

Universität Mannheim

Nutzen von Personalauswahlprogrammen.

**Diskussion der Nutzenmodelle und Bewertung eines
Assessment Centers.**

D i p l o m a r b e i t

Lehrstuhl Psychologie II

betreut von und vorgelegt bei Herrn Prof. Dr. Heinz Holling

Zweitbegutachtung durch Herrn Prof. Dr. Walter Bungard

vorgelegt von

Wolfram Reiners

im Januar 1994

Wolfram Reiners
Rheinhäuserstr. 33
68165 Mannheim

Für meine Eltern
Gerd und Inge Reiners

Man stelle sich einen Garten vor, mit
hunderterlei Bäumen, mit tausenderlei Blumen,
hunderterlei Obst, hunderterlei Kräutern. Wenn
nun der Gärtner dieses Gartens keine andre
botanische Unterscheidung kennt als 'eßbar' und
'Unkraut', dann wird er mit neun Zehnteln
seines Gartens nichts anzufangen wissen, er wird
die zauberhaftesten Blumen ausreißen, die
edelsten Bäume abhauen oder wird sie doch
hassen und scheel ansehen.

(Hermann Hesse im 'Tractat vom Steppenwolf')

Inhaltsverzeichnis

Verzeichnis der Abbildungen	X
Verzeichnis der Tabellen	XII
Vorwort	XIV
1 Einleitung	1
2 Personalauswahl als Diagnostik oder Intervention?	6
2.1 Psychologische Diagnostik als Messen und Entscheiden	6
2.1.1 Der Begriff Diagnostik	7
2.1.2 Psychologische Diagnostik	8
2.1.3 Berufliche Eignungsdiagnostik.....	10
2.2 Psychologische Diagnostik als Teil der Intervention.....	12
2.2.1 Personalauswahl.....	14
2.2.2 Personalauswahlprogramme als Intervention im System	15
2.2.3 Arten von Entscheidungen.....	21
2.2.3.1 Individuelle vs. institutionelle Entscheidungen	22
2.2.3.2 Univariate vs. multivariate Informationsrepräsentation	24
2.3 Überprüfung psychologischer Interventionsmaßnahmen.....	26
2.3.1 Evaluation	27
2.3.1.1 Das Konzept und verwandte Begriffe.....	27
2.3.1.2 Evaluation von Personalauswahlprogrammen.....	29
2.3.2 Die Validierungsstudie als spezielle Form der Evaluation	31
2.3.2.1 Kriterienorientierte Validitätsbestimmung eines eignungsdiagnostischen Verfahrens bei der Personalauswahl.....	33
2.3.2.2 Kritik der kriterienorientierten Validitätsbestimmung.....	36
2.3.3 Ansätze zur Validitätsgeneralisierung	37
2.3.4 Nutzenanalyse als spezielle Form der Evaluation	38
2.4 Funktion und Ziele von Nutzenmodellen.....	40

3	Historische Entwicklung der Nutzenmodelle	42
3.1	Der Validitätskoeffizient und verwandte Konzepte	42
3.1.1	Der Validitätskoeffizient.....	42
3.1.2	Funktionen des Validitätskoeffizienten	43
3.1.3	Kritik der Konzepte.....	45
	3.1.3.1 Interpretation des Validitätskoeffizienten durch Brogden (1946a) .	45
	3.1.3.2 Keine Bewertung ohne Beachtung der Situation.....	47
	3.1.3.3 Entscheidungstheoretische Interpretation	48
3.2	Das Taylor-Russel Modell	50
3.2.1	Darstellung des Modells	50
3.2.2	Kritik des Modells.....	53
3.3	Das Konzept von Brogden	54
3.3.1	Der durchschnittliche standardisierte Kriteriumswert der ausgewählten Bewerber.....	54
3.3.2	Das Dollarkriterium	57
3.3.3	Anmerkungen zum Modell	60
3.4	Der entscheidungstheoretische Ansatz und das Nutzenmodell von Cronbach und Gleser.....	60
3.4.1	Der entscheidungstheoretische Rahmen	61
3.4.2	Das Nutzenmodell.....	63
3.4.3	Annahmen des BCG-Personalauswahlnutzenmodells.....	65

4	Parameteroperationalisierung im BCG-Nutzermodell.....	67
4.1	Die Selektionsquote und Lambda	68
4.2	Die Validität	68
4.3	Die Kosten.....	70
4.3.1	Feste Kosten und variable Kosten.....	71
4.4	Der Skalierungsfaktor SDy als Achillesferse.....	73
4.4.1	Definitionsprobleme der Leistung y	75
4.4.2	Wie beeinflusst die Leistung y den Ertrag?	77
4.4.2.1	Ertrag der Leistung durch Kostenreduktion.....	77
4.4.2.2	Der Verkaufswert der Leistung erbringt den Ertrag	79
4.4.2.3	Der Deckungsbeitrag der Leistung erbringt den Ertrag.....	84
4.4.3	Die Leistung y - Ordnungsversuche	85
4.5	Methoden zur Ermittlung von SDy.....	89
4.5.1	Der Kostenrechnungsansatz.....	90
4.5.2	Die GLOBAL-Methode	92
4.5.3	Proportionale Schätzung: Die 40-70 Regel.....	95
4.5.4	Individuelle Schätzmethoden.....	96
4.5.3.1	Die CREPID-Methode	96
4.5.3.2	Individuelle Leistungsschätzung bei Verkaufspersonal.....	98
4.6	Fazit.....	100

5	Weiterentwicklung des BCG-Nutzernodells	101
5.1	Formalisierte Erweiterungen des BCG-Modells durch ökonomische Konzepte	102
5.1.1	Intervention als Investition	103
5.1.2	Variable Kosten der Leistung, Zinsen und Steuern	106
5.1.3	Das Mitarbeiterflußmodell.....	107
5.1.4	Integration von festen und variablen Investitionskosten	112
5.2	Nutzen der Personalauswahl jenseits der Leistungssteigerung.....	115
5.2.1	Nutzen innerhalb der Organisation	116
5.2.2	Der Arbeitsmarkt als limitiertes offenes System	117
5.3	Nutzen allgemeiner psychologischer Interventionsmaßnahmen.....	121
5.3.1	Klassifikation von Interventionsprogrammen im Personalbereich.....	121
5.3.2	BCG-Nutzernodelle bei der Veränderung des bestehenden Personals.....	123
5.4	Nebennutzen psychologischer Intervention	125
5.4.1	Rückschau der bisher betrachteten Nutzenformen	125
5.4.2	Nutzen der dritten Art.....	127
5.4.3	Nebennutzen am Beispiel eines Auswahl-AC	128

6	Nutzen der Personalauswahl durch ein Assessment Center: Eine empirische Untersuchung.....	132
6.1	Was ist ein Assessment Center?.....	132
6.2	Beschreibung des AC.....	134
6.2.1	Der Ablauf des AC und die vorliegenden Daten.....	134
6.2.1.1	Ablauf und Prädiktoren.....	134
6.2.1.2	Umsatzzahlen als Berufserfolgsindikatoren.....	137
6.2.2	Beschreibung der Stichprobe.....	139
6.2.3	Klinische vs. statistische Urteilsbildung.....	139
6.3	Evaluation der Personalauswahl durch das AC.....	141
6.3.1	Validierungsstudie.....	142
6.3.2	Diskussion der Validierungsstudie.....	146
6.4	Nutzenanalyse.....	148
6.4.1	Das verwendete Nutzenmodell.....	149
6.4.2	Operationalisierung der Parameter.....	149
6.4.2.1	Die Selektionsquote und Lambda.....	149
6.4.2.2	Behandlungsperioden, Anzahl der Bewerber und behandelten Mitarbeiter.....	150
6.4.2.3	Die Kosten.....	150
6.4.2.4	Die Validität.....	153
6.4.2.5	Steuersatz und Kalkulationszinsfuß.....	154
6.4.3	Operationalisierung von SD_y	154
6.4.3.1	Standardabweichung der Beitragseinnahmen zwischen den AM (SD_w).....	155
6.4.3.2	Anteil der variablen Kosten an den Beitragseinnahmen (V).....	159
6.4.4	Ergebnisse der Nutzenanalyse.....	160
6.5	Diskussion der Nutzenanalyse.....	162
6.5.1	Break-even-points.....	163
6.5.2	Sensitivitätsanalysen.....	164
7	Fazit und Ausblick.....	167
Anhang A:	Das Turbo-Pascal-Listing eines Programms zur Nutzenanalyse auf Grundlage des Nutzenmodells der Gleichung (29).....	170
Literatur.....		177

Verzeichnis der Abbildungen

Abbildung 1:	Vereinfachte Darstellung des Modells der psychologischen Interventionsstrategien von Pawlik (1976) mit einigen Interventionsbeispielen.....	14
Abbildung 2:	Prozeßmodell der Personalauswahl. Werden die jeweiligen Prozeßstufen konkretisiert, so spricht man von einem Personalauswahlprogramm.	17
Abbildung 3:	Die Erfolgsquote von Personalauswahlentscheidungen hängt von verschiedenen situativen Bedingungen ab. Dies ist die Validität r_{xy} , die Selektionsquote Q , und die Basisrate B , die hier Grundquote genannt wird (aus Schaler & Funke, 1993, S. 271).	51
Abbildung 4:	Wert eines Testverfahrens, ausgedrückt anhand der Erfolgsquote, bei der Basisrate $B=50\%$. Es sind verschiedene Validitätswerte und Selektionsquoten dargestellt (aus Taylor & Russel, 1939, S. 569).....	52
Abbildung 5:	Das Schaubild der Normalverteilungsfunktion.	56
Abbildung 6:	Darstellung des Zusammenhanges zwischen der Selektionsquote Q und \bar{z}_x^* . Das Schaubild sagt z.B. aus, daß die selektierte Stichprobe bei einer Selektionsquote von 30% einen durchschnittlichen standardisierten Prädiktorwert $\bar{z}_x^* \approx 1,2$ hat, sofern die Grundannahmen (Normalverteilung, Auswahl von oben nach unten) nicht verletzt sind.....	57
Abbildung 7:	Erwartete Nützlichkeit eines Auswahlverfahrens in Abhängigkeit der Validität, der Selektionsquote und der Testkosten. Die Nützlichkeit ist als Einsparung pro ausgewähltem Mitarbeiter ausgedrückt; die Bedeutung dieses Maßes kann dem Text entnommen werden (aus: Brogden, 1949 S. 174).	59

Abbildung 8:	Der Zusammenhang zwischen dem Anteil des Deckungsbeitrages y am Verkaufswert der Leistung w und dem Parameter V (Anteil der variablen Kosten s am Verkaufswert der Leistung w).....	88
Abbildung 9:	Der Arbeitsmarkt als limitiertes offenes System. Verschiedene Organisationen geben und entnehmen ihm Leistungsressourcen für bestimmte Berufstätigkeiten.....	119
Abbildung 10:	Der Scree-Test und das Kaiser-Guttman Kriterium schlagen vor, zwei Faktoren aus den sechs Übungsergebnissen zu extrahieren. Abgetragen sind auf der Ordinate die Eigenwerte der Faktoren und Einzelrestfaktoren nach der Hauptkomponentenanalyse.	145
Abbildung 11:	Die rotierten Faktorenladungen der sechs AC-Übungen. Faktor 2 wird durch die Planungsübung (Übung 6) repräsentiert, die anderen Übungen laden hoch auf Faktor 1 ($n=130$).....	146
Abbildung 12:	Verteilung der erzielten Umsatzpunkte im ersten Jahr nach Eintritt in das Versicherungsunternehmen (V12).....	157

Verzeichnis der Tabellen

Tabelle 1:	Beispiele für psychologische Diagnostik in einigen angewandten Disziplinen der Psychologie.....	9
Tabelle 2:	Klassifikation von Entscheidungen bei Cronbach & Gleser (1965). Es sind jeweils gegensätzliche Entscheidungsstrategien gegenübergestellt.....	22
Tabelle 3:	Verschieden Verfahren der Validitätsgeneralisierung in Anlehnung an Burke et al. (1986).....	38
Tabelle 4:	Die Zuordnung des Verkaufswertes der Leistung (w) bei der CREPID-Methode. Dargestellt ist das Beispiel eines fiktiven Mitarbeiters (in Anlehnung an Cascio & Ramos, 1986, S. 22).....	97
Tabelle 5:	Illustration des Mitarbeiterflußmodells. Erklärungen der Abkürzungen im Text (in Anlehnung an Boudreau, 1983b). Gekennzeichnete Angaben in Mio \$ (10^6 \$).	111
Tabelle 6:	Die Eigenschaftsdimensionen des AC. In Klammern finden sich die Mittelwerte (M) und Standardabweichungen (s) der Eigenschaftsurteile.	136
Tabelle 7:	Die Übungen des AC. In Klammern finden sich die Mittelwerte (M) und Standardabweichungen (s) der Übungsergebnisse.....	137
Tabelle 8:	Validitätskoeffizienten der durch Gruppendiskussion ermittelten Eigenschaftsurteile und der beobachtungsorientierten AC-Übungsergebnisse bezüglich des im Text beschriebenen Außenkriteriums V12. Angegeben ist die Produkt-Moment Korrelation (Pearson), die Rangkorrelation (Spearman) und die Stichprobengröße (n).....	143

Tabelle 9:	Die Mitarbeiterkennwerte in den einzelnen Behandlungsperioden. Die mit "*" gekennzeichneten Werte sind Schätzwerte.	151
Tabelle 10:	Die variablen Kosten des PAP in den 5 Anwendungsjahren des AC. Angegeben sind geschätzte Mehrkosten gegenüber einem fiktiven einfachen PAP, welches z.B. Auswahlinterviews beinhaltet. Die Berechnungsvorschriften sind im Text genannt.	152
Tabelle 11:	Die Parameterschätzwerte der Nutzenanalyse.	161
Tabelle 12:	Nutzenanalyse des PAP. In der letzten Zeile ist der Gesamtnutzen des PAP zu finden. Die Angaben sind ohne Einheit oder in DM.	162
Tabelle 13:	Die Break-even-points der Parameter. Es sind diejenigen Parameterwerte angegeben, bei welchen das PAP gerade noch einen positiven Nutzen erbringt, während jeweils alle anderen Parameter mit ihrem ursprünglichen Schätzwert (s. Tabelle 11) verrechnet werden.	164
Tabelle 14:	Die Sensitivität der Parameter des untersuchten Personalauswahlprogramms.	165

Vorwort

Wenn man zum erstenmal eine solche Arbeit schreibt, dann erwarten einen viele neue und unerwartete Erfahrungen. Man will viele Dinge, welche einen oft seit Jahren beschäftigen und faszinieren, mit dem eigentlichen Inhalt verflechten. Dies ist natürlich in den seltensten Fällen möglich und es fällt nicht leicht, angefangene oder bereits fertige Abschnitte wieder streichen zu müssen. Zum anderen erfordert die Arbeit aber auch die Behandlung von vermeintlich altbekannten und sicher beherrschten Themen. Müssen Gedanken jedoch schriftlich festgehalten werden, ist oft überraschend, wie unvollständig ihre Verarbeitung noch geblieben ist.

Ich bemühe mich, klar und verständlich zu schreiben. Manchmal fällt das sehr schwer, weil man dadurch deutlich Stellung bezieht und angreifbar wird. Darüber hinaus steht eine klare und eindeutige Sprache nicht immer im Einklang mit stilistischen Anforderungen. Es wird hier, wie bei jeder anderen wissenschaftlichen Arbeit auch, ersterem Priorität eingeräumt. Aus demselben Grund wird im Laufe der Arbeit nur die männliche Geschlechtsform verwandt. Ich bitte dies als technische Notwendigkeit zu verstehen.

Nutzenmodelle können deshalb faszinieren, weil sie zu den seltenen Theorien innerhalb der Psychologie gehören, die ihre Aussagen in formalisierter Form machen und Maßeinheiten benutzen. Dadurch sind die Modelle leicht kritisierbar. Nicht zuletzt diesem Umstand ist es wohl zu verdanken, daß sie immer wieder weiterentwickelt und verfeinert wurden. Popper schreibt (1984, S. 54) "Eine Theorie ist um so kühner, je größer ihr Gehalt ist. Sie ist auch um so riskanter: Es ist von Anfang an wahrscheinlicher, daß sie falsch ist." Nutzenmodelle muß man innerhalb der Psychologie sicherlich zu den 'kühnsten' Modellen zählen.

Schon im Vorfeld des Entstehens dieser Arbeit wurde immer wieder kritisch gefragt, warum ein Psychologe eine derart methodisch orientierte und inhaltlich fast schon 'wirtschaftswissenschaftliche' Arbeit schreibt. Darauf möchte ich zwei Antworten geben. Zum ersten beschäftigt sich die Organisationspsychologie u.a. mit den Vorgängen in wirtschaftlichen Organisationen. Der Inhalt dieser Arbeit verläßt also keineswegs Kernanwendungsgebiete der Psychologie. Zweitens halte ich die methodische Entwicklung innerhalb der Organisationspsychologie für elementar wichtig. Diese ist vielleicht für dieses junge Fachgebiet noch wichtiger als der inhaltliche Fortschritt, weil dadurch eine Abgrenzung von konkurrierenden Nachbardisziplinen erreicht werden kann. Die Organisationspsychologie sollte gewissermaßen neue Wege zur Lösung alter (inhaltlicher) Probleme finden. Auf diese

Weise können in ihr vielleicht Problemlösungen erarbeitet werden, die wirklich neu und in eher inhaltsorientierten Nachbardisziplinen nicht zu finden sind.

Eine Diplomarbeit hat nicht nur einen Inhalt. Sie ist auch ein Symbol. Ein Symbol für das Ende eines Lern- und Lebensabschnitts. In meiner Studienzeit habe ich viel Neues erfahren dürfen. Ich möchte an dieser Stelle den wichtigsten Personen danken, die mir diese Erfahrungen ermöglichten oder diese entscheidend beeinflussten. Prof. Dr. Wilhelm Glaser prägte mich in den ersten Semestern durch seine steten Bemühungen, seinen Studenten die psychologische Methodenlehre nahezubringen. Insbesondere aber seine wissenschaftstheoretischen Auffassungen, die er nicht nur lehrte sondern auch vertrat, machten und machen großen Eindruck auf mich. Meine ersten ausgiebigen praktischen Erfahrungen in der Organisationspsychologie machte ich bei Priv.-Doz. Dr. Rudolf Günther. Bei ihm mußte ich keinen 'Praxischock' erleben, weil er es versteht, wissenschaftliche Erkenntnisse und Methoden bei der Lösung praktischer Probleme gewinnbringend einzusetzen. Nach meinem Wechsel von der Uni Tübingen zur Uni Mannheim unterstützte und förderte mich Prof. Dr. Heinz Holling in meinen Studienangelegenheiten. Er riet mir auch zum Thema dieser Arbeit und betreute sie. Ich möchte mich an dieser Stelle bei ihm bedanken. Während meines gesamten Studiums begleitete mich mein Kollege und Freund cand. psych. Armin Trost. Ihm habe ich einen großen Teil meiner fachlichen und persönlichen Entwicklung zu verdanken. In ungezählten Gesprächen und Diskussionen wurde er nie müde, sich für die Psychologie und angrenzende Gebiete zu begeistern. Er war dabei genauso kreativ wie kritisch. Meinen Eltern Gerd und Inge Reiners danke ich dafür, daß sie mir mein Studium ermöglicht haben. Es würde Seiten füllen, deren Leistungen aufzuzählen. Ihnen ist diese Arbeit gewidmet.

Vorformen der Arbeit haben außer meinem Betreuer, Prof. Dr. Heinz Holling, die Herren cand. psych. Armin Trost und cand. psych. Alexander Wick gelesen. Ich danke ihnen für die wertvollen kritischen Anmerkungen, die sie mir gegeben haben. Frau cand. psych. Anke Bruns und Herrn Gerd Reiners danke ich für orthographische Korrekturen.

1 Einleitung

Personalauswahl und Berufseignungsdiagnostik sind etablierte Bestandteile der angewandten Psychologie. Sie gehören zu den Anwendungs- und Forschungszweigen der Psychologie, mit welchen große Teile der Bevölkerung mindestens einmal in ihrem Leben Kontakt haben. Sie begegnen dem Bundesbürger an vielen Stellen seines beruflichen Werdegangs, z.B. bei der Berufsberatung durch das Arbeitsamt, bei der Suche nach einer Lehrstelle, bei der militärischen Verwendungsauswahl und -plazierung für Wehrpflichtige und bei der Stellensuche für Hochschulabgänger. Diese Arbeit handelt vom Nutzen der Personalauswahl. Ich möchte zunächst einige allgemeine Fragen im Zusammenhang mit dem Begriff 'Nutzen' formulieren.

In der deutschen Umgangssprache ist der Nutzen eines Sachverhaltes oder eines Gegenstandes gleichbedeutend mit dem, was er zur persönlichen Zielerreichung beitragen kann. Weniger akademisch formuliert, etwas Nützliches 'bringt etwas', 'man kann etwas damit anfangen' oder 'hilft jemandem bei etwas'. Die Forschungsrichtung auf der diese Arbeit basiert, nennt sich Nutzenanalyse (NA). Welches Ziel ist aber bei der NA gemeint? Für mich war bei den ersten Kontakten mit der NA überraschend, daß die NA mit ökonomischen Zielen operiert, obwohl sie sich als eine psychologische Disziplin versteht. In der Ökonomie wird jedoch der Nutzen anders definiert als in der Umgangssprache. Hier bedeutet der Nutzen eines Sachverhaltes die Erträge oder Vorteile, die er erbringt, abzüglich der Kosten oder Nachteile die seinetwegen aufgewendet werden. Meist werden alle Variablen dieses Modells in monetären Einheiten ausgedrückt. Auch in den Modellen der NA stellen Psychologen ihre wertvollen, aber doch eher 'weichen' Leistungen auf den Prüfstand, der sonst nur 'harten' Investitionen zukommt. Können psychologische Leistungen überhaupt in D-Mark umgerechnet werden? Falls es überhaupt geht, kommen dabei auch ethische Bedenken zu ihrem Recht?

Ein zweiter Sachverhalt verdient nicht weniger Aufmerksamkeit. Man kann sich nämlich an dieser Stelle fragen, wessen Nutzen die Nutzenmodelle überhaupt betrachten. Erträge und Vorteile kann es bei der Personalauswahl für alle Beteiligten geben. Ist es das Unternehmen, der Psychologe, der Beurteilte oder die Gesellschaft, welche die vermuteten Vorteile der Maßnahme für sich verbuchen kann? Falls es mehrere Beteiligte sind: Womit beschäftigte sich die psychologische NA bisher vornehmlich?

Um vorab ein ungefähres Bild dessen zu vermitteln, was im weiteren Verlauf dieser Arbeit zu erwarten ist, mögen die folgenden Ausführungen dienlich sein.

Denn nichts, so wissen die Gedächtnispsychologen, ist dem Verständnis neuer Inhalte dienlicher, als eine vorab bestehende Struktur des Kommenden, ein Schema, in welches neue Lerninhalte eingeordnet werden können.

Im *zweiten Kapitel* sollen zunächst die Begriffe der psychologischen Diagnostik und der Eignungsdiagnostik, wie sie heute im allgemeinen verstanden werden, geklärt werden. Die Fachliteratur zeichnet ein teilweise inhomogenes und unscharf definiertes Bild von diesen Begriffen. Meist wird bei der psychologischen Diagnostik der Aspekt des Messens und der damit verbundenen Probleme in den Vordergrund gehoben. Symptomatisch für ein solches Fachverständnis ist die Tatsache, daß sich die universitäre Ausbildung im Fachgebiet der psychologischen Diagnostik nahezu ausschließlich mit Problemen der Testtheorie, Testkonstruktion, Testvorgabe, Meßgenauigkeit, Urteilsfehlern, etc. beschäftigt. Diese Sichtweise kann ein ganzheitliches Bild der Diagnostik behindern, wichtige Aspekte aus dem Aufmerksamkeitsfokus verdrängen und damit dysfunktional wirken. Aus diesem Grunde schließt sich meine Auffassung der Sichtweise an, die psychologische Diagnostik als Teil der psychologischen Intervention versteht. Dies hat den Vorteil, daß durch sie eine umfassendere Sichtweise des Inhalts, der Ziele und der Wirkung der Diagnostik gewonnen werden kann.

Wie der Titel der Arbeit anzeigt, beschränke ich mich hier auf einen Teilbereich der psychologischen Intervention, die Personalauswahl. Viele der nachfolgenden Fragen und Antworten sind auf andere Teilbereiche anwendbar. Aufgrund des Verständnisses der psychologischen Diagnostik als Teil der Intervention lassen sich die Inhalte dieser Arbeit vergleichsweise einfach auf andere Bereiche psychologischer Intervention (Weiterbildung, Training, Arbeitsplatzgestaltung, Psychotherapie, ...) anwenden. Dennoch bleibt die Beschränkung auf die Personalauswahl unumgänglich. Personalauswahl wird von mir als psychologische Intervention aufgefaßt. Intervention umfaßt mehr, als nur den Teilaspekt des Messens. Sie findet immer unter gewissen Umweltbedingungen statt. Mit anderen Worten, Intervention findet im System statt. Dieses System kann z.B. das sozio-technische System eines Unternehmens sein oder der Arbeitsmarkt. Es werden vom System Elemente entnommen und zurückgegeben. Dies sind z.B. Menschen, Qualifikation, Wissen, Fertigkeiten, Fähigkeiten, Einstellungen, Werte, Rohstoffe, Produkte, ..., um nur einen kleinen Teil der Systemaustauschbeziehungen zu skizzieren. Im Kernstück des zweiten Kapitels stelle ich ein Prozeßmodell der Personalauswahl vor. Dieses macht u.a. den systemischen Charakter der Personalauswahl deutlich. Noch viel wesentlicher aber ist, daß daran die

Unzulänglichkeiten der Sichtweise 'Diagnostik als Messen' demonstriert werden können. Dieses Prozeßmodell ist für die Verständlichkeit der folgenden Ausführungen wichtig.

Den Abschluß des zweiten Kapitels bildet die Besprechung verschiedener Überprüfungsmöglichkeiten psychologischer (und auch anderer) Interventionsmaßnahmen. Auch hier spielt das Prozeßmodell eine wichtige Rolle. Zunächst wird dabei ganz allgemein auf den Begriff der Evaluation eingegangen. Dies findet seine Rechtfertigung darin, daß Evaluationsmaßnahmen in der Praxis heute eine zunehmend gewichtige Rolle einnehmen. Evaluation wird hier als Überbegriff für viele mögliche Maßnahmen verstanden. Bei der Personalauswahl sind dies insbesondere die Validierungsstudie, Ansätze zur Validitätsgeneralisierung und die Nutzenanalyse. Zusammenfassend kann man sagen, daß hier die NA als spezielle Evaluationsmaßnahme psychologischer Interventionen verstanden wird. Werden Personalauswahlprogramme mittels der NA evaluiert, steht nicht nur das klassisch-diagnostische Meßproblem im Vordergrund. Vielmehr spielt 'messen' eine untergeordnete Rolle im Systemzusammenhang der Interventionsmaßnahme.

Kapitel 3 stellt die klassischen Ansätze der NA dar. Diese Ansätze finden ihren Ausgangspunkt in der Diskussion, woran der Wert eines Tests¹ erkannt werden kann. Eine zentrale Stellung nimmt dabei das Konzept des Validitätskoeffizienten ein. Nun findet man auch unter Psychologen die Auffassung, daß das wichtigste an einem Test eine hohe Validität¹ ist, andere Aspekte jedoch zweitrangig sind. Darüber hinaus gibt es einen enormen Dissens darüber, wo eine solche 'hohe Validität' anfängt, bzw. ab welcher Validität ein Test noch vertretbar ist. Doch damit nicht genug, oft werden verwandte Konzepte des Validitätskoeffizienten zur Interpretation herangezogen. Insbesondere der sog. Determinationskoeffizient und der Alienationskoeffizient spielen hierbei eine bedeutende Rolle. Daß diese Interpretationshilfen meist nur noch mehr zur Verwirrung beitragen, kann von vielen Psychologen sicher nicht bestritten werden. Die frühen nutzenanalytischen Ansätze versuchen diese Unklarheiten zu beseitigen.

Eine weitere Erkenntnis bei der Beurteilung des Testwertes besteht darin, daß die Umstände seiner Anwendung meist wichtiger sind als seine Validität. Zwei kleine Beispiele aus der Personalauswahl sollen dies verdeutlichen. Ein Psychologe hat die Aufgabe, drei Stellen für Führungskräfte im höheren Management zu besetzen. Im ersten Fall liegt ihm ein Test mit einer Validität von 0,7 bzgl. der geforderten Aufgaben vor. Allerdings ist die

¹ Der Begriff "Test" wird im weiteren Verlauf sehr weit gefaßt. Ich verstehe den Begriff hier als ein nach wissenschaftlichen Grundsätzen konstruiertes psychodiagnostisches Verfahren zur Informationsgewinnung über Testpersonen. Es fallen also auch Arbeitsproben, Assessment Centers, Leistungsbeurteilungssysteme, und sonstige Meßinstrumente darunter.

