begegnet werden.

Den Teilnehmern der Stichprobe 3 wurde zu zwei Messzeitpunkten jeweils die Originalversion des BfP vorgegeben. Um Erinnerungseffekte zu vermeiden, lag ein Zeitabstand von 8 Wochen zwischen den Datenerhebungen der ersten (t_1) und zweiten Vorgabe (t_2).

Zur Bewertung der erreichten Akzeptanzhöhe, wurde als für diesen Bereich einzig bekannte vorliegende Kriterium für ausreichende Akzeptanz, das Ergebnis der Akzeptanzbewertung in der Untersuchung von Stehle (1983) herangezogen. Nach inhaltlichen Überlegungen wurde zudem entschieden, dass nicht mehr als 15 % der Teilnehmer eine Frage als unzulässig für die Vorauswahl von Bewerbern bewerten sollten.

Als Fazit der vorliegenden Arbeit kann festgehalten werden, dass sich mit 16 der 56 Items des BfP die Gruppen erfolgreiche und nicht erfolgreiche Teilnehmer im Auswahlverfahren statistisch signifikant vorhersagen lassen. Hinsichtlich der Nettogewichte ließ sich eine prognostische Validität von $r = .52$ und hinsichtlich der Grobgewichte von $r = .50$ bestimmen. Beide für eine Personalauswahl außerordentlich zufriedenstellenden Ergebnisse konnten auch in der Kreuzstichprobe bestätigt werden (Nettogewichte: $r = .58$, Grobgewichte: $r = .56$).

Der durch Fishers Z-Transformation bestimmte durchschnittliche Retest-Reliabilitätskoeffizient zeigt mit $r = .69$ eine ausreichend hohe Stabilität. Damit ist eine ausreichend hohe Reliabilität als Grundlage für die Erschließung der Validität anzunehmen. Zudem kann die Retest-Reliabilität als ein Indiz für die Fälschungssicherheit des BfP gewertet werden. Die Höhe der Stabilität gilt sowohl für den gesamten biographischen Fragebogen wie auch allein für die 16 trennscharfen Items.

Im Rahmen der Erfassung der Akzeptanz des biographischen Fragebogens kann auf der Grundlage der Ergebnisse konzidiert werden, dass der BfP auch von erfahrenen Teilnehmern eines Auswahlverfahren als Vorauswahlinstrument akzeptiert wird. Die durchschnittliche Zustimmung bei allen Items des Fragebogens lag bei 95,24 %, die durchschnittliche Akzeptanz der 16 trennscharfen Items bei 94,78 % und kann damit in beiden Fällen als sehr zufriedenstellend bezeichnet werden.

7.2 Interpretation der Ergebnisse

Gemäß der von Lienert (1989) eingeführten allgemeinen Richtlinien über die erforderliche Höhe von Validitätskoeffizienten, kann die prädiktive Validität des entwickelten biographischen Bewerberfragebogens für den Polizeivollzugsdienst als zufriedenstellend hoch bewertet werden. Nach Schuler (1996) liegt die Höhe eines durchschnittlichen Korrelationskoeffizienten in der Berufseignungsdiagnostik im Bereich von $r = .30$. Ein Wert im Bereich von $r = .50$ ist für die Prognose beruflichen Erfolges schon als sehr gut zu bezeichnen. Sowohl in der Hauptstichprobe wie auch der Kreuzstichprobe liegen die Validitätswerte für den BfP über $r = .50$.

Grundsätzlich kommt im eignungsdiagnostischen Kontext der prädiktiven Validität ein entscheidender Stellenwert zu, da nur zeitlich Früheres verantwortlich sein kann für zeitlich Späteres. Insofern spricht besonders in der Berufseignungsdiagnostik vieles für die Nützlichkeit eines Auswahlinstrumentes, wenn dessen ermitteltes Testergebnis den Erfolg im Auswahlverfahren zu einem späteren Zeitpunkt vorherzusagen vermag.

Auch wenn sich nach Jäger (1986) die praktische Verwendbarkeit eines Auswahlinstrumentes nicht ausschließlich an der Höhe des Korrelationskoeffizienten festmacht, kann die hier erschlossene Höhe der Validität des Fragebogens als für den Zweck ausreichend bewertet

werden. Insbesondere die im Sinne des Autors zur weiteren Bewertung heranzuziehenden Abstände zu den alternativ zur Verfügung stehenden Instrumenten unterstreichen die Bedeutung des vorliegenden Ergebnisses. Da die prognostische Validität der gesamten Bewerbungsunterlagen mit $r = .18$ (Reilly & Chao, 1982) als unzureichend erachtet werden muss, findet alternativ die Vorauswahl von Bewerbern um einen Ausbildungsplatz überwiegend mittels des Einzelprädiktors Schulnoten statt. Auch wenn Hinweise über die nicht ausreichende Prädiktionskraft als Ergebnis einzelner Untersuchungen hierbei außer Acht gelassen bleiben (Althoff, 1986; Petersen, 2002), zeigen verschiedene Autoren, dass in Realschulzeugnissen ab den 80er Jahren ein deutlicher Abwärtstrend der prädiktiven Validität von Schulabschlüssen erkennbar ist. Die mittlere Korrelation lag nur noch bei $r = .26$ (Schuler, 1998). Zwar konnte dieser Trend für Abiturzeugnisse aus den 60er – 80er Jahren nicht festgestellt werden, in wie weit diese Aussage auch heute noch gilt, lässt sich jedoch angesichts vorliegender unterschiedlicher Ergebnisse nicht ohne Zweifel beantworten. Selbst jedoch im Vergleich zu der von Baron-Boldt, Funke und Schuler (1989) berichteten prädiktiven Validität von Schulnoten von $r = .41$ für den Ausbildungs- und $r = .46$ für den Studienerfolg spricht das Ergebnis der vorliegenden Arbeit für die praktische Nützlichkeit des BfP als ein alternatives Vorauswahlinstrument.

Auch ein Vergleich mit anderen Untersuchungen zur Erfassung der prognostischen Validität biographischer Verfahren aus dem Zeitraum von 1950 bis in die 80er Jahre stützt die positive Bewertung des Ergebnisses dieser Arbeit. So ergab bspw. die Metaanalyse von Schmitt, Gooding, Noe und Kirsch (1984) eine Validität von $r = .52$, wenn als Kriterium die Gehaltshöhe gewählt wurde, dagegen nur $r = .23$ wenn die Ausbildungsleistung herangezogen wurde. Bliesener (1992) kommt bei seiner Bewertung von 75 Primärstudien auf einen mittleren Koeffizienten von $r = .33$. Stehle (1983) erschloss nach Kreuzvalidierung eine prädiktive Validität von $r = .45$ bei der Auswahl zukünftiger Außendienstmitarbeiter für ein großes deutsches Versicherungsunternehmen.

Bei der Bewertung der überwiegend hohen prognostischen Validität biographischer Daten, sollten kritische Anmerkungen zu den Ergebnissen nicht außer Acht bleiben. Auch losgelöst von der in Kapitel 2.2 wiederholt diskutierten fehlenden Konstruktvalidität und damit der Frage, was eigentlich die Grundlage der für den Eignungsdiagnostiker erfreulich hohen Validitätskoeffizienten bildet, ist die prognostische Validität nicht per se anzunehmen. Zusammenfassende Studien von Holberg und Pugh (1978), Schmidt, Hunter & Caplan (1981)

und Reilly und Chao (1982) weisen auf eine starke Inkonsistenz der kriterienbezogenen Validität biographischer Daten in unterschiedlichen Berufsgruppen hin. Gerade dieser Hinweis unterstützt noch einmal die Notwendigkeit der vorliegenden Arbeit.

Da es in der eignungsdiagnostischen Praxis bei der Maximierung der prädiktiven Validität in der Regel darum geht, aus einer Vielzahl möglicher diagnostischer Instrumente diejenigen auszuwählen, die mit einem möglichst kleinen Aufwand und geringen Kosten eine größtmögliche Enge des Zusammenhanges zwischen gewählten Prädiktoren und Kriterien aufzeigen, erscheint der in dieser Studie entwickelte biographische Fragebogen geeignet, als Vorauswahlinstrument und Alternative zu Schulnoten angewandt zu werden. Der Wert der prognostischen Validität des BfP kann auch an seinem prozentualen Anteil an der Varianz des Kriteriums gemessen werden, der durch die Varianz des Prädiktors aufgeklärt werden kann. Dieser im Determinationskoeffizienten seinen Ausdruck findende Prozentwert wird determiniert durch das Quadrat von Prädiktor und Kriterium. Da der BfP zu $r = .50$ mit dem gewählten Kriterium, dem Ergebnis im Auswahlverfahren, korreliert, ermöglicht der Fragebogen die Aufklärung von 25 % der Kriteriumsvarianz.

Als Gesamtergebnis kann folglich konstatiert werden, dass sich der BfP als für den Zweck der Vorauswahl von Bewerbern für den Polizeivollzugsdienst für geeignet erwiesen hat. Darüber hinaus finden die Fragen eine hohe Zustimmung hinsichtlich ihrer Eignung für den vorgesehenen Zweck, d.h. der BfP darf in einem zufriedenstellenden Maße als prognostisch und sozial valide betrachtet werden.

Trotz dieser Einschätzung dürfen kritische Aspekte bei der Bewertung der vorliegenden Untersuchung und des Ergebnisses nicht unreflektiert bleiben.

Zu einem sind u.a. das Alter und die Verteilung in der gewählten Stichprobe zu nennen. Da es das Ziel der Untersuchung war, einen für die Praxis nützlichen Fragebogen zu entwickeln, war es notwendig, eine vergleichbare Population als Stichprobe heranzuziehen. Dabei lag das Alter der Bewerber zwischen dem 15. und dem 40. Lebensjahr, wobei über 96 % der Bewerber das 30. Lebensjahr noch nicht vollendet hatten. Insgesamt 83,1 % waren jünger oder gleich 25 Jahre, der Altersmittelwert lag bei 20,9 Jahren. Mit der Zusammensetzung der Stichprobe kann eine repräsentative Gruppe für Bewerber um einen Ausbildungsplatz im mittleren Polizeivollzugsdienst angenommen werden, doch eine Ausdehnung des Ergebnisses

auf andere Kontexte erscheint allein angesichts der Altersstruktur nicht möglich. Schon die Gruppe der Bewerber, die sich um eine Ausbildung für den gehobenen Polizeivollzugsdienst bewirbt, dürfte, da es sich hier erfahrungsgemäß in erster Linie um Schulabgänger mit einem deutlich geringeren Anteil an Berufswechslern handelt, nicht mehr hinsichtlich Alter, Bildung und Berufserfahrung vergleichbar sein. Diesbezüglich müsste der Fragebogen für die Vorauswahl für diese Laufbahn neu validiert werden.

Auch wenn versucht wurde, Stichprobeneffekten und damit einer situationsbedingten Überschätzung der Validität durch eine Kreuzvalidierung entgegenzuwirken, bleibt folgendes anzumerken: Die aus verfahrensökonomischen Gründen gewählte übliche Vorgehensweise, dass für die eine Teilstichprobe die Itemgewichte ermittelt werden, deren Validität dann an der zweiten Stichprobe überprüft wurde, eliminiert auf diese Weise zwar durch die Stichprobe bedingte Zufallsfehler, nicht jedoch die systematischen Besonderheiten der verwandten Stichprobe insgesamt. In diesem Kontext verweist Bliesener (1992) darauf, dass wenn der Item-schlüssel in einer Generierungsstichprobe entwickelt und anschließend in einer zweiten Stichprobe validiert wird, sich eine mittlere Validität von $r = .28$ erschließt. Wird dagegen ein externer Schlüssel aus einer anderen Untersuchung verwendet, sinkt die Validität auf $r = .22$. D.h. Schlüssel, die nicht speziell für die betrachtete Stichprobe entwickelt worden sind, zeigen deutlich schwächere Effektivitäten. Ein vergleichbarer Effekt ist auch für das Ergebnis der vorliegenden Studie anzunehmen. Alternativ hätte die Möglichkeit einer Doppelkreuzvalidierung bestanden. Hierbei werden Itemanalysen in zwei unabhängigen Stichproben durchgeführt. Die Itemauswahl und -gewichtung richtet sich nach den Antworten, die sich in den jeweiligen Stichproben als trennscharf erwiesen haben. Anschließend besteht dann die Möglichkeit, entweder einen Gesamtschlüssel aus den trennscharfen Items zu bilden und an der Gesamtstichprobe zu validieren oder es wird jeweils ein Schlüssel für jede Stichprobe berechnet und an der jeweiligen anderen Stichprobe kreuzvalidiert. Wegen der Abhängigkeit zwischen Generierungs- und Validierungsstichprobe wird hierbei die Validität des Verfahrens häufig überschätzt (Bliesener, 1992). Damit erschien eine einfache Kreuzvalidierung gerechtfertigt, da sie zumindest einen Schutz vor der Aggregation von Zufallseffekten bietet.

Zu berücksichtigen ist an dieser Stelle auch, dass die erschlossene Höhe des Validitätskoeffizienten wesentlich von folgenden Faktoren abhängt: dem Grad dessen, was an Gemeinsamkeit durch den Test und das Kriterium erfasst wird und der Reliabilität von

Prädiktor sowie der des Kriteriums. Das heißt, dass die erschlossene Reliabilität primär die Rolle hat, günstige Voraussetzungen für das Erreichen einer hohen Validität zu schaffen (Amelang & Zielinski, 1994, S. 113). Wenn auch die Erschließung der Reliabilität des zukünftigen Prädiktors (BfP) Gegenstand dieser Untersuchung war und als Retest-Reliabilität mit $r = .69$ als ausreichend hoch angenommen werden kann, blieb die Höhe der Reliabilität des Kriterium (Ergebnis im Auswahlverfahren) unerschlossen. Für einzelne eignungsdiagnostische Instrumente des Auswahlverfahrens, aus deren Gesamtergebnis sich das Kriterium zusammensetzt, kann zumindest für die standardisierten Testteile (CFT und IST) eine ausreichend hohe Reliabilität belegt werden. Dagegen wurde die Reliabilität nicht standardisierter Verfahrensteile (Diktat, Berichtsfertigung, Sporttest, Vorstellungsgespräch) bisher nicht erschlossen und kann allenfalls angenommen werden. Zur Erschließung der fehlenden Reliabilitäten wären weitere Untersuchung außerhalb der echten Auswahl-situation notwendig, die im Rahmen dieser Arbeit nicht realisierbar waren. Dagegen war es aus Gründen der Vergleichbarkeit zwingend, die Erschließung der Validität unter realen Durchführungsbedingungen umzusetzen.

Insbesondere die o.g. Abhängigkeiten des Korrelationskoeffizienten machen deutlich, dass die mögliche Höhe der Validität eines zur Eignungsdiagnostik herangezogenen Auswahlinstrumentes nur so hoch sein kann, wie es die gegebene Reliabilität von Prädiktor und Kriterium zulassen. An dieser Stelle erscheint es auch sinnvoll, noch einmal das Konzept der Bandbreite des konstruierten Prädiktors im Sinne Cronbach's (1949) zu diskutieren. Bei der Prognosestellung durch einen Biographischen Fragebogen geht ein großer und inhaltlich vielfältiger Bereich an Informationen die gewonnen werden (große Bandbreite) mit einer Einschränkung oder dem Verlust an Fidelität (geringe Messgenauigkeit) einher. Es liegt mit dem BfP kein anforderungsbezogenes und hochspezielles Instrument für die Auswahl von Polizeibeamten vor. Inwieweit es ein derartiges Instrument im Unterschied bspw. zu hochtechnisierten apparativen Messinstrumenten für die Auswahl von Flugzeugführern geben kann, soll an dieser Stelle offen bleiben. Insofern ist für derartig unspezifische Auswahlinstrumente grundsätzlich eine niedrigere Reliabilität anzunehmen, da eine möglichst breite Aussagemöglichkeit über die Eignung eines Bewerbers angestrebt wird. Nach Kleinevoss und Sonnenberg (1987) kann jedoch allgemein angenommen werden, dass der wahre Zusammenhang zwischen Prädiktor und Kriterium immer eher unterschätzt wird, da bei beiden mit Messfehlern gerechnet werden muss, so dass die Prognostizität von wissenschaftlich fundierten Auswahlverfahren generell eher unter- als überschätzt wird.

Kritisch hinsichtlich der erschlossenen Höhe der prädiktiven Validität ist außerdem die gewählte Vorgehensweise bei der Validierung des Fragebogens zu reflektieren. Einerseits belegen Vergleichsstudien zwischen dem empirischen und rationalen Ansatz die Überlegenheit des empirischen Vorgehens hinsichtlich der prädiktiven Validität (Mitchell & Klimoski, 1982; Hornick, James & Jones, 1977). Angesichts des dringenden Bedarfes an einem effektiven Vorauswahlinstrument, rechtfertigen diese Ergebnisse in der Praxis häufig den empirischen Ansatz. Dabei ist das Verfahren von England (1961, 1971), - „Weighted application blank approach“ - das bei Einsatz von Kontrastgruppen häufigste Verfahren. Andererseits ist einschränkend anzumerken, dass bis heute nur wenige vergleichende Studien vorliegen (Weinert, 1998). Zudem fanden Hough und Paullin (1994) in ihrer Metaanalyse eine minimal höhere Validität des deduktiven Ansatzes gegenüber dem empirischen und des empirischen gegenüber dem induktiven. In der vorliegenden Arbeit wurden aber zudem Aspekte verschiedener Ansätze berücksichtigt. Zur Generierung des Fragebogens wurde der puristische Ansatz von England (1961, 1971) durch eine induktive Fragenbogenkonstruktion ergänzt. Die allen Items der extrahierten 12 Faktoren inhärente Gemeinsamkeit liefert zumindest einen Aufschluss über die psychologische Struktur der fraglichen Dimensionen. Diese begriffliche Etikettierung der Konstrukte geht somit in ihrer Bedeutung deutlich über den rein empiristischen Ansatz hinaus. Die im BfP abgebildeten Faktoren (Vorhandenes Berufsbild, Sportliche Freizeitaktivitäten, Soziales und intellektuelles Engagement, Allgemeine Lebens- und Berufserfahrung, Gesellschaftliches Ansehen, Frühe finanzielle Selbstständigkeit, Passives Freizeitverhalten, Aktives Freizeitverhalten, Gemachte Erfahrungen mit dem Berufswunsch, Entspannendes Freizeitverhalten und negative Erfahrungen im Beruf oder Leben) können zumindest in weiteren Untersuchungen als Basis zur Erschließung der Konstruktvalidität herangezogen werden.