Bezahlung im Unternehmen vergleichsweise schlecht, und es herrscht gerade Vollbeschäftigung im Lande, so daß kaum mit der Bewerbung von Spitzenleuten gerechnet werden kann. Im zweiten Fall kann er nur mit einem Meßinstrument der Validität von 0,2 aufwarten. Allerdings sind aufgrund der schlechten Beschäftigungslage sehr viele gute Bewerber auf dem Arbeitsmarkt zu finden. In welchem der Fälle werden die Ausgewählten wohl eine höhere Berufseignung haben? Dies ist möglicherweise die Situation mit dem mindervaliden Test. Es lassen sich jedoch auch andere Fragen stellen: In welchem Fall wird mehr Bewerbern unrecht getan? Mit dem mindervaliden Test werden wahrscheinlich mehr geeignete Bewerber als solche nicht erkannt und fälschlicherweise abgelehnt. Diese und ähnliche Fragen greift der Ansatz von Taylor und Russel (1939) auf und gibt durch seine Systematik Lösungen vor.

Aufbauend auf diesen und anderen Überlegungen legen schließlich Cronbach und Gleser (1965) mit ihrer berühmt gewordenen Monographie "Psychological Tests and Personnel Decisions" den endgültigen Grundstein für die NA. Der Ansatz wird deswegen vergleichsweise ausführlich dargestellt und diskutiert werden. Das Nutzenmodell von Cronbach und Gleser (1965) beschäftigt sich ebenfalls vornehmlich mit der Personalauswahl. Die Autoren greifen einen Vorschlag von Brogden (1949) auf, welcher der NA zu ihrem besonderen Wesen verholfen hat, und bauen ihn aus: Der Erfolg der Intervention wird nicht durch schwerverständliche Maßzahlen, wie z.B. Korrelationskoeffizienten ausgedrückt, sondern durch eine leichtverständliche Maßeinheit: Geld.

Alle diese Vorteile bringen natürlich auch Nachteile mit sich. Das Nutzenmodell von Cronbach und Gleser enthält einige Parameter, die nur schwer operationalisierbar sind. Die Besprechung dieser Probleme werde ich in *Kapitel 4* vornehmen. Einen besonderen Platz darin nimmt die Schätzung der Standardabweichung des Berufserfolgskriteriums in Geldeinheiten ein. Dies ist nicht nur deswegen so, weil dieser Parameter besonders schwer zu schätzen ist. Er hat unglücklicherweise auch auf das Ergebnis der NA einen wichtigen Einfluß. Es ist daher nicht verwunderlich, wenn sich die Forschungsbemühungen der letzten 20 Jahre zur NA schwerpunktmäßig mit dieser Parameterschätzung beschäftigten. Ich werde dem Rechnung tragen müssen.

In *Kapitel 5* werden neuere Entwicklungen zur NA vorgestellt. Diese beziehen sich einerseits auf die Erweiterung des Brogden-Cronbach-Gleser-Modells. Andererseits werden auch weitere Auswirkungen von Personalauswahlentscheidungen unter Nutzenkriterien diskutiert, die nicht von den Brogden-Cronbach-Gleser-Modellen berücksichtigt werden. Hier finden auch qualitative und systemische Überlegungen ihren Platz.

Kapitel 6 schließlich wird diese Dinge an einem empirischen Beispiel umzusetzen versuchen. Ein Assessment Center (AC) - Personalauswahlverfahren, mit dem Versicherungsaußendienstvertreter ausgewählt wurden, soll auf seinen Nutzen geprüft werden. Hierzu dient uns eine Längsschnittstudie. Interessant für nutzen-analytische Betrachtungen ist hierbei, daß AC nicht nur sehr modern sind, sondern darüber hinaus zu den teuersten Personalauswahlverfahren gehören. Zum Zeitpunkt der Erstellung dieser Arbeit befindet sich die Bundesrepublik Deutschland in einer bedenklichen wirtschaftlichen Lage. Für viele Unternehmen ist es in solchen Zeiten nicht nur wichtig, sondern oft auch existenznotwendig, ihre Investitionsentscheidungen richtig zu fällen. Ob es sinnvoll ist, in teure Personalauswahlverfahren zu investieren, kann sicher nie generell beantwortet werden. Einen kleinen Hinweis dafür, ob es für das untersuchte Unternehmen innerhalb einer speziellen Situation sinnvoll war (oder nicht), wahrscheinlich weiterhin sein wird (oder nicht), und für ähnliche Unternehmen in ähnlichen Situationen sein wird (oder nicht), versucht diese Untersuchung zu geben.

2 Personalauswahl als Diagnostik oder Intervention?

In diesem Kapitel soll die Einordnung der Personalauswahl in die psychologische Diagnostik vorgenommen und der Zusammenhang zur psychologisch orientierten Intervention hergestellt werden. Wie der Titel schon andeutet, wird die klassische Zuordnung der Personalauswahl zur Diagnostik kritisch hinterfragt. Der übergeordnete Begriff wissenschaftlich-psychologischen Handelns stellt für mich die Intervention dar. Sie erweitert historisch gewachsene und daher eingeengte Fragestellungen der Diagnostik, ohne die früheren zu vernachlässigen.

Ein Schwerpunkt der folgenden Betrachtungen wird thematisieren, wie psychologische Interventionsmaßnahmen der Personalauswahl kritisch überprüft werden können. Die NA wird sich als spezielle Form dieser Überprüfungsöglichkeiten erweisen. Sicherlich ist sie heute in Deutschland noch ein selten eingesetztes Verfahren. Diese Arbeit hat jedoch auch das Ziel, die Vorteile der NA aufzuzeigen, welche in den USA bereits als fester Bestandteil der Organisationspsychologie angesehen werden kann. Es ist meine feste Überzeugung, daß die NA auch in Deutschland schon in einigen Jahren zum Standardrepertoire modern arbeitender Psychologen gehören wird.

2.1 Psychologische Diagnostik als Messen und Entscheiden

Bevor auf differenzierte nutzenanalytische Modelle eingegangen werden kann, muß ein anderes Problem bewältigt werden. Wie so oft in der Psychologie werden einige Begriffe der Fachliteratur, mit denen hier gearbeitet werden soll, in sehr verschiedenen Zusammenhängen und uneinheitlicher Bedeutung gebraucht. Dies liegt in diesem Fall allerdings nicht an der mangelnden theoretischen Durchdringung der fraglichen Konzepte, sondern eher an der übermäßigen, ja fast schon umgangssprachlichen Handhabung. Es handelt sich dabei um die Bedeutung der Begriffe psychologische Diagnostik, Eignungsdiagnostik und Personalauswahl. Ich möchte kurz betrachten, wie sie heute im allgemeinen verstanden werden, und jeweils klar stellen, wie sie im Verlauf dieser Darstellung von mir aufgefaßt werden.

2.1.1 Der Begriff Diagnostik

Bevor ich zum Begriff der psychologischen Diagnostik komme, soll kurz über den Tellerrand der eigenen Disziplin hinausgesehen werden. Viele wissenschaftliche Disziplinen kennen den Begriff der Diagnostik. Einen hohen Bekanntheitsgrad hat die medizinische Diagnostik erreicht. Sie bezeichnet die Lehre vom fachgerechten Erkennen von Krankheitssymptomen und deren Zuordnung zu Ursachen. Der diagnostische Prozeß endet in der Diagnose. Diese bildet die Grundlage für die Ableitung von Behandlungsmaßnahmen.

Ein ähnliches Denkmodell der Diagnostik liegt auch technischen Disziplinen zugrunde. Der Unterschied liegt im wesentlichen nur an der Art des Zielobjektes der Diagnose. Beim Arzt ist dies normalerweise ein menschlicher Organismus, beim Techniker ein technisches Objekt. Der Techniker oder Mechaniker betreibt demgemäß Diagnostik an seinem technischen Objekt. Dieser Vorgang, meist Störungsfindung genannt, besteht ebenfalls im Erkennen von Symptomen, der Feststellung des Zustandes des Objektes und dessen Ursachenzuordnung. Die daraus resultierende Diagnose ist wiederum Grundlage für die Auswahl einer geeigneten Behandlungs- oder Störungsbehebungsmethode.

Sowohl in der modernen Medizin, als auch in der Technik werden zur Symptomerkennung z.T. komplexe Meßinstrumente angewandt. Dies ist notwendig, weil die menschliche Fähigkeit zur Beobachtung von Sachverhalten begrenzt ist. Der Mechaniker kann mit seinen Sinnesorganen beispielsweise nicht direkt die Bremskräfte des Bremssystems eines Kraftfahrzeuges wahrnehmen. Er benötigt dazu apparative Hilfsmittel, über deren Anzeigen er einige Sachverhalte des Bremssystems wahrnehmen kann. Die Zuordnung von Sachverhalten zu Maßen, welche die Meßinstrumente liefern, erfolgt nach bestimmten Regeln. Diese Regeln werden Meßtheorie genannt. Ist die Meßtheorie 'gut', so kann die Messung ein genaues Abbild des Sachverhaltes ergeben. Sie kommt damit praktisch einer direkten Beobachtung des Sachverhaltes gleich. In unserem Beispiel wäre also zu Fragen, ob der Bremskraftmeßapparat gut funktioniert und ob er so konstruiert ist, daß er nicht andere Sachverhalte, wie z.B. das Gewicht des Fahrzeugs, fälschlicherweise auch mißt. Nach der Beobachtung kann auf den Objektzustand geschlossen werden. Was dieser Zustand bedeutet, bestimmt eine neue Theorie des Diagnostikers. Ich nenne diese die Inhaltstheorie. Diese Inhaltstheorie des Diagnostikers beschreibt die Zusammenhänge von beobachteten Sachverhalten und dem Zustand des Objektes. In unserem Beispiel ist die Inhaltstheorie das Wissen des Mechanikers über die normale Funktionsweise des Bremssystems. Ist diese sehr ausgefeilt, so können auch Ursachenzuordnungen gemacht werden. Wenn er feststellen sollte, daß die Bremskraft auf den beiden Fahrzeugseiten unterschiedlich ist, so kann er dem

Sachverhalt die Bedeutung eines nicht normalen Funktionierens des Bremssystems zuschreiben. Hat er mehr Wissen, so kann er die Ursache bei einem unterschiedlichen Zustand der Bremsbeläge oder einer Fehlfunktion des Bremskraftverstärkers vermuten. Dies ist die Diagnose. Auf Grundlage der Diagnose kann nun eine Entscheidung über eine Behandlung erfolgen. Die Auswahl der Behandlungsmethode erfordert dann schließlich noch eine Theorie darüber, wie der Sachverhalt durch eine geeignete Behandlung in gewünschter Weise geändert werden kann. Diese sog. Veränderungstheorie bestimmt die Mittel der Überführung eines Istzustandes in einen Sollzustand, was das letztendliche Ziel der Diagnostik ist. Paradoxe Weise ist die Veränderungstheorie aber nach allgemeinem Verständnis nicht mehr Bestandteil der Diagnostik.

Diese Grundprinzipien treffen auch auf die psychologische Diagnostik zu. Das Objekt der Diagnostik ist im Vergleich zu anderen Disziplinen ein anderes. Eines der wichtigsten Objekte der psychologischen Diagnostik ist die Gesamtheit der geistigen Funktionen des Menschen und deren Wirkungen. Wie wir jedoch gleich sehen werden, sind auch andere Objekte denkbar. Mit dieser Objektänderung gehen einige Besonderheiten der psychologischen Diagnostik einher. Der nächste Abschnitt soll darüber Klärung bringen.

2.1.2 Psychologische Diagnostik

In vielen angewandten Teildisziplinen der Psychologie wird psychologische Diagnostik betrieben. In Tabelle 1 sind einige Teildisziplinen mit ausgewählten Bereichen der psychologischen Diagnostik aufgeführt (siehe Tabelle 1).

Tabelle 1: Beispiele für psychologische Diagnostik in einigen angewandten Disziplinen der Psychologie.

Psychologische Teildisziplin	Diagnostische Maßnahmen
Arbeits- und Organisationspsychologie	Eignungsdiagnostik Organisationsdiagnostik Arbeitsanalyse
Klinische Psychologie	Persönlichkeitsdiagnostik Differentialdiagnostik
Pädagogische Psychologie	Schuleignungsdiagnostik Bildungsbedarfsdiagnostik
Verkehrspsychologie	Fahreignungsdiagnostik
Forensische Psychologie	Glaubwürdigkeitsdiagnostik von Zeugen Persönlichkeitsdiagnostik

Die Objekte, Ziele und Inhalte sind verschieden. Dennoch teilen diese Formen der psychologischen Diagnostik viele Gemeinsamkeiten. Diese in Worte zu fassen, ist nicht einfach. Einen Versuch, dies zu leisten, unternehmen Jäger und Petermann (1992, S. 11) in ihrer Definition von psychologischer Diagnostik:

Psychologische Diagnostik ist das systematische Sammeln und Aufbereiten von Informationen mit dem Ziel, Entscheidungen und daraus resultierende Handlungen zu begründen, zu kontrollieren und zu optimieren. Solche Entscheidungen und Handlungen basieren auf einem komplexen Informationsverarbeitungsprozeß. In diesem Prozeß wird auf Regeln, Anleitungen, Algorithmen usw. zurückgegriffen. Man gewinnt damit psychologisch relevante Charakteristika von Merkmalsträgern und integriert gegebene Daten zu einem Urteil (...).

Jäger und Petermann (1992) legen hierbei das Objekt der psychologischen Diagnostik nicht fest. Sie tragen damit dem Umstand Rechnung, daß diese Objekte sehr verschieden sein können (vgl. Tabelle 1). Häufige Objekte sind die geistigen Sachverhalte von Individuen. Es können aber auch die Anforderungen eines Arbeitsplatzes sein oder die formalen Strukturen von Unternehmen. Wesentlich ist die Informationssammlung über Sachverhalte. Diese Informationssammlung erfolgt, im Gegensatz zu den Wahrnehmungen des Menschen im Alltag, systematisch und regelgeleitet und ist zielorientiert. Ziele bestehen im Gewinn eines integrierten Urteils, in der Entscheidungsvorbereitung und in der Regulation nachfolgender Handlungen.

Die Definition von Jäger und Petermann (1992) ist vergleichsweise umfassend. Sie versteht, ähnlich wie Wottawa und Hossiep (1987), die psychologische Diagnostik als problemlösendes Handeln auf der Grundlage von Messungen. Einen weiteren Aspekt hebt Jäger (1986, S. 11) hervor. Neben der Rolle der Diagnostik als Ansammlung bestimmter Methoden zur Messung von Sachverhalten, betont er insbesondere ihre spezielle Bedeutung für nachfolgende Entscheidungen:

Wir verstehen unter 'diagnostischer Prozeß' den Ablauf von Maßnahmen, um mit deren Hilfe - unter Anwendung diagnostischer Methoden - eine mit diagnostischer Zielsetzung vorgegebene Fragestellung über eine Anzahl von Zwischenschritten so zu beantworten, daß für einen Auftraggeber eine Entscheidungshilfe bzw. eine Entscheidung herbeigeführt wird.

Wie bereits in der Einleitung bemerkt wurde, finden die Bereiche der Entscheidung und der darauf beruhenden Handlung innerhalb der psychologischen Diagnostik viel zu wenig Aufmerksamkeit. Dies ist unmittelbar feststellbar, indem Fachkundige befragt werden, was ihrer Meinung nach die Kernthemen der psychologischen Diagnostik sind. Andererseits verwundert diese Begriffsauffassung nicht, wenn die Inhalte der universitären Diagnostikausbildung im Studiengang Psychologie begutachtet werden. Schwerpunktthemen sind: Testtheorie, Testkonstruktion, Testanwendung und Gutachtenerstellung, Meßfehlerquellen und ähnliche Inhalte, die nahezu ausnahmslos auf den Meßaspekt innerhalb der Diagnostik abzielen. Der Urteilsbildungs- bzw. Entscheidungsaspekt wird meist noch berührt. Spätestens die nachfolgenden Handlungen, deren Bewertung und Konsequenzen bleiben jedoch so gut wie immer unberührt. Zusammenfassend möchte ich festhalten, daß der Begriff der psychologischen Diagnostik oft viel zu sehr, und meiner Meinung nach fälschlicherweise, auf das Messen und Entscheiden bezogen wird.

2.1.3 Berufliche Eignungsdiagnostik

Nutzenanalytische Modelle sind auf dem Gebiet der Organisationspsychologie entstanden. Aus diesem Grund soll unser Interesse auf die Diagnostik in diesem Bereich zentriert werden. Unser besonderes Augenmerk soll dabei auf der beruflichen Eignungsdiagnostik liegen, weil nutzenanalytische Modelle bisher vornehmlich in diesem Bereich Verwendung gefunden haben. Wir betrachten die Eignungsdiagnostik als speziellen Typ der psychologischen

Diagnostik, d.h. was über die psychologische Diagnostik gesagt wurde, soll auch für die berufliche Eignungsdiagnostik gelten.

Zunächst soll der Begriff der beruflichen Eignungsdiagnostik näher betrachtet werden. Eckardt & Schuler (1992, S. 534) schreiben zur Definition von Berufseignungsdiagnostik:

Unter Berufseignungsdiagnostik wird die Gesamtheit aller wissenschaftlichen und wissenschaftsgeleitet-praktischen Bemühungen verstanden, die auf dem Wege über eine gedankliche Zuordnung von beruflichen Situationen zu Personen oder von Personen zu beruflichen Situationen die Ziele 'Maximierung beruflicher Zufriedenheit' und 'Maximierung beruflicher Leistung' anstreben.

In dieser Definition sind neue Schwerpunkte ersichtlich. Die Ziele der Eignungsdiagnostik werden nun deutlich spezifiziert. Eckardt und Schuler betonen die Passung von beruflichen Situationen und der darin befindlichen Personen. Eine Diagnostik kann dann zur Optimierung der Passung beitragen, wenn zu einer gegebenen Person eine entsprechende Situation gefunden wird oder umgekehrt. Eine gute Passung bezeichnet man als Eignung. Diesen Begriff erläutert Kompa (1984, S. 35):

Als Eignung einer Person läßt sich dann eine Situation definieren, bei der die personellen Leistungsvoraussetzungen mit dem Anforderungskomplex einer Arbeitstätigkeit übereinstimmen.

Entscheidend für eine gute Person-Situation Passung ist die genaue Messung der Leistungsvoraussetzungen der Person und des Anforderungskomplexes der Situation. Mit anderen Worten muß eine so verstandene Eignungsdiagnostik zunächst zum Ziel haben, möglichst genau die relevanten Kennwerte beider Sachverhalte zu erfassen.

Die Grundlage für diese Zielerreichung stellen die reichlich vorhandenen und elaborierten Methoden der psychologischen Diagnostik dar. Die theoretische Basis dieses Methodenarsenals ist die klassische Testtheorie, die z.B. bei Gulliksen (1950) und Lord & Novick (1968) formalisiert dargestellt ist. Neueren Datums sind probabilistische Testtheorien, wie z.B. das logistische Modell von Rasch (1960), oder spätere Entwicklungen (eine Zusammenstellung findet sich bei Hambleton, Swaminathan & Rogers, 1991).

Das Ziel dieser Testtheorien besteht in der Optimierung der Meß- bzw. Testgütekriterien der Meßinstrumente bzw. psychologischer Tests. Auf Grundlage der klassischen Testtheorie wurden eine Vielzahl von psychologischen Tests konstruiert und kritisiert. Eine Darstellung der dabei verwendeten Maßnahmen findet sich bei Lienert (1989, erste Aufl. 1961).

2.2 Psychologische Diagnostik als Teil der Intervention

Die Sichtweise von Diagnostik als Messen und Entscheiden greift zu kurz. Im diesem Abschnitt soll versucht werden, dies zu begründen. Darüber hinaus soll eine ganzheitliche Sichtweise der Diagnostik skizziert werden. Diese Sichtweise begreift psychologische Diagnostik als Teil der psychologischen Intervention. Dies soll verdeutlichen, daß nicht nur Meßaspekte relevant sind, sondern auch Entscheidungen, darauf fußende Handlungen, deren Bewertung und die darauf folgenden Konsequenzen.

Pawlik (1976) sieht die psychologische Diagnostik im historisch gewachsenen Zusammenhang zur angewandten Psychologie. Diese wiederum charakterisiert er durch ihr Bemühen, praktische Problemlösungen im Hinblick auf psychologische Kriterien so zu optimieren, daß die Lösungen angemessen sind. Die Durchführung dieser Problemlösungen besteht im Eingreifen in ein System durch zielgerichtetes Handeln. Dieses Handeln nenne ich Behandlung (treatment) von Sachverhalten. Ein aufeinander bezogenes Bündel von Handlungen heißt Intervention. Problemlösendes Handeln wird immer durch mehr oder weniger intensive Informationssuche begleitet. Diesen Part übernimmt die Diagnostik. Wichtig dabei ist, daß die Diagnostik dadurch nicht mehr als losgelöstes Element betrachtet wird, sondern als Teil der übergeordneten Intervention. Sie stellt wie vorher auch Informationen für Entscheidungen bereit. Sie erfüllt jedoch auch andere Funktionen.

Wird ein Problem als eine unerwünschte Abweichung eines Istzustandes von einem Sollzustand definiert, so muß vorher ein Sollzustand bekannt sein. Diesen nennen wir auch das Handlungsziel. Problemlösende Handlungen setzen also die Bildung von Zielen voraus. Wenn problemlösende Handlungen in Gang gesetzt werden, so muß von Zeit zu Zeit die Ist-Soll-Abweichung neu überprüft werden, um Anhaltspunkte für die Wirksamkeit der Maßnahme bzw. über die Zielerreichung zu erhalten. Dieser informationsgewinnende Prozeß wird mit diagnostischen Methoden durchgeführt. An dieser Stelle wird deutlich, daß die ständig wechselnden Prozesse der Zielbildung, Zustands- und Abweichungsdiagnose, bzw. -evaluation, Handlung und Handlungskorrektur sich gegenseitig bedingen und gegenseitig aufeinander einwirken. Eine künstliche Trennung in die Ablaufphasen Diagnose - Intervention - Evaluation scheint mir kaum sinnvoll. Vielmehr besteht der gesamte Prozeß in einer Intervention innerhalb eines übergeordneten Systems. Die Intervention besteht u.a. aus Handlungen mit diagnostischen Methoden.

Interessant ist, daß bei dieser Betrachtungsweise auch keine (im klassischen Sprachgebrauch) Intervention zur Intervention wird. Besteht die Intervention aus der

Handlung des 'Nicht-Agierens', des 'Abwartens', so ist auch das eine Intervention, denn sie zieht Wirkungen nach sich. Sie ist gleichberechtigt mit derjenigen, welche andere Handlungsbündel umfaßt. Wesentlich am Interventionsbegriff ist für mich das Charakteristikum, daß die Intervention zeitlich vor einer Wirkung liegt. Aus Sicht dieser Wirkung ist die Intervention bzw. das Handlungsbündel, das sie umfaßt, klar definiert.

Werden beispielsweise die Bewerber eines Unternehmens nicht auf ihre berufliche Eignung untersucht, so zeigt diese Interventionsmaßnahme ebenso ihre Wirkung, wie jede andere. Die Behandlung des Objektes 'Bewerber' ist zwar unterblieben. Die Behandlung des Objektes 'Personalbestand' wurde jedoch durchaus vorgenommen (und wahrscheinlich nicht in Einklang mit den Unternehmenszielen). Die Sichtweise, daß Diagnostik hauptsächlich aus Messen und Entscheiden besteht, würde diesen Schluß nicht nahelegen.

Die Auffassung, daß der letztendliche Zweck psychologischer Diagnostik nicht in der genauen Messung und Vorhersage von Sachverhalten, sondern in der Entscheidung über Interventionsmaßnahmen begründet ist, wurde erstmals nachdrücklich von Cronbach und Gleser (1965, 1. Aufl. 1957, eine Zusammenfassung der Monographie in deutscher Sprache findet sich bei Michel & Mai, 1968) vertreten. Sie beklagen das Defizit der Testtheorie, über diese Entscheidungen keine Aussagen zu machen, und stellen eine Entscheidungstheorie vor. Sie beziehen außer den bekannten Testgütekriterien weitere situative Variablen in ihr Entscheidungsmodell mit ein. Das Ziel der Diagnostik besteht damit nicht mehr allein in der Optimierung von Testgütekriterien, sondern in der optimalen Wahl der richtigen Behandlung des Objektes. Die Behandlung sollte so erfolgen, daß Nutzenkriterien optimiert werden. Der Nutzen von Interventionsmaßnahmen bestimmt dabei, ob die Behandlung auf Grundlage der diagnostischen Information effizient ist. Der Kern der Modellvorstellungen von Cronbach & Gleser (1965) wird in Kap. 3 ausführlich diskutiert werden.

Pawlik (1976) schlägt eine Möglichkeit vor, Interventionsmaßnahmen zu systematisieren. Er beschreibt ein Interventionsstrategiekontinuum, an dessen einem Pol die reine Selektionsstrategie, am anderen die reine Modifikationsstrategie angesiedelt ist. Selektion besteht in der Klassifikation von Sachverhalten, Modifikation in der Veränderung von Sachverhalten. Selektionsstrategien der Intervention bestehen beispielsweise in der Auswahl oder Ablehnung von Personen für verschiedene Behandlungen (Schuleignungstestprogramm, Personalauswahl, ...) oder der Auswahl von Situationen für verschiedene Behandlungen (Berufsberatung, Studienberatung, ...). Im letzteren Fall kann die Behandlung der Situation durch das Ausfüllen der Situation mit einer Person verstanden werden. Dies ist für den menschlichen Geist mit seiner egozentrischen Weltanschauung zwar

kurz verwirrend, aber logisch denkbar. Modifikationsstrategien können sich ebenfalls wieder auf verschiedene Objekte richten. Die Modifikation von Personen kann beispielsweise durch Psychotherapie, Berufsausbildung oder Training erreicht werden. Eine Modifikation der Situation z.B. durch Arbeitsplatzgestaltung oder der Änderung von Therapiebedingungen. Werden die Interventionsobjekte in die Klassen Person und Situation/ Bedingung unterteilt, so läßt sich das Pawlik'sche Modell vereinfacht darstellen.

Psychologische Intervention	Selektionsstrategie	Modifikationsstrategie
Interventionsobjekt		
Person	Personalauswahl Personalplatzierung 'Sitzenbleiben'	Personalentwicklung Berufsausbildung Psychotherapie
Bedingung	Berufsberatung Zuordnung einer Person zu einer Therapieform	Arbeitsplatzgestaltung Lernmedienvverbesserung Wechsel der Therapieform

Abbildung 1: Vereinfachte Darstellung des Modells der psychologischen Interventionsstrategien von Pawlik (1976) mit einigen Interventionsbeispielen.

Diese Arbeit beschäftigt sich mit dem Nutzen von Personalauswahlprogrammen. Personalauswahl war bisher das Hauptgebiet für Forschung und Anwendung von Nutzenmodellen. Prinzipiell können diese Modelle auch bei allen anderen Formen der psychologischen Intervention zur Anwendung kommen. Ich werde diesen Punkt in Kapitel 5 näher betrachten. Bis dahin soll im weiteren Verlauf die Personalauswahl als Interventionsmaßnahme im Vordergrund stehen. Im folgenden Abschnitt wird der Begriff näher definiert.

2.2.1 Personalauswahl

Wie in Abbildung 1 ersichtlich ist, wird Personalauswahl im Lichte unserer bisherigen Überlegungen als Maßnahme psychologischer Intervention betrachtet. Intervention beinhaltet sowohl informationssuchende Handlungen (Diagnostik als Messen und Entscheiden), als auch zustandsverändernde Handlungen (Behandlungen). Durch diese Auffassung soll 'die Abgrenzung vom Begriff der Eignungsdiagnostik klargestellt werden. Eignungsdiagnostik als Maßnahme gehört der psychologischen Diagnostik (Messen und Entscheiden) an. Eignungsdiagnostik kann demnach ein Teil der Personalauswahl sein. Die Personalauswahl

fasse ich als Interventionsmaßnahme auf. Personalauswahl umfaßt mindestens die Prozesse der Informationssammlung mit wissenschaftlichen Methoden, die Entscheidung für eine der zur Verfügung stehenden Behandlungen und die Durchführung der Behandlung. Eine ähnliche Ansicht in diesem Punkt vertritt auch Kompa (1984, S. 3):

„...Personalauswahl ... läßt sich entsprechend dieser Systematik (von Pawlik, 1976; Anm. d. Autors) als eine Zuordnung von Personen zu vorgegebenen Arbeitsplätzen verstehen. “

Personalauswahl kann natürlich auch ohne Eignungsdiagnostik geschehen (Losverfahren, graphologische Gutachten, ...). Bei der Bewertung, inwieweit dies sinnvoll ist, geben nutzenanalytische Modelle wertvolle Hinweise.

Festgelegte Regelsysteme der Personalauswahl bezeichne ich als Personalauswahlprogramme (PAP). Diese PAP beinhalten über einen bestimmten Zeitraum festgelegte Regeln der Personalauswahl, d.h. Regeln für die angestrebten Ziele, für die Handlungsoperationen, für die Art und den Umfang der Eignungsdiagnostik, für die Entscheidungsfindung und die Bewertung der Konsequenzen, welche aus der Behandlung folgen.

2.2.2 Personalauswahlprogramme als Intervention im System

Wie bereits angedeutet, müssen Interventionsmaßnahmen im Kontext der sie umgebenden Sachverhalte betrachtet werden, um sie verstehen und bewerten zu können. Um diese abstrakte Aussage mit mehr Inhalt zu füllen, soll hier der Versuch unternommen werden, die Prozesse der Personalauswahl modellhaft darzustellen. Jedes Modell ist per definitionem eine starke Vereinfachung der tatsächlichen Sachverhalte. Ich bin mir wohl darüber im klaren, daß an meinem Prozeßmodell der Personalauswahl sicherlich Einzelheiten fehlen. Trotzdem hat ein einfaches Modell den unbestreitbaren Vorteil, sehr deutliche Strukturen vorzugeben, welche in einem komplexeren Modell in der Menge der Information verschwinden. Diese Strukturen sollen dazu benutzt werden, die bisherigen Begriffsklärungen noch einmal zu verdeutlichen. Darüber hinaus sollten sie dazu dienen, die NA in Beziehung zu bereits vorhandenen Modellen der Überprüfung psychologischer Interventionsmaßnahmen zu setzen (s. Abschnitt 2.3). Die postulierten Prozesse der Personalauswahl sind in Abbildung 2 dargestellt.

Im Prozeßmodell der Personalauswahl sind acht Stufen spezifiziert. Die erste Stufe ist etwas schwierig zu benennen. Sie macht deutlich, daß keine Organisation, welche Personalauswahl betreibt, absolut, sondern immer in übergeordnete Systeme eingebettet ist.

Stellvertretend für relevante Sachverhalte des übergeordneten Systems ist beispielhaft der Arbeitsmarkt genannt. Ein spezielles Personalauswahlprogramm kann verschiedene Zustände des Arbeitsmarktes berücksichtigen. Diese Zustände haben für eine Organisation bei der Personalauswahl erhebliche Konsequenzen. Es kann zum Beispiel Knappheit an bestimmtem Personal vorliegen. In diesem Fall entstehen z.B. der Organisation für die Personalwerbung Kosten, es kann nur eingeschränkt frei oder gar nicht unter den Bewerbern gewählt werden, usw. Eignungsdiagnostik als Messen und Entscheiden würde diese wichtigen Umstände unberücksichtigt lassen. Des weiteren können Personalauswahlprogramme in unterschiedlichem Maße den internen und externen Arbeitsmarkt berücksichtigen. Dies hat für die Personalwirtschaft erhebliche Konsequenzen. Zum einen wird die Informationssammlung mit sehr unterschiedlichen Methoden vor sich gehen, zum anderen kann die schwerpunktmäßige Berücksichtigung des internen Arbeitsmarktes durch die Betonung der beruflichen Aufstiegschancen eine Steigerung der Arbeitszufriedenheit, der Identifikation mit dem Unternehmen und damit eine erhöhte Arbeitsmotivation und Produktivität, einen verringerten Krankenstand und eine niedrigere Fluktuation zur Folge haben.

Stufe 1 soll auch andeuten, daß das Unternehmen auch selbst auf den Personalmarkt einwirken kann. Dies kann durch Einflußnahme auf die Bildungspolitik geschehen oder einfach durch mehr oder weniger aktive Suche nach geeigneten Bewerbern. Diese Suche beinhaltet z.B. das Anbieten von Praktikantenplätzen, inserierte Stellenangebote oder eine Bezahlung über dem üblichen Niveau.

Dies soll nur beispielhaft genannt sein. Das übergeordnete System ist natürlich sehr komplex und beinhaltet viele Elemente, welche bei der Personalauswahl berücksichtigt werden könnten oder sollten. Wichtig ist an dieser Stelle die Tatsache, daß diese Sachverhalte existent und relevant sind. Personalauswahl, verstanden als Intervention im System, berücksichtigt sie. Eignungsdiagnostik im klassischen Sinne läuft Gefahr, sie zu ignorieren. Eine vertiefende Besprechung der Systemauffassung erfolgt in Kapitel 5.

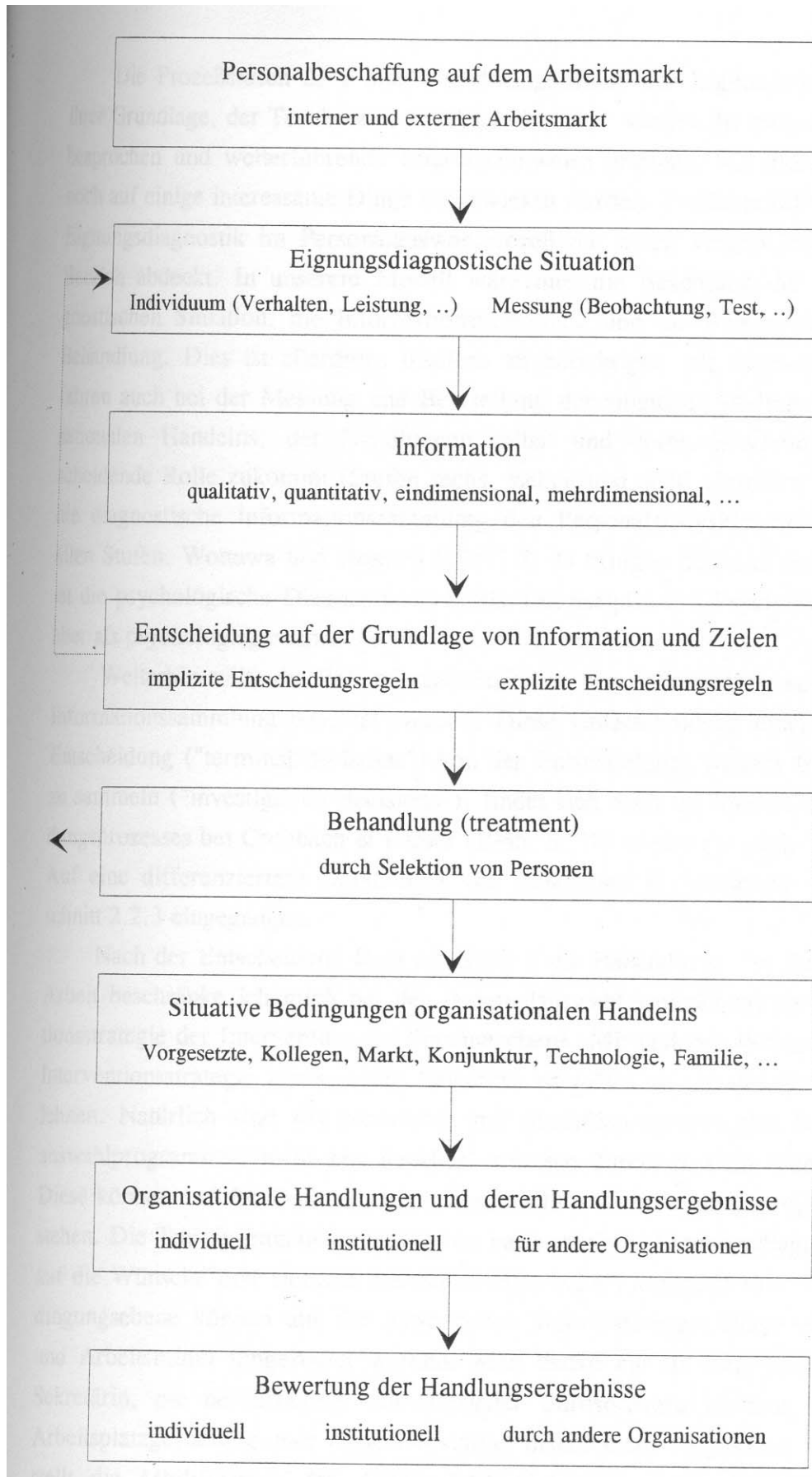


Abbildung 2: Prozeßmodell der Personalauswahl. Werden die jeweiligen Prozeßstufen konkretisiert, so spricht man von einem Personalauswahlprogramm.

Die Prozeßstufen 2, 3 und 4 sind Gegenstand der Eignungsdiagnostik und ihrer Grundlage, der Testtheorie. In Abschnitt 2.1.3 wurden die einzelnen Prozesse besprochen und weiterführende Literaturhinweise gegeben. An dieser Stelle soll noch auf einige interessante Dinge hingewiesen werden. Zunächst fällt auf, daß die Eignungsdiagnostik im Personalauswahlprozeß nur einen vergleichsweise kleinen Bereich abdeckt. In unserem Modell wäre dies die Beachtung der eignungsdiagnostischen Situation, die Informationssammlung und die Entscheidung für eine Behandlung. Dies ist allerdings insofern zu berichtigen, als diagnostischen Verfahren auch bei der Messung und Beurteilung der situativen Bedingungen organisationalen Handelns, der Handlungen selbst und deren Bewertung eine entscheidende Rolle zukommt (Stufen sechs, sieben und acht). Insofern 'durchwirkt' die diagnostische Informationssammlung den Personalauswahlprozeß auf nahezu allen Stufen. Wottawa und Hossiep (1987, S. 1) bringen dies auf den Punkt: „...ist die psychologische Diagnostik nicht als Teildisziplin der Psychologie, sondern eher als psychologiegestützte Technologie zu sehen.“

Weiterhin sollte die Rekursionsschleife von der Entscheidung auf die erneute Informationssammlung beachtet werden. Diese Unterscheidung einer endgültigen Entscheidung ("terminal decisions") von der Entscheidung, weitere Informationen zu sammeln ("investigatory decisions"), findet sich auch im Modell des Entscheidungsprozesses bei Cronbach & Gleser (1965, S. 18) wieder (s. auch Tack, 1976). Auf eine differenziertere Darstellung von möglichen Entscheidung wird in Abschnitt 2.2.3 eingegangen.

Nach der Entscheidung folgt auf Stufe 5 die Behandlung. Im Rahmen dieser Arbeit beschränke ich mich bei der Personalauswahl weitgehend auf eine Selektionsstrategie der Intervention auf Personenebene. Mit anderen Worten besteht die Interventionsstrategie darin, einen Bewerber entweder zu akzeptieren oder abzulehnen. Natürlich sind Kombinationen mit Modifikationsstrategien bei Personalauswahlprogrammen nicht nur denkbar, sondern durchaus auch wünschenswert. Diese können auf der Personenebene in speziellen Weiterbildungsprogrammen bestehen. Die Trainingsmaßnahmen können nach erfolgter Eignungsdiagnose speziell auf die Wünsche und Defizite des neuen Mitarbeiters angepaßt sein. Auf der Bedingungebene können auf der gewonnenen Informationsgrundlage Arbeitsplätze und Arbeitsmittel umgestaltet werden. Man denke nur an eine neu eingestellte Sekretärin, die hervorragend mit spezieller Bürosoftware vertraut ist, bei der Arbeitsplatzgestaltung aber diese Hilfsmittel bisher nicht vorgesehen waren. Hier stellt die Modifikation der Arbeitsplatzbedingungen durch Bereitstellung entsprechender Arbeitsmittel eine denkbare Behandlung dar. Inwiefern diese Investition sinnvoll ist, kann dann in einer

späteren Nutzenanalyse belegt werden. Die Behandlung mit Modifikationsstrategien wird in Kapitel 5 noch einmal kurz aufgegriffen.

Eine mögliche Behandlungsform besteht in der Zurückweisung des Bewerbers. Für ihn ist der Interventionsprozeß beendet, was der Pfeil auf der Stufe fünf andeuten soll. Das gleiche gilt jedoch nicht für die Organisation. Sie hat sich innerhalb ihrer Intervention für eine Behandlung entschieden und diese durchgeführt, und das hat für die Organisation weiterhin eine Wirkung. Ich will dies an einem Beispiel verdeutlichen. Auf ein Stelleninserat meldet sich ein Bewerber. Seine Bewerbungsunterlagen werden gesichtet, anschließend eine eintägige eignungsdiagnostische Untersuchung durchgeführt. Er hat fünf Mitbewerber, von welchen zwei ausgewählt werden. Der Bewerber wird trotz guter Eignung abgelehnt, weil nur zwei Stellen zu besetzen waren. Bis dahin bestehen die Wirkungen für die Organisation mindestens in den Kosten für die Personalwerbung und die eignungsdiagnostischen Maßnahmen. Weitere können aber in der Zukunft folgen. Man bedenke die Situation, daß der Abgelehnte bei einem Konkurrenzunternehmen eingestellt wird, und sehr erfolgreich seiner Tätigkeit nachgeht. In diesem Fall kann die falsche Entscheidung der Organisation einige negative Wirkungen einbringen.

Bei Annahme des Bewerbers (als Behandlung) wird er in Wechselwirkung mit der Organisation handeln. Diese Handlungen führen zu Ergebnissen für die Person selbst, für die Organisation und für andere Organisationen (Stufe 7). Das Konzept 'andere Organisationen' ist hier sehr weit gefaßt. Es kann sich z.B. auch um die Familie der Person handeln oder ihren Freundeskreis. Beispielsweise kann ein Ergebnis ihrer Tätigkeit aus der Sicht des Ehepartners darin bestehen, daß sie oft für längere Zeit abwesend ist. Im Rahmen der NA werden diese Handlungsergebnisse jedoch nicht näher thematisiert. Der Nutzenanalytiker interessiert sich hauptsächlich für Handlungsergebnisse der agierenden Person, welche in direktem Zusammenhang mit den Organisationszielen stehen. Bei einem Verkäufer interessiert beispielsweise dessen Umsatz innerhalb einer Zeiteinheit, dessen Verhalten gegenüber den Kunden, die langfristigen Maßnahmen zur Gewinnung neuer Kunden, usw. Dieses Ergebnis wird durch die Ziele, Eigenschaften und Verhaltensweisen der Person mitbestimmt. Aus einer systemischen Perspektive betrachtet, wird jedoch unmittelbar deutlich, daß nicht nur die handelnde Person das Ergebnis bestimmt. Der Umsatz des oben erwähnten Verkäufers hängt auch von seiner Umwelt ab, z.B. der Region, der Konjunktur, der Wirksamkeit der Marketingabteilung und den privaten Lebensumständen. Dies ist ein wichtiger Aspekt, wenn die Qualität der diagnostischen Entscheidung beurteilt werden soll. Eine weitere Eigenschaft des Prozeßmodells der Personalauswahl besteht in der Trennung des Handlungsergebnisses

von dessen Bewertung (Stufe 8). Diese Unterscheidung impliziert das Bestehen eines Wertsystems beim Beurteiler. Dieses kann beträchtlich zwischen verschiedenen Beurteilern variieren. Die Konsequenzen die hieraus entstehen sind so vielfältig, daß sie an dieser Stelle nur gestreift werden können. Wichtig in unserem Zusammenhang ist die Feststellung, daß bei der NA das Wertsystem der Organisation im Vordergrund steht. Handelt es sich bei der Organisation um ein gewinnorientiertes Unternehmen, so bewertet es relevante Sachverhalte schwerpunktmäßig durch Zuordnung einer Einheit: D-Mark. Nutzenanalytische Modelle tragen dem Rechnung, indem auch sie die Werteinheit DM verwenden. Dies ist insofern bemerkenswert, als innerhalb der Psychologie Skalen mit Einheiten extremen Seltenheitswert besitzen. Möglicherweise entsteht das oft grobe Unverständnis Fachfremder gegenüber psychologischer Argumentation aus diesem Umstand. Mit anderen Worten: Wer nicht mit Einheiten rechnet, muß auf Verteilungskonzepte (Varianz, Prozentrang, ...) zurückgreifen, um die Vorteile quantitativer Methoden nutzen zu können. Leider sind aber diese Konzepte den wenigsten Nichtpsychologen bekannt, so daß sie viele gewichtige Argumente nicht verstehen können. Die nutzenanalytische Einheit DM ist bekannt. Ihre Argumente sind daher leichter verständlich. Diesen Vorteil drückt Cascio (1982, S. V) kurz und prägnant mit den Worten aus: "... the language of business is dollars, not correlation coefficients." In Abschnitt 2.3 werden die Möglichkeiten der Bewertung psychologischer Interventionsmaßnahmen weiter besprochen.

Ich möchte an dieser Stelle noch einmal auf die Wichtigkeit des Prozeßmodells für die weitere Argumentation hinweisen. Als wichtigste Erkenntnis daraus möchte ich folgende Punkte stichwortartig auflisten:

- Personalauswahl ist eine psychologische Intervention und muß als mehrstufiger Prozeß gesehen werden.
- Die Sichtweise der Eignungsdiagnostik als möglichst genaues Messen von Sachverhalten und die darauf aufbauende Entscheidung greift zu kurz.
- Diagnostische Methoden spielen auf nahezu allen Prozeßstufen eine wichtige Rolle.
- Der Interventionsprozeß findet innerhalb seiner Umwelt statt, d.h. der Prozeß wird vom übergeordneten System beeinflusst und er beeinflusst dieses.
- Werden die Wirkungen von Interventionsmaßnahmen untersucht, so spielt das Wertsystem des Untersuchenden eine wesentliche Rolle.

Wenn Personalauswahl nicht nur im Lichte der Diagnostik gesehen wird, sondern als psychologische Interventionsmaßnahme, die sich im Kontext übergeordneter Systeme

abspielt, dann werden plötzlich neue Dinge sichtbar, die vorher im Verborgenen geblieben waren. Das Prozeßmodell der Personalauswahl soll diese neuen Aspekte sichtbar machen. Es wird im weiteren Verlauf der Arbeit noch häufig zur Klärung mancher Probleme herangezogen werden. Nutzenanalytische Modelle beachten alle Prozeßstufen, was enorme Vorteile mit sich bringt.

2.2.3 Arten von Entscheidungen

Wie bereits mehrfach angemerkt wurde, beschränkt sich diese Arbeit weitgehend auf Behandlungen der Personalselektion. Mit anderen Worten: Es werden Personalauswahlprogramme betrachtet, bei welchen die Entscheidung zwischen den Behandlungen 'annehmen eines Bewerbers' und 'ablehnen eines Bewerbers' gefällt wird. Diese Entscheidung kann auf sehr vielfältige Arten getroffen werden. Wichtiger aber ist, daß die Entscheidungen von unterschiedlichen Perspektiven zu bewerten sind. Welches Wertesystem wird angewandt, wenn ein Bewerber abgelehnt wird? Möglicherweise hat die Entscheidung für den Bewerber, der nicht mehr im günstigsten Alter ist und finanzielle Probleme hat, sehr weitreichende negative Konsequenzen. Aus der Sicht des Bewerbers wird die Entscheidung sehr negativ bewertet werden. Das Unternehmen hingegen hat ein ganz anderes Wertesystem. Auf dem Arbeitsmarkt finden sich weit qualifiziertere Bewerber. Die Einstellung des o.g. Bewerbers hätte negative Konsequenzen nach sich gezogen. Cronbach und Gleser (1965) unterscheiden verschiedene Arten von Entscheidungen. Geht es wie im Beispiel um die Frage nach dem Wertesystem, auf dessen Grundlage eine Entscheidung bewertet wird, so sprechen Cronbach und Gleser (1965) von einer 'individuellen Entscheidung', wenn das Wertesystem des Individuums im Vordergrund steht. Wird hingegen das organisationale Wertesystem zur Bewertung der Entscheidung herangezogen, so handelt es sich um eine 'institutionale Entscheidung'. Um einem Mißverständnis vorzubeugen sei darauf hingewiesen, daß es unerheblich ist, ob die Entscheidung von der Institution oder dem Individuum gefällt wird.

Cronbach und Gleser (1965) stellen neben dieser Unterscheidung ein differenziertes Klassifikationssystem von Entscheidungen vor, das hier kurz dargestellt werden soll. Danach möchte ich einige, im Zusammenhang mit der NA entscheidende Punkte kurz besprechen. Die Klassifikation von Cronbach & Gleser (1965) findet sich in Tabelle 2.

Tabelle 2: Klassifikation von Entscheidungen bei Cronbach & Gleser (1965). Es sind jeweils gegensätzliche Entscheidungsstrategien gegenübergestellt.

institutionelle Entscheidung	individuelle Entscheidung
mit Quotenbeschränkung (die Entscheidungen sind voneinander abhängig)	ohne Quotenbeschränkung (unabhängige Entscheidungen)
eine Behandlung pro Person	multiple Behandlungen pro Person
Ablehnung als Behandlung zugelassen	Ablehnung nicht zugelassen
Informationen sind in multivariater Form gegeben	Informationen in univariater Form
Entscheidungen sind endgültig (single stage testing)	Entscheidungen sind sequentiell aufgebaut (sequential strategy)

Diese 6 Entscheidungsformen sind frei kombinierbar, so daß $2^6=64$ Entscheidungsmuster entstehen können. Cronbach & Gleser (1965) besprechen in ihrem Buch hiervon einen Teil näher. Die Entscheidungen auf der linken Tabellenseite sind diejenigen, die innerhalb dieser Arbeit im Vordergrund stehen sollen.

2.2.3.1 *Individuelle vs. institutionelle Entscheidungen*

Die in unserem Zusammenhang wichtigste Unterscheidung ist die der individuellen vs. institutionellen Entscheidung. Entgegen der naheliegenden Bedeutungskonnotation dreht es sich dabei wie angesprochen nicht um die Frage, ob ein Individuum oder eine Institution die Entscheidung fällt oder ob über Individuen oder Institutionen entschieden wird. Vielmehr ist bedeutsam, welches Wertesystem der Entscheidung zugrunde liegt (vgl. Abb. 2, Stufe 8). Ein Diagnostiker, welcher darüber entscheidet, ob ein Verkehrssünder seinen Führerschein zurückerhalten soll, fällt normalerweise eine institutionelle Entscheidung, da er mögliche Konsequenzen aus der Sicht der Gesellschaft bewertet. (Gerade dieser Umstand macht es dem Betroffenen so schwer, die eventuell negative Entscheidung nachzuvollziehen. Ihm erscheint eine individuelle Entscheidung meist angemessener.)

Weshalb aber ist diese Unterscheidung so wesentlich? Die Antwort möchte ich zunächst in zwei Beispiele verpacken. Ein Großunternehmen stellt jedes Jahr mehrere tausend Arbeiter ein. Die Bewerber sind so zahlreich, daß man es sich leisten kann, nur jeden zweiten zu akzeptieren. Die abgelehnten Bewerber kommen nahezu immer irgendwo anders unter. Die Validität des Auswahlverfahrens bezüglich nachfolgender Berufszufriedenheit und bezüglich

der Leistung liegt bei $r_{xy}=0,1$ und seine Anwendung (Kurzfragebogen) kostet nahezu nichts. Ist die Anwendung des Auswahlverfahrens in diesem Fall trotz seiner niedrigen Validität gerechtfertigt? Eine vernünftige Antwort wäre: Die Anwendung ist gerechtfertigt, da das Auswahlverfahren immer noch eine bessere Vorhersage bieten kann als eine Zufallsauswahl. Darüber hinaus ist seine Anwendung billig. Es gewährleistet trotzdem eine leichte Steigerung der Zufriedenheit und der Leistung der Mitarbeiter.

Im zweiten Beispiel hat ein Psychologe die Aufgabe, vor Gericht ein Sachverständigengutachten über die Glaubwürdigkeit des Hauptbelastungszeugen in einem schwerwiegenden Fall abzugeben. Der Ausgang des Verfahrens wird maßgeblich davon beeinflusst, ob der Belastungszeuge für glaubwürdig gehalten wird. Dem Angeklagten droht trotz seiner Unschuldsbeteuerung eine langjährige Gefängnisstrafe. Die Validität des diagnostischen Verfahrens ist erstaunlich hoch. Es hat sich bei früheren Untersuchungen ergeben, daß sie bei über 0,8 liegt. Würde ein Richter dem Verfahren Glauben schenken, wenn dieses die Glaubwürdigkeit des Zeugen indiziert? Eine mögliche Antwort wäre: Es ist sehr wahrscheinlich, daß der Zeuge die Wahrheit sagt. Dennoch besteht ein Restrisiko, daß das diagnostische Urteil falsch ist. In diesem Fall wäre ein Verurteilung für den Angeklagten untragbar. Trotz der hohen Validität reicht das diagnostische Verfahren zu einer Verurteilung allein nicht aus.

Weshalb wird nun der Einsatz des niedrig validen Verfahrens in der einen Situation befürwortet, während das hoch valide Verfahren in der anderen Situation abgelehnt werden muß? Worin liegt der Unterschied? Im zweiten Fall beachten wir stärker das Wertesystem des Individuums. Es handelt sich um eine individuelle Entscheidung. Im ersten Fall hingegen ist das Wertesystem des Unternehmens wichtiger, eine institutionelle Entscheidung wird gefällt. Selbstverständlich kann man jeweils dagegen argumentieren und die Entscheidungsform als falsch ansehen. Ist jedoch erst einmal darüber ein Konsens gefunden, welches Wertesystem das dominante ist, dann zeigt sich ein klarer Effekt: Individuelle Entscheidungen bedürfen einer sehr hohen Validität, um sie zu rechtfertigen. Dies liegt daran, daß individuelle Kriteriumsvorhersagen sehr hohe Validitätskoeffizienten benötigen, damit die Schätzungen mit hoher Sicherheit gemacht werden können (vgl. hierzu das Konzept des Standardschätzfehlers der klassischen Testtheorie). Institutionelle Entscheidungen hingegen lassen sich oft gegen Alternativen rechtfertigen. Auch die zufällige Entscheidung kommt dabei als Alternative in Betracht. Die Validitätskoeffizienten sind meist dann hoch genug, wenn die Validität der Alternative niedriger ist. Wie wir sehen werden, stimmt diese Aussage noch nicht ganz. Bei der Betrachtung von Nutzenmodellen wird klar werden, daß noch andere

Größen bedeutsam sind. Weitere Beispiele, bei welchen der Unterschied zwischen individueller und institutioneller Entscheidung bedeutsam ist, sind z.B.: Aufnahmebedingungen für Studienplätze, medizinische Diagnosen für Therapien mit sehr wenigen bzw. sehr starken Nebenwirkungen, Raketenabwehrsysteme mit Gegenschlagautomatismus, ...

Personalauswahl in Unternehmen wird aus Sicht der Unternehmensziele bewertet. Die Entscheidungen, die bei der Intervention getroffen werden, sind institutionell. Personalauswahl wird als Interventionsform so optimiert, daß dem Unternehmen ein möglichst hoher Nutzen entsteht. Nutzenanalytische Modelle beruhen im wesentlichen auf dieser Sichtweise. Damit ist die Frage aus der Einleitung, wer den Nutzen von Personalauswahl haben soll, beantwortet. Dies mag für den oft auch humanistisch orientierten Psychologen zunächst unbefriedigend klingen. Es sollte jedoch bedacht werden, daß die Arbeits- und Organisationspsychologie seit einiger Zeit erfolgreich versucht, individuelle und organisationale Ziele in Einklang zu bringen bzw. gemeinsam zu optimieren (z.B. neue Arbeitsformen, Führungsforschung, Organisationsentwicklung, ...). Nutzenanalysen schließen diesen Weg nicht aus. Vielmehr liegt diese Denkweise implizit in ihnen verborgen, wenn man die Annahme trifft, daß beruflich erfolgreiche Mitarbeiter zufriedener mit ihrer Arbeit, ihren Aufstiegs- und Entwicklungsmöglichkeiten, ihrer Bezahlung und ihrem berufsgebundenen sozialen Status sind.

2.2.3.2 Univariate vs. multivariate Informationsrepräsentation

Eine weitere Unterscheidung im Klassifikationssystem von Cronbach und Gleser (1965) besteht darin, daß die diagnostische Information univariat oder in multipler Form vorliegen kann. Univariate Information hat die Eigenschaft, daß sie numerisch auf einem Kontinuum abgebildet werden kann. Unser Fahrzeugmechaniker kann dies, wenn er die Stärke der Bremsbeläge mißt. Dasselbe gilt für ein Personalauswahlprogramm, dessen Entscheidung nur auf der Grundlage des Testwertes eines (eindimensionalen) biographischen Fragebogens gefällt wird. Meist wird bei univariaten Entscheidungen explizit ein cut-off-Wert festgelegt. Liegt ein Objekt unter diesem Wert, dann erhält es Behandlung A, liegt es darüber, ist Behandlung B indiziert. Eine solche explizite Entscheidungsregel könnte also lauten: Bewerber mit einem Testrohwert von 60 oder darüber werden akzeptiert. Oder: Bremsbeläge

mit einer Stärke von 0,5 cm oder darunter werden ausgewechselt, über dieser Stärke bleiben sie erhalten.²

Die Entscheidungsfindung bei multipler Information ist dagegen schwieriger. Dies ist insofern paradox, da sie plausiblerweise informativer ist. Ein typisches Beispiel für multiple Informationen stellen die Ergebnisse von Assessment Centern dar. Hier können z.B. die Resultate verschiedener psychologischer Tests, von verschiedenen Rollenspielen, Planungsübungen und Auswahlinterviews vorliegen. Es gibt mehrere Möglichkeiten, mit dieser mehrdimensionalen Informationslage umzugehen. Eine wohl verbreitete ist die implizite Informationsintegration. Diese Vorgehensweise ist vergleichbar mit dem Vorgehen des normalen Alltagsmenschen bei der Urteils- und Entscheidungsbildung. Es wird die Information zum großen Teil automatisch und unbewußt ohne explizite Regeln zu einem Gesamturteil 'verrechnet'. Eine alternative Vorgehensweise besteht in der Explikation der Verrechnungsregeln. Beim Assessment Center könnte eine solche Regeln lauten: Akzeptiere die Bewerber, die in jedem Rollenspiel mindestens durchschnittliche Ergebnisse erzielen, in den psychometrischen Leistungstests mindestens Prozentrang 40 erreichen und beim Einstellungsinterview ihre hohe Motivation für die angebotene Tätigkeit betonen. Eine theoretisch und empirisch stringente Möglichkeit zur Bildung solcher expliziten Entscheidungsregeln bieten verschiedene multivariate statistische Verfahren (z.B. multiple Regression, Diskriminanzanalyse). Ihr Prinzip besteht darin, daß auf der Grundlage multipler Information eine eindimensionale Skala gebildet wird. Auf diese Weise kann dann, wie bei der univariaten Entscheidung, ein cut-off-Wert festgelegt werden, so daß eine eindeutige Zuordnung vom Prädiktorwert zur Entscheidung vorliegt. Auf diese Problematik wird in Kapitel 6 vertieft eingegangen, wenn ein reales Assessment Center-Verfahren empirisch überprüft wird.