Letztendlich dürfen aber mögliche verfahrensimmanente Schwierigkeiten mit dem gewählten weiteren Vorgehen nicht unbewertet bleiben. Gemäß den Vorgaben von England (1961, 1971) werden die errechneten Differenzen in den gebildeten Extremgruppen den von Strong (1926) tabellierten Nettogewichten zugeordnet bzw. in die von England (1961, 1971) empfohlenen Grobgewichte umgerechnet (s.a. Kapitel 2.2). Damit wird jedes Item aufgrund seiner Kriteriumsvalidität ausgesucht und entsprechend gewichtet. Dies entspricht, das Skalenniveau bei dem Vergleich nicht beachtend, einem regressionsanalytischen Ansatz. In der Itemanalyse wird für jeden Prädiktor (bei England jede Antwortalternative) anhand der Prognosegüte ein empirisches Gewicht ermittelt. Die empirische Itemgewichtung führt in der

dafür eingesetzten Stichprobe zu einer fast maximalen Kriteriumsvalidität. Diese Vorgehensweise führt allerdings dazu, dass der Itemschlüssel die Besonderheiten der Stichprobe berücksichtigt und kaum auf andere berufliche Kontexte übertragbar ist. Dies ist aber für den BfP auch nicht angedacht. Ein weiteres methodisches Problem ist dabei die sogenannte immanente kompensatorische Skalenbildung (Knoblauch, 1986). So können Bewerber mit gleicher Punktezahl Defizite in ganz unterschiedlich wichtigen Anforderungen aufweisen, ohne dass die Möglichkeit besteht, die Beurteilung des Summenscores zu bemerken oder zu berücksichtigen.

Wie in Kapitel 5 beschrieben, war die Retest-Reliabilitätsermittlung der in dieser Untersuchung gewählte Zugang zur Erschließung der Reliabilität. Unter Berücksichtigung der Voraussetzungen für die Anwendung dieser Methode (Lienert & Raatz, 1994) war die Grundlage der Entscheidung für dieser Vorgehensweise zum einen die generelle Empfehlung von Siegel (1956) für biographische Fragebogen und zum anderen die Möglichkeit eines direkten Vergleiches mit der Höhe der bestimmten Reliabilität in der Untersuchung von Stehle (1983).

Auch wenn die in Kapitel 6 dargestellte Höhe der Reliabilität mit $r = .69$ sowohl für den gesamten Fragebogen als auch für die 16 trennscharfen Items als ausreichend hoch bewertet wurde, wird deutlich, dass in der einzig bekannten vergleichbaren Untersuchung von Stehle eine deutliche Diskrepanz zu der von ihm berichteten Reliabilität von $r = .94$ gegeben ist. Allerdings muss in diesem Zusammenhang die von Stehle gewählte Methode hinterfragt werden. Bei den ihm insgesamt nur 20 für die Auswertung zur Verfügung stehenden Fragebogenpaaren wurden vom Autor im ersten Schritt auf der Itemebene Mittelwerte errechnet, die dann im zweiten Durchgang miteinander korreliert wurden. Diese Vorgehensweise führt dazu, dass die unterschiedlichen Antworten in der ersten und zweiten Erhebung durch Mittelwertbildung aggregieren und somit die Abweichung im Antwortverhalten relativiert wird. Die Folge ist eine Überschätzung der Reliabilität. Eine zur Kontrolle gerechnete vergleichbare Vorgehensweise ergäbe für den BfP eine überschätzte Reliabilität von $r = .91$.

Unter Berücksichtigung dieses Aspektes kann das Ergebnis der Reliabilitätsbestimmung als zufriedenstellend bewertet bleiben. Wie unter Kapitel 2.2.1.1 dargestellt, verweist Hollmann (1991) weiter auf die Bedeutung des Retest-Koeffizienten als Indiz für die

Fälschungssicherheit des Fragebogens. Unterschiedliche Autoren (Keating, Patterson & Stone, 1959; Kavanagh & York, 1972; Blum & Stumpf, 1983) konnten zudem die Fälschungssicherheit empirisch belegen. Auch Baehr und Williams (1967) verweisen darauf, dass die generelle Gefahr der Fälschung bei biographische Fragebogen wesentlich geringer sei als bei Persönlichkeitsfragebogen. Diese Annahme hat bei der geplanten Verwendung, d.h. der Auswahl von Bewerbungen, eine besondere Bedeutung. Gerade wenn in zukünftigen Untersuchungen dem Vorwurf einer rein empiristischen Vorgehensweise durch weitere Konstruktaufklärung entgegengewirkt werden soll, ist es entscheidend, dass die erhobenen Dimensionen Persönlichkeitsstrukturen zugeordnet werden können und nicht sozial erwünschtem Ankreuzverhalten entspringen.

Insofern wäre es sicherlich bereits in dieser Arbeit wünschenswert gewesen, einen konkreten Beitrag zur Konstruktaufklärung leisten zu können. Denn nach wie vor fehlt es an einer theoretischen Verknüpfung von gewonnenen Einzeldaten zu einem Konstrukt Persönlichkeit. So fasst Savage (1986, S.76) auch folgerichtig zusammen: „Wir mögen wissen, dass ein Item Individuen unterscheiden kann in Bezug auf irgendwelche Kriterien; aber solange wir nicht verstehen, warum es das tut und wie sich das zu anderen Items verhält, gibt es keine verlässliche Basis für die Behauptung, es sei irgend etwas Wissenschaftliches oder Rationales an diesem Prozess. Er ist reduziert auf a priori-Annahmen über statistische Methoden.“

So hätte bspw. der Hinweis von Schmidt und Hunter (1998), dass biographische Messinstrumente in beträchtlichem Umfang mit allgemeinen kognitiven Fähigkeiten korrelieren auch in der vorliegenden Studie aufgegriffen werden können. Es wäre naheliegend gewesen, die Skalen des BfB und weitere Einzelergebnisse der Teilnehmer im Auswahlverfahren miteinander in Beziehung zu setzen. Möglicherweise hätten die Ergebnisse einen weiteren Hinweis liefern können, ob biographische Daten wirklich ein indirektes Maß für mentale Fähigkeiten darstellen. Die dem Personalrat gegenüber gegebene Zusicherung (s.a. Kapitel 5), dass die erhobenen Daten ausschließlich zum dargestellten Zweck, d.h. der Erschließung der prädiktiven Validität, herangezogen und anschließend nicht mehr berücksichtigt werden, verwehrt jedoch weitere Analysen. Hinsichtlich einer notwendigerweise dauerhaften Akzeptanz des Vorauswahlinstrumentes und der eigenen Glaubwürdigkeit gegenüber den Interessenvertretern war es wichtiger, gemachte Absprachen als verbindlich zu akzeptieren. Da zu hoffen ist, dass dieses Verhalten sich förderlich auf die Einführung des Auswahlinstrumentes auswirkt, ergeben sich somit nur zukünftig

Möglichkeiten, anschließende Untersuchungen zur Erschließung der Konstruktvalidität zu initiieren.

Die Nützlichkeit des biographischen Fragebogens zur Vorauswahl von Bewerbern für den Polizeivollzugsdienst, wie sie in der vorliegenden Untersuchung belegt werden konnte, reicht allein nicht aus. Wie in Kapitel 2.3 belegt, ist es heutzutage fast genau so wichtig, dass Probanden, die in einem Verfahren ausgewählt bzw. nicht ausgewählt werden, das Instrument, mit dem diese Selektion stattfand, zumindest als dafür geeignet akzeptieren. Mit der Frage nach der Akzeptanz wird aber auch ein Dilemma der Eignungsdiagnostik deutlich: Verfahren, die hinsichtlich ihrer prognostischen Validität eine hohe Vorhersagegüte bieten, finden durch die Teilnehmer häufig nur eine geringe Wertschätzung. So werden standardisierte Testverfahren in der Regel schlechter beurteilt als zur Auswahl herangezogene Schulnoten. Dagegen erfreuen sich nicht standardisierte Interviews einer hohen Beliebtheit, auch wenn sie insgesamt keine hohe prognostische Validität besitzen.

Wie dargestellt, ergab die Erfassung der Akzeptanz der Items des BfP eine durchschnittliche Zustimmung bei allen Fragen von 95,24 %. Die durchschnittliche Zustimmung zu den 16 trennscharfen Items lag bei 94,78 %. In beiden Fällen kann das Ergebnis als sehr zufriedenstellend bezeichnet werden. Orientierung für die Bewertung des Ergebnisses bot erneut die Untersuchung von Stehle (1983). Er übersetzte für seine Untersuchung den „Experience and Background Inventory“ von Baehr und Williams (1967, 1968) mit insgesamt 151 Fragen, die 26 Mitarbeitern sowie 7 Betriebsratmitgliedern zur Bewertung vorgelegt wurden. Die Befragten beurteilten dabei 90 von 151 Fragen (59 %) als für einen biographischen Fragebogen nicht zulässig. Auch wenn der Einfluss durch Elimination von nicht akzeptierten Items auf die prognostische Validität sich bei Stehle nicht signifikant auswirkte, ist das im Vergleich sehr gute Ergebnis der vorliegenden Arbeit nur durch die initiierte frühzeitige Beteiligung von Betroffenen und Personalratsmitgliedern zu erklären. Dies war nur durch die Neukonzeption eines eigenen Fragebogens möglich und so auch intendiert. Schon bei der Sammlung von Itemkonzepten, Itemformulierung und abschließender Itemauswahl wurden die Ergebnisse von qualitativen Interviews mit erfahrenen Polizeibeamten berücksichtigt. Vor endgültiger Fertigstellung wurden die Fragen wiederholt mit Polizeibeamten, Auszubildenden und Personalratsvertretern abschließend diskutiert und bei fehlender Akzeptanz oder Verständlichkeit entfernt bzw. angepasst. Die Bedeutung des erreichten Ergebnisses kann weiterhin durch folgenden Umstand gewürdigt

werden: Der von Köchling und Körner (1996) erhobenen Kritik an anderen Studien konnte entgegengewirkt werden, denn es wurden in der vorliegenden Arbeit nur solche Personen befragt, welche die zu beurteilenden Verfahren und den Ernstfall von Personalauswahlentscheidungen aufgrund diagnostischer Informationen aus eigener Erfahrung kannten.

Wichtig für eine hohe Akzeptanz scheint auch die Länge des Fragebogens zu sein. Denn wie Bliesener (1992) konstatiert, besitzen auch kleine und mittlere Verfahren mit etwa 50 Items eine gute prädiktive Validität und führen zudem zu einer größeren Akzeptanz bei Bewerbern und zur ökonomischeren Handhabung. Da die Validität keineswegs linear mit dem Umfang des Fragebogens steigt, wurde mit der endgültigen Länge des BfP (56 Items) sowohl hinsichtlich des Einflusses auf die erschlossene prädiktive Validität wie auch auf die soziale Akzeptanz ein gutes Maß gefunden.

Nicht weiter beleuchtet werden konnte die Frage, in wie weit die hohe Akzeptanz des biographischen Fragebogens dazu beitragen kann, dass die erfolgreichen Teilnehmer des Auswahlverfahrens die angebotenen Ausbildungsplätze auch wirklich annehmen. Mit der Beantwortung dieser Fragestellung könnte zumindest ein Hinweis für die längerfristige Stabilität des Eindrucks im Auswahlprozess und die mittelbare positive Wirkung auf die Entscheidung hinsichtlich der Annahme bzw. Ablehnung eines Ausbildungsplatzes durch geeignete Kandidaten gesehen werden. Letztendlich kann mit der Erhebung und Sicherstellung der Akzeptanz eines Vorauswahlinstrumentes im Sinne einer Verbesserung der Sozialen Validität nur ein erster Schritt in eine notwendige Richtung gesetzt werden. Eine positive Selbstdarstellung des Unternehmens gegenüber seinen potenziellen Nachwuchskräften ist bspw. eine weitere wichtige Voraussetzung, um Interessierte an ein Unternehmen binden zu können. Bei der Durchführung von Auswahlverfahren sind weiteren Bedürfnissen der Bewerber Rechnung zu tragen. Neben tätigkeits- und firmenbezogenen Informationen, Situationskontrolle, Nachvollziehbarkeit des Bewertungsprozesses und detailliertem Ergebnisfeedback ist es im Sinne von Köchling und Körner (1998) wichtig, dass das Ausmaß der Befriedigung dieser Bedürfnisse zu einem Kriterium wird, an dem sich schon die Gestaltung des gesamten Auswahlverfahrens orientiert. Es bleibt dabei auch von Bedeutung, dass die positive Präsentation der Organisation nicht nur ein Mittel zur Erhöhung der Annahmequote bei Stellenangeboten an geeignete Bewerber sein kann, sondern auch über

diesen Kreis hinaus in das soziale Umfeld eines Bewerbers wirkt, in dem neue Quellen für potenziellen beruflichen Nachwuchs zu finden sein dürften.

Allerdings bleibt die Aussage von Kersting (1998) in Erinnerung zu rufen, nach der die soziale Akzeptanz diagnostischer Instrumente nicht mit utopischen Erwartungen überfrachtet werden sollte. Auswahlverfahren dienen der Leistungsmessung und zur Auswahl einzelner Bewerber. Es sollen sinngemäß jene ausgewählt werden, bei denen aufgrund ihrer erhobenen Leistungen im Verfahren eine höhere Wahrscheinlichkeit empirisch belegt wurde, dass sie den Anforderungen der Ausbildung oder der Berufsausübung eher gerecht werden. Akzeptanz ist kein Ersatz für die Erschließung der Validität, denn die Frage der Akzeptanz stellt sich nur innerhalb derjenigen Verfahren, welche die Voraussetzungen der Validität erfüllen.

Zudem sollten keine utopischen Erwartungen erhoben werden, denn eine hundertprozentige Akzeptanz von Auswahlverfahren wird es nicht geben. Leistungsmessungen sind anstrengend und stressig und mit Hilfe der Auswahlverfahren werden Personen zudem zurückgewiesen. Teilweise gehen Bewerber auch mit unrealistischen Erwartungen in die Auswahlverfahren. Außerdem sind die Interessen von Bewerber und Organisation im Auswahlverfahren nicht grundsätzlich deckungsgleich. Es wird immer Bewerber geben, die den Anforderungen nicht im ausreichenden Maße gerecht werden. Vor dem Hintergrund der Theorie der kognitiven Dissonanz (Festinger, 1957) ist zu erwarten, dass negative Entscheidungen bei Bewerbern eine relevante Dissonanz zwischen Kognitionen auslösen. So ist einerseits die Teilnahme am Auswahlverfahren als bedeutsam für den erstrebten Ausbildungsplatz zu bewerten und andererseits zeigt das Ergebnis des Auswahlverfahrens, dass der Bewerber den hier abgebildeten Anforderungen nicht gewachsen ist. Die Folge ist eine möglicherweise selbstwertdienliche Abwertung des Verfahrens.

Einen besonderen Stellenwert dürfte der wahrgenommenen Akzeptanz von Auswahlverfahren bei unternehmensinternen Auswahlentscheidungen zu kommen. Hier sind die abgelehnten Bewerber anschließend weiterhin an ihrer bisherigen Funktion für die Organisation tätig. Neben jeder Demotivierung, die durch das Erleben im Auswahlverfahren zu vermeiden ist, da sie sich auch auf die Arbeitseffektivität des Mitarbeiters auswirken kann, sind, insbesondere im öffentlichen Dienst, durch nicht akzeptierte Verfahren rechtliche Auseinandersetzungen zwischen Bewerber und Organisation eine häufigere Folge. Insofern sollten die durch die Studie initiierten Überlegungen zur allgemeinen Gestaltung von Auswahlverfahren auch

Berücksichtigung in internen Auswahlverfahren für den Aufstieg in eine höhere Laufbahn finden.

Defizitär bleibt in dieser Untersuchung aber die bisher nicht ausreichend systematisch untersuchte Abhängigkeit der Akzeptanzeinschätzung von Personenfaktoren. Denn wie Kersting (1998) darstellt, könnte die Berücksichtigung differentialpsychologischer Aspekte ein neues Licht auf die vermeintlich niedrige oder hohe Akzeptanz, die einzelnen Verfahren zugeschrieben wird, werfen und insgesamt zu einer realistischeren Einschätzung der maximal erzielbaren Akzeptanzwerte für einzelne Verfahren führen. Allerdings ergeben sich gerade bei Auswahlverfahren im Öffentlichen Dienst und den damit einhergehenden Mitsprachemöglichkeiten und Vorbehalten bei Personalratsvertretern gegen die Erhebung von Persönlichkeitsmerkmalen u.U. Schwierigkeiten. Hier wären vermutlich allenfalls Untersuchungen möglich, wenn Daten erhoben würden, die in keinem Zusammenhang mit Entscheidungen über die Annahme oder Ablehnung von Bewerbern stehen. Damit ergeben sich aber wiederum Schwierigkeiten bei der Vergleichbarkeit der Situationen.

Auch einer weiteren berechtigten Kritik von Kersting (1998) konnte im Rahmen dieser Erhebung nicht entgegengewirkt werden. Es fehlt nach wie vor an theoriegeleitet konstruierten mehrdimensionalen und validierten Messinstrumenten zur Erfassung der sozialen Akzeptanz. Die Akzeptanz von Auswahlverfahren sollte nicht uni-, sondern mehrdimensional erfasst werden. Ein Verfahren hat nicht nur einen Akzeptanzwert, sondern mehrere, die zusätzlich noch in Abhängigkeit von Person- und Situationsmerkmalen variieren. Es ist durchaus möglich, dass bestimmte Verfahren im Rahmen spezifischer eignungsdiagnostischer Fragestellungen von Personen akzeptiert werden, die das gleiche Verfahren bei einer anderen eignungsdiagnostischen Fragestellung ablehnen würden. Da für den Bereich der biographischen Daten kaum vergleichbare Untersuchungsergebnisse zur Bewertung der Akzeptanzhöhe vorlagen, war es vorrangiges Ziel, eine Vergleichbarkeit des erzielten Resultates mit dem in der Arbeit von Stehle (1983) zu gewährleisten. Insofern wurde dem zur Anwendung kommenden Messinstrument nicht die notwendige Bedeutung beigemessen und die Vorgehensweise, analog wie bei Stehle, auf der single-item Ebene angesiedelt und Daten nur uni-dimensional erhoben. Da die eingesetzten Messinstrumente grundsätzlich den Anforderungen an die allgemeinen Gütekriterien entsprechen sollten, ist das erzielte Ergebnis unter Berücksichtigung dieses Defizits und im Vergleich zur Arbeit von Stehle zu bewerten.