In meinem Prozeßmodell der Personalauswahl habe ich auf Stufe 4 diesem Problem Rechnung getragen. Hier wird zwischen expliziten und impliziten Entscheidungen unterschieden. Wie noch gezeigt werden wird, spielt dieser Punkt bei der Überprüfung und Optimierung psychologischer Interventionsmaßnahmen eine entscheidende Rolle.

Ich möchte die wichtigsten Punkte bis hierher noch einmal kurz zusammenfassen. Ich habe zunächst zu zeigen versucht, daß ein Verständnis der Diagnostik als Messen und Entscheiden zu kurz greift. Wenn über Gütekriterien der Personalauswahl diskutiert wird, darf

² Interessant bei diesem Beispiel ist, daß eine Behandlungsform darin besteht, nichts zu machen. Hier wird wieder deutlich, daß Diagnostik als Intervention gesehen werden muß, um ihre Konsequenzen begreifen zu können. Denn wenn keine Diagnostik betrieben wird, folgt trotzdem eine Behandlung - nämlich keine. Eine Meß- und Entscheidungsdiagnostik vernachlässigt diese wichtige Folge.

man sich nicht nur auf die Testgütekriterien der diagnostischen Verfahren beschränken. Vielmehr fasse ich Personalauswahl als Interventionsmaßnahme in Organisationen auf und zwar völlig analog zu anderen Interventionsformen, wie z.B. der Anschaffung von neuen Maschinen und der Durchführung von Führungstrainings. Auch hierbei muß nach den voraussichtlichen Konsequenzen der Maßnahme gefragt werden - in Analogie zu den Konsequenzen von PAP. Um die Frage nach den Konsequenzen zu beantworten, sind informationssuchende Handlungen einzuleiten - Diagnostik ist zu betreiben. Auch bei Interventionen der o.g. Art sind die Zustände des übergeordneten Systems bedeutsam. Die wichtigsten Fragen und Probleme vor der Durchführung der Intervention sollten in einem Modell, das dem Prozeßmodell der Personalauswahl (Abb. 2) ähnlich ist, deutlich werden, weil auch dieses ein Modell eines Interventionsprozesses ist.

Danach habe ich mich den Entscheidungsformen nach Cronbach und Gleser (1965) zugewandt. Hierbei wurde insbesondere auf die Unterscheidung institutionelle vs. individuelle Entscheidung hingewiesen, da diese Unterscheidung bei der Bewertung der Angemessenheit von Testverfahren essentiell ist. Darüber hinaus wurde auf die Problematik der multiplen Informationsrepräsentation hingewiesen.

Im nächsten Abschnitt werde ich mich der Frage zuwenden, wie psychologische Interventionsmaßnahmen überprüft werden können. Beispielhaft für allgemeine Interventionsmaßnahmen betrachte ich auch hier PAP. Das Prozeßmodell der Personalauswahl wird dabei gute Dienste leisten. Wie nämlich gezeigt werden wird, setzen verschiedene Verfahren an bestimmten Prozeßstufen an und vernachlässigen andere. Auch die Entscheidungsarten spielen wieder eine bedeutende Rolle.

2.3 Überprüfung psychologischer Interventionsmaßnahmen

Menschen haben Ziele. Organisationen haben Ziele. In wirtschaftlichen Unternehmen treffen die Ziele verschiedener Menschen und die des Unternehmens aufeinander. Wenn nun einer der Beteiligten etwas verändern will oder eine Veränderung beibehalten will, so kann seine Intervention nur dann von Bestand sein, wenn er andere

- von der Qualität seiner Ziele überzeugt und
- die Wirksamkeit seiner Maßnahmen zur Zielerreichung feststellt.

Der erste Punkt entzieht sich empirischer Forschung. Er ist Gegenstand geisteswissenschaftlicher Ethik und soll hier, ungeachtet seiner schwerwiegenden Bedeutung, nicht weiter behandelt werden. Die zweite Forderung dagegen führt zu Fragen, die mit

empirischer Wissenschaft durchaus beantwortbar sind. Von den Beantwortungsmöglichkeiten handelt dieser Abschnitt.

Den Begriff der Evaluation verstehe ich als Überbegriff aller Maßnahmen zur Bewertung und Überprüfung (psychologischer) Interventionsmaßnahmen. Einige spezielle Evaluationsmöglichkeiten sollen am Beispiel von Personalauswahlprogrammen näher betrachtet werden. Dies wird die Validierungsstudie sein und damit zusammenhängend die neueren Ansätze zur Validitätsgeneralisierung. Unser eigentliches Kernthema, die Nutzenanalyse, betrachte ich als spezielle Form der Evaluation.

2.3.1 Evaluation

2.3.1.1 Das Konzept und verwandte Begriffe

Der Begriff der Evaluation ist sehr umfassend und wird von vielen wissenschaftlichen Disziplinen, auch innerhalb der Psychologie, in sehr verschiedenen Zusammenhängen gebraucht. Ich schließe mich der Meinung von Wottawa und Thierau (1990) an, die auf eine Definition des Begriffes verzichten. Stattdessen beschreiben sie allgemeine Kennzeichen wissenschaftlicher Evaluation. Evaluation beinhaltet die Bewertung von Handlungsalternativen mit dem Ziel, Planungs- und Entscheidungshilfen zu erhalten. Sie ist eine praktische Maßnahme um Interventionen zu überprüfen, zu verbessern oder über ihre Durchführung zu entscheiden. Des weiteren betonen Wottawa und Thierau (1990) die Wichtigkeit der Verwendung wissenschaftlicher Methoden und Techniken bei der Evaluation.

Thierau, Stangel-Meseke und Wottawa (1992) nennen einige notwendige Voraussetzungen sinnvoller wissenschaftlicher Evaluation. Demnach müssen die Ziele der Intervention auf der Konstruktebene geklärt sein, und die operationalisierten Kriterien der Zielerreichung und der Zeitpunkt ihres erwarteten Eintretens festgelegt sein. Darüber hinaus müssen die Kosten der Maßnahme feststehen und die Kosten gegenüber ihrem Nutzen verrechenbar sein.

Einige Klassifikationsmöglichkeiten der Evaluation schlagen Wottawa und Thierau (1990) vor. Zwei davon haben im Zusammenhang mit der NA eine größere Bedeutung: Die Beachtung der Zeitperspektive bei der Evaluation und die Unterscheidung zwischen summativer und formativer Evaluation.

Mit der Zeitperspektive sind die unterschiedlichen Möglichkeiten bei der Festsetzung des Zeitpunktes der Evaluation in bezug auf die Intervention gemeint. Prospektive Evaluation wird auf der Basis verschiedener sozialwissenschaftlicher Techniken vor der Durchführung einer Interventionsmaßnahme betrieben (z.B. durch innerbetriebliche Interviews vor der

Einführung eines Weiterbildungskonzeptes). Prognostische Evaluation ist die empiriegestützte Bewertung von Interventionsfolgen auf der Grundlage von Pilotprojekten oder Prototypen (z.B. Überprüfung der Konsequenzen eines neuen Marketingkonzeptes auf einem Testmarkt). Werden abgeschlossene Interventionsmaßnahmen bewertet, sprechen die Autoren von statistischer bzw. vergangenheitsorientierter Evaluation. Obwohl bei Wottawa und Thierau (1990) nicht ausdrücklich erwähnt, spielt der Terminus 'statistische Evaluation' wohl auf die Methoden an, mit der diese vergangenheitsorientierten Maßnahmen oft durchgeführt werden. Diese Form der Evaluation dient selbstverständlich nicht nur der Interpretation von Vergangenen, sondern wesentlich der Neugestaltung und Verbesserung ähnlicher zukünftiger Maßnahmen.

Nicht ganz unabhängig von der Zeitperspektive ist die Unterscheidung zwischen summativer und formativer Evaluation. Formative Evaluation bezieht sich auf die Bewertung und Verbesserung der Intervention während ihrer Durchführung. Eine abschließende, globale und zusammenfassende Bewertung der Maßnahmen wird summative Evaluation genannt.

Wenn Interventionsmaßnahmen evaluiert werden sollen, ist die Definition von Kriterien für den Maßnahmenerfolg (bzw. -mißerfolg) notwendig. Bewertung ist wertgebunden. In einem Unternehmen wird beispielsweise im Verwaltungsbereich die 'Gleitzeit' im Zuge der Flexibilisierung von Arbeitszeiten eingeführt. Darüber hinaus erhalten die Mitarbeiter eine Einweisung, unter welchen Umständen Absprachen erfolgen sollten, damit keine unnötigen Koordinationsverluste aus der Maßnahme entstehen, weil voneinander abhängige Mitarbeiter zu verschiedenen Zeiten anwesend sind. Die Planer dieser Behandlung (als zielorientierte Handlung) erwarten bestimmte Folgen der Maßnahme für die Betroffenen und das Unternehmen. Dies könnten in diesem Fall eine höhere Zufriedenheit mit den Arbeitsbedingungen, eine höhere Produktivität, eine niedrigere Abwesenheitsquote, nur geringfügige Störungen des Arbeitsablaufs, usw. sein. Dies sind die Kriterien der Interventionsmaßnahme. Die Erreichung dieser Kriterien entscheiden über den Erfolg oder Mißerfolg der Intervention.

Bei einer Evaluation müssen diese Erfolgskriterien operationalisiert, d.h. beobachtbar gemacht werden. Dies ist auf verschiedenen Ebenen möglich (Kirkpatrick, 1960, zit. nach Holling & Liepmann, 1993): Auf den Ebenen der Reaktionen, des Lernens, des Verhaltens und der Ergebnisse für die betroffenen Individuen oder für die Organisation. Die Reaktionen sind subjektive Meinungen, Einstellungen und Bewertungen. Diese können beispielsweise per Fragebogen bei der Evaluation der Zufriedenheit der Mitarbeiter mit der Maßnahme erhoben werden. Die Lernebene bezieht sich auf die Wissensaneignung und auf

Einstellungsveränderungen. Die Wirkung der Einweisung kann z.B. durch 'Abfrage' der Lerninhalte bewertet werden. Besser wäre die Beobachtung, ob sich die Mitarbeiter an die Anweisungen halten. Dies wäre ein Erfolgskriterium auf der Verhaltensebene. Werden die Auswirkungen der Maßnahme bzgl. der Produktivität oder der Abwesenheitsquote untersucht, so steht das Kriterium der Ergebnisse für das Unternehmen im Vordergrund. Hamblin (1974, zit. nach Neuberger, 1991) erweitert das Kirkpatrick'sche Modell noch um die Ebene der übergeordneten Werte im Sinne der Organisationskultur. Diese Ebene soll hier nicht gesondert betrachtet werden, sondern wird zur Ergebnisebene gezählt.

Der Begriff 'Evaluation' ist seit einiger Zeit in unserem Sprachraum innerhalb der Sozial- und Verhaltenswissenschaften in Mode gekommen. Trotzdem sind dessen Inhalte nicht ganz so neu. Innerhalb der Wirtschaftswissenschaften wird schon seit einiger Zeit die Evaluation als managementunterstützendes Instrument eingesetzt, dies allerdings unter dem Namen Controlling³. Ziele des Controlling sind nach Schildbach (1992) die Gestaltung eines unternehmensinternen Informationssystems, welches jeden Entscheidungsträger mit Informationen unterstützt. Diese Informationen können zum Beispiel die Ziele eines Unternehmens, die Handlungsmöglichkeiten des Entscheidungsträgers, die Entscheidungen anderer Entscheidungsträger im Unternehmen, mögliche Umweltentwicklungen oder mögliche Ergebnisse von Handlungsalternativen sein (vgl. auch Horváth, 1990). Bei Nieschlag, Dichtl und Hörschgen (1991) findet sich eine weitere Funktion des Controlling. Sie betonen seinen Wert bei der Früherkennung von Risiken und Fehlentwicklungen von Maßnahmen. Interessant ist im Zusammenhang mit der Nutzenanalyse, daß Schildbach (1992) ausdrücklich auf die Verständlichkeit der Informationen für den Entscheidungsträger und die Vergleichbarkeit der Informationen untereinander hinweist. Möglicherweise erfüllen sozial- und Verhaltenswissenschaftliche Evaluationsergebnisse diese Forderungen nicht sehr häufig - die Nutzenanalyse kann dieses Defizit weitgehend überwinden, da sie durch die DM-Einheit ein allgemeinverständliches und oft vergleichbares Wertmaß benutzt.

2.3.1.2 Evaluation von Personalauswahlprogrammen

Bei der Evaluation von PAP bietet es sich an, das Prozeßmodell der Personalauswahl, wie es in Abbildung 2 skizziert ist, zu Hilfe zu nehmen. Die Wirkung von PAP können grundsätzlich

³ Der Begriff Controlling basiert auf dem englischen Wort 'to control', welches einen wesentlich weiter gefaßten Bedeutungsumfang als das deutsche Wort 'kontrollieren' besitzt. Neben dem Aspekt des Kontrollieren als überwachen und beaufsichtigen, gehören dazu auch steuern, lenken, regulieren, planen, usw. Dies gilt auch, wenn im deutschen Sprachraum statt Controlling das Wort 'Kontrolle' gelegentlich verwendet wird.

auf allen acht beschriebenen Prozeßstufen evaluiert werden. Die Evaluationsmaßnahmen können gemäß Abschnitt 2.2.1.1 prospektiv, prognostisch, vergangenheitsorientiert, formativ oder summativ angelegt sein. Bei der Wahl der Erfolgskriterien stößt man, je nach Evaluationsobjekt, auf theoretische und praktische Grenzen. Dennoch kann man auch hier oft zwischen verschiedenen Ebenen der Erfolgskriterien wählen: Reaktion, Lernen, Verhalten und Ergebnis. Eine umfassende Besprechung der Evaluationsmöglichkeiten von PAP anhand des Prozeßmodells soll hier allerdings nicht angestrengt werden. Vielmehr möchte ich es bei einigen Anregungen belassen.

Wie in Abschnitt 2.2.2 bereits angemerkt wurde, kann die Einbeziehung des internen Arbeitsmarktes erhebliche Auswirkungen auf den Erfolg von PAP haben. Die Entscheidung über eine Berücksichtigung dieses Punktes kann z.B. durch eine prospektive Evaluationsstudie der vorhandenen Potentiale interner und externer Bewerber optimiert werden. Wäre diese Studie auf der Verhaltens- oder sogar (langfristigen) Ergebnisebene angelegt, wäre jedoch ein aufwendiger (quasi-) experimenteller Versuchsplan erforderlich (z.B. Cook & Campbell, 1979).

Die Bewertung der eignungsdiagnostischen Situation und der Informationsgewinnung wird weitgehend durch die Testtheorie abgedeckt. Aber auch hier können weitere wichtige Maßnahmen durchgeführt werden. Beispielsweise legt das Konzept der sozialen Validität von Schuler und Stehle (1983) nahe, die eignungsdiagnostische Situation nicht nur gemäß den Testgütekriterien der Testtheorien zu gestalten. Kurz gesagt sollten auch die Informationsfunktion der Auswahl-situation, die Partizipationsmöglichkeiten der Führungskräfte, die Transparenz der Situation und die Kommunizierbarkeit des Ergebnisses Beachtung finden und damit bewertet werden.

Die Art der Entscheidungsfindung bei multipler Information bedarf der besonderen Aufmerksamkeit. Stellt sich durch eine Evaluation der Entscheidungen heraus, daß die impliziten Entscheidungsregeln einer Optimierung bedürfen, so kann diese durch eine Validierungsstudie erfolgen. Sie ist eine spezielle Form der Evaluation unter Verwendung der Statistik und wird im nächsten Abschnitt näher betrachtet.

Die Erfassung und Bewertung der situativen Bedingungen und der Ergebnisse organisationalen Handelns (Stufen sechs, sieben und acht) stellen eine besondere Herausforderung an die Techniken der Evaluation. Hier wird deutlich, daß Evaluationsforschung nicht nur mit quantitativen Methoden auskommen kann. Oft sind qualitative oder sogar geisteswissenschaftliche Ansätze gefragt, wenn z.B. Werte der betroffenen Mitarbeiter ins Spiel kommen.

Eine Kernfrage bei der Beurteilung von PAP ist die Bewertung der Vorhersagekraft der eignungsdiagnostischen Information für die Qualität späterer organisationaler Handlungen der Bewerber. Diesen speziellen Punkt der Evaluation, die Validierungsstudie, will ich ausführlich im nächsten Abschnitt behandeln.

2.3.2 Die Validierungsstudie als spezielle Form der Evaluation

Die Validierungsstudie ist eine spezielle Form der Evaluation von Interventionsmaßnahmen. Sie ist ein quantitatives Verfahren. Ihr Ziel besteht darin, die Validität von Konstrukten, Informationen oder Maßnahmen zu bestimmen. Der Begriff der Validität stammt aus der klassischen Testtheorie und wird in sehr verschiedenen Zusammenhängen gebraucht, so daß eine Definition nahezu unmöglich ist. Cook & Campbell (1979, S. 37) fassen den Begriff sehr weit:

We shall use the concepts of validity and invalidity to refer to the best available approximation to the truth or falsity of propositions ...
(Hervorhebungen im Original).

Durch die Rekursion auf das erkenntnistheoretische Konzept der Wahrheit bleiben allerdings die Voraussetzungen der Begriffe in der Definition weiterhin ungeklärt. Dennoch ist das Verständnis der Validität von Cook und Campbell sehr elegant, weil sie sehr verschiedene Auffassungen vereinen kann.

Michel und Conrad (1982, S. 52) definieren die Validität eines Test als den "Grad der Genauigkeit, mit dem ein Test das mißt, was er messen soll". Auch diese Sichtweise hilft hier nicht wesentlich weiter, die begriffliche Unschärfe bleibt bestehen. Ein weiterer Lösungsansatz, diesen unbefriedigenden Umstand zu beseitigen, besteht in der Aufteilung in verschiedene Validitätskonzepte. Die beiden wichtigsten, die Konstruktvalidität und die Kriteriumsvalidität, sollen hier kurz besprochen werden.

Die Konstruktvalidität ist ein stark theoretisch orientiertes Konzept. Es problematisiert die Relation von beobachtbaren Elementen und einem theoretischen Konstrukt. Beispielsweise behauptet jemand, daß in der Realität etwas vorhanden wäre, das er 'beruflichen Erfolg' nennt. Die Frage, ob seine Behauptung absurd oder (teilweise) 'wahr' ist, kann zu diesem Zeitpunkt nicht beantwortet werden. Sie ist äquivalent mit der Frage 'gibt es einen Gott?'. Eine zulässige Frage an den 'Konstruktgläubigen' wäre diejenige, durch welche Beobachtungen er zu seiner Behauptung gekommen ist. Wir verlassen mit der Antwort die Konstruktebene, und begeben uns auf die Ebene der beobachtbaren Sachverhalte, die

Indikatorebene. Eine mögliche Antwort wäre, daß er beobachtet habe, daß Menschen verschiedene qualitative und quantitative Ergebnisse in ihrem Beruf erbringen, daß sie eine unterschiedliche Erfüllung bei ihrer Tätigkeit erfahren und daß ihre Vorgesetzten und Mitarbeiter ihre Arbeit unterschiedlich beurteilen. Diese Beobachtungen können mit einer Reihe von Methoden gemacht werden (Verhaltensbeobachtung, Befragung, Tests, quantitative Erfassung der Berufsergebnisse, ...). Die Konstruktvalidität macht Aussagen über den Zusammenhang der beobachtbaren Indikatoren und den theoretischen Konstrukten. Ein beobachtendes Verfahren und seine Indikatoren ist dann konstruktvalid, wenn ein solcher Zusammenhang zu bestehen scheint und daher auf die Existenz des Konstruktes geschlossen werden kann (vgl. es gibt einen Gott, weil die wunderbare Ordnung der Natur, die ich beobachtet habe, nicht ohne ihn existieren könnte'). In der Sprache von Cook & Campbell (1979) heißt das, daß die Indikatoren sich der Realität annähern, bzw. daß die Schlüsse, die aus ihnen gezogen werden, gültig sind (zur Kritik der Validitätskonzepte von Cook & Campbell, siehe Gadenne, 1984; zur Kritik ähnlicher Konzepte von Campbell & Stanley, 1963 bzw. Campbell & Stanley, 1966, siehe Gadenne, 1976).

Die Konstruktvalidität von Indikatoren ist empirisch schwer zu fassen. Dennoch wurden eine Reihe von Verfahren vorgeschlagen, von denen einige hier genannt werden sollen. Eine wissenschaftstheoretische Begutachtung der nachfolgenden Verfahren zur Konstruktvalidierung soll hier unterbleiben. Trotzdem soll angemerkt sein, daß die Konstruktvalidierung hinsichtlich dieses Punktes äußerst problematisch ist.

- inhaltlich-logische Analyse
- experimentelle Überprüfung durch systematische Variation der (vermeintlichen) Konstrukte auf der Grundlage einer Theorie
- Konstruktvalidierung über den Umweg der Kriterienvolidierung (s. unten)
- Korrelation der Indikatoren mit solchen, die mit anderen Methoden gewonnen wurden. Zum Beispiel die Korrelation von Tests mit anderen Tests, welche die gleichen Konstrukte messen sollten Multitrait-Multimethod-Matrix-Verfahren nach Campbell und Fiske (1959), zum Nachweis der konvergenten und diskriminanten Validität
- Faktorenanalyse (exploratorisch) bei verwandten und theoretisch fremden Meßverfahren und verschiedenen Konstrukten
- Konfirmatorische Faktorenanalyse mit LISREL (Jöreskog und Sorböm, 1988)

Die *kriterienorientierte Validität*, oder kurz Kriteriumsvalidität, kann als spezielle Form der Konstruktvalidität aufgefaßt werden. Sie kann verstanden werden als eine technologieorientierte Validierungsstrategie. Sie ist aus diesem Grunde auch theoretisch

wesentlich einfacher zu fassen und vor allem empirisch klar operationalisiert. Sie behandelt den Fall, daß Indikatoren für ein (oder mehrere) Prädiktorkonstrukt(e) sowie für ein vorherzusagendes Kriterium vorliegen und folgt der Logik der (multiplen) Regressionsanalyse. Das heißt, daß auf Grundlage der Ausprägung der Prädiktorvariable eines Objektes die Ausprägung der Kriteriumsvariable geschätzt werden soll. Die Information der Ausprägung des Prädiktor- und Kriteriumsindikators liegt in quantifizierter Form vor. Wenn beide Informationen gleichzeitig verfügbar sind, so spricht man von konkurrender Validität. Fällt die Kriterieninformation erst zeitlich später an, spricht man von prädiktiver oder prognostischer Validität (z.B. Vorhersage von Schulnoten (Kriterium) auf der Grundlage von Intelligenztests, Leistungsmotivationstests, ..).

Als Maßzahl für die Kriteriumsvalidität dient der Korrelationskoeffizient (r_{xy}), bzw. im Falle des Vorliegens mehrerer Prädiktoren der multiple Korrelationskoeffizient (R)⁴. Diese Maßzahlen werden auch Validitätskoeffizienten genannt. Liegt ein gesicherter Zusammenhang zwischen der Prädiktorinformation und dem Kriterium vor, so ist diese Information bezüglich des Kriteriums valide. Dies bedeutet, es läßt sich auf der Grundlage der Prädiktorinformation eines Objektes eine gültige Schätzung der Ausprägung eines Objektes auf der Kriteriumsdimension machen.

2.3.2.1 Kriterienorientierte Validitätsbestimmung eines eignungsdiagnostischen Verfahrens bei der Personalauswahl

Die Kriterien bei der kriterienorientierten Validierung eignungsdiagnostischer Verfahren sind die Handlungen eines eingestellten Bewerbers in der Organisation, deren Ergebnisse und deren Bewertung. Diese finden sich auf den Stufen 7 und 8 des Prozeßmodells der Personalauswahl (Abb. 2) wieder. Wie die Differenzierung im Modell andeutet, sind dabei vielfältige Kriterienkonstrukte wählbar. Das Verfahren kann zum Beispiel an den bewerteten individuellen oder organisationalen Handlungsergebnissen validiert werden. Die Operationalisierung der Kriterienkonstrukte kann auf ebenso vielfältige Art geschehen. Die bewerteten individuellen Handlungsergebnisse können zum Beispiel durch Befragung des Mitarbeiters erfolgen (z.B. 'Glauben sie, daß diese Arbeit zu Ihnen paßt?'). Die bewerteten Ergebnisse für die Organisation könnte durch Vorgesetztenurteile, Messung der produzierten Stückzahlen, der Ausschußquote, die der Mitarbeiter erzeugt, usw. erhoben werden.

⁴ Statistische Maßzahlen werden im Verlauf der Arbeit mit Buchstaben benannt und einheitlich gebraucht.

Die 'Knackpunkte' auf der Prädiktorseite bei der kriterienorientierten Validierung eines eignungsdiagnostischen Verfahrens liegen darin, ob

- die Information ein- oder mehrdimensional vorliegt (Abb. 2, Stufe 3), ob
- die Entscheidungsregeln auf Grundlage der diagnostischen Information in expliziter Form gegeben sind oder nicht (Stufe 4) und ob
- eine Behandlung der Bewerber durch Zurückweisung bei der Personalauswahl erlaubt ist (Stufe 5).

Abgelehnte Bewerber erbringen leider keine Handlungen und Handlungsergebnisse in der Organisation. Es ist also kein Kriterium verfügbar. Für die Validierung eines eignungsdiagnostischen Verfahrens wäre es damit wünschenswert, wenn zumindest einige der Bewerber, bei welchen die Entscheidung normalerweise auf Ablehnung lauten müsste, trotzdem in die Organisation aufgenommen würden. Diese Vorgehensweise widerspricht jedoch meist massiv den organisationalen Interessen und ist deswegen in der Praxis kaum durchzusetzen.

Kann sie trotzdem realisiert werden, ist die Validität relativ einfach durch die Korrelation zwischen der eindimensionalen Information und dem operationalisierten Kriterium zu ermitteln. Liegt die Information mehrdimensional vor, so können explizite Entscheidungsregeln geschaffen oder optimiert werden. Hierzu wird eine multiple Regressionanalyse des Kriteriums auf die Prädiktoren durchgeführt. Die Regressionsgleichung erbringt eine Schätzung des eindimensionalen Kriteriums. Die Eindimensionalität wird durch eine Linearkombination der gewichteten Prädiktoren geleistet. Bevor eine Regressionsgleichung als explizite Entscheidungsregel zum Einsatz kommen sollte, ist jedoch unbedingt eine Kreuzvalidierung notwendig. Diese überprüft die regressionsanalytisch modellierten Zusammenhänge zwischen den Prädiktoren und dem Kriterium an einer Stichprobe, welche nicht der Regressionsmodellbildung zugrundegelegt wurde.

Werden die Bewerber abgelehnt, die auf der Grundlage der diagnostischen Information (trotz nichterfolgter Prüfung des Verfahrens) für ungeeignet befunden wurden, ist die praktische Durchführung der Validierung erheblich schwieriger. Hier kommt es entscheidend darauf an, ob die Entscheidungsregeln (Stufe 4) explizit oder implizit gegeben sind.

Im impliziten Fall ist eine Kriterienvalidierung streng genommen nicht möglich. Die Entscheidungsfindung innerhalb von Assessment Centern ist leider zu oft von diesem Mangel gekennzeichnet. Man kann versuchen, die Untersuchung dadurch zu retten, daß die

ursprüngliche Information, auf der die implizite Entscheidung basiert, herangezogen wird, und ihr immanenter Vorhersagegehalt überprüft wird. Bei Assessment Centern werden dann z.B. die protokollierten Übungsergebnisse auf ihre Validität überprüft. Zeigen diese eine prognostische Validität, dann können analog zur oben beschriebenen Vorgehensweise explizite Entscheidungsregeln mittels der multiplen Regressionsanalyse geschaffen werden.⁵ Eine weitere Möglichkeit zur Explikation der Entscheidungsregeln findet sich bei Wottawa und Hossiep (1987). Es handelt sich um das Verfahren der Hypothesenagglutination (HYPAG). Ziel der HYPAG ist es, ein explizites Modell der zuvor impliziten diagnostischen Entscheidungsregeln zu erstellen. Dabei wird das tatsächliche Verhalten der Diagnostiker erfaßt. Ein Modell, welches dieses Verhalten abbilden soll, wird vorläufig aufgestellt, sukzessive überprüft und weiter angepaßt.

Ist die Behandlung 'Ablehnung des Bewerbers' zugelassen, und sind die Entscheidungsregeln explizit vorhanden, kann eine kriterienorientierte Validierung angestrengt werden. Logisch konsistente, explizite Entscheidungsregeln ordnen die Bewerber in eine (eindimensionale) Rangreihe. Über den Zusammenhang der Prädiktorrangreihe und der Kriteriumsranreihe kann der Validitätskoeffizient berechnet werden.

Können empirisch keine Validitätskoeffizienten erstellt werden, oder soll dies z.B. aus Kostengründen nicht versucht werden, so kann die Validität auch mittels der einschlägigen Forschung geschätzt werden. Experten, welche diese Forschung kennen, scheinen gut in der Lage sein, Validitätskoeffizienten zu schätzen (z.B. Hirsh, Schmidt & Hunter, 1986). Eine andere Möglichkeit beruht auf den Verfahren zur Validitätsgeneralisierung, den sogenannten Metaanalysen. Diese ergeben Schätzungen für Validitätskoeffizienten auf der Grundlage von mehreren Einzeluntersuchungen. Die Koeffizienten werden dabei für die Fehler der Einzeluntersuchungen korrigiert. Darüber hinaus liefern sie Hinweise auf situationsspezifische Validitäten (fiktives Beispiel: kognitive Leistungstests haben nur bei älteren Menschen eine mittlere Validität bzgl. beruflicher Leistungen im Verwaltungsbereich).

⁵ In diesem Zusammenhang ist allerdings auf die Varianzeinschränkung der Prädiktoren und des Kriteriums hinzuweisen. Es gibt einige Vorschläge, wie die auf varianzeingeschränkten Stichproben beruhenden Korrelationskoeffizienten korrigierbar sind (z.B. Hunter & Schmidt, 1990; Lord & Novick, 1968). Diese Korrekturen basieren jedoch auf strengen, praktisch jedoch kaum haltbaren und prüfbar Annahmen, so daß ich ihre Praktikabilität in diesem Zusammenhang für fraglich erachte.

2.3.2.2 Kritik der kriterienorientierten Validitätsbestimmung

Die Definition der kriterienorientierten Validität ist zweifelsohne praktisch gut handhabbar und erfüllt eine wesentliche Funktion innerhalb der psychologischen Diagnostik. Ein grundlegender Hinweis auf die Tauglichkeit eines eignungsdiagnostischen Verfahrens gibt der Validitätskoeffizient bezüglich verschiedener Kriterien. Dennoch ist diese Vorgehensweise auch kritisierbar. Vom Standpunkt des Konzeptes der Konstruktvalidität ist die Kriteriumsvalidität sehr empirisch orientiert. Durch die technologieorientierte Überprüfung eignungsdiagnostischer Verfahren innerhalb spezifischer Bedingungen wird kein weiterer wissenschaftlicher Fortschritt erzielt. Anstatt theoretische Konstrukte und deren komplexe Zusammenhänge zu erforschen, was zur wissenschaftlichen Theorienbildung und damit zur Erfahrungsakkumulation beiträgt, ermittelt die Kriteriumsvalidität lediglich Zusammenhänge zwischen beobachtbaren Sachverhalten und dies meist in sehr spezifischen Situationen. Allerdings benötigen die neueren Verfahren der Validitätsgeneralisierung (s. Abschnitt 2.3.3) die quantitative Information der kriterienorientierten Validität. Ziel dieser Verfahren ist es, eben eine solche Erfahrungsakkumulation über viele Einzelstudien zu erreichen. Damit können sie einige der Nachteile der Validitätsüberprüfung mittels Prädiktor-Kriteriumskorrelation kompensieren.

Ein anderer Kritikpunkt ist aus dem Prozeßmodell der Personalauswahl (Abb. 2) ersichtlich. Wie oben gezeigt, werden bei der kriterienorientierten Validierung hauptsächlich und meist ausschließlich die eignungsdiagnostische Information (Stufe 3), die Entscheidung (Stufe 4) und die organisationalen Handlungen und deren Ergebnisse (Stufe 7) berücksichtigt. So sinnvoll die Validierung eignungsdiagnostischer Verfahren ist, sie kann nicht verbergen, daß sie wichtige, weitere Information unberücksichtigt läßt, wenn die Qualität des Personalauswahlprozesses als ganzheitliche Intervention im System beurteilt werden soll. Die Entscheidung, ein PAP und seine eignungsdiagnostischen Verfahren für unbrauchbar und nutzlos zu halten, nur weil es niedrig valide ist, kann völlig falsch sein, wenn weitere Gegebenheiten des Personalauswahlprozesses betrachtet werden. Beispielsweise könnte dies in einer Situation der Fall sein, in welcher das zu beurteilende PAP sehr kostengünstig ist, die höher valide Alternative allerdings sehr teuer ist. Die Nutzenanalyse ist ein Ansatz, welcher diese Unzulänglichkeiten mildern kann. Wir wollen uns noch kurz die heute immer wichtiger werdenden Ansätze zur Validitätsgeneralisierung ansehen, bevor ich auf die Erweiterungen der Nutzenanalyse zur Überprüfung psychologischer Interventionsmaßnahmen eingehe.

2.3.3 Ansätze zur Validitätsgeneralisierung

Jeder Wissenschaftler und wissenschaftlich arbeitende Praktiker steht heute vor dem Problem der Informationsüberflutung. Es ist angesichts der Menge der veröffentlichten theoretischen und empirischen Arbeiten kaum mehr möglich, den Überblick über Forschungsgebiete zu behalten. Ein weiteres Problem besteht in der oftmals anzutreffenden Sachlage, daß sich empirische Befunde zu widersprechen scheinen.

Eine Möglichkeit, mit dem Informationsüberfluß umzugehen, besteht darin, daß sich wenige fachkompetente Forscher von Zeit zu Zeit die Mühe machen, die Forschungsaktivitäten der letzten Jahre umfassend darzustellen. Diese Überblicksarbeiten ('narrative Reviews') haben, neben ihren vielen Vorteilen, einige entscheidende Nachteile. Die Zusammenstellung der besprochenen Arbeiten beruht auf der subjektiven Auswahl des Zusammenfassenden. Die Auswahlkriterien werden selten offengelegt. Auf inkonsistente Befunde wird meist nur hingewiesen, eine Erklärung bleibt so gut wie immer aus. Darüber hinaus besteht die Paradoxie, daß einzelne Forschungsarbeiten mit differenzierten mathematisch-statistischen Techniken operieren, die Überblicksarbeiten, welche letztendlich den status quo einer Forschungsdisziplin festlegen, aber mit 'Feld-Wald-und-Wiesen-Methoden' arbeiten. Symptomatisch für die Leistungen der Überblicksarbeiten ist für Hunter, Schmidt & Jackson (1982, S. 10) die folgende Lage: "Wich as many as one hundred studies on a relationship (zwischen zwei Variablen, Anm. des Autors), one might think that there would be a resolution on the issue. Yet most review studies do not conclude with resolution but with a call for more research on the question."

Eine andere Möglichkeit, die Informationsüberflutung zu bewältigen, sind die Ansätze zur Validitätsgeneralisierung. Sie leisten eine quantitative Integration von mehreren Forschungsbefunden auf der Basis ihrer statistischen Kennwerte. Der Gedanke der Analyse von Analysen, der Metaanalyse, wird von Glass (1976) propagiert. Neben einer Reihe von weniger entwickelten Verfahren, wie zum Beispiel das Auszählen signifikanter Forschungsbefunde oder die Durchschnittsbildung von Korrelationskoeffizienten oder Effektgrößen, gibt es eine Reihe von ausgeklügelten metaanalytischen Verfahren, die für die Bewältigung dieser Aufgaben entwickelt wurden. Tabelle 3 stellt die wichtigsten von ihnen in Anlehnung an Burke, Raju und Pearlman (1986) vor. Die Verfahren wurden mehreren empirischen Vergleichsstudien unterzogen. Burke et al. (1986) kommen zu dem Schluß, daß sie vergleichbare korrigierte Mittelwerte und Varianzen des Effektmaßes (in der Regel des Korrelationskoeffizienten r_{xy} oder der standardisierten Effektgröße d) liefern. Natürlich können auch Metaanalysen nicht die Verzerrungen der wissenschaftlichen Befundlage

kompensieren, die dadurch entstehen, daß die veröffentlichten Studien eine selektive Stichprobe aus allen durchgeführten Studien darstellen.

Tabelle 3: Verschieden Verfahren der Validitätsgeneralisierung in Anlehnung an Burke et al. (1986).

Prozedurbenennung nach Burke et al. (1986)	Literaturhinweise
noninteractive procedure	Pearlman, Schmidt & Hunter (1980)
interactive procedure	Schmidt, Gast-Rosenberg & Hunter (1980)
independent/ dependent multiplicative procedures	Callender & Osburn (1980)
Taylor Series Approximations 1 and 2	Raju & Burke (1983)

Die derzeit wohl bekannteste metaanalytische Technik ist die von Hunter et al. (1982) und deren Erweiterung von Hunter & Schmidt (1990). Ihr Grundprinzip besteht darin, daß die durchschnittlichen Effektkenngrößen (Korrelationskoeffizienten oder standardisierte Effektgrößen) zunächst von Artefakten bereinigt werden (Stichprobenfehler, Meßfehler, Varianzeinschränkung). Erklären diese Artefakte noch nicht die Unterschiede zwischen diesen Effektkenngrößen über die Einzelstudien, so wird nach situationsspezifischen Moderatorvariablen gesucht, welche die Unterschiede zu erklären vermögen.

Bei der Beurteilung der Wirksamkeit und der Effektgröße psychologischer Interventionsmaßnahmen können metaanalytische Untersuchungen eine große Hilfe sein. Sie liefern zum einen den Anhaltspunkt, wie valide eignungsdiagnostische Maßzahlen bezüglich verschiedener Kenngrößen bei früheren Untersuchungen durchschnittlich waren. Zum anderen geben sie auch Hinweise für eine eventuelle Situationsspezifität der diagnostischen Maßnahmen, indem Moderatorvariablen identifiziert wurden, welche den Prädiktor-Kriterium-Zusammenhang beeinflussen.

2.3.4 Nutzenanalyse als spezielle Form der Evaluation

Die NA fasse ich als spezielle Form der Evaluation psychologischer (und anderer) Interventionsmaßnahmen auf. Sie greift den Gedanken der kriterienorientierten Validierungsstudie auf, Evaluation quantitativ zu betreiben. Allerdings kann sie einige Nachteile der Kriteriumsvalidierung vermeiden. Nutzenmodelle beinhalten neben anderen

Parametern weiterhin Validitätskoeffizienten. Darüber hinaus besitzen sie mindestens zwei weitere positive Eigenschaften.

Zum ersten explizieren sie einen klaren Bewertungsmaßstab. Die Bewertung von Interventionsmaßnahmen wird in DM vorgenommen. Dies hat zunächst den Vorteil, daß die Ergebnisse einer Nutzenanalyse auch gegenüber Nicht-Psychologen leicht kommunizierbar sind. Darüber hinaus ist eine nutzenanalytisch untersuchte psychologische Interventionsmaßnahme jederzeit leicht mit anderen, ebenso untersuchten konkurrierenden Maßnahmen vergleichbar. Organisationale Investitionsmaßnahmen, welche nicht aus dem Bereich der Psychologie kommen, werden nahezu immer auf ihren Nutzen untersucht. Auch hier wird mit derselben Einheit gerechnet, so daß die Vergleichbarkeit gegeben ist.

Zum zweiten leistet der Skalierungsfaktor, daß Parameter, welche verschiedene Sachverhalte repräsentieren aber mit der gleichen Einheit DM operieren, einfach in das Nutzenmodell integriert werden können. Ein schwerwiegender Nachteil der Validierungsstudie, die Mißachtung wesentlicher Stufen des Prozeßmodells der Personalauswahl, kann dadurch gemildert werden. Bei der NA können Teile von Stufe 1 des Prozeßmodells berücksichtigt werden, indem zum Beispiel der Personalbeschaffung auf dem Arbeitsmarkt ein Wert in DM-Einheiten zugewiesen wird. Z.B. ist es bei Vollbeschäftigung schwieriger und damit teurer (Personalwerbung; Untersuchung von mehr Personen, bis die richtige gefunden wird; höhere Löhne als Anreiz; ...), sich geeignetes Personal zu beschaffen. Dieser Umstand wird von Validierungsstudien gänzlich vernachlässigt. Nutzenanalytische Modelle können diese Tatsache in Form von Kosten verrechnen. Stufe 8 zeigt, daß die Ergebnisse des organisationalen Handelns der Mitarbeiter Bewertungen erfahren. Wie noch ausführlich gezeigt werden wird, explizieren die Modelle der NA die Wertskala der Handlungsergebnisse und lassen diese so mit den Kosten vergleichen.

In mehrparametrischen Modellen stellt sich immer die Frage, welche Parameterkonstellation der Interventionsmaßnahmen die günstigste ist. Was ist zu bevorzugen: Ein billiges PAP, dessen Entscheidungen niedrig valide sind, oder ein teures mit höherer Validität? Fragen dieser Art können nutzenanalytische Modelle unter Beachtung spezifischer Randbedingungen oft beantworten. Eine solche wichtige Randbedingung ist z.B., wie hoch die bewerteten Leistungsunterschiede zwischen guten und schlechten Mitarbeitern einer bestimmten Position sind. Es ist intuitiv einleuchtend, daß die Unterschiede bei Managern gravierender sein werden als bei ungelerten Arbeitern. Aus diesem Grunde ist es auch ebenso verständlich, daß bei ersteren teurere PAP zum Einsatz kommen als bei letzteren.

Wie alle Evaluationsmaßnahmen sollen auch Nutzenmodelle Entscheidungen über Interventionsmaßnahmen unterstützen. Durch die Aufnahme mehrerer Parameter in quantitativer Form sind sowohl die Nutzenmodelle selbst, als auch die darauf basierenden Entscheidungen, vergleichsweise leicht kritisierbar. Dies schafft Transparenz und kann dazu beitragen, die Rationalität von organisationalen Entscheidungen zu steigern.

2.4 Funktion und Ziele von Nutzenmodellen

Funke, Schuler und Moser (in Vorbereitung) definieren die NA folgendermaßen: "Nutzenanalysen stellen eine in der Organisationspsychologie entwickelte Methode dar zur

- Evaluation leistungsbeeinflussender personalbezogener Maßnahmen, z.B. Personalauswahl- oder Personalentwicklungsprogramme
- durch Schätzung des mit ihrer Anwendung erzielbaren finanziellen Nutzens
- auf der Grundlage der durch die Personalprogramme erreichten leistungsrelevanten Verhaltensänderung bzw. Leistungssteigerung der betroffenen Mitarbeiter."

Stichwortartig möchte ich die Funktionen und Ziele von Nutzenmodellen zusammenfassen. Die aufgenommenen Punkte sind alle im voranstehenden Text mehr oder weniger betont schon zu finden. Im weiteren Verlauf, insbesondere in den Kapiteln 3 und 5, werden einige Punkte noch deutlicher zum Ausdruck kommen.

- Überprüfung von Entscheidungen in Zusammenhang mit Personalauswahl- und Personalentwicklungsprogrammen.
- Verbesserung von Personalauswahl- und Personalentwicklungsprogrammen.
- Evaluationsfunktion unter Berücksichtigung situativer und betriebswirtschaftlicher Größen; nutzenanalytische Modelle zwingen zur Explikation aller Einflußgrößen und schaffen damit Transparenz.
- Berücksichtigung des Kontextes bei der Bewertung von psychologischen Interventionsmaßnahmen.
- Ökonomische Demonstrations- und Legitimationsfunktion von PAP. Dies gilt insbesondere, wenn PAP als Investitionsvorhaben betrachtet werden und in Konkurrenz mit anderen Investitionsvorhaben stehen.
- Hilfe bei der Ressourcenbemessung im Zusammenhang mit Interventionsmaßnahmen.

-
- Kommunikationsfunktion durch Darstellung von Personalauswahlinvestitionen in Geldeinheiten und damit Schaffung von Sprachkompatibilität zu anderen, evtl. disziplinfremden Investitionsmöglichkeiten.
 - Beschreibung, Erklärung und Vorhersage von Entscheidungsprozessen bei Personalauswahl- und Personalentwicklungsprogrammen (Boudreau, 1991).
 - Die explizite Aufnahme der Entscheidungsgrundlagen soll die Rationalität von Entscheidungen verbessern. Ansätze in dieser Richtung sind insbesondere deswegen wichtig, weil die Psychologie die nicht bewußte Irrationalität vieler Entscheidungen vielfältig gezeigt hat (z.B. Dörner, 1992; Kahneman & Tversky, 1972, 1973; Simon, 1957, 1976).

3 Historische Entwicklung der Nutzenmodelle

In diesem Kapitel wird die grundlegende Entwicklung der Nutzenmodelle skizziert, auf welche die NA aufbaut. Diese nimmt einerseits ihren Ausgangspunkt bei der Diskussion um die Interpretationsschwierigkeiten des Validitätskoeffizienten. Zum anderen betonten viele Forscher die Wichtigkeit der Besonderheiten der diagnostischen Situation, wenn der Wert von diagnostischen Instrumenten beurteilt werden soll. Im Zusammenhang mit der Personalauswahl habe ich schon darauf hingewiesen, daß dieser Aspekt unmittelbar aus dem Prozeßmodell (Abb. 2) erkennbar ist.

Zunächst werde ich das Konzept des Validitätskoeffizienten und Maße, welche von diesem ableitbar sind, vorstellen. Da diese Maßzahlen bei der Beurteilung psychodiagnostischer Verfahren eine weite Verbreitung gefunden haben, werde ich ausführlich auf die Kritik eingehen, die an ihnen geübt wurde. Dabei werden die Arbeiten dreier Autorengruppen zur Sprache kommen, die in der NA-Diskussion grundlegend sind. Diese Arbeiten von Taylor und Russel (1939), Brogden (1946a, 1946b, 1949), Brogden und Taylor (1950) und Cronbach und Gleser (1965) werden dann in den Abschnitten 3.2, 3.3 und 3.4 ausführlich dargestellt.

Der Abschnitt endet mit dem Nutzenmodell von Cronbach & Gleser (1965). Dieses Modell stellt bis heute die Grundlage der meisten Nutzenmodelle dar. Im Anschluß wird dann in Kapitel 4 auf die Schwierigkeiten der Parameter-operationalisierung beim Cronbach-Gleser-Modell und deren Lösungsversuche eingegangen. Spätere Erweiterungen von Nutzenmodellen und eigene Vorschläge werden in Kapitel 5 dargestellt.

Die mathematischen Variablen der Nutzenmodelle tragen in jeder Veröffentlichung der NA-Literatur eine andere Bezeichnung. Dies kann insbesondere beim schnellen Nachschlagen zur Verwirrung führen. In dieser Arbeit werden die Bezeichnungen so weit als möglich konsequent durchgeführt.

3.1 Der Validitätskoeffizient und verwandte Konzepte

3.1.1 Der Validitätskoeffizient

Das grundlegende und bekannteste Konzept bei der Beurteilung des Wertes diagnostischer Instrumente ist der Validitätskoeffizient r_{xy} . Es handelt sich dabei meist um die Pearson'sche

Produkt-Moment-Korrelation zwischen dem Prädiktor x und dem Kriterium y . So verbreitet seine Verwendung innerhalb der Sozial-, Verhaltens- und Wirtschaftswissenschaften jedoch ist, so unscharf bleibt in manchen Fällen seine Interpretation. Würde man heute einige Sozialwissenschaftler fragen, was den nun genau eine Validität von 0,5 bedeutet, so würde man sicherlich sehr verschiedene Antworten erhalten. Einige unter ihnen werden auf andere Maßzahlen verweisen, die aus dem Validitätskoeffizienten ableitbar sind. Andere wiederum werden auf die Spezifität und das Ziel der diagnostischen Situation verweisen. Beide haben unter bestimmten Randbedingungen recht.

Zur Kritik des Validitätskoeffizientenkonzeptes bei der Beurteilung der Nützlichkeit eines Tests werde ich die drei Ansätze von Brogden (1946a), Taylor und Russel (1939) sowie Cronbach und Gleser (1965) heranziehen. Nach der kurzen Besprechung der Kritik dieser Autoren am Konzept des Validitätskoeffizienten in diesem Abschnitt (3. 1), werden ihre nutzenanalytischen Ansätze vertieft in den folgenden Abschnitten 3.2 bis 3.4 besprochen.

3.1.2 Funktionen des Validitätskoeffizienten

In den frühen Monographien zur Statistik finden sich bereits statistische Maßzahlen, die direkt aus dem Validitätskoeffizienten ableitbar sind (z.B. Hull, 1928; Kelley, 1923). Die wichtigsten unter ihnen sind

- der Alienationskoeffizient $k=(1-r_{xy}^2)^{1/2}$ (1a)
- das Vorhersageeffizienzmaß (Index of forecasting efficiency) $E = 1 - k$
- der Determinationskoeffizient r_{xy}^2

Alle diese Maßzahlen sind Funktionen des Validitätskoeffizienten. Sie wurden dazu konzipiert, die Interpretation des Validitätskoeffizienten zu unterstützen. Beim *Alienationskoeffizienten* ist die Affinität zum Standardschätzfehler (SSF) der klassischen Testtheorie unverkennbar. Dieser wird definiert als

$$SSF=SD_y(1-r_{xy}^2)^{1/2} \quad (1b)$$

Damit kann man den Alienationskoeffizienten als den Teil der Standardabweichung der Kriteriumswerte SD_y interpretieren, welche die Schätzung der Vorhersage des Kriteriums y aufgrund der (nicht perfekten) Validität des Prädiktors x hat.

In Abschnitt 2.2.3.1 habe ich die Unterscheidung zwischen der individuellen vs. institutionellen Entscheidung vorgestellt (Cronbach & Gleser, 1965). An dieser Stelle wurde festgestellt: Will man im Falle einer individuellen Entscheidung den Kriterienwert eines

Objektes auf der Grundlage seines Prädiktorwertes einigermaßen sicher vorhersagen, so benötigt man in der Praxis unrealistisch hohe Validitätskoeffizienten. Die Konzepte des Standardschätzfehlers bzw. des Alienationskoeffizienten machen dies deutlich. Es sei beispielsweise der Mittelwert eines Kriteriums $M_y=100$ und die Standardabweichung $SD_y=10$. Ein Proband wird auf grund eines Test auf den Kriteriumswert $y=110$ geschätzt. Die Validität des Test bzgl. des Kriteriums liegt bei $r_{xy}=0,8$ sehr hoch. Der Alienationskoeffizient berechnet sich nach (1a) auf $k=0,6$ und der Standardschätzfehler ist nach (1b) $SSF=6$. Das 95 %-Konfidenzintervall (KI_{95}) für den Probanden berechnet sich nach

$$KI_{95} y \pm 1,96 * SSF$$

$$KI_{95} = 110 \pm 1,96 * 6 = 110 \pm 11,8$$

Die Werte $-1,96$ und $+1,96$ sind dabei die z-Werte der Standardnormalverteilung, welche die mittleren 95% ihrer Fläche begrenzen. Damit liegt die Kriteriumsschätzung unseres Probanden mit 95-prozentiger Wahrscheinlichkeit zwischen $y=98,2$ und $y=121,8$. Anders ausgedrückt liegt sein Prozentrang zwischen 42 und 98. Daß dies ein äußerst unbefriedigendes Ergebnis ist, liegt auf der Hand. Damit wird die Behauptung, daß individuelle Entscheidungen sehr hohe Validitäten benötigen, eindrucksvoll bestätigt.

Das *Vorhersageeffizienzmaß* ist lediglich eine einfache Funktion des Alienationskoeffizienten. Es vergleicht den Standardfehler (SF) der durch den Test vorhergesagten Werte SF_{y^*} mit dem Standardfehler der Kriteriumswerte SF_y , wenn diese nur aufgrund des Mittelwertes der Kriteriumswerte (M_y) vorhergesagt werden, also keine individuelle Testinformation vorliegt. Man kann also sagen, daß das Vorhersageeffizienzmaß den Wert eines Tests dadurch festlegt, inwieweit der Test den Standardfehler der Kriteriumswerte SF_y proportional reduziert.

Der *Determinationskoeffizient* ist eine statistische Maßzahl, die auch heute noch oft zur Interpretation des Korrelationskoeffizienten herangezogen wird. Er gibt den Anteil der Varianz der Schätzwerte des Kriteriums an, welche diese an der Gesamtvarianz der tatsächlichen Kriteriumswerte haben. Diesen Zusammenhang kann man sich leicht anhand eines Streudiagramms verdeutlichen. An dieser Stelle soll aus Platzgründen darauf verzichtet werden.

Der *Validitätskoeffizient* gehört als Korrelationskoeffizient zu den bekanntesten Konzepten der sozialwissenschaftlichen Statistik. Die Interpretation dieser Maßzahl bereitet jedoch oft Schwierigkeiten, was die Existenz der Funktionen des Korrelationskoeffizienten bereits andeutet. Diese Problematik versucht Brogden (1946a) aufzulösen.

3.1.3 Kritik der Konzepte

Das Konzept des Validitätskoeffizienten soll hier unter dem Gesichtspunkt betrachtet und kritisiert werden, inwieweit es zur Beurteilung der Nützlichkeit eines diagnostischen Verfahrens herangezogen werden kann.

3.1.3.1 Interpretation des Validitätskoeffizienten durch Brogden (1946a)

Brogden (1946a) beschäftigt sich mit der Bewertung von Tests zur Vorhersage von Merkmalen. Insbesondere Personalauswahlsituationen stehen dabei im Zentrum seines Interesses. Zunächst beschreibt Brogden den Zustand, daß bei der Beurteilung eines Tests bzgl. seiner Kriterienvorhersagekraft die Unsicherheit besteht, welche Maßzahl Verwendung finden soll. Da neben dem Validitätskoeffizienten auch der Alienationskoeffizient oder der Determinationskoeffizient zur Interpretation der Testnützlichkeit herangezogen werden, hat dies zu dem Zustand geführt, daß diese Nützlichkeit, unter sonst gleichen Bedingungen, unterschiedlich eingeschätzt wird. Darüber hinaus führt die Verwendung der beiden letztgenannten Maßzahlen zu einer sehr gering eingeschätzten Testnützlichkeit. Dies liegt daran, daß die Werte dieser Koeffizienten insbesondere im Bereich einer niedrigen bis mittleren Validität wesentlich kleiner sind, als die des Validitätskoeffizienten (bei $r_{xy}=0,4$ ist $r_{xy}^2=0,16$, $k=0,92$ bzw. $E=0,08$).

Der Kern der Arbeit von Brogden (1946a) besteht darin, daß er eine direkte lineare Beziehung der Vorhersageeffizienz eines Tests mit seinem Validitätskoeffizienten zeigt. Dies bedeutet, daß der Validitätskoeffizient ein direktes Maß bei der Beurteilung der Vorhersagekraft eines Tests darstellt, wenn die Regression von x auf y linear ist, und die Verteilung der beiden Maße gleich ist. Ein Test mit einer Validität von 0,5 ist demnach doppelt so nützlich bei der Vorhersage eines Kriteriums, wie eine Validität von 0,25 und halb so nützlich wie eine Validität von 1. Eine Steigerung der Validität von 0,4 auf 0,45 bedeutet die gleiche Verbesserung der Vorhersageleistung, wie eine Steigerung von 0,9 auf 0,95.

Brogden beschäftigt sich in seinem Aufsatz hauptsächlich mit der Personalauswahl. Ihm geht es dabei weniger um den Zusammenhang von gemessenen und wahren Werte, wie dies die klassische Testtheorie nahelegt, sondern vielmehr um die sog. 'Einsparung' (savings), die durch die Verwendung von Tests bei der Personalauswahl geleistet werden kann. Die Einsparung ist die Differenz zwischen dem mittleren Kriteriumswert (z.B. der mittleren Arbeitsleistung) der angenommenen Bewerber und dem mittleren Kriteriumswert, den die

Population aller Bewerber erzielen würde. Brogden behauptet plausiblerweise, daß eine maximale Einsparung dann geleistet werden kann, wenn die Personalselektion auf der Grundlage des Kriteriumswertes selbst vorgenommen werden könnte. Mit anderen Worten heißt dies beispielsweise, daß die bestmögliche Einsparung dann erzielt werden kann, wenn alle Bewerber zunächst angenommen würden und aus diesen dann diejenigen ausgewählt würden, welche die höchste Arbeitsleistung erbrächten. Der Validitätskoeffizient ist der Anteil der Einsparung, die auf der Grundlage eines Tests erzielt werden kann, an der Einsparung, die auf der Grundlage des Kriteriums selbst erzielt werden könnte. Nun ist natürlich klar, daß der Kriteriumswert (z.B. die Arbeitsleistung) bei der Auswahl noch nicht vorliegt. Dennoch ist diese theoretisch denkbare Begründung beeindruckend. Das ist um so mehr der Fall, wenn man bedenkt, daß die Alternative der klassischen Testtheorie, die maximale Einsparung durch Kenntnis des wahren Wertes zu erreichen, zu den oben beschriebenen Mißverständnissen bei der Interpretation des Validitätskoeffizienten geführt haben.