7.3 Ausblick

In den letzten Jahren gab es kaum neue Veröffentlichungen, die einen zusätzlichen Beitrag zur Klärung der Konstruktvalidität biographischer Daten liefern konnten. Angesichts der häufigen Anwendung in der Praxis bleibt hier ein Forschungsdefizit, das dringender Klärung bedarf. Auch wenn weiterhin offen bleiben muss, was letztendlich genau durch die Anwendung biographischer Fragebogen zur Personalauswahl erfasst wird, rechtfertigt der auch in dieser Untersuchung wieder belegte hohe praktische Nutzen die Anwendung. Dessen ungeachtet bleibt die weitergehende Konstruktaufklärung in zukünftigen Untersuchungen wünschenswert und notwendig, um auch die Wissenschaftlichkeit eines nützlichen Verfahrens zu reflektieren.

Wie Schuler (1996) anmerkt, werden die in Lebensläufen enthaltenden Daten und Fakten bei der Bewertung von vorliegenden Bewerbungen als Belege verarbeitet und gelten als Indikatoren künftiger Bereitschaft und Verhaltens. Hollmann und Reitzig (1995) verweisen bspw. auf Indikatoren im Lebenslauf (Auslandsaufenthalt, Vielfalt der bisherigen Stellen, außerberufliche Aktivitäten) die Analogieschlüsse auf Faktoren wie Selbstsicherheit, Selbstständigkeit oder Durchsetzungs- und Entscheidungsvermögen erlauben sollen. Statistisch sind solche Zusammenhänge jedoch nicht abgesichert. Hier könnte in zukünftigen Untersuchungen über Zusammenhänge zwischen den Dimensionen im BfP und herkömmlichen Datengebern reflektiert werden. Weitere Untersuchungsfelder bieten sich im Zusammenhang mit dem von Hossiep (1995) dargestellten Überblick möglicher Determinanten beruflichen Erfolges, die als mögliche Prädiktoren dem Bereich biographischer Daten zugeordnet werden können. Die Beziehung der hier genannten Prädiktoren (Persönliches Wertesystem / Werthaltung, Kultureller Hintergrund, Sozioökonomische Sozialisation, Erfahrung, Ausbildung, Erziehung, Interesse / Motivation, Physische Eigenschaften, Alter und Geschlecht, Charaktereigenschaften, Kognitive / motorische Fähigkeiten, Persönliches Lebensschicksal, Sonstige Merkmale der Person) zu den Konstrukten des BfP sowie weiteren Personenvariablen könnten unter Umständen zur weiteren Konstruktaufklärung beitragen.

Schmidt und Rothstein (1994) postulieren, dass einige eignungsdiagnostische Verfahren Konstruktkombinationen messen und die tatsächlichen Konstrukte unbekannt sind. Als ein Beispiel nennen sie Vorstellungsgespräche, die vermutlich eine Kombination aus früheren

Erfahrungen, mentalen Fähigkeiten und einer Reihe von Persönlichkeitsmerkmalen erfassen. So ist nach den Autoren ein wesentlicher Grund, warum allgemeine kognitive Fähigkeiten einen guten Prädiktor darstellen, im vorhandenen Berufswissen zu finden. Intelligenter Menschen eignen sich schneller und mehr Berufswissen an als andere. In diesem Zusammenhang wären weitere Erkenntnisse über das Ankreuzverhalten der Teilnehmer im BfP bzw. die Bewertung des Ankreuzverhaltens im Zusammenhang mit den erhobenen Skalen und ihren erreichten Ergebnissen in anderen Teilen des Auswahlverfahrens, insbesondere dem Vorstellungsgespräch, den kognitiven Leistungstests oder weiteren Persönlichkeitstests wünschenswert, um weiteren Aufschluss über mögliche Konstrukte gewinnen zu können.

Knoblauch (1986) verweist darauf, dass biographische Fragebogen, im Gegensatz zu Leistungs- und Persönlichkeitstests, parallel die Prognose der Ausbildungs- wie auch der Berufsbewährung gewährleisten können. Damit bietet sich an, den BfP auch für die Prognose des Berufserfolges zu validieren. Allerdings sind die in Kapitel 2.1 dargestellten Schwierigkeiten bei der Definition von Kriterien für den Berufserfolg zu antizipieren. Gerade bei einer sehr vielseitigen und mit vielfachen Anforderungen einhergehenden Tätigkeit wie der des Polizeivollzugsdienstes, dürfte es nicht einfach sein, reliable und valide Kriterien zu finden.

Albright und Glennon (1961) konnten feststellen, dass sich forschungsinteressierte und managementinteressierte Wissenschaftler in einem biographischen Fragebogen bezüglich Arbeits- und Sozialverhalten signifikant unterscheiden. Ähnliche Vergleich wären bspw. zwischen Polizeibeamten und Vertretern anderer Berufe (z.B. helfenden Berufen wie Krankenschwester oder sonstigen wie bspw. Bankkaufleute) möglich. Die Ergebnisse könnten zudem in Beziehung zu zusätzlich erhobenen Personenvariablen bewertet werden. Auch ließe sich der BfP zusammen mit einem zeitgleich erhobenen Persönlichkeitsfragebogen faktorenanalysieren.

Kritisch bewertet bleibt in diesem Zusammenhang allerdings die Frage, in wie weit die im Auswahlverfahren erfassten Anforderungen denen in der späteren Berufsausübung entsprechen. Eine notwendige Evaluation hat bisher nicht stattgefunden. Die Prognosegüte des Auswahlverfahrens bezieht sich in erster Linie auf die Ausbildungsleistung und weniger auf den späteren Berufserfolg. Insofern wäre es von besonderer Bedeutung, weitere Erkenntnisse über mögliche Zusammenhänge zwischen den Skalen des BfP und anderen

biographischen Fragebogen sowie den Anforderungen, die an die erfolgreiche Ausübung des Polizeiberufes gestellt sind, zu erhalten. So könnte man die Hypothese aufstellen, dass bei dem Antwortverhalten eines erfolgreichen Polizeibeamten nur ein geringer Zusammenhang mit dem psychologischen Konstrukt „Introversion“ zu erwarten wäre. Gleichzeitig dürfte zu erwarten sein, dass hinsichtlich des Konstruktes „Stressfähigkeit“ hohe Zusammenhänge gegeben sein sollten.

Hinsichtlich der erfassten Akzeptanz ist für weitere Untersuchungen eine wichtige Fragestellung, inwieweit akzeptierte eignungsdiagnostische Instrumente dazu beitragen können, dass die Bewerber die ihnen angebotenen Ausbildungsplätze auch annehmen.

Auch die Abhängigkeit der Akzeptanzeinschätzung von Personenfaktoren wurde bislang noch nicht ausreichend systematisch untersucht. Die Berücksichtigung differentialpsychologischer Aspekte könnte ein neues Licht auf die vermeintlich niedrige oder hohe Akzeptanz, die einzelnen Verfahren zugeschrieben wird, werfen und insgesamt zu einer realistischeren Einschätzung der maximal erzielbaren Akzeptanzwerte für einzelne Verfahren führen.

Dabei bleibt die teilweise eklatante Vernachlässigung des Messproblems bei der Erhebung der Daten als Hauptschwierigkeit bisheriger empirischer Untersuchungen weiter berücksichtigungswert. Bei der Vorbereitung kommender Untersuchungen ist zu bedenken, dass Auswahlverfahren nicht nur einen Akzeptanzwert haben, sondern mehrere, die zusätzlich noch in Abhängigkeit von Person- und Situationsmerkmalen variieren. Dies muss sich in der mehrdimensionalen Erfassung des Akzeptanzwertes ausdrücken.

Trotz der belegten Nützlichkeit des BfP für die Vorauswahl im Auswahlverfahren für den Polizeivollzugsdienst und der hohen Akzeptanz als eignungsdiagnostisches Instrument darf abschließend nicht unerwähnt bleiben, was der vorliegende biographische Fragebogen nicht in der Lage zu leisten ist. Mit biographischen Daten können keine kognitive Strukturen oder andere potentielle Fähigkeiten abgebildet werden, wie dies mit anderen Testverfahren möglich wäre. Interaktionen von unterschiedlichen Probanden, wie es im Assessment Center zu beobachten ist, oder die Beobachtung von Verhaltensweisen sind nicht möglich. Für eine Platzierung im Rahmen eines Auswahlverfahren eignen sich derartige Datengeber nur begrenzt, denn für die Beurteilung des optimalen Einsatzes von Ressourcen sind die Verfahren in der Regel nicht konzipiert. Neues, in einer bisherigen Tätigkeit nicht

aufgetretenes Verhalten, kann durch biographische Fragebogen nicht beurteilt werden. Im Gegensatz zu standardisierten Leistungstests ist analog den Persönlichkeitsfragebogen auch hier eine Form der sozialen Erwünschtheit von Antworten, wenn auch in einem geringeren Ausmaß, zu erwarten.

Die Generalisierbarkeit auch dieses biographischen Fragebogens ist nicht nur nicht intendiert, sondern der BfP ist ausschließlich für eine spezifische Population – die Bewerber für den mittleren Polizeivollzugsdienst – konzipiert. Für diesen Zweck erschließt sich allerdings für den BfP eine ausreichend hohe Reliabilität und eine gute prädiktive Validität. Damit liegt ein Instrument vor, mit dem zwischen erfolgreichen und nicht erfolgreichen Teilnehmern im Auswahlverfahren diskriminiert werden kann. Zudem wird der Fragebogen als Instrument zur Vorauswahl von Bewerbern für den Polizeivollzugsdienst von erfahrenen Teilnehmern am Auswahlverfahren akzeptiert.

8 Literatur

- Althoff, K. (1986).** Zur Aussagekraft von Schulzeugnissen im Rahmen der Eignungsdiagnostik. *Psychologie und Praxis, Zeitschrift für Arbeits- und Organisationspsychologie*, 30, 2, 77-85
- Amelang, M. & Zielinski, W. (1994).** *Psychologische Diagnostik und Intervention*. Berlin, Heidelberg: Springer
- Amthauer, R. (1970).** *Intelligenz-Struktur-Test 70 (IST 70)*. Göttingen: Hogrefe
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2000).** *Intelligenz-Struktur-Test 2000 (IST 2000)*. Göttingen: Hogrefe
- Anastasi, A. (1982).** *Psychological Testing*. 5th Edition. New York: Macmillan
- Anastasi, A. (1985).** The use of personality assessment in industry. Methodological and interpretive problems. In J.W. Atkins & J.O. Rynor (Eds.), *Personality, motivation, and achievement*, 221-242. Washington: Hemisphere
- Anders, P.E. (1992).** *Betriebswirtschaftslehre humoris causa*. 2. Aufl., Wiesbaden: Gabler
- Asher, J.J. (1972).** The biographical item: Can it be improved? *Personal Psychology*, 25, 251-269
- Baehr, M.E. & Williams, G.B. (1967).** Underlying demensions of personal background data and their relation to occupational classification. *Journal of Applied Psychology*, 52, 481 – 490
- Baehr, M.E. & Williams, G.B. (1968).** Prediction of sales success from factorially determined dimensions of personal background data. *Journal of Applied Psychology*, 52, 98 – 103
- Baron-Boldt, J., Funke, U. & Schuler, H (1989).** Prognostische Validität von Schulnoten. Eine Metaanalyse der Prognose des Studien- und Ausbildungserfolgs. In R.S. Jäger und K. Ingenkamp (Hrsg.), *Tests und Trends* 7, 11-39. Weinhheim: Beltz
- Barrick, M. R. & Mount, M.K. (1991).** The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology*, 44, 1 – 26
- Barthel, E. & Schuler, H. (1989).** Nutzenkalkulation eignungsdiagnostischer Verfahren am Beispiel eines biographischen Fragebogens. *Zeitschrift für Arbeits- und Organisationspsychologie*, 33, 73 – 83
- Bartussek, D., Ratz, U. & Stapf, K.H. (1986).** *Die Evaluation des Tests für medizinische Studiengänge*. Bonn: Institut für Test- und Begabungsforschung
- Behrens, R. & Merkel, R. (1989).** *Erfolgreiche Personalauswahl. Leitfaden für Unternehmer und Führungskräfte*. Köln: TÜV-Rheinland
- Benziman, H. (1986).** The Psychodianostic Experience: A Call for Systematic Feedback Procedures. In B. Nevo & R.S. Jäger (Eds.), *Psychological Testing: The Examinee Perspective*, 147 – 163. Göttingen: Hogrefe
- Berthel, J. (1995).** *Personal-Management*. Stuttgart: Schäffer-Poeschel
- 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. (1993).** Der Einfluss der Forschungsqualität auf das Forschungsergebnis. Zur Evaluation der Validierung biographischer Daten in der Eignungsdiagnostik. Unveröffentlichte Habil.schrift, Universität Erlangen-Nürnberg

- Blum, F. & Stumpf, H. (1983).** Zur Zuverlässigkeit von Fragebogenangaben zu edukativen und soziodemographischen Merkmalen. *Psychologie in Erziehung und Unterricht*, 30, 2, 135 – 143
- Borkenau, P. & Ostendorf, F. (1993).** NEO-Fünf-Faktoren-Inventar (NEO-FFI). Göttingen: Hogrefe
- Bortz, J. (1984).** Lehrbuch der empirischen Forschung für Sozialwissenschaftler. Berlin: Springer Verlag.
- Bortz, J. (1989).** Statistik für Sozialwissenschaftler. 3.Auflage. New York, Berlin, Heidelberg: Springer Verlag
- Bortz, J. & Döring, N. (1995).** Forschungsmethoden und Evaluation für Sozialwissenschaftler. 2.Auflage. Berlin, Heidelberg, New York: Springer
- Boudreau, J. (1989).** Economic considerations in estimating the utility of human resource productivity improvement programs. *Personnel Psychology*, 36, 551 – 576
- Brambring, M. (1983).** Spezielle Eignungsdiagnostik. In K.D. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik. Enzyklopädie der Psychologie*, B,II,2, 414 – 481. Göttingen: Hogrefe
- Bray, D.W. & Howard, A. (1983).** Personality and the assessment center method. In: C.D. Spielberger & J.N. Butcher (Eds.). *Advances in Personality Assessment*. Hillsdale: Erlbaum
- Brengelmann, L. (1983).** Personal-Auslese-Test der Polizei Nordrhein-Westfalen (PatPol). Landespolizei NRW
- Brickenkamp, R. (1972).** Aufmerksamkeits-Belastungstest (d2). Göttingen: Hogrefe
- Brickenkamp, R. (1981).** Aufmerksamkeits-Belastungstest (d2). Göttingen: Hogrefe
- Brickenkamp, R. (1975).** Handbuch psychologischer und pädagogischer Tests. Göttingen: Hogrefe
- Brickenkamp, R. (1983).** Handbuch psychologischer und pädagogischer Tests. Erster Ergänzungsband. Göttingen: Hogrefe
- Brush, D.H. & Owens, W.A. (1979).** Implementation and evaluation of an assessment classification model for manpower utilization. *Personnel Psychology*, 32, 369 – 383
- Campbell, D. T. & Fiske, D. W. (1959).** Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81 - 105
- Cascio, W.F. (1991).** Costing human resources: The financial impact of behavior in organizations, 3rd. Kent: Publishing Company
- Cassens, F.P. (1966).** Cross cultural dimensions of executive life history antecedents (biographical informations). Institute for Behavioral Research in Creativity (Eds.), Greensboro: IBRIC.
- Comer, J.R. (1995).** Klinische Psychologie. Heidelberg: Spektrum
- Cronbach, L.J. (1970).** Essentials of psychological testing. New York: Harper & Row
- Cronbach, L.J. & Gleser, G.C. (1965).** Psychological tests and personnel decisions. Urbana: University of Illinois press
- Dickinson, T.L. (1993).** Attitudes about performances appraisal process. In H. Schuler, J.L. Farr & M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives*, 141 – 161. Hillsdale: Erlbaum
- Dill, W.R., Hilton, T.L. & Reimann, W.R. (1960).** How aspiring managers promote their own careers. *California Management Review*, 2, 9 – 15

- DuBios, P.H. (1970).** A history of psychological testing. Boston: Allyn & Bacon.
- Dunette, M.D. & Kirchner, W.K. (1960).** A checklist for defferentiating different kinds of sales jobs. *Personal Psychology*, 44, 2, 121-125
- Eberhardt, B.J. & Muchinski, P.M. (1982).** Biodata determinants of vocational typology. An integration of two paradigms. *Journal of Applied Psychology*, 67, 714 – 727
- Edwards, A.L. (1957).** Techniques of attitude scale constructions. New York: Appleton
- Ehlers, A., Stangier, U., Dohn, D. & Gieler, U. (1993).** Kognitive Faktoren beim Juckreiz: Entwicklung und Validierung eines Fragebogens. *Verhaltenstherapie*, 3, 112-119
- England, G.W. (1971).** Development and Use of Weighted Application Blanks. Industrial Relations Center. University of Minnesota
- Fahrenberg, J., Selg, H. & Hampel, R. (1978).** Das Freiburger Persönlichkeitsinventar FPI. 3. ergänzte Auflage. Göttingen: Hogrefe
- Fay, E. (1992).** Über die Übertragbarkeit der Leistung in einem Durchstreichverfahren zur Messung der Konzentrationsfähigkeit. *Diagnostica*, 38, 301 – 311
- Feger, H. (1983).** Planung und Bewertung von wissenschaftlichen Beobachtungen. In H.Feger & J. Bredenkamp (Hrsg.), *Datenerhebung. Enzyklopädie der Psychologie. Bd.1, Forschungsmethoden*, 1-75. Hogrefe: Göttingen
- Festinger, L. (1957).** A theory of cognitive dissonance. Stanford: Stanford University Press
- Flanagan, J. C. (1949).** Critical requirements for research personnel: A study of observed behaviors of personnel in research laboratories. Pittsburgh: American Institute for Research
- Forsyth, D. R. (1987).** Social Psychology. Monterey: Books / Cole
- Frey, D. & Gaska, A. (1993).** Die Theorie der kognitiven Dissonanz. In D. Frey und M. Irle (Hrsg.), *Theorien der Sozialpsychologie, Band I: Kognitive Theorien*, 275 – 324. Huber: Bern
- Frieling, E. & Hoyos, C. Graf (1978).** Fragebogen zur Arbeitsanalyse. (FAA) Bern: Huber
- Fruhner, R. & Schuler, H. (1991).** Gibt es Unterschiede in der Bewertung von Assessment-Center Aufgaben in Abhängigkeit vom eigenen Abschneiden? In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis*, 313 – 319. Stuttgart: Verlag für Angewandte Psychologie
- Fruhner, R., Schuler, H., Funke, U. & Moser, K. (1991).** Einige Determinanten der Bewertung von Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*. 35, 4, 170 – 178
- Fuller, C. (1974).** Effect of anonymity on return rate and response bias in a mail survey. *Journal of Applied Psychology*, 59, 292-296
- Funke, U. & Schuler, H. (1986).** Weiterentwicklung biographischer Fragebogen durch Konstruktaufklärung. Grundlagen und erste Ergebnisse. In H. Schuler & W. Stehle (Hrsg.), *Biographische Fragebogen als Methode der Personalauswahl. Beiträge zur Organisationspsychologie, 2*. Stuttgart: Verlag für Angewandte Psychologie
- Funke, U., Krauß, J., Schuler, H. & Stapf, K.H. (1987).** Zur Prognostizierbarkeit wissenschaftlich technischer Leistungen mittels Personenvariablen: Eine Metaanalyse der Validität diagnostischer Verfahren im Bereich Forschung und Entwicklung. *Gruppendynamik*, 18, 407-428
- Ghiselli, E.E. (1955).** The measurement of occupational aptitude. *University of California Publications in Psychology*, 8, 101 – 216

- Gilliland, S.W. (1993).** The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, 18, 694 – 734
- Goldsmith, D.B. (1922).** The use of the personal history blank as a salesmanship test. *Journal of Applied Psychology*, 48, 26 – 34
- Graudenz, H. (1987).** Eignungsuntersuchung aus der Sicht der Bewerber. *DGP Information*, 38, 1 - 28
- Greif, S. (1995).** Geschichte der Organisationspsychologie. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie*, 15-48. Bern: Huber
- Guion, R. M. & Gottier, R. F. (1965).** Validity of Personality Measures in Personnel Selection. *Personnel Psychology*, 18, 135 - 164
- Guski, R., Wichmann, U., Rohrmann, B. & Finke, H.O. (1978).** Konstruktion und Anwendung eines Fragebogens zur sozialwissenschaftlichen Untersuchung der Auswirkung von Umweltlärm. *Zeitschrift für Sozialpsychologie*, 9, 50-65
- Guthke, J., Böttcher, H.R. & Sprung, L. (1990).** (Hrsg.) *Psychodiagnostik Band 1*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Hathaway, S. R. & McKinley, J. C. (1951).** The Minnesota Multiphasic Personality Inventory manual revised. New York: The Psychological Corporation
- Heinze, B. (1995).** Graphologie. In W. Sarges (Hrsg.), *Management-Diagnostik*, 470-475. Göttingen: Hogrefe
- Henry, E.R. (1965).** Research conference on the use of biographical data as psychological predictors. Greensboro, N.C.: Richardson Foundation
- Hesse, J. & Schrader, H. C. (1991).** Das neue Test-Training-Programm: Einstellungs- und Eignungstests erfolgreich bestehen; die wichtigsten Testaufgaben - und wie man sie löst. Frankfurt: Eichborn
- Hesse, J. & Schrader, H. C. (1992).** Testtrainings für Ausbildungsplatzsuchende der 90er Jahre. Frankfurt: Eichborn
- Höft, S. & Funke, U. (2001).** Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe
- Holberg, A. & Pugh, W.M. (1978).** Predicting navy effectiveness: Expectations, motivation, personality, aptitude and background variables. *Personnel Psychology*, 31, 841 - 852
- Holling, H. & Leippold, W. (1991).** Zur sozialen Validität von Assessment Centern. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis*, 10, Stuttgart: Verlag für Angewandte Psychologie
- Hollmann, H. (1991).** Validität in der Eignungsdiagnostik. Göttingen: Hogrefe
- Hollmann, H. & Reitzig, G. (1995).** Referenzen und Dokumentenanalyse. In W. Sarges (Hrsg.), *Management-Diagnostik*, 463-470. Göttingen: Hogrefe
- Holmes, T.H. & Rahe, R.H. (1967).** The Social Readjustment Rating Scale. *Journal of Psychomatic Research*, 11, 213-218
- Horn, W. (1983).** Leistungsprüfsystem (LPS). Göttingen: Hogrefe
- Hornick, C., James, L. & Jones, A. (1977).** Empirical item keying vs. a rational approach to analyze a psychological climate questionnaire. *Applied Psychological Measurement*, 1, 489 – 500

- Hornke, L.F. (1991).** Neue Itemformen für computergestütztes Testen. In H. Schuler & U. Funke (Hrsg), *Eignungsdiagnostik in Forschung und Praxis*, 10, Stuttgart: Verlag für Angewandte Psychologie
- Hossiep, R. (1995).** Berufseignungsdiagnostische Entscheidungen. Zur Bewährung eignungsdiagnostischer Ansätze. Göttingen: Hogrefe
- Hossiep, R., Paschen, M. & Mühlhaus, O. (2000).** Persönlichkeitstests im Personalmanagement. Grundlagen, Instrumente und Anwendungen. Göttingen: Verlag für Angewandte Psychologie
- 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*, 109 – 145. Palo Alto: Consulting Psychologists Press
- Hoyos, C. (1956).** Zur Methodik der Bewährungskontrolle in der Berufseignungsprüfung *Psychologische Rundschau*, 7, 204 – 223
- Hunter, J.E. (1986).** Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340-362
- Hunter, J.E. und Hunter, R.F. (1984).** Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98
- Ingenkamp, K. (1969).** Möglichkeiten und Grenzen des Lehrerurteils und der Schultests. In Roth, H. (Hrsg.), *Begabung und Lernen*, 407 – 431. Stuttgart: Klett
- Irle, M. (1955).** Berufs-Interessen-Test (BIT). Göttingen: Hogrefe
- Janz, T. (1982).** The patterned behavior descriptions interview versus unstructured interviews. *Journal of Applied Psychology*, 67, 577-580
- Jäger, A.O. (1986).** Validität von Intelligenztests. *Diagnostica*, 32 (4), 272-289
- Jäger, R. (1974).** Methoden zur Mittelung von Korrelationen. *Psychologische Beiträge*, 16, 417-427
- Jäger, R. (1976).** Moderatoren, differenzielle Diagnostizierbarkeit und Vorhersage. In Michel, L. (Hrsg.), *Psychologische Diagnostik*. Göttingen: Hogrefe
- Jäger, R., Berbig, E., Gossler, H., Hagen, J., Lieblich, W. & Schafheutle, R. (1973).** Mannheimer Biographisches Inventar (MIB), Göttingen: Hogrefe
- Jäger, R., Berbig, E., Gossler, H., Hagen, J., Lieblich, W. & Schafheutle, R. (1976).** Biographisches Inventar zur Diagnose von Verhaltensstörungen (BIV), Göttingen: Hogrefe
- Jäger, R.S. (1992).** Biographische Daten. In Jäger, R.S. und Petermann, F. (Hrsg.), *Psychologische Diagnostik*, 350-362. Weinheim: PVU
- Jäger, R.S. (1995).** Eignungsdiagnostik aus der Sicht der Kandidaten. In W. Sarges (Hrsg.), *Management-Diagnostik*, 102 – 108, Göttingen: Hogrefe
- Jäger, R.S & Nord-Rüdiger, D. (1983).** Biographische Analyse. In Minsel, W.R. und Scheller, R. (Hrsg.), *Diagnostik*, 62-82. München: Kösel
- Jeserich, W. (1991).** Mitarbeiter auswählen und fördern. Assessment Center Verfahren. *Handbuch der Weiterbildung für die Praxis in Wirtschaft und Verwaltung*. 6. Auflage. München: Hanser
- Kallus, K. W. & Jahnke, W. (1992).** Klassenzuordnungen. In S. Jäger und F. Petermann (Hrsg.), *Psychologische Diagnostik*, 2.Aufl., 170 – 186

- Kannheiser, W. & Frieling, E. (1992).** Arbeitsstrukturierung und Arbeitsanalyse. In D. Frey, Graf C. Hoyos & D. Stahlberg (Hrsg.), *Angewandte Psychologie*, 129-146. Weinheim: PVU
- Kaufmann, W. (1984).** Wie man die "beste" Führungskraft auswählt. Die Assessment-Center-Methode zur Auswahl des Führungskräfte Nachwuchses. *Management Zeitschrift*, 53, 474 - 476
- Kavanagh, M.J. & York, D.R. (1972).** Biographical correlates of middle managers` performance. *Personnel Psychology*, 25, 319 – 332
- Keating, E., Patterson, D.G. & Stone, C.H. (1950).** Validity of work histories obtained by interviews. *Journal of Applied Psychology*, 34, 1 – 5
- Kelly, G.A. (1965).** Der Motivationsbegriff als irreführendes Konstrukt. In H. Thomae (Hrsg.), *Die Motivation menschlichen Handelns*. Köln: Kiepenheuer & Witsch
- Kepes, S.Y. & True, J.E. (1967).** Anonymity and attitudes toward work. *Psychological Reports*, 21, 353-356
- Kersting, M. (1998).** Differentielle Aspekte der sozialen Akzeptanz von Intelligenztests und Problemlöseszenarien als Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 42, 61 – 75
- Keßler, B. (1982).** Biographische Diagnostik. In K.J. Groffmann und L. Michel (Hrsg.), *Persönlichkeitsdiagnostik*, 1-56. Göttingen: Hogrefe
- Kleinevoss, R. & Sonnenberg, H. G. (1987).** Die prognostische Validität einer eignungsdiagnostischen Prozedur – dargestellt mit einem konfiguralen Diskriminanzmodell. *Psychologie und Praxis. Zeitschrift für Arbeits- und Organisationspsychologie*, 31, 15 – 21
- Kleinmann, M. (1997).** Assessment Center, Stand der Forschung – Konsequenzen für die Praxis. Göttingen: Verlag für Angewandte Psychologie
- Kleinmann, M. & Strauß, B. (1998).** (Hrsg.), *Potentialfeststellung und Personalentwicklung*. Göttingen: Verlag für Angewandte Psychologie
- Klimoski, R. & Brickner, M. (1987).** Why do assessment centers work ? The puzzle of assessment center validity. *Personnel Psychology*, 40, 243 - 260
- Knoblauch, R. (1986).** 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. Beiträge zur Organisationspsychologie*, 2. Stuttgart: Verlag für Angewandte Psychologie
- Köchling, A. C. (1999).** Bewerberorientierte Personalauswahl. Der Vorstellungstermin als Instrument des Personalmarketing. In A. Thiele & B. Eggers (Hrsg.) *Innovatives Personalmarketing für High-Potentials*, 135 – 148. Göttingen: Hogrefe
- Köchling, A. C. & Körner, S. (1996).** Personalauswahl aus der Sicht der Betroffenen: Zur bewerberorientierten Gestaltung von Beurteilungssituationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 40, 22 – 37
- Köhnken, G. (1990).** *Glaubwürdigkeit*. München: Psychologie Verlags Union
- Kompa, A. (1989).** Assessment-Center und Eliteherrschaft in Organisationen: Analyse und Kritik einer Methode der Führungskräfteauswahl. *Augsburger Beiträge zu Organisationspsychologie und Personalwesen*, 5, 1-76
- Kubinger, K.D. (1993).** Testtheoretische Probleme der Computerdiagnostik. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 130 – 137

- Lammers, F. (1994).** Einfluss der Interraterreliabilität und Beobachterrotation auf die Konstruktvalidität des Assessment Centers. Vortrag, 39. Kongress der deutschen Gesellschaft für Psychologie, Hamburg
- Lang von Wins, T. & Rosenstiel von, L. (1998).** Potentialfeststellungsverfahren. In M. Kleinmann & B. Strauß (Hrsg.), Potentialfeststellung und Personalentwicklung, 71-95. Göttingen: Verlag für Angewandte Psychologie
- Latham, G.P., Saari, L.M., Pursell, E.D. & McDaniel, M.A. (1980).** The situational interview. *Journal of Applied Psychology*, 65, 422-427
- Lienert, G.A. (1969).** Testaufbau und Testanalyse. Weinheim: Beltz
- Lienert, G.A. (1989).** Testaufbau und Testanalyse. 4.Auflage. Weinheim: PVU
- Lienert, G.A. & Raatz, U. (1994).** Testaufbau und Testanalyse. 5. Auflage. Weinheim: PVU
- Mael, F. A. (1991).** A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology*, 44, 763 – 792
- Markgraf, J. & Lieb, R. (1995).** Was ist Verhaltenstherapie? Versuch einer zukunftsorientierten Neucharakterisierung. *Zeitschrift für klinische Psychologie*, 24, 1-7
- Marschner, G. (1980).** Möglichkeiten und Grenzen betriebspsychologischer Eignungsuntersuchungen. In R. Neubauer & L. v. Rosenstiel (Hrsg.), *Handbuch der angewandten Psychologie*, Bd. I, 774 – 794. München: Moderne Industrie
- Maukisch, H. (1978).** Einführung in die Eignungsdiagnostik. In A. Mayer (Hrsg.), *Organisationspsychologie*, 105-136. Stuttgart: Poeschel
- Mead, A. D. & Drasgow, F. (1993).** Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449 – 458
- Melfsen, S, Florin, I. & Walter, H-J. (1999).** Die deutsche Fassung des social phobia and anxiety inventory for children (SPAI-C-D): Psychometrische Eigenschaften und Normierung. *Diagnostica*, 2, 95-103
- Michel, L. & Conrad, W. (1982).** Theoretische Grundlagen psychometrischer Tests. In K.J. Groffmann & L. Michel (Hrsg.) *Enzyklopädie der Psychologie*, II: Psychologische Diagnostik, 1, 1 – 129
- Mitchell, T.W. & Klimoski, R.J. (1982).** Is it rational to be empirical? A test of methods for scoring biographical data. *Journal of Applied Psychology*, 67, 411 - 418
- Mohr, E. (1990).** Methoden und Kriterien der Personalvorauswahl in bundesdeutschen Unternehmen für die Zielgruppe Hochschulabsolventen. Unveröffentlichte Diplomarbeit, Universität Kiel
- Moser, K. (1995).** Planung und Durchführung organisationspsychologischer Untersuchungen. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie*, 2. Aufl., 71 – 107, Bern: Huber
- Moser, K. & Rhyssen, D. (2001).** Referenzen als eignungsdiagnostische Methode. *Zeitschrift für Arbeits- und Organisationspsychologie*, 45, 40 – 46
- Moser, K., Stehle, W. & Schuler, H. (1992).** Personalmarketing. Göttingen: Verlag für Angewandte Psychologie
- Moser, K. & Zempel, J. (2001).** Personalmarketing. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe
- Mummendey, H.D. (1987).** Die Fragebogenmethode. Göttingen: Hogrefe
- Neter, E. & Ben-Shakhar, G. (1989).** The predictive validity of graphological inferences: A meta-analytic approach. *Personality and Individual Differences*, 10, 737-745

- Neubauer, A. & Urban, E. (1991).** Der Vergleich von computergestützter Testdarbietung und Standardvorgabe am Beispiel von Ravens Advanced Progressive Matrices. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis*, 10, Stuttgart: Verlag für Angewandte Psychologie
- Neuberger, O. & Allerbeck, M. (1978).** *Messung und Analyse von Arbeitszufriedenheit*. Bern: Huber
- Nevo, B. (1986).** Face validity and others related variables. In B. Nevo und R.S. Jäger (Eds.), *Psychological Testing: The examinee perspective*, 49 – 68. Göttingen: Hogrefe
- Owens, W.A. (1976).** Background data. In M.D. Dunette (Ed.), *Handbook of industrial and organizational psychology* 609 - 644. Chicago: Rand McNally
- Owens, W.A. & Schoenfeldt, L.F. (1979).** Toward a classification of persons. *Journal of Applied Psychology*, 64, 569 - 607
- Pawlik, K. (1976).** Modell- und Praxisdimensionen psychologischer Diagnostik. In K. Pawlik (Hrsg.), *Diagnose der Diagnostik. Beiträge zur Diskussion der psychologischen Diagnostik in der Verhaltensmodifikation* 13 – 43. Stuttgart: Klett
- Petersen, R. (2002).** Sind Schulnoten als Instrument zur Vorauswahl noch verlässlich? In *Vorbereitung*.
- Pook, M., Tuschen-Caffier, B., Schnapper, U., Speiger, K., Krause, W. & Florin, I. (1999).** Kognitionen bei Infertilität: Entwicklung und Validierung eines Fragebogens (KINT). *Diagnostica*, 2, 104-113.
- Poortinga, Y.H., Coetsier, P., Meuris, G., Miller, K.M., Samsonowitz, V., Seisededos, N. & Schlegel, E. (1982).** A survey of attitudes towards tests among psychologists in six Western European countries. *International Review of Applied Psychology*, 31, 7 – 34
- Pulakos, E.D. & Schmitt, N. (1995).** Experience-based and situational interview questions: Studies of validity. *Personnel Psychology*, 48, 289-308
- Rastetter, D. (1999).** Bewerbungsunterlagencreening. Analyse eines hochselektiven, kaum erforschten Auswahlinstrumentes. *Zeitschrift für Arbeitswissenschaft*, 53 (25 NF), 37-44.
- Raven, J.C. (1958).** *Advanced Progressive Matrices*. London: Lewis
- Rawls, D.J. (1968).** The use of the BIB in the identification of college seniors with managerial potential. *Annual Convention of the American Psychological Association*, 3, 579 – 580
- Rawls, J.R. & Rawls, D.J. (1974).** Recent trends in management selection. *Personnel Journal*, 53, 104 – 109
- Reilly, R.R. & Chao, G.R. (1982).** Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, 35, 1-62
- Richter, H. & Stehle, W. (1986).** Die Entwicklung eines firmenspezifischen Auswahlsystems auf der Basis biographischer Daten. In H. Schuler & W. Stehle (Hrsg.), *Biographische Fragebogen als Methode der Personalauswahl. Beiträge zur Organisationspsychologie*, 2. Stuttgart: Verlag für Angewandte Psychologie
- Rösler, F. (1992).** Personalauslese, Training und Personalentwicklung in Organisationen. In D. Frey, Graf C. Hoyos & D. Stahlberg (Hrsg.), *Angewandte Psychologie*, 66-91. Weinheim: PVU
- Rohmert, W. & Landau, K. (1979).** *das arbeitswissenschaftliche Erhebungsverfahren zur Tätigkeitsanalyse (AET)*. Bern: Huber