Bei der Personalauswahl mit einem positiv validen Test sollte die selektierte Gruppe höhere standardisierte Kriteriumswerte erzielen als die Gesamtgruppe. Im Sprachgebrauch Brogdens sollte also eine positive Einsparung vorliegen. Die höchsten Werte würde wiederum eine Gruppe erreichen, die mittels des Kriteriums selbst ausgewählt wurde. Eine durch den Test selektierte Gruppe erbringt einen Teil der Steigerung der Gruppe, welche durch Selektion mittels des Kriteriums selbst erzielt werden könnte. Diesen Teil gibt der Validitätskoeffizient an. Nach Brogden gilt, daß die Auswahl durch einen Test mit $r_{xy} = 0,4$ 40% der mittleren Kriteriumswertsteigerung erbringt, welche die Auswahl durch das Kriterium selbst leisten könnte. Ebenso gilt, daß eine Validitätssteigerung von 0,1 eine gleich große Steigerung des mittleren standardisierten Kriteriumswertes der selektierten Gruppe bedeutet, ganz gleich, ob die Validität von 0,4 auf 0,5 oder von 0,8 auf 0,9 angehoben wird.

Ich will dieses Verständnis noch einmal am Prozeßmodell der Personalauswahl (Abb. 2) darstellen. Nehmen wir an, daß der Validitätskoeffizient die Korrelation zwischen der Information auf Stufe 3 und einem bewerteten Handlungsergebnis auf Stufe 8 ist. Die optimale Entscheidung, welche Behandlung (Stufe 5) einem Objekt zuteil werden sollte, würde auf der Information über die bewerteten Handlungsergebnisse selbst basieren. Die Entscheidung für eine Behandlung, welche auf der Information auf Stufe 3 basiert, erbringt nur $r_{xy} * 100$ % der Einsparung, bzw. $r_{xy} * 100$ % so gute bewertete Handlungsergebnisse, wie die optimale Entscheidung.

Wie schon bemerkt wurde, beinhalten auch Nutzenmodelle Validitätskoeffizienten. Die Interpretation des Validitätskoeffizienten gemäß der Sichtweise von Brogden ist dabei die Grundlage für nahezu alle modernen Nutzenmodelle. Dennoch gelten auch bei der richtigen Interpretation weiterhin die anderen Kritikpunkte, die Validierungsstudien auf sich vereinigen.

3.1.3.2 Keine Bewertung ohne Beachtung der Situation

Der Wert eines Tests ist ohne Zweifel stark von seiner Validität abhängig. Diese Kenngröße ist jedoch nicht allein ausschlaggebend. Darauf verweisen mit Nachdruck Taylor und Russel (1939). Auch hier wird zunächst auf die Interpretationsschwierigkeiten des Validitätskoeffizienten und dessen Funktionen eingegangen. Taylor und Russel vertreten jedoch im Gegensatz zu Brogden (1946a) die Ansicht, daß keines dieser Konzepte die Nützlichkeit eines Tests beschreiben kann. Stattdessen verweisen sie auf situative Bedingungen, welche bei der Testbewertung großen Einfluß haben. Auch Taylor und Russel behandeln das Problem der Personalauswahl. Hierbei dichotomisieren sie vereinfachend das Berufserfolgskriterium in 'geeignete' bzw. 'ungeeignete Mitarbeiter'. Zwei situative Bedingungen werden von den Autoren diskutiert. Dies sind die Basisrate (B), welche den Anteil der tatsächlich geeigneten Bewerber an allen Bewerbern in der unausgelesenen Stichprobe darstellt, und die Selektionsquote (Q). Die Selektionsquote ist der Quotient aus der Anzahl der angenommenen Bewerber und der Gesamtzahl der Bewerber.

Ausschlaggebend für den Wert eines Testprogrammes ist nun für Taylor und Russel (1939), inwieweit die Applikation des Testverfahrens gegenüber der Zufallsauswahl den Anteil der geeigneten an den angenommenen Bewerbern erhöht. Nehmen wir beispielsweise eine Basisrate von $B=0,5$ an, die Hälfte der Bewerber sind also a priori geeignet. Die Validität eines Tests sei $r_{xy}=0,5$ im Vergleich zur Nullvalidität der Zufallsauswahl. Werden aus den Bewerbern die 30% besten ausgewählt ($Q=0,3$), so würde dies bei einer Zufallsauswahl bedeuten, daß 50% der Angenommenen geeignet sind. Bei der Personalauswahl durch unseren Test ist dagegen zu erwarten, daß 74% der Ausgewählten geeignet sind. Werden lediglich die besten 10% ausgewählt, so wächst der Anteil der Geeigneten auf 84% an. Diese Zahlen sind aus Tafeln ablesbar, die bei Taylor und Russel (1939) zugänglich sind. Das Ergebnis kann weder aus Sicht des Validitätskoeffizienten, noch aus der Perspektive seiner zahlreichen Funktionen antizipiert werden.

Allgemein gesagt steigt der Anteil der Geeigneten an den ausgewählten Bewerbern mit der Erhöhung der Basisrate und der Validität und der Minderung der Selektionsquote.⁶ Das Modell von Taylor und Russel (1939) wird in Abschnitt 3.2 vertieft dargestellt. An dieser Stelle soll es nur um die Kritik am Konzept des Validitätskoeffizienten gehen. Der Kern dieses Ansatzes ist die Betonung der Wichtigkeit des Kontextes, in dem Personalauswahl stattfindet. Ohne diesen kann keine Bewertung der Nützlichkeit eines PAP stattfinden. Dieser Aspekt wurde in etwas abgewandelter Form bereits am Prozeßmodell der Personalauswahl (Abb. 2) verdeutlicht.

3.1.3.3 Entscheidungstheoretische Interpretation

Der Ansatz von Cronbach und Gleser (1965) beruht auf entscheidungstheoretischen Betrachtungen. Er kritisiert das Konzept des Validitätskoeffizienten als Nützlichkeitsmaßzahl für Tests von ganz anderer Seite als die vorhergehenden Autoren, obwohl er z.T. auf deren Erkenntnissen beruht. Den Kern der Kritik treffen Cronbach und Gleser schon auf der ersten Seite ihres Buches (1965, S.1), wenn sie schreiben:

Hull (1928, p.268) voiced a principle that has been the root of nearly all work on test theory: 'The ultimate purpose of using aptitude tests is to estimate or forecast aptitudes from test Scores.' It is this view that we propose to abandon.

Cronbach und Gleser schlagen also vor, das Credo vieler Psychodiagnostiker aufzugeben, nämlich 'wahre Werte' zu schätzen und daraus Vorhersagen zu machen. Ich denke, es ist den Autoren bis heute noch nicht vollständig gelungen, diese Denkhaltung zu durchbrechen. Noch immer finden ihre Gedanken innerhalb der akademischen Ausbildung wenig Gehör. Immer noch träumen Psychologen davon, daß ihre diagnostischen Verfahren eines Tages die Güte von physikalischen Meßinstrumenten erreichen. Bei der Besprechung des Standardschätzfehlers habe ich gezeigt, wie verheerend schlecht unsere individuellen Kriterienvorhersagen sind, wenn dieses Ziel nicht erreicht wird. Obwohl diese Sachlage auch Psychologen in der Praxis bekannt ist, verhalten sich viele immer noch so, als ob ihre Meßinstrumente und ihre Vorhersagen auch bei individuellen Entscheidungen tauglich sind.

⁶ Eine kleinere Selektionsquote heißt inhaltlich, daß der Anteil der akzeptierten Bewerber an allen Bewerbern kleiner wird. Der numerische Wert von Q sinkt damit, die personalauswählende Organisation verhält sich 'wählerischer'.

Was aber ist das Neue bei Cronbach und Gleser? Sie sehen den letztendlichen Sinn und Zweck von Testverfahren nicht in der möglichst genauen Schätzung von hypothetischen Parametern oder zukünftigen beobachtbaren Sachverhalten, sondern im Fällen von möglichst richtigen Entscheidungen. Bei der Bewertung von Entscheidungen ist aber nicht nur deren objektive Qualität, sondern insbesondere ihre Güte in Relation zu Alternativentscheidungen zu beachten.

Im Hinblick auf die Interpretation des Validitätskoeffizienten zeigt diese Sichtweise deutlich, daß die schätzfehlerorientierten Funktionen der Korrelation einen wesentlichen Umstand vergessen. Schon Wesman (1953) hat darauf hingewiesen, daß bei der Personalauswahl die Überschätzung der ausgewählten Bewerber und die Unterschätzung der abgelehnten Bewerber nichts ausmachen, im Grunde gar keine Fehler sind. Eine entscheidungstheoretische Begutachtung macht deutlich, warum dies so ist: Es wurde trotz der objektiven fehlerbehafteten Messung bei diesen Individuen die richtige Entscheidung getroffen. Der Nutzen des Tests wird dadurch also nicht gemindert. Nur in Fällen, bei welche die falsche positive oder negative Entscheidung getroffen wurde, wird die Nützlichkeit des Testverfahrens beeinträchtigt.

Welche Variablen beeinflussen aber die Qualität der Entscheidung? Die Validität des Tests gehört sicher dazu, sie allein ist es jedoch nicht. Ebenso habe ich bereits die Basisrate und die Selektionsquote als wichtige Parameter im Zusammenhang mit dem Modell von Taylor und Russel (1939) besprochen. Cronbach und Gleser bleiben jedoch an dieser Stelle nicht stehen. Die Art der Entscheidung spielt bei ihnen eine zentrale Rolle. Im Abschnitt 2.2.3.1 habe ich deren Unterscheidung zwischen individuellen und institutionellen Entscheidungen vorge- stellt. Hierbei wurde in einem etwas anderen Zusammenhang deutlich, daß die Beurteilung eines Tests ganz wesentlich davon abhängig ist, welches Wertsystem der Beurteilung zugrundeliegt. Die Frage, ob die Höhe der Validität zufriedenstellend ist oder nicht, kann erst beantwortet werden, nachdem die Frage geklärt wurde, um welche Entscheidungsart es sich handelt.

Wenn PAP bezüglich ihres Wertes beurteilt werden sollen, ist des weiteren wichtig, welche Handlungsergebnisse die eingestellten Individuen erzielen und welche bewerteten Ergebnisunterschiede dabei auftreten (Stufe 7 und 8 des Prozeßmodells, Abb. 2). Was nützt das beste Auswahlverfahren, bei dem die schnellsten Montagearbeiter sicher ausfindig gemacht werden können, wenn die Arbeitsgeschwindigkeit am Fließband durch die langsamsten Arbeiter bestimmt wird. In diesem Fall besteht keine Varianz zwischen den Handlungsergebnissen der guten und der schlechten Arbeiter. Das PAP erbringt in diesem

Fall keinen Nutzen. Das PAP mit seinen hochvaliden Diagnostica kann sogar, wenn es sehr teuer ist und dabei doch nichts bringt, sogar einen negativen Nutzen haben und damit als sehr schlecht beurteilt werden. Damit ist eine weitere Situationsvariable identifiziert, welche bei der Bewertung von PAP beachtet werden muß: Die Kosten des Tests.

Diese Andeutungen sollen zunächst genügen. Der Ansatz von Cronbach & Gleser (1965) wird in Abschnitt 3.4 genauer dargestellt. An dieser Stelle ist nur das Verständnis wichtig, warum der Validitätskoeffizient alleine nicht tauglich ist, ein Testverfahren zu bewerten. Ich denke, daß die Argumente dafür zahlreich und schwer widerlegbar sind.

3.2 Das Taylor-Russel Modell

3.2.1 Darstellung des Modells

Die Grundzüge des Modells von Taylor und Russel (1939) wurden bereits im Zusammenhang mit der Kritik des Validitätskoeffizienten als Maß der Testbewertung vorgestellt. Im Modell werden neben der Validität auch situative Bedingungen bei der Personalauswahl berücksichtigt. Die Autoren sehen den Nutzen eines Tests dadurch charakterisiert, daß er den Anteil der Geeigneten unter den ausgewählten Bewerbern gegenüber einer Zufallsauswahl steigert. Relevante Größen sind die Basisrate B , die Selektionsquote Q , die Validität r_{xy} und der Anteil der Geeigneten unter den Eingestellten (EG). Das Merkmal 'geeignet' bzw. 'ungeeignet' wurde von Taylor und Russel als künstlich dichotom angenommen. Die Zusammenhänge zwischen EG, der Validität, der Selektionsquote und der Basisrate veranschaulicht Abbildung 3. Entscheidend bei Taylor und Russel ist das Verhältnis von EG zu den eingestellten Ungeeigneten (EU). Im weiteren Verlauf der Arbeit wird dieses Verhältnis die 'Erfolgsquote' genannt. Wie zu erkennen ist, wird die Erfolgsquote mit steigender Validität, steigender Basisrate und fallender Selektionsquote günstiger. Nicht berücksichtigt wird von den Autoren der Anteil der AG (abgelehnte Geeignete), der ebenfalls eine falsche Entscheidung darstellt und negative Konsequenzen für eine Organisation haben kann.

Bei Taylor und Russel finden sich mehrere Tabellen, aus denen die erwartete Erfolgsquote in Abhängigkeit der Validität und der situativen Randbedingungen B und Q ablesbar sind. Solche sog. Taylor-Russel Tabellen für echt dichotome Kriteriumsvariablen, also für solche Kriterien, die nicht künstlich dichotomisiert wurden und welchen keine Verteilungsannahme zugrunde liegt (z.B. männlich-weiblich), sind bei Abrahams, Alf und Wolfe (1971) abgedruckt.

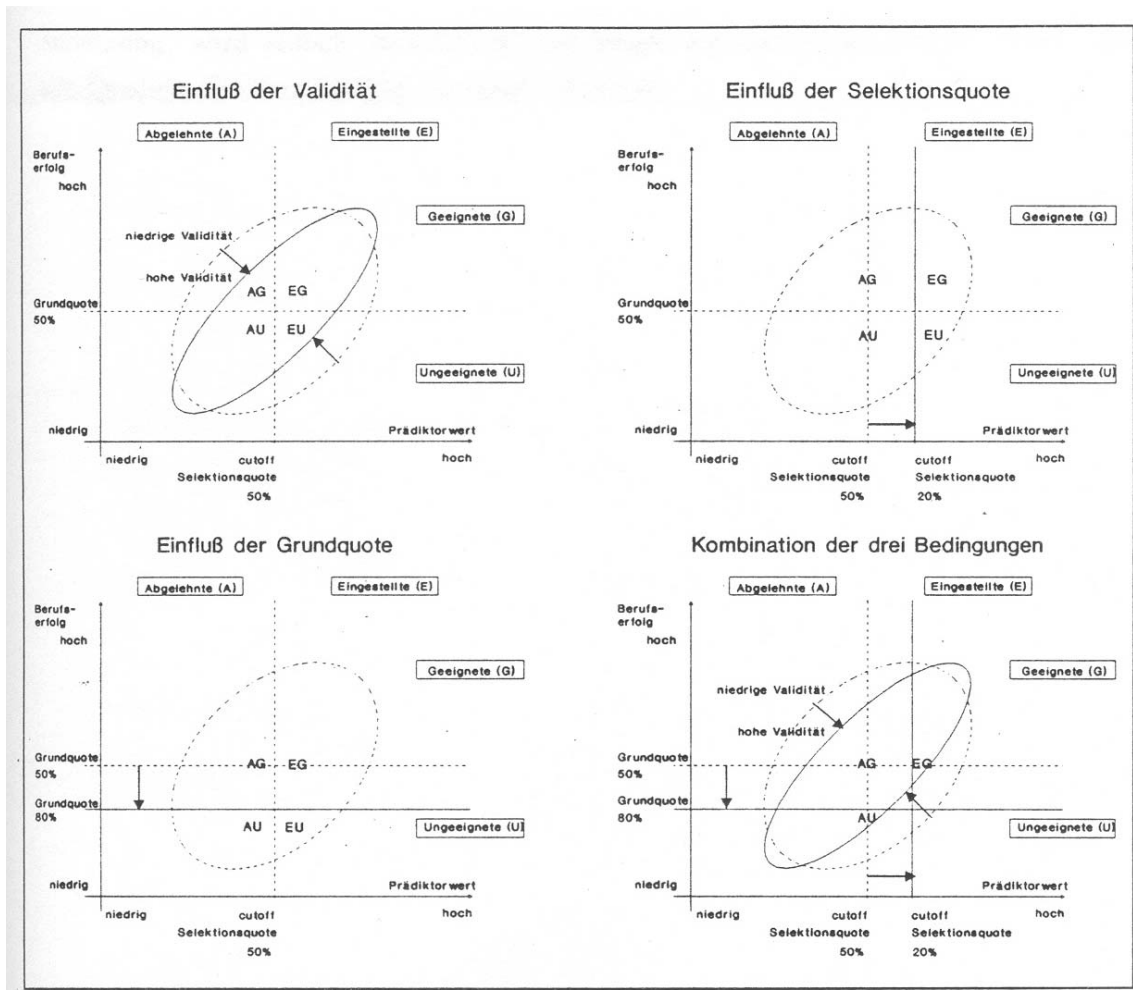


Abbildung 3: Die Erfolgsquote von Personalauswahlentscheidungen hängt von verschiedenen situativen Bedingungen ab. Dies ist die Validität r_{xy} , die Selektionsquote Q , und die Basisrate B , die hier Grundquote genannt wird (aus Schaler & Funke, 1993, S. 271).

Anhand ihrer Überlegungen zeigen Taylor und Rassel interessante Zusammenhänge auf. Wie aus Abbildung 4 zu entnehmen ist, ist die Zunahme der Erfolgsquote mit abnehmender Selektionsquote bei Validitäten unter 0,7 positiv beschleunigt. Da in den meisten praktischen Situationen die Validität deutlich unter dieser Marke liegen wird, ist dieser Umstand bemerkenswert.

Aus der Abbildung 4 wird außerdem deutlich, daß hohe Selektionsquoten auch hoch valide Testverfahren unwirksam machen. Oft werden aus scheinbar zwingenden Rahmenbedingungen hohe Selektionsquoten gefordert. Dies kann z.B. der Fall sein, wenn an die Personalabteilung die Forderung gestellt wird, vakante Stellen schnell zu besetzen. Der verantwortliche Psychologe ist in dieser Situation meist gut beraten, wenn er durch verstärkte Personalwerbung mehr Bewerber gewinnt, um dann aus diesen die Geeigneten ausfindig zu machen. Eine Anhebung der Selektionsquote statt dessen befriedigt zwar kurzfristig die

Wünsche nach Vollbesetzung, wird jedoch mittelfristig und langfristig nachteilige Folgen haben, da die Qualität des eingestellten Personals abnimmt.

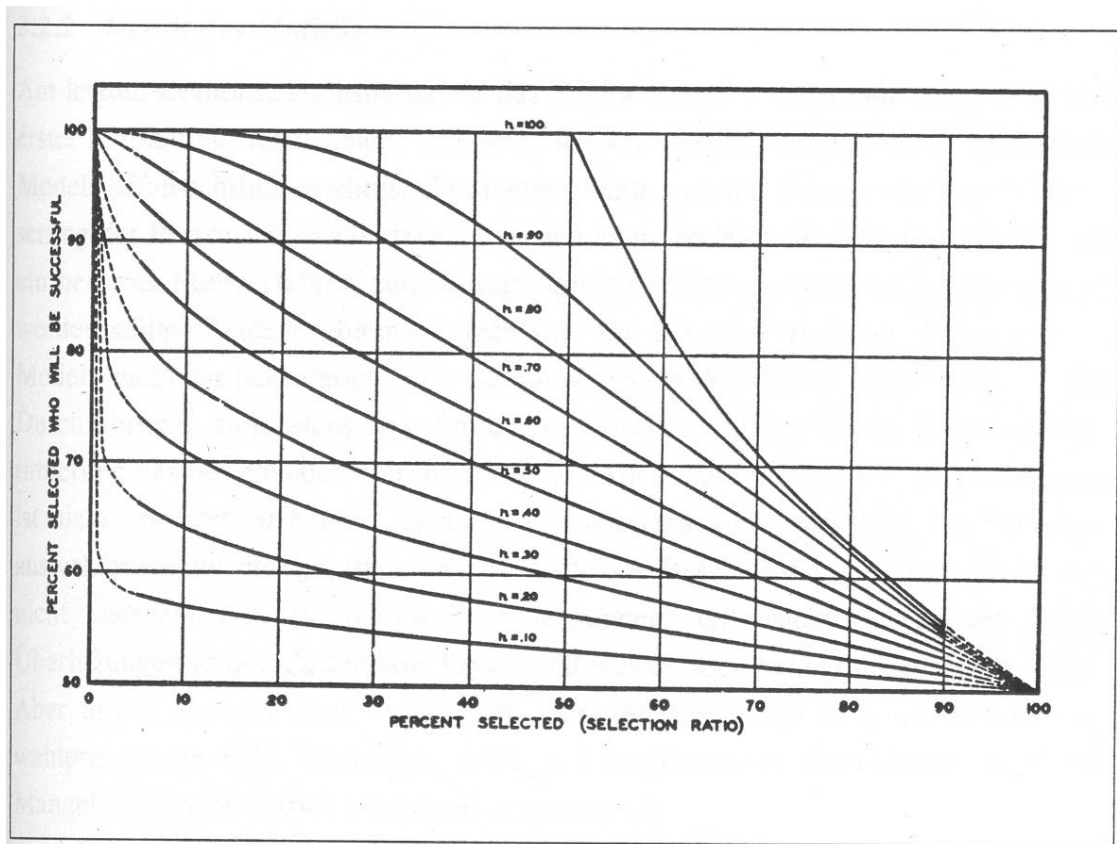


Abbildung 4: Wert eines Testverfahrens, ausgedrückt anhand der Erfolgsquote, bei der Basisrate $B=50\%$. Es sind verschiedene Validitätswerte und Selektionsquoten dargestellt (aus Taylor & Russel, 1939, S. 569).

Aus Abbildung 4 ist drittens ablesbar, daß auch der Wunsch eines jeden Psychologen, ein perfekt valider Test, nur unter bestimmten situativen Randbedingungen auch eine perfekte Erfolgsquote erbringen kann. Dieser Umstand ist allerdings trivial, da auch ein perfekter Test nicht die Qualifikationen der Bewerber anheben kann.

Viertens sei darauf hingewiesen, daß eine Steigerung der Validität um einen bestimmten Betrag eine unterschiedliche Änderung der Erfolgsquotensteigerung erbringt. Bei niedrigen Selektionsquoten bringt eine Validitätssteigerung des Tests eine höhere Erfolgsquotensteigerung als bei hohen Selektionsquoten.

Wie uns Abbildung 4 vorführt, ist eine Steigerung der Erfolgsquote nicht nur auf dem Wege der Testvaliditätssteigerung erreichbar. In manchen Fällen wird es einfacher und weniger kostenaufwendig sein, die Selektionsquote zu senken, als die Validität zu steigern. So wird z.B. eine Validität von $r_{xy} = 0,5$ bei einer Selektionsquote $Q=0,7$ ähnlich viele erfolgreiche Mitarbeiter erbringen, wie ein Test mit $r_{xy} = 0,2$ bei $Q=0,25$.

3.2.2 Kritik des Modells

Am letzten Demonstrationspunkt für das Taylor-Russel Modell läßt sich gleich sein erster Kritikpunkt festmachen. Wie beim letzten Beispiel deutlich wird, fehlen dem Modell offensichtlich wichtige Parameter. Stellt sich die Frage, wie eine Verbesserung der Personalauswahl erreicht werden kann, so ist noch nicht beantwortet, ob ein besseres Meßverfahren zum Einsatz kommen oder die Selektionsquote gesenkt werden sollte. Beide Problemlösungen sind mit Kosten verbunden. Diese sind im Modell nicht berücksichtigt. Oft ist ein valideres Verfahren auch teurer in der Durchführung, mindestens aber kostet die wissenschaftliche Überprüfung und Optimierung des bestehenden Verfahrens Geld. Auch die Senkung der Selektionsquote ist nicht umsonst zu haben. Es müßten mindestens mehr Bewerber die Personalauswahlprozedur durchlaufen. Oft ist auch die Beschaffung von Bewerbern mit nicht unerheblichen Werbungskosten verbunden. Offensichtlich spielten solche Überlegungen in den Zeiten von Taylor und Russel noch keine dominierende Rolle. Aber in der heutigen Zeit, in der z.B. teure AC-Verfahren mit einfacheren Auswahlprozeduren (z.B. Interview, Tests, ...) verglichen werden müssen, ist dieser Mangel im Taylor-Russel Modell doch gravierend.

Ein zweiter Kritikpunkt bezieht sich auf die legitime und vereinfachende, aber doch stark realitätsverzerrende Dichotomisierung der Kriteriumsvariable. Ich möchte zu diesem Punkt nicht zu viele Worte verlieren, sondern nur andeuten, daß ein wesentlicher Bestandteil moderner Nutzenmodelle stetige Berufserfolgskriteriumsvariablen sind. Aus diesem Grunde darf die befruchtende Wirkung des Taylor-Russel Modells nicht unterschätzt werden.

Ein Ziel der Anstrengung der beiden Autoren war, eine angemessene Interpretationsmöglichkeit des Validitätskoeffizienten zu schaffen. Sie haben dieses Ziel auch erreicht, allerdings mit einem kleinen Schönheitsfehler. Die Prozentwerte der Erfolgsquote sind zwar leicht interpretierbar, allerdings muß eine solche Deutung durch die Dichotomisierung der Kriteriumsvariable sofort in Frage gestellt werden. Letztendlich denke ich, daß die Autoren das Problem nicht gelöst, sondern nur verlagert haben. Wie im nächsten Abschnitt gezeigt werden wird, hatte auch hier Brogden (1949) einen entscheidenden Gedanken.

3.3 Das Konzept von Brogden

Die Arbeiten von Brogden und seinen Kollegen sind Grundlage aller modernen Nutzenmodelle. In Abschnitt 3.1.3.1 wurde bereits ein wichtiger Punkt seiner Überlegungen dargestellt. Dies war im Zusammenhang mit der Kritik des Validitätskoeffizienten als Maßzahl des Testwertes. In diesem Abschnitt werde ich auf die Arbeiten von Brogden vertieft eingehen.

3.3.1 Der durchschnittliche standardisierte Kriteriumswert der ausgewählten Bewerber

Einer der Kernpunkte der Brogden-Konzepte findet sich im Aufsatz von 1946 (Brogden, 1946a). Hierbei geht es um die Interpretation des Korrelationskoeffizienten zwischen Testwert und Kriterium (Validitätskoeffizient). Entgegen vielen anderen Behauptungen zeigt der Autor, daß eine direkte lineare Beziehung der Vorhersageeffizienz eines Tests und dem Validitätskoeffizienten besteht.

Nach einer Personalauswahlsituation gibt es eine ausgewählte Gruppe und die ursprüngliche Bewerberpopulation. Der durchschnittliche standardisierte Kriteriumswert der Bewerberpopulation ist per definitionem $\bar{z}_y=0$. Dagegen bezeichne ich mit \bar{z}_y^* den Durchschnitt der standardisierten Kriteriumswerte der anhand des Tests ausgewählten Gruppe.⁷ Liegt \bar{z}_y^* über dieser Marke, so ist der Test positiv valide. Dies bedeutet, daß die durch den Test ausgewählten Bewerber bessere bewertete Handlungsergebnisse (vgl. Abb. 2) für die Organisation erbringen, als zufällig ausgewählte. Diese Verbesserung nennt Brogden 'Einsparung' (saving).

Im Kontext der Personalauswahl interpretiert Brogden die Höhe von r_{xy} als den Anteil der Kriteriumsvorhersagekraft durch den Prädiktor, welcher aufgrund des Kriteriums selbst geleistet werden könnte.

$$r_{xy} = \bar{z}_y^* / \bar{z}_y'$$

\bar{z}_y^* ist der Durchschnitt der standardisierten Kriteriumswerte, welche durch die Gruppe erreicht werden, die durch den Prädiktor ausgewählt wurde. \bar{z}_y' ist der (hypothetische)

⁷ Maßzahlen, die mit einem hochgestellten Stern gekennzeichnet sind, beziehen sich auf die Population der akzeptierten Bewerber, solche ohne Stern auf die Population aller Bewerber.

Durchschnitt der standardisierten Kriteriumswerte der selektierten Gruppe, wenn es möglich wäre, die Mitglieder durch das Kriterium selbst auszuwählen.