- Rosenfeld, M. & Thornton, R.F. (1976).** A case study in job analysis methodology. Results obtained from analyzing two positions: Employment service interviewer and employment service local office manager. Princeton: Educational Testing Service, Center for Occupational and Professional Assessment
- Ryan, A. & Sackett, P. (1987).** Pre-employment honesty testing: fakability, reactions of test takers, and company image. Human Sciences Press, 248 – 256
- Salgado, J. F. (1997).** The Five Factor Model of Personality and Job Performance in the European Community. *Journal of Applied Psychology*, 82, 1, 30 - 43
- Sarges, W. (1995a).** (Hrsg.), *Management-Diagnostik*. Göttingen: Hogrefe
- Sarges, W. (1995b).** Interviews. In W. Sarges (Hrsg.), *Management-Diagnostik*, 475 – 489, Göttingen: Hogrefe
- Sarges, W. (1996).** Weiterentwicklung der Assessment Center-Methode. Göttingen: Verlag für Angewandte Psychologie
- Sarges, W. & Weinert, A. B. (1991).** Früherkennung von Management-Potentialen. In W.E. Feix (Hrsg.) *Personal 2000 – Visionen und Strategien erfolgreicher Personalarbeit*, 267 – 301, Frankfurt, Wiesbaden: Gabler
- Sarges, W. & Wottawa, H. (2001).** *Handbuch wirtschaftspsychologischer Testverfahren*. Lengerich: Pabst Science Publishers
- Savage, A.W. (1986).** Biographische Fragebogen : Neuere Ergebnisse aus England. In H. Schuler & W. Stehle (Hrsg.), *Biographische Fragebogen als Methode der Personalauswahl. Beiträge zur Organisationspsychologie*, 2. Stuttgart: Verlag für Angewandte Psychologie
- Schmale, H. & Schmidtke, H. (1966).** *Berufseignungstest (BET)*. Bern: Huber
- Schmitt, N., Gooding, R.Z., Noe, R.D. & Kirsch, M. (1984).** Metaanalysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personel Psychology*, 37, 407-422
- Schmidt, F.L. & Hunter, J.E. (1977).** Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540
- Schmidt, F.L. & Hunter, J.E. (1981).** Employment Testing: Old theories and new research findings. *American Psychologist*, 36, 1128-1137
- Schmidt, F.L. & Hunter, J.E. (1998).** Meßbare Personenmerkmale: Stabilität, Variabilität und Validität zur Vorhersage zukünftiger Berufsleistung und berufsbezogenen Lernens. In M. Kleinmann & B. Strauß (Hrsg.), *Potentialfeststellung und Personalentwicklung*, 15-43. Göttingen: Verlag für Angewandte Psychologie
- Schmidt, F.L., Hunter, J.E., & Caplan, J.R. (1981).** Validity generalization results for two job groups in the petroleum industry. *Journal of Applied Psychology*, 66, 261 - 273
- Schmidt, F.L., Hunter, J.E., Pearlmann, K. & Shane, G.S. (1979).** Further tests of the Schmidt-Hunter Bayesian Validity Generalization Model. *Personel Psychology*, 32, 257-281
- Schmidt, F.L. & Rothstein, H.R. (1994).** Application of validity generalization methods of metaanalysis to biographical data scores in employment selection. In G.S. Stokes, M.D. Mumford & W.A. Owens (Eds.), *The biodata handbook: Theory, research and applications*, 237-260. Palo Alto: CPP Books
- Schmidtke, A. (1992).** Diagnostik als Inventarisieren. In R.S. Jäger und F. Petermann (Hrsg.), *Psychologische Diagnostik*, 263-267. Weinheim: PVU

- Schmitt, N. & Coyle, B. W. (1976).** Applicant decisions in the employment interview. *Journal of Applied Psychology*, 61, 184 – 192
- Schmitt, N., Gooding, R.Z., Noe, R.D. & Kirsch, M. (1984).** Metaanalysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422
- Schneewind, K.A., Schröder, G. & Cattell, R.B. (1994).** Der 16-Persönlichkeitsfaktoren-Test (16 PF). 3. Aufl.. Bern: Huber
- Scholz, G. & Schuler, H. (1993).** Das nomologische Netzwerk des Assessment Centers: eine Metaanalyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 73 - 85
- Schorr, A. (1991).** Diagnostische Praxis in der Arbeits- und Organisationspsychologie. Teilergebnisse aus einer repräsentativen Umfrage zur diagnostischen Praxis. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis*, 10, Stuttgart: Verlag für Angewandte Psychologie
- Schuler, H. (1986).** Der Einsatz biographischer Fragebogen zur Prognose des Berufserfolges: Einleitende Überlegungen und Überblick. In H. Schuler & W. Stehle (Hrsg.), *Biographische Fragebogen als Methode der Personalauswahl. Beiträge zur Organisationspsychologie*, 2. Stuttgart: Verlag für Angewandte Psychologie
- Schuler, H. (1990).** Personenauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 34, 184-191
- Schuler, H. (1992a).** Das Multimodale Einstellungsinterview. *Diagnostica*, 38, 281-300
- Schuler, H. (1992b).** Assessment-Center als Auswahl und Entwicklungsinstrument: Einleitung und Überblick. In H. Schuler & W. Stehle, *Assessment-Center als Methode der Personalentwicklung, Beiträge zur Organisationspsychologie*. 2. Auflage. Göttingen, Stuttgart: Verlag für Angewandte Psychologie
- Schuler, H. (1996).** Psychologische Personalauswahl. Einführung in die Berufseignungsdiagnostik. Göttingen: Verlag für Angewandte Psychologie
- Schuler, H. (1998).** Noten und Studien- und Berufserfolg. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie*, 370 – 374. Weinheim: Psychologische Verlags Union
- Schuler, H. (2001).** (Hrsg.), *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe
- Schuler, H. & Funke, U. (1991).** *Eignungsdiagnostik in Forschung und Praxis*. Stuttgart: Verlag für Angewandte Psychologie
- Schuler, H. & Funke, U. (1993).** Diagnose beruflicher Eignung und Leistung. In H. Schuler (Hrsg.), *Lehrbuch der Organisationspsychologie*. Bern: Huber
- Schuler, H., Frier, D. & Kauffmann, M. (1993).** Personalauswahl im europäischen Vergleich. Göttingen: Verlag für Angewandte Psychologie
- Schuler, H. & Höft, S. (2001).** Konstruktorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe
- Schuler, H. & Moser, K. (1992).** Entscheidung von Bewerbern. In K. Moser, W. Stehle & H. Schuler (Hrsg.), *Personalmarketing*, 51 - 75. Göttingen: Hogrefe
- Schuler, H. & Moser, K. (1995).** Geschichte der Management-Diagnostik. In W. Sarges (Hrsg.), *Management-Diagnostik*, 32 - 42. Göttingen: Hogrefe
- Schuler, H. & Marcus, B. (2001).** Biographieorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe

- Schuler, H. & Stehle, W. (1983).** Neuere Entwicklungen des Assessment-Center-Ansatzes beurteilt unter dem Aspekt der sozialen Validität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 27, 33-44
- Schuler, H. & Stehle, W. (1985).** Soziale Validität eignungsdiagnostischer Verfahren: Anforderungen für die Zukunft. In H. Schuler & W. Stehle (Hrsg.), *Organisationspsychologie und Unternehmenspraxis: Perspektiven der Kooperation*. Stuttgart: Hogrefe
- Schuler, H. & Stehle, W. (1986).** (Hrsg.) *Biographische Fragebogen als Methode der Personalauswahl*. Beiträge zur Organisationspsychologie, 2. Stuttgart: Verlag für Angewandte Psychologie
- Schuler, H. & Stehle, W. (1992).** (Hrsg.) *Assessment Center als Methode der Personalentwicklung*. Beiträge zur Organisationspsychologie, 2. Göttingen, Stuttgart: Verlag für Angewandte Psychologie
- Seibt, H. & Kleinmann, M. (1991).** Personalauswahl von Bewerbern: Derzeitiger Stand und Alternativen. In H. Schuler & U. Funke (Hrsg.), *Eignungsdiagnostik in Forschung und Praxis*, 10, Stuttgart: Verlag für Angewandte Psychologie
- Siegel, L. (1956).** A biographical inventory for students: I. Construction and standardization of the instrument. *Journal of Applied Psychology*, 40, 5 – 10
- Sieverding, M. (2000).** Die Bedeutung von Selbstdarstellungsregeln im Bewerbungsgespräch. *Zeitschrift für Arbeits- und Organisationspsychologie*, 44, 3, 152-156
- Stehle, W. (1983).** Zur Konzeption eines Personalauswahlverfahrens auf der Basis biographischer Daten. Unveröffentlichte Dissertation. Universität Hohenheim
- Stehle, W. (1986).** Personalauswahl mittels biographischer Fragebogen. In H. Schuler & W. Stehle (Hrsg.), *Biographische Fragebogen als Methode der Personalauswahl*. Beiträge zur Organisationspsychologie, 2. Stuttgart: Verlag für Angewandte Psychologie
- Stehle, W. (1995).** Biographische Fragebogen. In W. Sarges (Hrsg.), *Management-Diagnostik*, 2. Göttingen: Hogrefe, 526 - 530
- Steiner, D.D. & Gilliland, S.W. (1996).** Fairness reactions to personnel selection techniques in France and United States. *Journal of Applied Psychology*, 81, 134 – 141
- Strasser, W. (1975).** Untersuchung zur Validität von Persönlichkeitsfragebögen und biographischen Informationsbögen. Unveröffentlichte Dissertation. Universität München
- Strong, E.K. (1926).** An Interest Test for Personal Managers. *Journal of Personnel Research*, 5, 194 – 203
- Taylor, H. C. & Russel, J. T. (1939).** The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 23, 565-578
- Thomae, H. & Jüttemann, G. (1987).** (Hrsg.) *Biographie und Psychologie*. Heidelberg: Springer
- Thornton, G. & Byham, W.C. (1982).** *Assessment Center and managerial performance*. New York: Academic Press
- Tränkle, U. (1983).** Fragebogenkonstruktion. In H. Feger & J. Bredenkamp (Hrsg.), *Enzyklopädie der Psychologie*, B,1,2 (222-301). Göttingen: Hogrefe
- Weinert, A.B. (1998).** *Organisationspsychologie*. 4. Auflage. Weinheim: PVU
- Wegener, H. Köhnken, G. & Steller, M. (1988).** Recht. In D. Frey, C. Graf Hoyos, C. & D. Stahlberg (Hrsg.), *Angewandte Psychologie*, 342-362; München: Psychologie Verlags Union
- Wendt, D. (1989).** *Allgemeine Psychologie. Eine Einführung*. Stuttgart: Kohlhammer

Westhoff, K. (2000). Das psychologisch-diagnostische Interview. Ein Überblick über die Forschung für die Praxis. Report Psychologie, 25, 1/2000, 18 – 24

Westin, A. F. (1970). Privacy and freedom. New York

Wiendieck, G. (1994). Arbeits- und Organisationspsychologie. Berlin, München: Quintessenz

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, 72, 484-487

Wottawa, H. (1980). Grundriß der Testtheorie. München: Juventa

Wottawa, H. & Amelang, M. (1985). Testknacker. Report Psychologie, 10, 6 - 9

Zerssen, D. v. (1975). Die Befindlichkeits-Skala. Manual. München: Beltz

Items des Bewerberfragebogen für den Polizeivollzugsdienst (BfP)

Die ersten vier Items wurden nicht bewertet und dienten dem warm up.

- 01 1) Wie alt sind Sie?
- 02 2) Geschlecht ?
- 03 3) Welcher Dienstzweig interessiert Sie am ehesten?
- 04 4) Wie viel Einwohner hatte der Ort / die Stadt, in dem / der Sie in den letzten fünf Jahren gelebt haben?

Die folgenden 56 Items waren Grundlage der Bewertung.

- 1 Wie wichtig ist Ihnen, was andere über Sie denken?
- 2 Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf im Sozialbereich zu engagieren?
- 3 Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf in der Politik zu engagieren?
- 4 Wie wichtig ist Ihnen ein großer Freundeskreis für Ihre Lebensplanung?
- 5 Wie wichtig ist Ihnen gesellschaftliches Ansehen für Ihre Lebensplanung?
- 6 Welche Vorstellung von Verantwortung teilen Sie am ehesten?
- 7 Welche der folgenden Aufgaben liegt Ihnen am ehesten?
- 8 Wie häufig verdienen Sie parallel zur Schule eigenes Geld?
- 9 Wie häufig verdienen Sie während der Ferien eigenes Geld?
- 10 Wie häufig waren Sie im Dienstleistungsbereich beschäftigt?
- 11 Wie häufig waren Sie im Verkauf beschäftigt?
- 12 Wie häufig waren Sie in Verwaltung / Büro beschäftigt?
- 13 Wie häufig waren Sie in sonstigen Bereichen beschäftigt?
- 14 Wie viele Hobbys haben Sie?
- 15 In wie viel Vereinen / Gruppen sind oder waren Sie Mitglied?
- 16 Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Tageszeitungen?
- 17 Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Wochenzeitungen?
- 18 Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Zeitschriften?
- 19 Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Büchern?
- 20 Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Fernsehen?
- 21 Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Sport?
- 22 Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Hobbys?
- 23 Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Freundschaften?
- 24 Wie oft betreiben Sie Mannschaftssportarten?
- 25 Wie oft betreiben Sie Einzelsportarten?
- 26 Haben Sie durch Freizeitaktivitäten Zusatzqualifikationen erworben?
- 27 Wie wichtig ist Ihnen das Interessengebiet Wirtschaft?
- 28 Wie wichtig ist Ihnen das Interessengebiet Politik?

- 29 Wie wichtig ist Ihnen das Interessengebiet Geschichte?
- 30 Wie wichtig ist Ihnen die Freizeitaktivität Lesen?
- 31 Wie wichtig ist Ihnen die Freizeitaktivität TV / Radio?
- 32 Wie wichtig sind Ihnen Aktivitäten im Verein als Freizeitaktivität?
- 33 Welche Sportform bevorzugen Sie?
- 34 Wie häufig sehen Sie im Fernsehen Spielfilme / Serien?
- 35 Wie häufig sehen Sie im Fernsehen Dokumentarfilme?
- 36 Wie häufig sehen Sie im Fernsehen Sportsendungen?
- 37 Wie häufig sprachen Sie mit Ihren Geschwistern über Ihre berufliche Zukunft?
- 38 Wie häufig sprachen Sie mit Ihrer(m) PartnerIn über Ihre berufliche Zukunft?
- 39 Wie häufig sprachen Sie mit Anderen über Ihre berufliche Zukunft?
- 40 Mit wie viel Jahren haben Sie sich für Ihren jetzigen Berufswunsch entschieden?
- 41 Wie viel Jahre sind Sie schon berufstätig?
- 42 Hatten Sie parallel zur Schule / Ausbildung / Beruf Funktionen inne?
- 43 Wie wichtig war für Sie das Vorhandensein von freien Stellen bei dem Wunsch, den Polizeiberuf zu ergreifen?
- 44 Wie wichtig waren für Sie die Verdienstmöglichkeiten bei dem Wunsch, den Polizeiberuf zu ergreifen?
- 45 Wie wichtig war für Sie das gesellschaftliche Ansehen bei dem Wunsch, den Polizeiberuf zu ergreifen?
- 46 Wie wichtig war für Sie der Rat von Eltern / Freunden bei dem Wunsch, den Polizeiberuf zu ergreifen?
- 47 Wie wichtig war für Sie die Berufsdarstellung in den Medien bei dem Wunsch, den Polizeiberuf zu ergreifen?
- 48 Wie wichtig war für Sie ein Praktikum bei der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?
- 49 Wie wichtig waren für Sie Verwandte bei der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?
- 50 Wie wichtig waren für Sie eigene Erfahrungen mit der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?
- 51 Welche Art der Bekanntgabe besonderer Leistungen bevorzugen Sie am ehesten?
- 52 Welche Vorstellungen von Teamarbeit teilen Sie am ehesten?
- 53 Welche Arbeitssituation bevorzugen Sie am ehesten?
- 54 Welche Zeiteinteilung bevorzugen Sie am ehesten?
- 55 Wie häufig haben Sie Erfahrungen mit mangelnder Zusammenarbeit in Schule / Ausbildung / Beruf gemacht?
- 56 Wie häufig haben Sie Erfahrungen mit Widerstand gegen neue Ideen in Schule / Ausbildung / Beruf gemacht?

Häufigkeitsverteilung der Items des Bewerberfragebogen für den Polizeivollzugsdienst (BfP)

Wie wichtig ist, was andere über Sie denken?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
gar nicht	18	2,9	2,9	2,9
kaum	181	29,4	29,4	32,4
ziemlich	369	59,9	60,0	92,4
außerordentlich	47	7,6	7,6	100,0
Gesamt	615	99,8	100,0	
Fehlend	99	1	,2	
Gesamt	616	100,0		

Wie wichtig ist Ihnen, sich im Sozialbereich zu engagieren?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
gar nicht	17	2,8	2,8	2,8
kaum	167	27,1	27,1	29,9
ziemlich	331	53,7	53,7	83,6
außerordentlich	100	16,2	16,2	99,8
99	1	,2	,2	100,0
Gesamt	616	100,0	100,0	

Wie wichtig ist Ihnen, sich in der Politik zu engagieren?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
gar nicht	49	8,0	8,0	8,0
kaum	310	50,3	50,4	58,4
ziemlich	232	37,7	37,7	96,1
außerordentlich	24	3,9	3,9	100,0
Gesamt	615	99,8	100,0	
Fehlend	99	1	,2	
Gesamt	616	100,0		

Wie wichtig ist Ihnen ein großer Freundeskreis für Ihre Lebensplanung?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
gar nicht	9	1,5	1,5	1,5
kaum	143	23,2	23,3	24,8
ziemlich	326	52,9	53,1	77,9
außerordentlich	136	22,1	22,1	100,0
Gesamt	614	99,7	100,0	
Fehlend	99	2	,3	
Gesamt	616	100,0		

Wie wichtig ist Ihnen gesellschaftliches Ansehen für Ihre Lebensplanung?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
gar nicht	12	1,9	2,0	2,0
kaum	167	27,1	27,2	29,2
ziemlich	368	59,7	59,9	89,1
außerordentlich	67	10,9	10,9	100,0
Gesamt	614	99,7	100,0	
Fehlend	99	1	,2	
System	1	,2		
Gesamt	616	100,0		

Welche Vorstellung von Verantwortung teilen Sie am ehesten?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
trage gem Verantwortung	417	67,7	67,9	67,9
trage gem Verantwortung, wenn Dritter Gesamtverantwortung	177	28,7	28,8	96,7
Verantwortung tragen liegt mir nicht	17	2,8	2,8	99,5
möchte keine Verantwortung tragen	3	,5	,5	100,0
Gesamt	614	99,7	100,0	
Fehlend	99	2	,3	
Gesamt	616	100,0		

Welche der folgenden Aufgaben liegt Ihnen am ehesten?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
Berichte schreiben	133	21,6	21,7	21,7
Ideen vermitteln	322	52,3	52,5	74,2
Konzepte entwickeln	112	18,2	18,3	92,5
Vorträge halten	46	7,5	7,5	100,0
Gesamt	613	99,5	100,0	
Fehlend	99	3	,5	
Gesamt	616	100,0		

Wie häufig verdienen Sie parallel zur Schule eigenes Geld?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
eher nie	140	22,7	23,0	23,0
eher selten	216	35,1	35,5	58,6
eher oft	150	24,4	24,7	83,2
eher immer	102	16,6	16,8	100,0
Gesamt	608	98,7	100,0	
Fehlend	99	8	1,3	
Gesamt	616	100,0		

Wie häufig verdienen Sie während der Ferien eigenes Geld?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
eher nie	112	18,2	18,5	18,5
eher selten	192	31,2	31,6	50,1
eher oft	192	31,2	31,6	81,7
eher immer	111	18,0	18,3	100,0
Gesamt	607	98,5	100,0	
Fehlend	99	9	1,5	
Gesamt	616	100,0		

Wie häufig waren Sie im Dienstleistungsbereich beschäftigt?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
eher nie	310	50,3	50,4	50,4
eher selten	103	16,7	16,7	67,2
eher oft	128	20,8	20,8	88,0
eher immer	74	12,0	12,0	100,0
Gesamt	615	99,8	100,0	
Fehlend	99	1	,2	
Gesamt	616	100,0		

Wie häufig waren Sie im Verkauf beschäftigt?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
eher nie	346	56,2	56,3	56,3
eher selten	105	17,0	17,1	73,3
eher oft	126	20,5	20,5	93,8
eher immer	38	6,2	6,2	100,0
Gesamt	615	99,8	100,0	
Fehlend	99	1	,2	
Gesamt	616	100,0		

Wie häufig waren Sie in Verwaltung / Büro beschäftigt?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
eher nie	404	65,6	65,6	65,6
eher selten	113	18,3	18,3	83,9
eher oft	67	10,9	10,9	94,8
eher immer	32	5,2	5,2	100,0
Gesamt	616	100,0	100,0	

Wie häufig waren Sie in sonstigen Bereichen beschäftigt?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
eher nie	236	38,3	38,4	38,4
eher selten	140	22,7	22,8	61,1
eher oft	158	25,6	25,7	86,8
eher immer	81	13,1	13,2	100,0
Gesamt	615	99,8	100,0	
Fehlend	99	1	,2	
Gesamt	616	100,0		

Wieviel Hobbies haben Sie?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
keine	5	,8	,8	,8
1 - 2	236	38,3	38,3	39,1
3 - 4	330	53,6	53,6	92,7
mehr als 5	45	7,3	7,3	100,0
Gesamt	616	100,0	100,0	

Wie wichtig ist Ihnen das Interessengebiet Geschichte?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
gar nicht	6	1,0	1,0	1,0
kaum	211	34,3	34,3	35,2
ziemlich	299	48,5	48,5	83,8
außerordentlich	100	16,2	16,2	100,0
Gesamt	616	100,0	100,0	

In wieviel Vereinen / Gruppen sind oder waren Sie Mitglied?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
keinen	79	12,8	12,8	12,8
1 - 2	375	60,9	60,9	73,7
3 - 4	140	22,7	22,7	96,4
mehr als 5	22	3,6	3,6	100,0
Gesamt	616	100,0	100,0	

9.1 Anhang A Durchführung: Häufigkeitsverteilung der Items des Bewerberfragebogens (BfP)

Wieviel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Tageszeitungen?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig keine	41	6,7	6,7	6,7
1 - 3	382	62,0	62,3	69,0
4 - 6	151	24,5	24,6	93,6
7 - 9	32	5,2	5,2	98,9
mehr als 10	7	1,1	1,1	100,0
Gesamt	613	99,5	100,0	
Fehlend 99	3	,5		
Gesamt	616	100,0		

Wieviel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Zeitschriften?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig keine	113	18,3	18,4	18,4
1 - 3	365	59,3	59,5	78,0
4 - 6	101	16,4	16,5	94,5
7 - 9	30	4,9	4,9	99,3
mehr als 10	4	,6	,7	100,0
Gesamt	613	99,5	100,0	
Fehlend 99	3	,5		
Gesamt	616	100,0		

Wieviel Stunden verwenden Sie pro Woche durchschnittlich für Fernsehen?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig keine	17	2,8	2,8	2,8
1 - 3	202	32,8	33,0	35,7
4 - 6	210	34,1	34,3	70,0
7 - 9	131	21,3	21,4	91,4
mehr als 10	53	8,6	8,6	100,0
Gesamt	613	99,5	100,0	
Fehlend 99	3	,5		
Gesamt	616	100,0		

Wieviel Stunden verwenden Sie pro Woche durchschnittlich für Hobbies?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig keine	16	2,6	2,6	2,6
1 - 3	131	21,3	21,4	24,0
4 - 6	209	33,9	34,1	58,1
7 - 9	154	25,0	25,1	83,2
mehr als 10	103	16,7	16,8	100,0
Gesamt	613	99,5	100,0	
Fehlend 99	3	,5		
Gesamt	616	100,0		

Wie oft betreiben Sie Mannschaftssportarten?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig gar nicht	167	27,1	27,3	27,3
1 mal monatlich	64	10,4	10,5	37,7
2 - 3 mal monatlich	77	12,5	12,6	50,3
1 mal wöchentlich	111	18,0	18,1	68,5
2 - 3 mal wöchentlich	164	26,6	26,8	95,3
mehr als 4 mal wöchentlich	29	4,7	4,7	100,0
Gesamt	612	99,4	100,0	
Fehlend 99	4	,6		
Gesamt	616	100,0		

Haben Sie Zusatzqualifikationen erworben?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig nein, keine	260	42,2	42,7	42,7
keine, außer Führerschein	170	27,6	27,9	70,6
begonnen, aber wieder abgebrochen	15	2,4	2,5	73,1
längerfristig, aber ohne Abschlussprüfung	32	5,2	5,3	78,3
längerfristig, mit Abschlussprüfung	132	21,4	21,7	100,0
Gesamt	609	98,9	100,0	
Fehlend 99	7	1,1		
Gesamt	616	100,0		

Wie wichtig ist Ihnen das Interessengebiet Politik?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig gar nicht	9	1,5	1,5	1,5
kaum	152	24,7	24,8	26,2
ziemlich	372	60,4	60,6	86,8
außerordentlich	81	13,1	13,2	100,0
Gesamt	614	99,7	100,0	
Fehlend 99	2	,3		
Gesamt	616	100,0		

Wie wichtig ist Ihnen die Freizeitaktivität Aktivität im Verein?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig gar nicht	77	12,5	12,5	12,5
kaum	167	27,1	27,1	39,6
ziemlich	218	35,4	35,4	75,0
außerordentlich	154	25,0	25,0	100,0
Gesamt	616	100,0	100,0	

Wieviel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Wochenzeitungen?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig keine	160	26,0	26,1	26,1
1 - 3	382	62,0	62,3	88,4
4 - 6	64	10,4	10,4	98,9
7 - 9	5	,8	,8	99,7
mehr als 10	2	,3	,3	100,0
Gesamt	613	99,5	100,0	
Fehlend 99	3	,5		
Gesamt	616	100,0		

Wieviel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Büchern?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig keine	107	17,4	17,5	17,5
1 - 3	279	45,3	45,5	63,0
4 - 6	148	24,0	24,1	87,1
7 - 9	56	9,1	9,1	96,2
mehr als 10	23	3,7	3,8	100,0
Gesamt	613	99,5	100,0	
Fehlend 99	3	,5		
Gesamt	616	100,0		

Wieviel Stunden verwenden Sie pro Woche durchschnittlich für Sport?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig keine	8	1,3	1,3	1,3
1 - 3	111	18,0	18,1	19,4
4 - 6	213	34,6	34,7	54,2
7 - 9	161	26,1	26,3	80,4
mehr als 10	120	19,5	19,6	100,0
Gesamt	613	99,5	100,0	
Fehlend 99	3	,5		
Gesamt	616	100,0		

Wieviel Stunden verwenden Sie pro Woche durchschnittlich für Freundschaften?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig keine	7	1,1	1,1	1,1
1 - 3	72	11,7	11,7	12,9
4 - 6	190	30,8	31,0	43,9
7 - 9	165	26,8	26,9	70,8
mehr als 10	179	29,1	29,2	100,0
Gesamt	613	99,5	100,0	
Fehlend 99	3	,5		
Gesamt	616	100,0		

Wie oft betreiben Sie Einzelsportarten?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig gar nicht	50	8,1	8,1	8,1
1 mal monatlich	56	9,1	9,1	17,3
2 - 3 mal monatlich	63	10,2	10,3	27,5
1 mal wöchentlich	104	16,9	16,9	44,5
2 - 3 mal wöchentlich	237	38,5	38,6	83,1
mehr als 4 mal wöchentlich	104	16,9	16,9	100,0
Gesamt	614	99,7	100,0	
Fehlend 99	2	,3		
Gesamt	616	100,0		

Wie wichtig ist Ihnen das Interessengebiet Wirtschaft?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig gar nicht	9	1,5	1,5	1,5
kaum	224	36,4	36,5	37,9
ziemlich	349	56,7	56,8	94,8
außerordentlich	32	5,2	5,2	100,0
Gesamt	614	99,7	100,0	
Fehlend 99	2	,3		
Gesamt	616	100,0		

Wie wichtig ist Ihnen die Freizeitaktivität Lesen?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig gar nicht	10	1,6	1,6	1,6
kaum	242	39,3	39,3	41,0
ziemlich	307	49,8	49,9	90,9
außerordentlich	56	9,1	9,1	100,0
Gesamt	615	99,8	100,0	
Fehlend 99	1	,2		
Gesamt	616	100,0		

Wie wichtig ist Ihnen die Freizeitaktivität TV / Radio?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig gar nicht	17	2,8	2,8	2,8
kaum	253	41,1	41,1	43,9
ziemlich	305	49,5	49,6	93,5
außerordentlich	40	6,5	6,5	100,0
Gesamt	615	99,8	100,0	
Fehlend 99	1	,2		
Gesamt	616	100,0		

9.1 Anhang A Durchführung: Häufigkeitsverteilung der Items des Bewerberfragebogens (BfP)

Wie häufig sehen Sie Spielfilme / Serien?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	eher nie	31	5,0	5,0	5,0
	eher selten	250	40,6	40,7	45,8
	eher oft	315	51,1	51,3	97,1
	eher immer	18	2,9	2,9	100,0
	Gesamt	614	99,7	100,0	
Fehlend	99	2	,3		
Gesamt		616	100,0		

Wie häufig sehen Sie Sportsendungen?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	eher nie	75	12,2	12,2	12,2
	eher selten	211	34,3	34,3	46,4
	eher oft	250	40,6	40,6	87,0
	eher immer	80	13,0	13,0	100,0
	Gesamt	616	100,0	100,0	

Wie häufig sprachen Sie mit Ihren Geschwistern über Ihre berufliche Zukunft?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	eher nie	126	20,5	20,5	20,5
	eher selten	206	33,4	33,6	54,1
	eher oft	203	33,0	33,1	87,1
	eher immer	79	12,8	12,9	100,0
	Gesamt	614	99,7	100,0	
Fehlend	99	2	,3		
Gesamt		616	100,0		

Wie häufig sprachen Sie mit anderen Erwachsenen über Ihre berufliche Zukunft?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	eher nie	54	8,8	8,8	8,8
	eher selten	350	56,8	57,1	65,9
	eher oft	188	30,5	30,7	96,6
	eher immer	21	3,4	3,4	100,0
	Gesamt	613	99,5	100,0	
Fehlend	99	3	,5		
Gesamt		616	100,0		

Wieviel Jahre sind Sie schon berufstätig?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gar nicht	291	47,2	48,5	48,5
	1 Jahr oder weniger	34	5,5	5,7	54,2
	1 - 3	76	12,3	12,7	66,8
	4 - 6	85	13,8	14,2	81,0
	7 oder mehr Jahre	114	18,5	19,0	100,0
	Gesamt	600	97,4	100,0	
Fehlend	99	16	2,6		
Gesamt		616	100,0		

Hatten Sie parallel zur Ausbildung etc. Posten inne?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	keine	291	47,2	49,1	49,1
	einen über kürzeren Zeitraum	85	13,8	14,3	63,4
	einen über längeren Zeitraum	184	29,9	31,0	94,4
	mehrere über längeren Zeitraum	33	5,4	5,6	100,0
	Gesamt	593	96,3	100,0	
Fehlend	99	23	3,7		
Gesamt		616	100,0		

Wie wichtig waren die Verdienstmöglichkeiten für Ihren Wunsch, den Polizeiberuf zu ergreifen?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gar nicht	22	3,6	3,8	3,8
	kaum	244	39,6	41,7	45,5
	ziemlich	301	48,9	51,5	96,9
	außerordentlich	18	2,9	3,1	100,0
	Gesamt	585	95,0	100,0	
Fehlend	99	31	5,0		
Gesamt		616	100,0		

Wie wichtig war ein Praktikum bei der Polizei für Ihren Wunsch, den Polizeiberuf zu ergreifen?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gar nicht	278	45,1	47,8	47,8
	kaum	78	12,7	13,4	61,2
	ziemlich	103	16,7	17,7	78,9
	außerordentlich	123	20,0	21,1	100,0
	Gesamt	582	94,5	100,0	
Fehlend	99	34	5,5		
Gesamt		616	100,0		

Mie wichtig waren Erfahrungen im Umgang mit der Polizei für Ihren Wunsch, den Polizeiberuf zu ergreifen?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gar nicht	152	24,7	26,1	26,1
	kaum	214	34,7	36,7	62,8
	ziemlich	149	24,2	25,6	88,3
	außerordentlich	68	11,0	11,7	100,0
	Gesamt	583	94,6	100,0	
Fehlend	99	33	5,4		
Gesamt		616	100,0		

Welche Sportarten bevorzugen Sie?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Mannschaftssportarten	423	68,7	68,8	68,8
	Einzel sportarten	192	31,2	31,2	100,0
	Gesamt	615	99,8	100,0	
Fehlend	99	1	,2		
Gesamt		616	100,0		

Wie häufig sehen Sie Dokumentarfilme?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	eher nie	54	8,8	8,8	8,8
	eher selten	284	46,1	46,2	55,0
	eher oft	242	39,3	39,3	94,3
	eher immer	35	5,7	5,7	100,0
	Gesamt	615	99,8	100,0	
Fehlend	99	1	,2		
Gesamt		616	100,0		

Wie häufig sprachen Sie mit Ihrem Partner über Ihre berufliche Zukunft?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	eher nie	112	18,2	18,2	18,2
	eher selten	96	15,6	15,6	33,9
	eher oft	209	33,9	34,0	67,9
	eher immer	197	32,0	32,1	100,0
	Gesamt	614	99,7	100,0	
Fehlend	99	2	,3		
Gesamt		616	100,0		

Mit wieviel Jahren haben Sie sich für den jetzigen Berufswunsch entschieden?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	jünger als 15	276	44,8	45,8	45,8
	16 - 17	110	17,9	18,3	64,1
	18 - 19	67	10,9	11,1	75,2
	älter als 20	149	24,2	24,8	100,0
	Gesamt	602	97,7	100,0	
Fehlend	99	14	2,3		
Gesamt		616	100,0		

Wie wichtig war das Vorhandensein freier Stellen für Ihren Wunsch, den Polizeiberuf zu ergreifen?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gar nicht	97	15,7	16,6	16,6
	kaum	263	42,7	45,0	61,6
	ziemlich	193	31,3	33,0	94,7
	außerordentlich	31	5,0	5,3	100,0
	Gesamt	584	94,8	100,0	
Fehlend	99	32	5,2		
Gesamt		616	100,0		

Wie wichtig war das gesellschaftliche Ansehen für Ihren Wunsch, den Polizeiberuf zu ergreifen?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gar nicht	22	3,6	3,8	3,8
	kaum	232	37,7	39,7	43,4
	ziemlich	280	45,5	47,9	91,3
	außerordentlich	51	8,3	8,7	100,0
	Gesamt	585	95,0	100,0	
Fehlend	99	31	5,0		
Gesamt		616	100,0		

Wie wichtig war der Rat von Eltern / Freunde für Ihren Wunsch, den Polizeiberuf zu ergreifen?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gar nicht	124	20,1	21,3	21,3
	kaum	258	41,9	44,3	65,5
	ziemlich	167	27,1	28,6	94,2
	außerordentlich	34	5,5	5,8	100,0
	Gesamt	583	94,6	100,0	
Fehlend	99	33	5,4		
Gesamt		616	100,0		

Wie wichtig war die Berufsdarstellung in den Medien für Ihren Wunsch, den Polizeiberuf zu ergreifen?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gar nicht	69	11,2	11,9	11,9
	kaum	300	48,7	51,5	63,4
	ziemlich	173	28,1	29,7	93,1
	außerordentlich	40	6,5	6,9	100,0
	Gesamt	582	94,5	100,0	
Fehlend	99	34	5,5		
Gesamt		616	100,0		

Wie wichtig waren Verwandte bei der Polizei für Ihren Wunsch, den Polizeiberuf zu ergreifen?