Der so verstandene Validitätskoeffizient erlaubt nun, auf Grundlage des Durchschnitts der standardisierten Prädiktorwerte der selektierten Gruppe (\bar{z}_x^*) die durchschnittlichen standardisierten Kriteriumswerte der selektierten Gruppe (\bar{z}_y^*) zu schätzen. Für die ausgewählte Gruppe gilt

$$\bar{z}_y^* = r_{xy} \bar{z}_x^* \quad (2)$$

Das heißt aber nichts anderes, als daß der Durchschnitt der standardisierten Kriteriumswerte der selektierten Stichprobe \bar{z}_y^* geschätzt werden kann, wenn der durchschnittliche standardisierte Prädiktorwert der durch den Test ausgewählten Stichprobe \bar{z}_x^* und der Validitätskoeffizient r_{xy} bekannt sind. \bar{z}_y^* und \bar{z}_x^* sind, wie der Stern indiziert, die jeweiligen durchschnittlichen standardisierten Kriteriums- bzw. Prädiktorwerte der selektierten Gruppe. Diese Behauptung gilt bei Vorliegen einer linearen Regression des Kriteriums auf den Prädiktor.⁸

Um diese Kernaussage der Gleichung (2) deutlicher zu machen, will ich ein kleines Beispiel geben, welches den mathematisch formulierten Inhalt veranschaulichen soll. Es existiere eine Bewerberpopulation. Würden alle Individuen der Population eingestellt, so wäre ein durchschnittlicher standardisierter Kriteriumswert von $\bar{z}_y=0$ zu erwarten. Der Bewerberpopulation wird ein Test vorgegeben. Auf der Informationsgrundlage der Testergebnisse wird zuerst der Beste ausgewählt, danach der Zweitbeste usw. bis die Anzahl der zu besetzenden Stellen ausgefüllt ist. Die so ausgewählten Bewerber nenne ich auch die selektierte Stichprobe. Ihr durchschnittlicher standardisierter Prädiktorwert wird aufgrund der Auswahlprozedur 'von oben nach unten' über 0 liegen, sagen wir beispielsweise bei $\bar{z}_x^*=1$. Dies bedeutet, daß der Testwert der selektierten Stichprobe im Durchschnitt eine Standardabweichung über dem der Bewerberpopulation liegt. Nach Gleichung (2) läßt sich nun \bar{z}_y^* schätzen. Hat der Test x bzgl. dem Kriterium y eine Validität von $r_{xy}=0,3$, so wird die selektierte Stichprobe im Durchschnitt $\bar{z}_y^*=0,3$ erreichen. D.h. sie erreicht Kriteriumswerte, die durchschnittlich 0,3 Standardabweichungen über dem Kriteriumswert der Bewerberpopulation liegen würde, wenn eine Zufallsstichprobe aus dieser Bewerberpopulation ausgewählt würde.

Nun hat diese Vorgehensweise in der Praxis einen gravierenden Nachteil. Aus datenschutzrechtlichen oder anderen Gründen sind die Testwerte der abgelehnten Bewerber

⁸ Die Haltbarkeit der Annahme des linearen Zusammenhangs wurde oft in Frage gestellt. Neuere empirische Arbeiten (z.B. Coward & Sackett, 1990; zit. nach Schmidt, Ones & Hunter, 1992) bestätigen jedoch diese Hypothese.

meist nicht mehr verfügbar, so daß \bar{z}_x^* nicht bekannt ist. In diesem Fall kann jedoch aufgrund der Selektionsquote Q indirekt auf \bar{z}_x^* geschlossen werden. Kelley (1923, zit. nach Boudreau, 1991) gibt hier statistische Hilfe. Werden die Bewerber auf der Grundlage der Prädiktorwerte von oben nach unten selektiert, und sind die Prädiktorwerte normalverteilt, dann ist der durchschnittliche standardisierte Prädiktorwert der ausgewählten Stichprobe eine Funktion des Anteils der Bewerberpopulation, die über dem cut-off (welcher durch Q definiert ist) liegen.

$$\bar{z}_x^* = \lambda_{(Q)}/Q \quad (3)$$

$\lambda_{(Q)}$ ist dabei die Ordinatenhöhe der Normalverteilungsfunktion an der Stelle Q . Abbildung 5 zeigt das Schaubild Normalverteilungsfunktion. Man kann sich die beschriebenen Zusammenhänge daran anschaulich machen, indem man sich vorstellt, daß bei einer gegebenen Selektionsquote der Teil der Bewerber abgelehnt wird, die links vom cut-off Punkt liegen. Der Durchschnitt der standardisierten Prädiktorwerte der verbleibenden Personen liegt dann über 0.

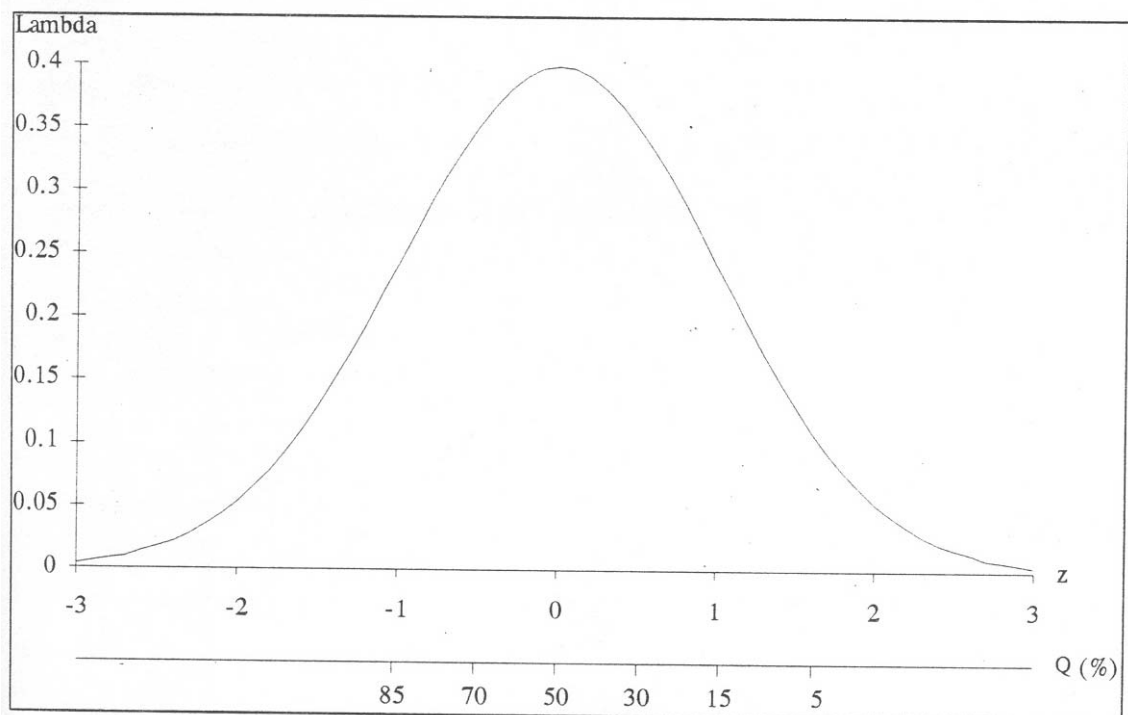


Abbildung 5: Das Schaubild der Normalverteilungsfunktion.

Der Zusammenhang zwischen der Selektionquote Q und dem z -Wert auf der Abszisse des Schaubildes ist nicht unmittelbar erkennbar. Naylor und Shine (1965) haben deswegen eine Tabelle entwickelt, aus welcher bei Kenntnis von Q direkt k bzw. \bar{z}_x^* ablesbar ist. Für heuristische Zwecke läßt sich \bar{z}_x^* auch aus Abbildung 6 entnehmen.

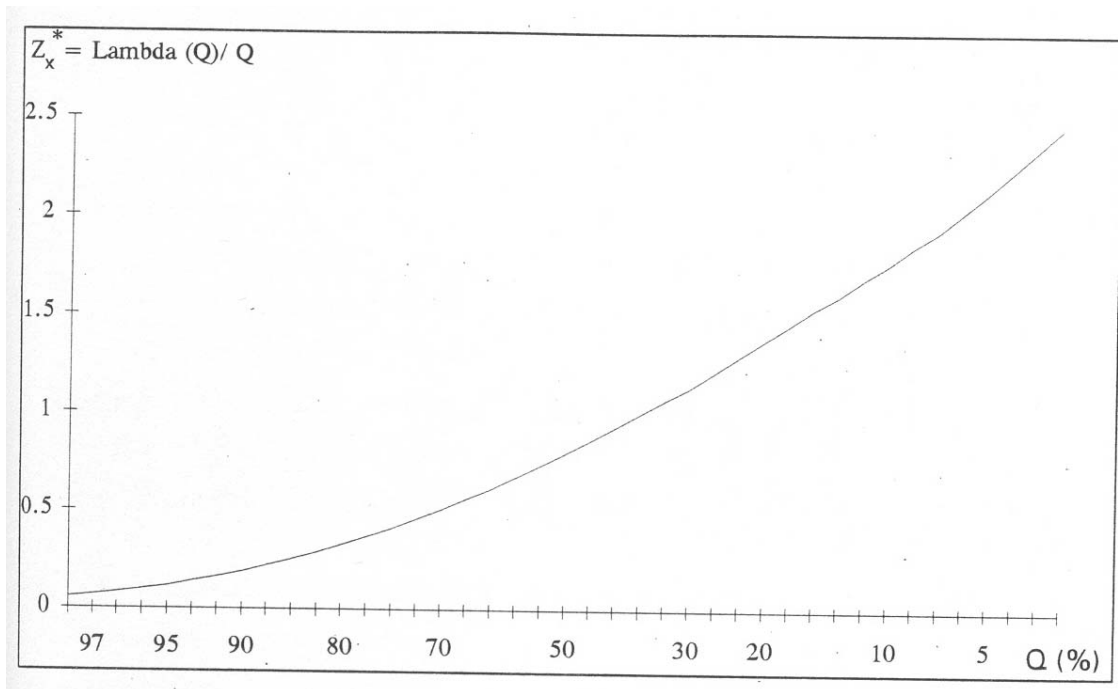


Abbildung 6: Darstellung des Zusammenhanges zwischen der Selektionsquote Q und \bar{z}_x^* . Das Schaubild sagt z.B. aus, daß die selektierte Stichprobe bei einer Selektionsquote von 30% einen durchschnittlichen standardisierten Prädiktorwert $\bar{z}_x^* \approx 1,2$ hat, sofern die Grundannahmen (Normalverteilung, Auswahl von oben nach unten) nicht verletzt sind.

In Gleichung (3) läßt sich in (2) einsetzen, so daß gilt

$$\bar{z}_y^* = r_{xy} \lambda_{(Q)}/Q \quad (4)$$

Diesen Zusammenhang erkannte Brogden (1949). Man kann mit Gleichung (4) die Steigerung des durchschnittlichen standardisierten Kriteriumswertes der ausgewählten Gruppe gegenüber der Bewerberpopulation berechnen, wenn die Validität des Selektionsinstrumentes und die Selektionsquote bekannt sind. Die Gleichung wird uns auch bei Cronbach und Gleser (1965) wieder begegnen. Die Herleitung ist an dieser Stelle allerdings nicht gerade einfach dargestellt, so daß Verständnisschwierigkeiten entstehen können. Ich hoffe, die Herleitung dieser Beziehung ist mir anschaulicher gelungen.

3.3.2 Das Dollarkriterium

Bei Brogden (1946a, 1946b) wird bereits auf die Möglichkeit hingewiesen, die bewerteten Handlungsergebnisse der Eingestellten nicht in standardisierten Kriteriumswerten auszudrücken, sondern in Geldeinheiten. Die konsequente Aufnahme von Geldeinheiten in ein Nutzenmodell nahm meines Wissens erstmals Brogden (1949) vor. Sein bekannt gewordener Aufsatz "When testing pays off" beginnt damit, die relevanten Variablen des Taylor-Russel Modells darzustellen. Wie in Abschnitt 3.2 deutlich wurde, zeigt das Taylor-

Russel Modell, daß außer der Validität die Selektionsquote sehr erheblichen Einfluß auf den Wert eines Testverfahrens hat. Eine niedrige Selektionsquote, d.h. sehr wählerisches Verhalten bei der Bewerberauswahl, sollte nach dem Modell zur Folge haben, daß die Ausgewählten vergleichsweise gut geeignet sind. Bis hierhin stimmt auch Brogden zu.

Wie allerdings von mir bei der Kritik des Taylor-Russel Modells schon angemerkt wurde, kostet eine Senkung der Selektionsquote auch Geld. Mit anderen Worten: Nicht die absolute Qualität eines Auswahlverfahrens ist wichtig, sondern seine relative Qualität, d.h. seine positiven Erträge in Relation zu den Kosten des Verfahrens. Diese Eigenschaft wird in der Ökonomie Nutzen genannt. Brogden (1949) diskutiert die Wirkung dieser Kosten sehr ausführlich. Ich möchte nicht alle seine Gedankengänge hier nachzeichnen. Wichtiger sind die Ergebnisse seiner Gedanken. Zum ersten zeigt er, daß die Testkosten in kompensatorischer Beziehung zu den Einsparungen ('savings') stehen, die durch, die Testanwendung entstehen. Da die Testkosten in Geldeinheiten ausgedrückt werden, fordert Brogden, auch die Einsparungen in derselben Einheit 'Dollar'⁹ zu beschreiben. Diese Denkhaltung führt letztlich zum sog. 'Dollarkriterium', mit dem sich Brogden und Taylor (1950) ausführlich beschäftigen.

Brogden (1949) zieht aus diesen Überlegungen weitere Schlüsse. Der Punkt der maximalen Einsparungen, die durch Personalauswahlverfahren erzielt werden können, ist nicht mehr nur abhängig von der Validität und der Selektionsquote, sondern auch von den Testkosten. Im Taylor-Russel Modell wurde das Auswahlverfahren als das nützlichste bewertet, welches möglichst valide war und die niedrigste Selektionsquote hatte. Sobald die Kosten des Verfahrens mitberücksichtigt werden, stimmt diese Aussage nicht mehr. Allgemein gesagt verschiebt sich der Punkt der maximalen Nützlichkeit mit zunehmenden Testkosten in Richtung einer höheren Selektionsquote. Diese Verschiebung ist um so stärker, je niedriger die Validität des Verfahrens ist. Bei sehr hohen Testkosten nähert sich der Punkt der maximalen Einsparungen der Selektionsquote $Q=50\%$ an. Abbildung 7 verdeutlicht diese Zusammenhänge. Es ist zu beachten, daß die Testkosten in diesem Schaubild für heutige Verhältnisse als sehr niedrig angenommen werden.

⁹ Die Bezeichnung 'Dollar' steht in diesem Zusammenhang selbstverständlich nicht als Symbol für ein bestimmtes Währungssystem. Sie ist durch alle Währungssystembezeichnungen ersetzbar.

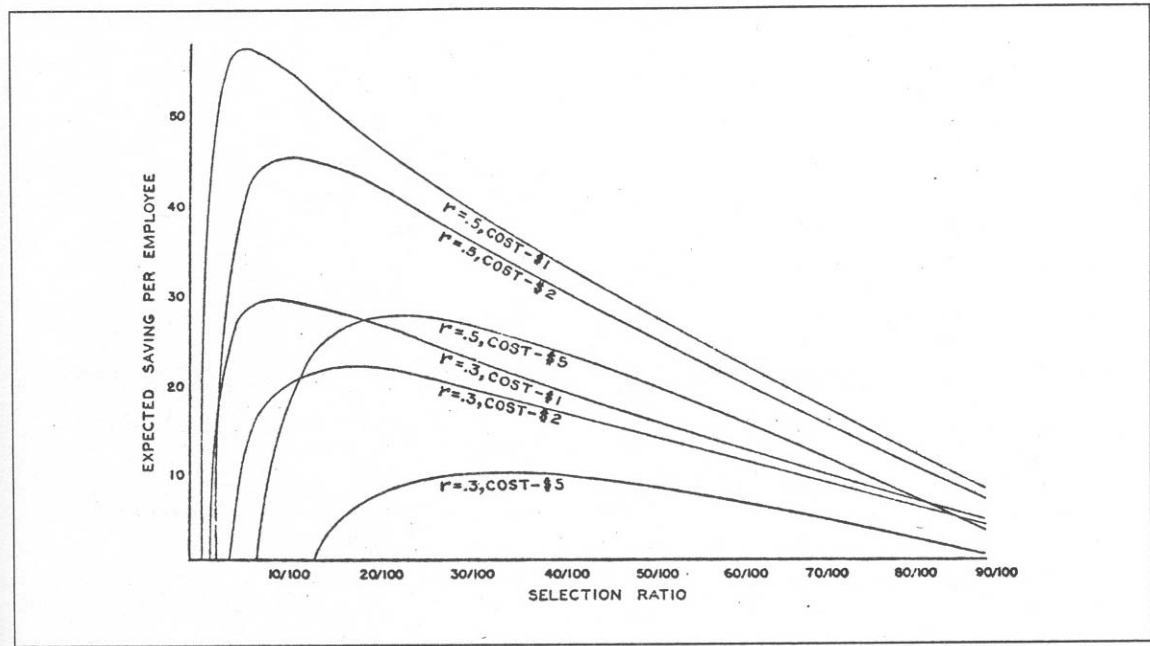


Abbildung 7: Erwartete Nützlichkeit eines Auswahlverfahrens in Abhängigkeit der Validität, der Selektionsquote und der Testkosten. Die Nützlichkeit ist als Einsparung pro ausgewähltem Mitarbeiter ausgedrückt; die Bedeutung dieses Maßes kann dem Text entnommen werden (aus: Brogden, 1949 S. 174).

Eine wichtige Implikation des erweiterten Modells besteht darin, daß ein PAP auch einen negativen Nutzen haben kann. Dies ist dann der Fall, wenn die Testkosten die positiven Erträge des Verfahrens übersteigen.

Wie aber werden die positiven Erträge des Auswahlverfahrens in Geldwerteinheiten übersetzt? Hierzu benötigt man Berufserfolgskriterien, die in Geldeinheiten übersetzbar sind. Brogden und Taylor (1950) beschäftigen sich ausführlich mit diesem Vorgang. Die Kriterien sollten den Beitrag des Individuums zu den Organisationszielen widerspiegeln. Nach der Identifikation dieser Kriterien kommt die Hauptschwierigkeit: Die Übersetzung in die Geldeinheit (z.B. Dollar, DM, ...). Der gesamte bewertete Berufserfolg kann dann additiv aus den Unterkriterien zusammengesetzt werden, da diese die gemeinsame Metrik (Dollar, ...) besitzen. Diese Vorgehensweise nennen Brogden und Taylor (1950) den 'Kostenrechnungsansatz der Kriteriumskonstruktion'.

Ist ein solches Kriterium einmal konstruiert, so läßt sich die Standardabweichung der Mitarbeiter einer Berufsgruppe bzgl. diesem ermitteln. Diese wird in Übereinstimmung mit der neueren NA-Literatur mit SD_y bezeichnet. Wie das Kriterium selbst ist natürlich auch seine Standardabweichung mit der Geldeinheit versehen. Inhaltlich bedeutet SD_y die Standardabweichung des Beitrages, den die Mitarbeiter zu den Organisationszielen erbringen. Unter Beachtung von

$$(2) \quad \bar{z}_x^* = r_{xy} \bar{z}_x^*$$

berechnet sich der Ertrag E , der sich durch Anwendung des Testverfahrens auf einen ausgewählten Mitarbeiter ergibt, zu

$$E = r_{xy} \bar{z}_x * SD_y. \quad (5)$$

Die Einheit von E ist [DM]. Wird in Gleichung (3) in (5) eingesetzt, so ergibt sich

$$E = r_{xy} SD_y \lambda_{(Q)}/Q \quad (6)$$

Um den Nutzen der Personalauswahlintervention pro ausgewähltem Individuum (U) zu erhalten, sind von der Einsparung die Kosten des Verfahrens abzuziehen.

$$U = r_{xy} SD_y \lambda_{(Q)}/Q - C_a/Q \quad (7)$$

C_a stellt dabei die Testkosten pro ausgewähltem Bewerber dar. Die Testkosten werden durch die Selektionsquote Q geteilt, damit auch die Testkosten der Abgelehnten ihre Berücksichtigung finden. Dieses Modell wurde von Cronbach und Gleser (1965) aufgegriffen und noch etwas verfeinert. Es stellt die Grundlage der meisten modernen Nutzenmodelle dar. Wir nennen es in Anlehnung an Boudreau (1991) Brogden-Cronbach-Gleser (BCG) Nutzenmodell.

3.3.3 Anmerkungen zum Modell

Bei dieser Art der Modellierung findet die Basisrate der geeigneten Bewerber, die bei Taylor und Russel (1939) eine wichtige Variable darstellt, keine Berücksichtigung mehr. Dies ist aber nur scheinbar der Fall. Taylor und Russel verwenden in ihrem Modell ein künstlich dichotomisiertes Kriterium. Im Modell von Brogden (1949) ist dies nicht mehr der Fall. Der durchschnittliche Kriteriumswert orientiert sich durch die z-Standardisierung an der Population der Bewerber. Der Populationsmittelwert \bar{z}_y nimmt dadurch den Wert 0 an. Die stetige und normalverteilte 'Basisrate' ist in diesem Konzept implizit enthalten.

3.4 Der entscheidungstheoretische Ansatz und das Nutzenmodell von Cronbach und Gleser

In diesem Abschnitt soll vergleichsweise ausführlich auf die Arbeit von Cronbach und Gleser (1965) eingegangen werden. Wesentliche Grundlagen zum Verständnis der Cronbach und Gleser'schen Denkweise wurden bereits durch die vergangenen Abschnitte vorbereitet.

3.4.1 Der entscheidungstheoretische Rahmen

Als im Abschnitt 3.1.3.3 die Kritik von Cronbach und Gleser am Konzept des Validitätskoeffizienten zur Bewertung eines Tests besprochen wurde, habe ich versucht deutlich zu machen, daß die Autoren eine neuartige Sichtweise der Diagnostik propagieren. Es kommt ihnen bei der Beurteilung von Tests nicht so sehr darauf an, wie fehlerfrei sie ein Kriterium vorhersagen können, sondern ob sie innerhalb ihres Anwendungskontextes in der Lage sind, richtige Entscheidungen zu indizieren. Diese Sichtweise zeigt der klassischen Diagnostik ihre Unzulänglichkeiten auf. Während sich diese schwerpunktmäßig mit der Schätzung 'wahrer Werte' und der dabei auftretenden Fehler beschäftigt, erkennen Cronbach und Gleser diese gutgemeinte, aber oft sinnlose Beschäftigung als Irrweg. Ich möchte kurz auf das Kraftfahrzeugmechanikerbeispiel aus dem zweiten Kapitel zurückkommen. Es kann uns hervorragend die Gegensätzlichkeit der Sichtweisen verdeutlichen. Der erste Mechaniker, der dem kontextfreien Diagnostizieren als Messen und Entscheiden 'verfallen ist', wird auf die berechnete diagnostische Fragestellung, ob die Zündkerzen schon abgenutzt sind, viel von seinem Wissen und seinen beruflichen Fertigkeiten zum Einsatz bringen. Er mißt Abnutzungsindikatoren, sammelt die Information, faßt sie zusammen und kommt schließlich zu einem Urteil. Auf dieser Grundlage entscheidet er sich für die geeignete Behandlung. Wir wollen unserem Mechaniker keine Fähigkeiten absprechen. Er hat sauber gearbeitet und versteht etwas von seinem Handwerk. Seine Messungen sind hochvalide und seine Entscheidungen wohlbegründet. Allerdings hat soviel Tun und Können auch seinen Preis. Er liegt möglicherweise höher, als die Behandlung (z.B. wechseln der Kerzen) selbst.

Der andere Mechaniker sieht Diagnostik als Teil der Intervention im System. Er beachtet alle Stufen der Intervention und kommt nach kurzem Überlegen zu dem Schluß, daß nur die Entscheidung zum Wechseln der Zündkerzen auf der Grundlage ihrer Betriebsdauer zu rechtfertigen ist, weil die Meßkosten für eine andere Vorgehensweise zu hoch liegen.

Aber hat er sich tatsächlich nur für das Wechseln der Kerzen entschieden? Nein. Die Entscheidung war vielmehr eine zwischen verschiedenen Interventionsprogrammen. Die Entscheidungsgrundlage war nicht nur die Meßqualität (z.B. Validität) oder andere einzelne Parameter, sondern ein Verbund von Parametern. Die Entscheidungsgrundlage war der Nutzen der Interventionsprogramme.

Um den Nutzen eines Interventionsprogrammes auszudrücken, benötigt man ein Wertesystem. Im Prozeßmodell der Personalauswahl (Abb. 2) wird dieser Sachverhalt dadurch verdeutlicht, daß auf Stufe 8 explizit die bewerteten Handlungsergebnisse aufgenommen sind. Des weiteren können die Träger des Wertesystems unterschiedliche Individuen bzw.

Institutionen sein. Auf Stufe 8 des Prozeßmodells wird diesem Umstand durch die Anmerkungen 'individuell, institutionell, durch andere Organisationen' Rechnung getragen. Wie bei der Unterscheidung zwischen individuellen vs. institutionellen Entscheidungen schon des öfteren deutlich wurde, spielt der Ort des Wertsystems eine bedeutende Rolle.

Cronbach und Gleser benutzen bei der Bewertung von Entscheidungen auch die Dollar-Skala. Sie wurde, wie beschrieben, schon bei Brogden (1946a, 1946b, 1949) propagiert. Die Dollar-Skala hat verschiedene Vorzüge. So wie die Kosten einer Maßnahme in Dollars angegeben werden, sollen auch die Ergebnisse (outcomes) nicht in standardisierten Maßzahlen, sondern durch Erträge (payoffs) in Dollar angegeben werden. Dadurch lassen sich nicht nur Aufwendungen und Ergebnisse miteinander vergleichen, welches das Hauptanliegen von Brogden (1949) war. Es sind auch Interventionsprogramme untereinander vergleichbar. Die Vergleichbarkeit entsteht durch die Verwendung derselben Nutzeinheit. Der zweite Mechaniker hat sich diesen Vorteil zunutze gemacht und sich für die richtige Intervention entschieden.

Auf diesen entscheidungstheoretischen Betrachtungen aufbauend entwickeln Cronbach und Gleser (1965) ihr Nutzenmodell und betonen dabei, daß dieses im wesentlichen der Aufgabe dienlich sei, über Interventionsprogramme (z.B. PAP) zu entscheiden. Bei der Personalauswahl gibt es natürlich nicht nur die Programmalternativen Zufallsauswahl oder PAP mit Eignungsdiagnostik. Auch unterschiedlich gestaltete PAP (z.B. PAP mit psychologischen Tests vs. Einstellungsinterview) können miteinander konkurrieren. Der Verantwortliche kann mit Hilfe der Nutzenanalyse rational zwischen diesen Programmen unterscheiden. Interessant ist auch hier die Parallele zur Sichtweise der Diagnostik als Intervention im System, wie sie in Kapitel 2 dargestellt wurde. Auch Cronbach und Gleser machen zwischen Personalauswahl mittels diagnostischen Instrumenten und einer Personalauswahl ohne solche Instrumente keinen qualitativen Unterschied. Man könnte dies überspitzt so formulieren: Auch keine Intervention ist Intervention. Ein Unternehmen, das keine Personalauswahl betreibt, 'behandelt' trotzdem sein Personal.

Das grundlegende Nutzenmodell wird im nächsten Abschnitt vorgestellt. Streng genommen gilt es nur für einen Spezialfall der Personalauswahl: Beim einstufigen Testen mit fester Behandlung. Dies bedeutet, daß nach dem Meßvorgang eine endgültige Entscheidung getroffen wird, und die Behandlung nicht adaptiv zu individuellen Meßergebnissen gestaltet wird. Eine klassische feste Behandlung wäre die Annahme oder Ablehnung von Bewerbern. Adaptive Behandlungen könnten z.B. darin bestehen, daß spezielle Trainingsmaßnahmen auf der Grundlage der Meßergebnisse stattfinden, wie dies oft nach AC der Fall ist. Diese

Erweiterungen behandeln Cronbach und Gleser im weiteren Verlauf ihres Buches. Sie wurden von der Fachwelt nicht in dem Maße rezipiert wie ihr spezielles Nutzenmodell und sollen auch hier nicht weiter beachtet werden.

3.4.2 Das Nutzenmodell

Die Herleitung des Modells folgt der Vorgehensweise von Brogden (1946, 1949). Sie ist in Abschnitt 3.3 beschrieben und soll hier nicht mehr wiederholt werden. Der einzige Unterschied des Modells besteht darin, daß Cronbach und Gleser den Nutzen nicht je Eingestelltem angeben, sondern je getestetem Bewerber. Hierzu unterscheiden sie zwischen der Anzahl der Bewerber (N_b) und der Anzahl der Angenommenen (N_a). Der Gesamtnutzen eines PAP bezeichne ich mit ΔU . In Anlehnung an Gleichung (5) gilt:

$$\Delta U = N_a SD_y r_{xy} \bar{z}_x^* - C_b N_b \quad (8)$$

C_b bezeichnet dabei die Testkosten pro getesteten Bewerber. Wird Gleichung (7) durch N_b dividiert, und \bar{z}_x^* durch $\lambda_{(Q)}/Q$ (3) ersetzt, so ergibt sich

$$\Delta U/N_b = N_a/N_b SD_y r_{xy} \lambda_{(Q)}/Q - C_b \quad (9)$$

$\lambda_{(Q)}$ läßt sich dadurch ermitteln, daß die Selektionsquote Q als Prozentrang aufgefaßt wird und mit Hilfe gängiger statistischer Tafeln ein z -Wert zugeordnet wird. Anschließend kann $\lambda(Q)$ als die Ordinatenhöhe der Normalverteilungsfunktion an der Stelle z berechnet oder aus Tabellen abgelesen werden. Zu heuristischen Zwecken bediene man sich wieder Abbildung 6. N_a/N_b entspricht der Selektionsquote. Damit vereinfacht sich Formel (9) zu

$$\Delta U/N_b = SD_y r_{xy} \lambda_{(Q)} - C_b \quad (10)$$

Dieses Modell gibt den relativen Nutzen eines PAP je Bewerber an. Multipliziert mit N_b erhält man das Modell, welches sich auf den relativen Nutzen der gesamten Personalauswahlmaßnahme bezieht.

$$\Delta U = N_b SD_y r_{xy} \lambda_{(Q)} - N_b C_b \quad (11)$$

Ich nenne dieses Nutzenmodell in Anlehnung an Boudreau (1991) das Brogden-Cronbach-Gleser (BCG) Nutzenmodell. Man kann es als das Grundmodell heutiger nutzeranalytischer Ansätze ansehen.