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	gar nicht	235	38,1	40,4	40,4
	kaum	134	21,8	23,0	63,4
	ziemlich	116	18,8	19,9	83,3
	außerordentlich	97	15,7	16,7	100,0
	Gesamt	582	94,5	100,0	
Fehlend	99	34	5,5		
Gesamt		616	100,0		

9.1 Anhang A Durchführung: Häufigkeitsverteilung der Items des Bewerberfragebogens (BfP)

Welche Vorstellungen von Teamarbeit teilen Sie am ehesten?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
macht Arbeit angenehmer	204	33,1	36,1	36,1
hält von effektiver Arbeit ab	11	1,8	1,9	38,1
verspricht eher Erfolg	248	40,3	43,9	81,9
macht keinen Unterschied	12	1,9	2,1	84,1
hängt von den Mitgliedern ab	90	14,6	15,9	100,0
Gesamt	565	91,7	100,0	
Fehlend	99	51	8,3	
Gesamt	616	100,0		

Welche Zeiteinteilung bevorzugen Sie am ehesten?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
freie Zeiteinteilung	337	54,7	60,2	60,2
vorgegebene Zeiteinteilung	223	36,2	39,8	100,0
Gesamt	560	90,9	100,0	
Fehlend	99	56	9,1	
Gesamt	616	100,0		

Wie häufig haben Sie Erfahrungen mit Widerstand gegen neuen Ideen gemacht?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
eher nie	119	19,3	21,4	21,4
eher selten	305	49,5	54,8	76,1
eher oft	123	20,0	22,1	98,2
eher immer	10	1,6	1,8	100,0
Gesamt	557	90,4	100,0	
Fehlend	99	59	9,6	
Gesamt	616	100,0		

Welche Art der Bekanntgabe von Leistungen bevorzugen Sie am ehesten?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
vor einer Gruppe	118	19,2	20,6	20,6
anonym	35	5,7	6,1	26,7
öffentlich	56	9,1	9,8	36,5
persönlich	363	58,9	63,5	100,0
Gesamt	572	92,9	100,0	
Fehlend	99	44	7,1	
Gesamt	616	100,0		

Welche Arbeitssituation bevorzugen Sie am ehesten?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
Gruppe leiten und überwachen	8	1,3	1,4	1,4
Gruppe leiten und mit ihr arbeiten	279	45,3	49,6	51,0
in einer Gruppe zu arbeiten	263	42,7	46,7	97,7
allein arbeiten	13	2,1	2,3	100,0
Gesamt	563	91,4	100,0	
Fehlend	99	53	8,6	
Gesamt	616	100,0		

Wie häufig haben Sie Erfahrungen mit mangelnder Zusammenarbeit gemacht?

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig				
eher nie	78	12,7	14,0	14,0
eher selten	281	45,6	50,5	64,6
eher oft	183	29,7	32,9	97,5
eher immer	14	2,3	2,5	100,0
Gesamt	556	90,3	100,0	
Fehlend	99	60	9,7	
Gesamt	616	100,0		

Beispielhafte Darstellung des Fragebogens zur Bestimmung der Retest-Reliabilität des Bewerberfragebogens für den Polizeivollzugsdienst (BfP)

Geburtsdatum: _____

Datum: _____

1. Allgemeines

1) Wie alt sind Sie?

≤ 18	19 - 20	21 - 22	23 - 24	25- 26	27 - 28	29 - 30	≥ 31
------	---------	---------	---------	--------	---------	---------	------

2) Geschlecht ?

männlich	weiblich
----------	----------

3) Welcher Dienstzweig interessiert Sie am ehesten?

Schutzpolizei	Kriminalpolizei	Wasserschutzpolizei
---------------	-----------------	---------------------

4) Wie viel Einwohner hatte der Ort / die Stadt, in dem / der Sie in den letzten fünf Jahren gelebt haben?

≤ 20000	21000 - 50000	51000 - 100000	100000 - 500000	≥ 500000
---------	---------------	----------------	-----------------	----------

5) Wie wichtig ist Ihnen, was andere Menschen über Sie denken?

gar nicht	kaum	ziemlich	außerordentlich
-----------	------	----------	-----------------

6) Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf im Sozialbereich zu engagieren?

gar nicht	kaum	ziemlich	außerordentlich
-----------	------	----------	-----------------

7) Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf in der Politik zu engagieren?

gar nicht	kaum	ziemlich	außerordentlich
-----------	------	----------	-----------------

8) Wie wichtig ist Ihnen ein großer Freundeskreis für Ihre Lebensplanung?

gar nicht	kaum	ziemlich	außerordentlich
-----------	------	----------	-----------------

9) Wie wichtig ist Ihnen gesellschaftliches Ansehen für Ihre Lebensplanung?

gar nicht	kaum	ziemlich	außerordentlich
-----------	------	----------	-----------------

Beispielhafte Darstellung des Fragebogens zur Bestimmung der Sozialen Validität (Akzeptanz) des Bewerberfragebogens für den Polizeivollzugsdienst (BfP)

Datum: _____

Die nachfolgende Frage würde ich in einem Bewerberfragebogen für die Vorauswahl zum Einstellungsverfahren für den Polizeivollzugsdienst	betrachten als
4) Wie viel Einwohner hatte der Ort / die Stadt, in dem / der Sie in den letzten fünf Jahren gelebt haben?	zulässig
	unzulässig
5) Wie wichtig ist Ihnen, was andere Menschen über Sie denken?	zulässig
	unzulässig
6) Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf im Sozialbereich zu engagieren?	zulässig
	unzulässig
7) Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf in der Politik zu engagieren?	zulässig
	unzulässig
8) Wie wichtig ist Ihnen ein großer Freundeskreis für Ihre Lebensplanung?	zulässig
	unzulässig

Netto- und Grobgewichte der Items des Bewerberfragebogens für den Polizeivollzugsdienst (BfP)

Extremgruppe	erfolgreich n = 143		nicht erfolgreich n = 302		Prozentdifferenz	Nettogewicht	Grobgewicht
	absolute Häufigkeiten	relative Häufigkeiten	absolute Häufigkeiten	relative Häufigkeiten			
0:Wie alt sind Sie?							
jünger gleich 18	10	10,5	99	49,3	- 38,8	-10	-
19 - 20	22	23,2	20	10,0	13,2	3	1
21 - 22	23	24,2	23	11,4	12,8	3	1
23 - 24	18	18,9	15	7,5	11,4	2	1
25 - 26	8	8,4	17	8,5	- 0,1	-	-
27 - 28	3	3,2	10	5,0	- 1,8	-1	1
29 - 30	7	7,4	7	3,5	3,9	2	1
älter gleich 31	4	4,2	10	5,0	- 0,8	-	1
0:Geschlecht?							
männlich	63	66,3	125	62,2	4,1	1	1
weiblich	32	33,7	76	37,8	- 4,1	-1	1
0:Welcher Dienstzweig interessiert Sie am ehesten?							
Schutzpolizei	68	71,6	126	63,0	8,6	2	1
Kriminalpolizei	14	14,7	56	28,0	- 13,3	-3	1
Wasserschutzpolizei	13	13,7	18	9,0	4,7	1	1
0:Wie viel Einwohner hat der Ort, in dem Sie in den letzten 5 Jahren überwiegend gelebt haben?							
kleiner gleich 20000	39	41,1	75	37,9	3,2	1	1
21000 - 50000	23	24,2	30	15,2	9,0	2	1
51000 - 100000	9	9,5	12	6,1	3,4	1	1
100001 - 500000	7	7,4	18	9,1	- 1,7	-1	1
größer 500000	17	17,9	63	31,8	- 13,9	-3	1
1:Wie wichtig ist Ihnen, was andere über Sie denken?							
gar nicht	1	1,1	4	2,0	- 0,9	-1	1
kaum	31	32,6	57	28,6	4,0	1	1
ziemlich	60	63,2	131	65,8	- 2,6	-1	1
außerordentlich	3	3,2	7	3,5	- 0,3	-	1
2:Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf im Sozialbereich zu engagieren?							
gar nicht	0	0,0	6	3,0	- 3,0	-2	1
kaum	29	30,9	57	28,5	2,4	-	1
ziemlich	59	62,8	115	57,5	5,3	1	1
außerordentlich	6	6,4	22	11,0	- 4,6	-2	1
3:Wie wichtig ist Ihnen, sich außerhalb von Schule / Beruf in der Politik zu engagieren?							
gar nicht	16	16,8	26	13,0	3,8	1	1
kaum	64	67,4	132	66,0	1,4	0	1
ziemlich	13	13,7	34	17,0	-3,3	-1	1
außerordentlich	2	2,1	8	4,0	-1,9	-2	1
4:Wie wichtig ist Ihnen ein großer Freundeskreis für Ihre Lebensplanung?							
gar nicht	3	3,2	1	0,5	2,7	2	1
kaum	39	41,1	49	24,4	16,7	4	2
ziemlich	40	42,1	108	53,7	-11,6	-3	1
außerordentlich	13	13,7	43	21,4	-7,7	-2	1
5:Wie wichtig ist Ihnen gesellschaftliches Ansehen für Ihre Lebensplanung?							
gar nicht	7	7,4	5	2,5	4,9	2	1
kaum	37	38,9	56	27,9	11,0	2	1
ziemlich	48	50,5	121	60,2	-9,7	-2	1
außerordentlich	3	3,2	19	9,5	-6,3	-2	1
6:Welche Vorstellung von Verantwortung teilen Sie am ehesten?							
Trage gern alleinige Verantwortung	77	81,1	163	81,9	-0,8	0	1
Trage gern Verantwortung, wenn ein Dritter die Gesamtverantwortung trägt	18	18,9	36	18,1	0,8	0	1
7:Welche der folgenden Aufgaben liegt Ihnen am ehesten?							
Berichte schreiben	26	27,4	44	22,0	5,4	1	1
Anderen eigene Ideen vermitteln	69	72,6	155	77,5	-4,9	-1	1
8:Wie häufig verdienen Sie parallel zur Schule eigenes Geld?							
eher nie	9	9,5	42	21,0	-11,5	-3	1
eher selten	21	22,1	89	44,5	-22,4	-5	0
eher oft	50	52,6	47	23,5	29,1	7	2
eher immer	15	15,8	22	11,0	4,8	1	1
9:Wie häufig verdienen Sie während der Ferien eigenes Geld?							
eher nie	8	8,4	41	20,5	-12,1	-3	1
eher selten	38	40,0	73	36,5	3,5	1	1
eher oft	34	35,8	56	28,0	7,8	2	1
eher immer	15	15,8	30	15,0	0,8	0	1
10:Wie häufig waren Sie im Dienstleistungsbereich beschäftigt?							
eher nie	17	17,9	91	45,5	-27,6	-6	0
eher selten	32	33,7	53	26,5	7,2	1	1
eher oft	28	29,5	38	19,0	10,5	2	1
eher immer	18	18,9	18	9,0	9,9	2	1
11:Wie häufig waren Sie im Verkauf beschäftigt?							
eher nie	32	33,7	121	60,2	-26,5	-6	0
eher selten	24	25,3	46	22,9	2,4	0	1
eher oft	28	29,5	22	10,9	18,6	4	2
eher immer	11	11,6	12	6,0	5,6	2	1
12:Wie häufig waren Sie in Verwaltung / Büro beschäftigt?							
eher nie	47	49,5	135	67,2	-17,7	-4	0
eher selten	24	25,3	48	23,9	1,4	0	1
eher oft	18	18,9	15	7,5	11,4	2	1
eher immer	6	6,3	3	1,5	4,8	3	1
13:Wie häufig waren Sie in sonstigen Bereichen beschäftigt?							
eher nie	15	15,8	74	37,2	-21,4	-5	0
eher selten	26	27,4	75	37,7	-10,3	-2	1
eher oft	49	51,6	39	19,6	32,0	7	2
eher immer	5	5,3	11	5,5	-0,2	0	1
14:Wie viele Hobbies haben Sie?							
keine	0	0,0	0	-	0,0	0	1
1 bis 2	39	41,1	103	51,5	-10,4	-2	1
3 oder mehr	56	58,9	97	48,5	10,4	2	1
15:In wie viel Vereinen / Gruppen sind oder waren Sie Mitglied?							
keine	4	4,2	17	8,5	-4,3	-2	1
1 bis 2	52	54,7	138	68,7	-14,0	-3	1
3 oder mehr	39	41,1	46	22,9	18,2	4	2
16:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Tageszeitungen?							
keine	7	7,4	16	8,0	-0,6	0	1
1 bis 3	70	73,7	153	76,1	-2,4	0	1
4 oder mehr	18	18,9	32	15,9	3,0	1	1
17:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Wochenzeitungen?							
keine	25	26,3	95	47,3	-21,0	-5	0
1 bis 3	68	71,6	97	48,3	23,3	5	2
4 oder mehr	2	2,1	9	4,5	-2,4	-2	1

9.2 Anhang B Ergebnisse: Netto- und Grobgewichte der Items des Bewerberfragebogens (BfP)

18:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Zeitschriften?							
keine	20	21,1	45	22,4	-1,3	0	1
1 bis 3	61	64,2	138	68,7	-4,5	-1	1
4 oder mehr	14	14,7	18	9,0	5,7	1	1
19:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für das Lesen von Büchern?							
keine	17	17,9	62	30,8	-12,9	-3	1
1 bis 3	41	43,2	118	58,7	-15,5	-4	0
4 oder mehr	37	38,9	21	10,4	28,5	7	2
20:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Fernsehen?							
keine	0	0,0	3	1,5	-1,5	-2	1
1 bis 3	34	35,8	90	44,8	-9,0	-2	1
4 oder mehr	61	64,2	107	53,2	11,0	2	1
21:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Sport?							
keine	2	2,1	5	2,5	-0,4	0	1
1 bis 3	13	13,7	47	23,4	-9,7	-2	1
4 oder mehr	80	84,2	149	74,1	10,1	2	1
22:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Hobbies?							
keine	0	0,0	5	2,5	-2,5	0	1
1 bis 3	18	18,9	67	33,3	-14,4	-3	1
4 oder mehr	77	81,1	129	64,2	16,9	4	2
23:Wie viel Stunden verwenden Sie pro Woche durchschnittlich für Freundschaften?							
keine	0	0,0	2	1,0	-1,0	-1	1
1 bis 3	16	16,8	42	20,9	-4,1	-1	1
4 oder mehr	79	83,2	157	78,1	5,1	1	1
24:Wie oft betreiben Sie Mannschaftssportarten?							
gar nicht	18	18,9	52	25,9	-7,0	-1	1
1 mal monatlich	8	8,4	9	4,5	3,9	2	1
2 - 3 mal monatlich	13	13,7	20	10,0	3,7	1	1
1 mal wöchentlich	25	26,3	52	25,9	0,4	-	1
2 - 3 mal wöchentlich	28	29,5	51	25,4	4,1	1	1
4 oder mehr mal wöchentlich	3	3,2	17	8,5	-5,3	-2	1
25:Wie oft betreiben Sie Einzelsportarten?							
gar nicht	2	2,1	13	6,5	-4,4	-3	1
1 mal monatlich	5	5,3	13	6,5	-1,2	-	1
2 - 3 mal monatlich	5	5,3	19	9,5	-4,2	-2	1
1 mal wöchentlich	15	15,8	37	18,4	-2,6	-1	1
2 - 3 mal wöchentlich	45	47,4	83	41,3	6,1	1	1
4 oder mehr mal wöchentlich	23	24,2	36	17,9	6,3	1	1
26:Haben Sie durch Freizeitaktivitäten Zusatzqualifikationen erworben?							
nein, keine oder nur den Führerschein	52	54,7	112	56,6	-1,9	-1	1
ja, begonnen oder erworben	43	45,3	86	43,4	1,9	1	1
27:Wie wichtig ist Ihnen das Interessengebiet Wirtschaft?							
gar nicht	1	1,1	5	2,5	-1,4	-1	1
kaum	31	32,6	99	49,5	-16,9	-4	0
ziemlich	55	57,9	80	40,0	17,9	4	2
außerordentlich	8	8,4	16	8,0	0,4	0	1
28:Wie wichtig ist Ihnen das Interessengebiet Politik?							
gar nicht	2	2,1	6	3,0	-0,9	-1	1
kaum	36	37,9	79	39,3	-1,4	0	1
ziemlich	52	54,7	106	52,7	2,0	0	1
außerordentlich	5	5,3	10	5,0	0,3	0	1
29:Wie wichtig ist Ihnen das Interessengebiet Geschichte?							
gar nicht	0	0,0	4	2,0	-2,0	-2	1
kaum	42	44,2	82	40,8	3,4	1	1
ziemlich	49	51,6	105	52,2	-0,6	0	1
außerordentlich	4	4,2	10	5,0	-0,8	-1	1
30:Wie wichtig ist Ihnen die Freizeitaktivität Lesen?							
gar nicht	3	3,2	9	4,5	-1,3	0	1
kaum	21	22,1	74	37,0	-14,9	-3	1
ziemlich	52	54,7	107	53,5	1,2	0	1
außerordentlich	19	20,0	10	5,0	15,0	5	2
31:Wie wichtig ist Ihnen die Freizeitaktivität TV / Radio?							
gar nicht	1	1,1	6	3,0	-1,9	-2	1
kaum	36	37,9	75	37,3	0,6	0	1
ziemlich	53	55,8	102	50,7	5,1	1	1
außerordentlich	5	5,3	18	9,0	-3,7	-2	1
32:Wie wichtig sind Ihnen Aktivitäten im Verein als Freizeitaktivität?							
gar nicht	2	2,1	14	7,0	-4,9	-3	1
kaum	19	20,0	40	19,9	0,1	0	1
ziemlich	41	43,2	82	40,8	2,4	0	1
außerordentlich	33	34,7	65	32,3	2,4	0	1
33:Welche Sportform bevorzugen Sie?							
Mannschaftssport	66	69,5	134	67,0	2,5	-1	1
Einzelsport	29	30,5	66	33,0	-2,5	-1	1
34:Wie häufig sehen Sie im Fernsehen Spielfilme / Serien?							
eher nie	1	1,1	9	4,5	-3,4	-2	1
eher selten	56	58,9	94	46,8	12,1	3	1
eher oft	36	37,9	94	46,8	-8,9	-2	1
eher immer	2	2,1	4	2,0	0,1	0	1
35:Wie häufig sehen Sie im Fernsehen Dokumentarfilme?							
eher nie	8	8,4	9	4,5	3,9	2	1
eher selten	39	41,1	120	59,7	-18,6	-4	0
eher oft	44	46,3	64	31,8	14,5	3	1
eher immer	4	4,2	8	4,0	0,2	0	1
36:Wie häufig sehen Sie im Fernsehen Sportsendungen?							
eher nie	15	15,8	14	7,0	8,8	3	1
eher selten	37	38,9	84	41,8	-2,9	-1	1
eher oft	40	42,1	83	41,3	0,8	0	1
eher immer	3	3,2	20	10,0	-6,8	-3	1
37:Wie häufig sprachen Sie mit Ihren Geschwistern über Ihre berufliche Zukunft?							
eher nie	7	7,4	21	10,4	-3,0	-1	1
eher selten	24	25,3	60	29,9	-4,6	-1	1
eher oft	36	37,9	82	40,8	-2,9	-	1
eher immer	17	17,9	18	9,0	8,9	2	1
Einzelkind	11	11,6	20	10,0	1,6	-	1
38:Wie häufig sprachen Sie mit Ihrer(m) PartnerIn über Ihre berufliche Zukunft?							
eher nie	2	2,1	13	6,5	-4,4	-3	1
eher selten	2	2,1	16	8,0	-5,9	-3	1
eher oft	47	49,5	74	37,0	12,5	3	1
eher immer	24	25,3	36	18,0	7,3	1	1
keine(n) PartnerIn	20	21,1	61	30,5	-9,4	-2	1
39:Wie häufig sprachen Sie mit Anderen über Ihre berufliche Zukunft?							
eher nie	1	1,1	2	1,0	0,1	0	1
eher selten	29	30,5	48	23,9	6,6	1	1
eher oft	58	61,1	130	64,7	-3,6	-1	1
eher immer	7	7,4	21	10,4	-3,0	-1	1

9.2 Anhang B Ergebnisse: Netto- und Grobgewichte der Items des Bewerberfragebogens (BfP)

40: Mit wie viel Jahren haben Sie sich für Ihren jetzigen Berufswunsch entschieden?							
jünger gleich 15	27	28,4	102	50,7	-22,3	-5	0
16 – 19	32	33,7	49	24,4	9,3	2	1
älter gleich 20	36	37,9	50	24,9	13,0	3	1
41: Wie viel Jahre sind Sie schon berufstätig?							
gar nicht	29	30,5	107	53,2	-22,7	-5	0
1 bis 3	24	25,3	29	14,4	10,9	2	1
4 oder mehr	42	44,2	65	32,3	11,9	3	1
42: Hatten Sie parallel zur Schule / Ausbildung / Beruf Funktionen inne?							
keine	17	17,9	77	38,3	-20,4	-4	0
gelegentlich über einen kürzeren Zeitraum	38	40,0	72	35,8	4,2	1	1
über einen längeren Zeitraum	40	42,1	52	25,9	16,2	4	2
43: Wie wichtig war für Sie das Vorhandensein von freien Stellen bei dem Wunsch, den Polizeiberuf zu ergreifen?							
gar nicht	20	21,1	35	17,5	3,6	1	1
kaum	43	45,3	45	22,5	22,8	5	2
ziemlich	21	22,1	63	31,5	-9,4	-2	1
außerordentlich	11	11,6	57	28,5	-16,9	-4	0
44: Wie wichtig waren für Sie die Verdienstmöglichkeiten bei dem Wunsch, den Polizeiberuf zu ergreifen?							
gar nicht	8	8,4	18	9,0	-0,6	0	1
kaum	45	47,4	107	53,2	-5,8	-1	1
ziemlich	39	41,1	74	36,8	4,3	1	1
außerordentlich	3	3,2	2	1,0	2,2	2	1
45: Wie wichtig war für Sie das gesellschaftliche Ansehen bei dem Wunsch, den Polizeiberuf zu ergreifen?							
gar nicht	8	8,4	29	14,4	-6,0	-1	1
kaum	43	45,3	85	42,3	3,0	1	1
ziemlich	40	42,1	76	37,8	4,3	1	1
außerordentlich	4	4,2	11	5,5	-1,3	0	1
46: Wie wichtig war für Sie der Rat von Eltern / Freunden bei dem Wunsch, den Polizeiberuf zu ergreifen?							
gar nicht	14	14,7	37	18,4	-3,7	-1	1
kaum	50	52,6	80	39,8	12,8	3	1
ziemlich	25	26,3	71	35,3	-9,0	-2	1
außerordentlich	6	6,3	13	6,5	-0,2	0	1
47: Wie wichtig war für Sie die Berufsdarstellung in den Medien bei dem Wunsch, den Polizeiberuf zu ergreifen?							
gar nicht	16	16,8	38	18,9	-2,1	0	1
kaum	59	62,1	115	57,2	4,9	1	1
ziemlich	17	17,9	43	21,4	-3,5	-1	1
außerordentlich	3	3,2	5	2,5	0,7	0	1
48: Wie wichtig war für Sie ein Praktikum bei der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?							
gar nicht	58	61,1	82	41,0	20,1	4	2
kaum	17	17,9	22	11,0	6,9	1	1
ziemlich	8	8,4	52	26,0	-17,6	-4	0
außerordentlich	12	12,6	44	22,0	-9,4	-2	1
49: Wie wichtig waren für Sie Verwandte bei der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?							
gar nicht	38	40,4	101	50,5	-10,1	-2	1
kaum	27	28,7	45	22,5	6,2	1	1
ziemlich	23	24,5	46	23,0	1,5	0	1
außerordentlich	6	6,4	8	4,0	2,4	1	1
50: Wie wichtig waren für Sie eigene Erfahrungen mit der Polizei bei dem Wunsch, den Polizeiberuf zu ergreifen?							
gar nicht	8	8,4	37	18,4	-10,0	-2	1
kaum	38	40,0	75	37,3	2,7	1	1
ziemlich	40	42,1	62	30,8	11,3	2	1
außerordentlich	9	9,5	27	13,4	-3,9	-1	1
51: Welche Art der Bekanntgabe besonderer Leistungen bevorzugen Sie am ehesten?							
Bekanntgabe vor einer Gruppe	36	37,9	81	40,7	-2,8	-1	1
persönliche Bekanntgabe	59	62,1	118	59,3	2,8	1	1
52: Welche Vorstellungen von Teamarbeit teilen Sie am ehesten?							
macht Arbeit eher angenehmer	18	18,9	52	25,9	-7,0	-1	1
verspricht eher Erfolg als Einzelarbeit	66	69,5	101	50,2	19,3	4	2
Erfolg hängt von den einzelnen Mitgliedern ab	11	11,6	48	23,9	-12,3	-3	1
53: Welche Arbeitssituation bevorzugen Sie am ehesten?							
eine Gruppe zu leiten und mit ihr zu arbeiten	53	55,8	116	58,0	-2,2	0	1
in einer Gruppe zu arbeiten, ohne diese zu führen	42	44,2	83	41,5	2,7	1	1
54: Welche Zeiteinteilung bevorzugen Sie am ehesten?							
freie Zeiteinteilung	56	58,9	98	49,2	9,7	2	1
vorgegebene Zeiteinteilung	39	41,1	101	50,8	-9,7	-2	1
55: Wie häufig haben Sie Erfahrungen mit mangelnder Zusammenarbeit in Schule / Ausbildung / Beruf gemacht?							
eher nie	6	5,3	21	10,4	-5,1	-2	1
eher selten	66	69,5	124	61,7	7,8	2	1
eher oft	24	25,3	55	27,4	-2,1	-1	1
eher immer	0	0,0	1	0,5	-0,5	-1	1
56: Wie häufig haben Sie Erfahrungen mit Widerstand gegen neue Ideen in Schule / Ausbildung / Beruf gemacht?							
eher nie	6	6,3	18	9,0	-2,7	-1	1
eher selten	52	54,7	141	70,5	-15,8	-4	0
eher oft	35	36,8	37	18,5	18,3	4	2
eher immer	2	2,1	4	2,0	0,1	0	1

Ergebnisse der Bestimmung der Retest-Reliabilität der Items des Bewerberfragebogens für den Polizeivollzugsdienst (BfP)

Fragen des BfP		Mittelwert	Standardabweichung	Korrelation
1	Wichtig was Menschen über Sie denken	2,80	,46	,5926
2	Engagement Sozialbereich	2,61	,53	,7312
3	Engagement Politik	1,89	,56	,6490
4	Freundeskreis	3,06	,67	,7781
5	gesellschaftliches Ansehen	2,80	,39	,5065
6	Verantwortung	1,31	,47	,7063
7	Aufgaben	1,91	,28	,5209
8	Parallel Schule Geld verdienen	2,38	1,03	,7848
9	Parallel Ferien Geld verdienen	2,38	,96	,7028
10	Dienstleistungsbereich beschäftigt	2,70	,97	,6593
11	Verkauf beschäftigt	2,12	,96	,6133
12	Verwaltung beschäftigt	1,65	,81	,7256
13	Sonstig beschäftigt	2,21	,85	,6518
14	Wie viele Hobbies	2,40	,53	,6761
15	Mitglied in Vereinen	2,29	,54	,7512
16	Lesen von Tageszeitungen	1,95	,46	,6022
17	Lesen von Wochenzeitungen	1,70	,54	,5552
18	Lesen von Zeitschriften	1,87	,61	,6341
19	Lesen von Büchern	1,95	,72	,7234
20	Fernsehen	2,59	,53	,7741
21	Sport	2,78	,46	,6359
22	Hobbies	2,78	,46	,4394
23	Freundschaften	2,89	,31	,7090
24	Mannschaftssportarten	3,19	1,54	,7601
25	Einzel sportarten	4,36	1,25	,6746
26	Zusatzqualifikationen	1,51	,50	,6178
27	Interessengebiet Wirtschaft	2,42	,54	,6419
28	Interessengebiet Politik	2,34	,56	,7038
29	Interessengebiet Geschichte	2,53	,54	,7681
30	Lesen als Freizeitaktivität	2,74	,76	,7268
31	TV als Freizeitaktivität	2,76	,63	,5270
32	Verein als Freizeitaktivität	2,97	,76	,9611
33	Mannschafts- oder Einzelsport	1,27	,45	,5747
34	Fernsehen Spielfilme	2,51	,54	,7457
35	Fernsehen Dokumentarfilme	2,53	,71	,6974
36	Fernsehen Sportsendungen	2,19	,79	,7672
37	Geschwistern berufliche Zukunft	3,04	1,12	,6560
38	Partner berufliche Zukunft	3,29	,99	,6546
39	Andere berufliche Zukunft	2,65	,56	,7401
40	Alter Berufswunsch	2,25	,82	,7132
41	Jahre berufstätig	2,14	,88	,8829
42	Parallel Ausbildung Funktionen	2,00	,78	,7382
43	Freie Stellen Polizeiberuf	2,36	,96	,7852
44	Verdienst Polizeiberuf	2,61	,57	,7330
45	Gesellschaftliches Ansehen	2,55	,61	,6432
46	Rat Eltern	2,29	,71	,5743
47	Mediendarstellung	2,10	,63	,6717
48	Praktikum	1,68	,91	,7985
49	Verwandte	1,91	,92	,6028
50	Eigene Erfahrungen	2,57	,77	,6964
51	Bekanntgabe Leistungen	1,57	,49	,6587
52	Teamarbeit	1,87	,61	,5385
53	Arbeitssituation	1,55	,50	,6612
54	Zeiteinteilung	1,34	,47	,8609
55	Zusammenarbeit	2,13	,55	,5679
56	Widerstand Erfahrungen	2,40	,64	,6426

Ergebnisse der Bestimmung der Sozialen Validität der Items des Bewerberfragebogens für den Polizeivollzugsdienst (BfP)

Item		Bewertet wurde als zulässig		unzulässig	
		Anzahl der Teilnehmer	Prozentsatz der Zustimmung	Anzahl der Teilnehmer	Prozentsatz der Ablehnung
1	Wichtig was Menschen über Sie denken	65	100	--	--
2	Engagement Sozialbereich	63	96,9	2	3,1
3	Engagement Politik	62	95,4	3	4,6
4	Freundeskreis	60	94,3	5	7,7
5	gesellschaftliches Ansehen	62	95,4	3	4,6
6	Verantwortung	64	98,5	1	1,5
7	Aufgaben	62	95,4	3	4,6
8	Parallel Schule Geld verdienen	57	87,7	8	12,3
9	Parallel Ferien Geld verdienen	59	90,7	6	9,2
10	Dienstleistungsbereich beschäftigt	62	95,4	3	4,6
11	Verkauf beschäftigt	60	94,3	5	7,7
12	Verwaltung beschäftigt	61	93,8	4	6,2
13	Sonstig beschäftigt	60	94,3	5	7,7
14	Wie viele Hobbies	63	96,9	2	3,1
15	Mitglied in Vereinen	59	90,7	6	9,2
16	Lesen von Tageszeitungen	64	98,5	1	1,5
17	Lesen von Wochenzeitungen	61	93,8	4	6,2
18	Lesen von Zeitschriften	62	95,4	3	4,6
19	Lesen von Büchern	63	96,9	2	3,1
20	Fernsehen	61	93,8	4	6,2
21	Sport	65	100	--	--
22	Hobbies	63	96,9	2	3,1
23	Freundschaften	62	95,4	3	6,2
24	Mannschaftssportarten	65	100	--	--
25	Einzel sportarten	64	98,5	1	1,5
26	Zusatzqualifikationen	62	95,4	3	4,6
27	Interessengebiet Wirtschaft	60	94,3	5	7,7
28	Interessengebiet Politik	62	95,4	3	4,6
29	Interessengebiet Geschichte	65	100	--	--
30	Lesen als Freizeitaktivität	64	98,5	1	1,5
31	TV als Freizeitaktivität	65	100	--	--
32	Verein als Freizeitaktivität	63	96,9	2	3,1
33	Mannschafts- oder Einzelsport	59	90,7	6	9,2
34	Fernsehen Spielfilme	59	90,7	6	9,2
35	Fernsehen Dokumentarfilme	59	90,7	6	9,2
36	Fernsehen Sportsendungen	56	86,2	9	13,8
37	Geschwistern berufliche Zukunft	60	94,3	5	7,7
38	Partner berufliche Zukunft	59	90,7	6	9,2
39	Andere berufliche Zukunft	60	94,3	5	7,7
40	Alter Berufswunsch	60	94,3	5	7,7
41	Jahre berufstätig	63	96,9	2	3,1
42	Parallel Ausbildung Funktionen	64	98,5	1	1,5
43	Freie Stellen Polizeiberuf	59	90,7	6	9,2
44	Verdienst Polizeiberuf	61	98,5	4	6,2
45	Gesellschaftliches Ansehen	61	98,5	4	6,2
46	Rat Eltern	62	95,4	3	4,6
47	Mediendarstellung	62	95,4	3	4,6
48	Praktikum	61	98,5	4	6,2
49	Verwandte	57	87,7	8	12,3
50	Eigene Erfahrungen	61	98,5	4	6,2
51	Bekanntgabe Leistungen	55	84,6	10	15,4
52	Teamarbeit	65	100	--	--
53	Arbeitssituation	60	94,3	5	7,7
54	Zeiteinteilung	60	94,3	5	7,7
55	Zusammenarbeit	61	98,5	4	6,2
56	Widerstand Erfahrungen	59	90,7	6	9,2

Lebenslauf

Name	Rainer Petersen	
Familienstand	verheiratet, zwei Kinder	
Schulbildung	1966 – 1970	Grundschule in Husum
	1970 – 1977	Realschule Husum Süd, Realschulabschluss
	1977 – 1980	Fachgymnasium der Beruflichen Schulen des Kreises Nordfriesland, Allgemeine Hochschulreife
Studium	1990 – 1991	Studium der Rechtswissenschaft und Soziologie an der CAU zu Kiel
	1991 – 1997	Studium der Psychologie an der CAU zu Kiel Abschluss der Diplomprüfung am 21.10. 1997
Berufsausbildung /	1981 – 1983	Zivildienst auf einer interdisziplinären Intensivstation
Berufserfahrung	1983 – 1984	Verschiedene Tätigkeiten
	1984 – 1987	Ausbildung zum Krankenpfleger im Kreiskrankenhaus Heide,
	1987 – 1998	Examiniertes Krankenpfleger im Klinikum der CAU zu Kiel
	WS 98 /99	Lehrbeauftragter an der CAU zu Kiel Studienfachfach Rechtspsychologie
	01.04.1998 – z.Zt.	Leiter des Personal-Auswahl-Centers der Polizei Hamburg
	WS 99 – z.Zt.	Lehrbeauftragter an der FHÖV Hamburg - Fachbereich Polizei - Studienfach Psychologie