Ich habe oben vom *relativen Nutzen* gesprochen. Dieser Ausdruck ist erklärungsbedürftig. Es ist der Nutzen einer Maßnahme im Vergleich zu einer anderen Maßnahme. Der Vergleichsmaßstab des relativen Nutzens einer Interventionsmaßnahme hängt vom Validitätskoeffizienten r_{xy} ab. Wird dieser an einer un ausgelesenen Stichprobe ermittelt, so gibt das Modell den inkrementellen Nutzen gegenüber der Zufallsauswahl an.

Diese Denkweise verfolgt schon Brogden (1949). Benennt der Validitätskoeffizient aber die 'zusätzliche' Validität eines Auswahlverfahrens nach einer Vorauswahlmaßnahme, so gibt das Modell den inkrementellen Nutzen gegenüber der Vorauswahl an. Die Bewerberpopulation besteht, bei positiver Validität der Vorauswahl, aus geeigneteren Individuen, als die Bevölkerungspopulation. Der Validitätskoeffizient bezieht sich in diesem Fall auf die Beziehung zwischen dem Prädiktor und dem Kriterium innerhalb der bereits ausgelesenen Bewerberpopulation. Cronbach und Gleser nennen die Population, auf die sich die Nutzenfunktion bezieht, die 'a priori Population'. Wichtig an dieser Logik ist, daß bei einer empirischen Bestimmung der Validität der Prädiktoren innerhalb von PAP meist schon eine Vorauswahl stattgefunden hat und der berechnete Validitätskoeffizient die zusätzliche Validität angibt. Aus diesem Grunde ist es nicht verwunderlich, daß bei empirischen Validierungsstudien oft nur geringe Validitäten gefunden werden. Die Bewerber stellen eben nur selten eine unausgelesene Stichprobe dar, da allein schon eine Stellenausschreibung für die Vorselektion sorgt. Für Zahnarztstellen bewerben sich nun einmal meist nur Zahnärzte. Alle Bewerber haben demnach eine Eignung für den Arbeitsplatz, die weit über der einer unausgelesenen Stichprobe liegt. Die Stellenausschreibung allein stellt damit schon ein Teil der Intervention dar; sie ist eine valide Behandlung. Eine weitere Auslese, z.B. mittels einer Arbeitsprobe, wird deshalb - kaum überraschend - nur eine geringe zusätzliche Validität erbringen. Der relative Nutzen der diagnostischen Maßnahme mittels der Arbeitsprobe ist in diesem Fall der inkrementelle Nutzen gegenüber dem Nutzen, der durch die Stellenausschreibung allein geleistet wird.

Einer der Diskussionspunkte bei Cronbach und Gleser ist, wie schon bei Brogden (1949), die Wahl der optimalen Selektionsquote Q . Wie aus Gleichung (10) ableitbar ist, steigt mit abnehmender Selektionsquote der Ertrag, da $\lambda(Q)$ größer wird. Gleichzeitig steigt aber auch die Anzahl der Bewerber bei gleichbleibender Anzahl der Ausgewählten; mit anderen Worten nehmen die Auswahlkosten je eingestelltem Mitarbeiter zu. Die Erträge der Maßnahme und die Auswahlkosten sind also komplementär und können über die Selektionsquote zu einem Nutzenoptimum geführt werden.

Die optimale Selektionquote hängt nach Cronbach und Gleser von den sog. relativen Kosten $C_b/SD_y r_{xy}$ ab. Sind die relativen Kosten niedrig, wie z.B. bei einfachen Gruppentestverfahren, so lohnt sich eine Senkung der Selektionsquote. Sind die relativen Kosten aber hoch, so nähert sich die optimale Selektionsquote der 50%-Marke an. Dies wird zum Beispiel bei kostenaufwendigen AC-Verfahren der Fall sein. Diesen Zusammenhang fand schon Brogden (1949). Er wurde durch Abbildung 7 veranschaulicht.

3.4.3 Annahmen des BCG-Personalauswahlnutzenmodells

Das BCG-Nutzenmodell macht verschiedene Annahmen. Diese sollen hier noch einmal zusammenfassend veranschaulicht werden. Die Darstellung orientiert sich teilweise an Boudreau (1991) und an Cascio (1980). Viele der Punkte sind einfach in das Prozeßmodell der Personalauswahl einzuordnen. Aus Gründen der einfachen Darstellung soll dies jedoch hier nicht geschehen.

- Das BCG-Nutzenmodell beschäftigt sich mit dem Nutzen von Personalauswahlprogrammen. Nutzen wird hier verstanden als die Gesamtheit aller monetären Erträge der Interventionsmaßnahme abzüglich der Kosten der Maßnahme. Damit wird deutlich, daß der Beurteilung der Interventionsmaßnahme das Wertsystem der Organisation zugrundeliegt.
- Die Testwerte sind an der Bewerberpopulation mit Mittelwert 0 und der Standardabweichung 1 standardisiert.
- Die Entscheidung über die Behandlung wird durch Rangreihenbildung der Bewerber getroffen. Von dieser Reihe wird der Beste akzeptiert, danach der Zweitbeste, usw., bis zum Erreichen des Guttoff-Punktes. Damit wird deutlich, daß die endgültige Informationsrepräsentation univariat sein und mindestens Rangskalenniveau haben muß.
- Innerhalb des Personalauswahlprogrammes sind die Entscheidungen, die über ein Individuum getroffen werden, dichotom. Es handelt sich entweder um die Behandlung 'Annahme' oder 'Ablehnung' des Bewerbers.
- Die Regression der bewerteten Handlungsergebnisse eines akzeptierten Bewerbers für die Organisation auf den Prädiktorwert ist linear.
- Die Ertrag für die Organisation bei abgelehnten Bewerbern ist 0 und hat keinen Zusammenhang zum Prädiktor.
- Die durchschnittlichen Kosten der Anwendung des Auswahlverfahrens je Bewerber ist C_b .

An dieser Stelle wird deutlich, daß das BCG-Nutzenmodell nur für ganz bestimmte PAP Gültigkeit besitzt. Es werden wohl nur wenige PAP so konzipiert sein, daß sie nur die genannte dichotome Behandlung vorsehen. Da über neueingestellte Mitarbeiter aufgrund der diagnostischen Maßnahmen vielfältige Bildungsbedarfsinformationen vorliegen, bietet es sich geradezu an, diese auch zu nutzen.

Darüber hinaus ist im Modell nichts darüber gesagt, wie lange ein akzeptierter Bewerber in der Organisation verbleibt. Nun ist es gemäß der Logik des Modells aber nahezu zwingend, diese Variable zu beachten. Ein Mitarbeiter, der nur vorübergehend in der Organisation verbleibt, hat nun einmal zu wenig Zeit, 'seine Personalauswahlkosten' zu amortisieren.

Die Annahme, daß der Ertrag bei abgelehnten Bewerbern für die Organisation Null sei, wird in manchen Fällen nicht haltbar sein. Eine Fehlentscheidung, bei welcher ein sehr gut geeigneter Bewerber abgelehnt wird, kann durchaus einen negativen Ertrag für die Organisation nach sich ziehen. Man denke nur an konkurrierende Unternehmen. Ein Unternehmen A, welches an sich geeignete Bewerber ablehnt, muß damit rechnen, daß der Bewerber in eine Konkurrenzunternehmung eintritt. Ab diesem Punkt sind für das Unternehmen A nicht nur die Auswahlkosten des Bewerbers wirksam, sondern auch dessen Beitrag zum Ertrag der Konkurrenz.

Diese kritischen Anmerkungen erheben keinen Anspruch auf Vollständigkeit. Sie sollen lediglich demonstrieren, daß das BCG-Modell die komplexe Realität natürlich alles andere als perfekt abzubilden vermag. Um die Realitätsanpassung besser zu leisten, wurden eine Reihe von Erweiterungen des Modells vorgeschlagen. Den Erweiterungen des BCG-Nutzenmodells ist ein eigenes Kapitel gewidmet. In Kapitel 5 werde ich sowohl publizierte Neuerungen vorstellen, als auch eigene Vorschläge in diese Richtung bringen. Bevor es jedoch soweit ist, möchte ich zunächst einen großen Vorteil des ursprünglichen BCG-Modells ausnutzen. Es ist seine Einfachheit, die es erlaubt, exemplarisch die Schwierigkeiten zu besprechen, die bei der Operationalisierung seiner Parameter entstehen. Wie sich über die Jahre der Forschungsbemühungen gezeigt hat, besteht die Hauptschwierigkeit in der Schätzung der Standardabweichung des Kriteriums (SD_y). Nahezu die Hälfte der Literatur zur Nutzenanalyse beschäftigt sich mit diesem Problem ausführlich. Ich werde dem durch eine umfassende Besprechung des Problems Rechnung tragen. Der Aufwand ist nicht zuletzt auch deswegen zu rechtfertigen, da eine derartig weitreichende Besprechung des Problems meines Wissens bisher nicht veröffentlicht wurde.

4 Parameteroperationalisierung im BCG-Nutzermodell

Dieses Kapitel beschäftigt sich mit der Operationalisierung der Parameter des BCG-Nutzenmodells. Die Grundstruktur des BCG-Nutzenmodells soll zunächst zur Verdeutlichung in anderer Form dargestellt werden. Das Modell ist nach folgendem Muster aufgebaut (Cronshaw & Alexander, 1985):

$$\Delta U = E - C \quad (12)$$

Dies bedeutet, daß sich der Nutzen einer Interventionsmaßnahme dadurch berechnet, daß von dem Ertrag der Maßnahme (E) ihre Kosten (C) abgezogen werden. Alle Variablen der Gleichung (12) sind mit der Einheit DM versehen.

Gleichung (11) präziserte diesen Zusammenhang durch die nähere Spezifikation sowohl des Ertrags als auch der Kosten.

$$(11) \quad \Delta U = N_b SD_y r_{xy} \lambda_{(Q)} - N_b C_b$$

AU ist der relative Nutzen der gesamten Personalauswahlmaßnahme. Rechts vom Gleichheitszeichen steht die subtraktive Verknüpfung der Elemente E und C. Beide Elemente sind mit der Einheit Dollar bzw. DM versehen. Die Erträge entstehen durch die bewerteten Handlungsergebnisse aller eingestellten Mitarbeiter für die Organisation. Es sind die Nettobeiträge der Mitarbeiter zum Organisationsziel. Hiervon werden die Kosten für die Personalauswahlmaßnahme abgezogen. Die Parameter bedeuten im einzelnen:

- N_b Anzahl der Bewerber
- $\lambda_{(Q)}$ Ordinatenabschnitt der Normalverteilung als Funktion der Selektionsquote
- C_b durchschnittliche Kosten der Auswahlmaßnahmen je Bewerber in DM
- r_{xy} Validität des Testverfahrens (Prädiktor-Kriteriums-Korrelation)
- SD_y Standardabweichung des Kriteriums y in der Bewerberpopulation in DM

Der Kostenterm ist einfach und bedarf keiner erneuten Erklärung. Im Ertragsterm ist SD_y Träger der Skalierungseinheit DM. Der Ausdruck $\lambda_{(Q)}$ entstand im Zusammenhang mit der Schätzung des durchschnittlichen standardisierten Prädiktorwertes Z der ausgewählten Bewerbergruppe. Dieser Wert ist eine Funktion der Selektionsquote. Auf der Grundlage der Annahmen von Brogden (1946a) kann der durchschnittliche standardisierte Kriteriumswert der Ausgewählten durch (2) $\bar{z}_y^* = r_{xy} \bar{z}_x^*$ geschätzt werden. Ziel der Nutzerfunktion ist jedoch nicht, standardisierte Kriteriumswerte zu bekommen, sondern leicht interpretierbare Werte, die mit der Einheit DM versehen sind. Aus diesem Grunde wird \bar{z}_x^* mit dem

Skalierungsfaktor SD_y multipliziert. Man erhält nach der Aufnahme der Bewerberzahlen Gleichung (11). Es gilt die Parameter, welche sie enthält, zu schätzen.

4.1 Die Selektionsquote und Lambda

Die Ermittlung der Selektionsquote Q bereitet die wenigsten Schwierigkeiten. Man erhält sie in der Praxis durch Quotientenbildung zwischen der Anzahl der angenommenen Bewerber und der Gesamtzahl der Bewerber $Q = N_a / N_b$. Auf dieser Grundlage kann k berechnet werden oder aus der Tabelle von Naylor und Shine (1965) abgelesen werden. Die Vorgehensweise wurde in Abschnitt 3.4.2 beschrieben. Wichtig bei Nutzung der Aussage von Gleichung (3)

$$(3) \quad \bar{z}_x^* = \lambda_{(Q)}/Q$$

ist die Beachtung der Grundannahmen für deren Gültigkeit. Es müssen die Bewerber auf der Grundlage der Prädiktorwerte von oben nach unten selektiert werden und die Prädiktorwerte müssen normalverteilt sein. Insbesondere die erste Forderung dürfte in der Praxis oft verletzt sein. Durch verschiedene situative Zwänge werden die Bewerber nicht immer nach der strikten Regel 'von oben nach unten' ausgewählt (z.B. durch die bevorzugte Berücksichtigung von Kindern langjähriger Betriebsangehöriger, Härtefällen, Bewerbern mit 'guten Beziehungen' zum Unternehmen, etc.). Dadurch vermindert sich das tatsächliche \bar{z}_x^* gegenüber dem Wert, der aufgrund der Selektionsquote theoretisch zu erwarten ist. Aufgrund dieser und ähnlicher Schwierigkeiten sollte, wenn möglich, \bar{z}_x^* empirisch ermittelt werden, anstatt den Parameter mit Hilfe der Gleichung (3) zu schätzen.

Murphy (1986) behandelt das Problem, daß durch die Ablehnung von Stellenangeboten (z.B. aufgrund attraktiver Konkurrenzangebote) die Auswahl der Bewerber auf der Grundlage der Prädiktoren von oben nach unten gestört wird. Die Berechnung von \bar{z}_x^* durch Gleichung (3) führt in diesem Fall zu einer Überschätzung des Parameters. Murphy gibt eine Korrekturformel an, um auch bei abgelehnten Stellenangeboten den Parameter schätzen zu können.

4.2 Die Validität

Das altbekannte Konzept des Validitätskoeffizienten findet auch beim BCG-Nutzenmodell Verwendung. Aus diesem Grunde muß der Validitätskoeffizient entweder durch eine empirische Validierungsstudie ermittelt werden oder durch andere Verfahren geschätzt

werden. Die Validierungsstudie wurde ausführlich in Abschnitt 2.3.2 besprochen. Ein Verfahren zur Schätzung der Validität, die Metaanalyse, wurde in Abschnitt 2.3.3 vorgestellt. In diesem Abschnitt soll auf bereits Gesagtes nicht mehr eingegangen werden. Vielmehr soll hier der Versuch unternommen werden, eine Besonderheit im Umgang mit Validitätskoeffizienten zu diskutieren. Einige Autoren unterscheiden zwischen dem empirisch ermittelten Validitätskoeffizienten und der "wahren" Validität. Letztere ist ein theoretischer Populationsparameter, für den der empirische Validitätskoeffizient kein erwartungstreuer Schätzwert darstellt. Die empirische Validität wird vielmehr systematisch gemindert, so daß Validierungsstudien Gefahr laufen, die wahre Validität zu unterschätzen. Für die systematische Minderung sind zwei Ursachen verantwortlich: Die Varianzeinschränkung des Prädiktors oder des Kriteriums und die nicht perfekt reliable Messung des Kriteriums. Es gibt Vorschläge, wie die auf varianzeingeschränkten Stichproben beruhenden Validitätskoeffizienten korrigierbar sind (z.B. Hunter & Schmidt, 1990). Diese Korrekturen basieren allerdings auf strengen und schwer prüfbar Annahmen, so daß ich ihre Praktikabilität in diesem Zusammenhang für fraglich erachte.

Die Annahmen über die Reliabilität der Kriteriumsmessung hingegen sind meist wesentlich leichter zu verteidigen. Auch hier- wurden Verfahren entwickelt, die wahre Validität aus dem empirischen Koeffizienten und dem Kriteriumsreliabilitätskoeffizienten zu schätzen. Sie sind unter dem Namen Attenuationskorrektur bzw. Minderungskorrektur bekannt und finden sich z.B. bei Lienert (1989).

Inwieweit die Korrektur des empirischen Validitätskoeffizienten sinnvoll ist, ist nicht leicht und schon gar nicht allgemeingültig zu beurteilen. Grundsätzlich hängt es von der Entscheidungssituation ab, welches Risiko minimiert werden sollte: Die Validität (und damit den Nutzen) fälschlicherweise zu unterschätzen, oder ihre irrtümliche Überschätzung. Soll letzteres vermieden werden, so empfiehlt es sich, keine Korrekturen des Validitätskoeffizienten vorzunehmen, da solche immer zur Vergrößerung der Maßzahl beitragen.

Wenn die Messungen des Prädiktors oder des Kriteriums mit einer niedrigen Reliabilität erfolgen, so besteht die Gefahr, daß in der Stichprobe 'Ausreißerwerte' vorkommen. In diesem Fall kann bei einer Validitätsermittlung aufgrund der niedrigen Reliabilität des Kriteriums die wahre Validität auch überschätzt sein. Wird danach noch eine Minderungskorrektur wegen der niedrigen Reliabilität des Kriteriums vorgenommen, so erhöht sich die geschätzte Validität weiter. Dieses Beispiel soll zeigen, daß jeder Einzelfall sorgsam zu prüfen ist, ob derartige Korrekturen mit dem Ziel einer genaueren

Validitätsschätzung vorzunehmen sind. Bei großen Stichproben, bei welchen die Annahmen der Korrekturformeln plausibel vertretbar sind, kann eine Korrektur sinnvoll sein. Eine routinemäßige Anwendung der Verfahren resultiert sicherlich darin, daß die durchschnittlichen Validitätskoeffizienten eine systematische Tendenz nach oben erhalten. Burke und Frederick (1986) erhalten bei der Validierung eines AC zunächst einen Validitätskoeffizienten von $r_{xy}=0,41$. Nach den Korrekturen schätzen sie die wahre Validität auf $\rho_{xy}=0,59$. Bei allem Optimismus würde ich diesen Wert als hoffnungslos zu hoch beurteilen. Dies ist insbesondere dann nachvollziehbar, wenn berücksichtigt wird, daß die Bewerberpopulation in dieser Studie keine Zufallsstichprobe aus der Bevölkerung darstellt, sondern bereits stark vorselektiert ist (Selektion durch das Reagieren auf das Stellenangebot, Vorauswahl auf der Grundlage der Bewerbungsunterlagen, ...). Eine Validierungsstudie ergibt dann natürlich einen niedrigeren Validitätskoeffizienten, als dies bei einer Zufallsstichprobe aus der Bevölkerung der Fall wäre; er gibt die inkrementelle Validität des AC gegenüber allen anderen Maßnahmen der Personalauswahl an, die zur Stichprobenselektion beigetragen haben. Es wäre absurd, diese empirisch ermittelte Validität so zu korrigieren, daß sie sich auf eine nicht selektierte (d.h. nicht varianzeingeschränkte) Stichprobe bezieht. Bei der NA interessiert gerade die inkrementelle Validität des Verfahrens, mit dem auch die (inkrementellen) Kosten verbunden sind.

Es wird deutlich, daß auch der Validitätskoeffizient vor dem Hintergrund der spezifischen Situation interpretiert werden muß. Nahezu jedes berufseignungsdiagnostische Verfahren kann eine nahezu perfekte Validität erreichen, indem in die Validierungsstichprobe ausschließlich kochqualifizierte Bewerber und Schwachsinnige aufgenommen werden. Ob allerdings der Sinn einer solchen Studie darin besteht, möglichst hohe Koeffizienten zu erzielen oder vernünftige Entscheidungen zu treffen, ist eine andere Frage.

4.3 Die Kosten

Das Kostenkonzept nach Cronbach und Gleser (1965) ist in der gegebenen Form leicht nachvollziehbar. Allerdings dürfte es für praktische Fälle leider kaum angemessen sein. Dies bemerkte bereits Roche (1961, 1965), in dessen Dissertationsschrift von der ersten empirischen Anwendung des Cronbach-Gleser'schen Ansatzes berichtet wird.

Hilfe erhalten wir in diesem Punkt von den wirtschaftswissenschaftlichen Nachbardisziplinen. In der betriebswirtschaftlichen Kostenrechnung werden feste und variable Kosten getrennt betrachtet. Die festen Kosten eines Personalauswahlprogrammes

bestehen z.B. aus den anteiligen Gemeinkosten der Personalabteilung (Verwaltung, Miete, Strom, ...), den Personalkosten, welche durch das PAP anfallen, den Entwicklungs- und Evaluationskosten des PAP, etc. Alle Kosten, welche in direktem Zusammenhang mit der Häufigkeit der Durchführung des Programms bzw. mit der Anzahl der Bewerber stehen, sind variable Kosten. Darunter fallen Kosten für Testmaterial, Lizenzgebühren, Unterkunft und Verpflegung der Bewerber (z.B. bei AC), Fahrtkostenerstattungen, etc. Werden die Kosten eines PAP derartig aufgeschlüsselt, so ist die Realität dadurch leichter faßbar. Die Aufnahme des Konzeptes der festen und variablen Kosten in das Nutzenmodell trägt zur erhöhten Transparenz des Modells bei. Die Entscheidungen, welche auf dem Modell basieren sind damit leichter nachvollziehbar und kritisierbar.

Wenn die getrennte Aufnahme der beiden Kostenkonzepte ohne große Veränderung des BCG-Modells geleistet werden soll, dann ist ein kleiner gedanklicher Umweg hilfreich. Unter der Annahme, daß bei der Kostenberechnung der Zeitraum und die Häufigkeit der Anwendung des. PAP feststeht, können die Gesamtkosten geschätzt werden. C_b kann berechnet werden, indem diese Gesamtkosten durch die Anzahl der Bewerber geteilt werden.

Die elegantere Möglichkeit besteht allerdings in der getrennten Aufnahme der Konzepte 'variable Kosten' und 'feste Kosten'. Diese ökonomischen Konzepte spielen bei Boudreau (1983a) eine zentrale Rolle. Cronshaw und Alexander (1985) behalten den Gedanken der getrennten Kostenkonzepte bei. Sie betrachten alle Maßnahmen, welche auf die Veränderung der personellen Ressourcen eines Unternehmens abzielen, aus der Perspektive von Investitionsentscheidungen. Die ökonomischen Beiträge zur Nutzenanalyse von Boudreau (1983a, 1983b) und Cronshaw und Alexander (1985) werden in Kapitel 5 ausführlich vorgestellt werden. An dieser Stelle möchte ich nur grob auf Operationalisierungsprobleme der Kostenparameter eingehen.

4.3.1 Feste Kosten und variable Kosten

Wird das Kostenkonzept von Cronbach und Gleser in die Komponenten feste und variable Kosten aufgeteilt, dann verändert sich die Struktur des Grundnutzenmodells (Gleichung 12). Das neue Modell läßt sich so darstellen:

$$\Delta U = T E - C_1 - C_0 \quad (13)$$

In diesem Modell steht E wieder für den Ertrag, der durch die Interventionsmaßnahme entsteht, C_1 bezeichnen die variablen Kosten der Maßnahme je Behandeltem, und C_0 steht für die festen Kosten der Maßnahme. Diese sind von der Anzahl der Behandelten unabhängig.

Eine weitere Variable muß nach dieser Differenzierung in das Modell aufgenommen werden. Die Variable T bezeichnet die Wirkdauer der Interventionsmaßnahme (tenure). Diese zu beachten ist deshalb wichtig, weil die Kosten der Maßnahme von der Wirkdauer unabhängig sind, jedoch nicht die Höhe des Ertrages. Innerhalb der Personalauswahl kann die Wirkdauer T als der durchschnittliche Zeitraum angenommen werden, den ein eingestellter Mitarbeiter beim Unternehmen verbringt. Ein Unternehmen, welches ein PAP anwendet und dadurch qualifizierte Mitarbeiter gewinnt, die durchschnittlich 10 Jahre beim Unternehmen verbleiben, hat trotz gleicher Kosten mehr Nutzen von der Maßnahme, als ein anderes Unternehmen mit dem identischen PAP, aber hohen Fluktuationskennwerten.

Nach diesen theoretischen Überlegungen soll nun auf die Schätzungen der Parameter C_1 und C_0 im Rahmen von PAP eingegangen werden. Die variablen Kosten C_1 beinhalten alle die Kosten in DM, die direkt proportional zu der Anzahl der Bewerber bzw. der Anzahl der Eingestellten sind. Prinzipiell ist es egal, ob die variablen Kosten auf die Bewerberzahl oder die Anzahl der akzeptierten Bewerber bezogen werden. Die folgende Liste gibt einen Eindruck über die möglichen variablen Kostenquellen im Rahmen von PAP. Sie erhebt allerdings keinen Anspruch auf eine erschöpfende Darstellung.

Personalkosten für Vorbereitung, Betreuung, Interviews, Auswertung, Rückmeldung, etc.

Materialkosten für Fragebögen, Testmaterial, etc.

Lizenzgebühren für Tests, insbesondere bei computergestützten Verfahren

Nebenkosten für Unterbringung, Verpflegung, Räumlichkeiten, ...

Alle die Kosten, die nicht bei den variablen Kosten berücksichtigt worden sind, aber durch das PAP verursacht werden, werden feste Kosten oder Fixkosten genannt. Diese sind oft erheblich schwieriger zu schätzen. Wenn beispielsweise ein Personalauswahlverfahren in einer Organisation intern entwickelt wird, dann sind die Entwicklungskosten nur grob schätzbar. Eine Möglichkeit der Schätzung besteht darin, diejenigen Kosten als Schätzgröße zu übernehmen, die entstanden wären, wenn eine organisationsexterne Institution die Entwicklungsaufgabe übernommen hätte. Aber auch in diesem Fall sind die Personalkosten für die Organisation schwer kalkulierbar, die z.B. durch die vorbereitende Planung und Koordination der externen Kräfte entstehen.

Noch schwieriger ist die Schätzung der festen Personal- und Materialkosten, die der Personalabteilung anteilig durch das PAP entstehen. Dies wird insbesondere dann deutlich, wenn man sich noch einmal vor Augen führt, daß der Prozeß der Personalauswahl sich nicht in der Testdurchführung und Entscheidung erschöpft. Wie Abbildung 2 verdeutlicht, beginnt

das Personalauswahlprogramm schon bei der Beobachtung des Arbeitsmarktes und setzt sich in dessen aktiver Beeinflussung fort (z.B. durch Werbung, Praktikantenausbildung, Abstellen von Lehrbeauftragten, Veranstaltungen an Berufsschulen und Hochschulen, ...). Nach der Handlungsapplikation sollten umfangreiche Evaluationsmaßnahmen die Qualität des PAP sichern und weiterentwickeln. Die hierzu notwendigen finanziellen Aufwendungen können nur grob abgeschätzt werden. Um mit diesem Problem zweckmäßig umgehen zu können, läßt Boudreau (1983a) diese Art der Fixkosten außerhalb der Betrachtung. Er begründet diese Haltung damit, daß diese Fixkosten nicht mit der Art des Interventionsprogramms variieren. Mit anderen Worten heißt das, daß diese Kosten unabhängig von den konkreten Personalauswahlmaßnahmen ohnehin aufgebracht werden müßten. Im Gegensatz zu vielen anderen Autoren spricht Boudreau (1983a) dieses schwierige Problem offen an und schlägt eine Lösungsmöglichkeit vor. Ob seine Annahmen sinnvoll sind, kann sicherlich in vielen Fällen stark bezweifelt werden. Wenn die empirischen Arbeiten zur Nutzenanalyse einmal genauer bzgl. dieses Punktes untersucht werden, fällt auf, daß nahezu alle Autoren die anteiligen Personal- und Materialkosten des 'Überbaus' des PAP nicht in ihre Kostenüberlegungen aufnehmen. Diese Vorgehensweise bringt natürlich Schönwetterergebnisse. Schlimmer jedoch als die Nichtverrechnung ist die Tatsache, daß viele Autoren diesen wunden Punkt diskussionslos übergehen.

Außer den besprochenen anteiligen Personal- und Materialkosten der Personalabteilung, die im Zusammenhang mit der Personalauswahl stehen, können die festen Personal- und Materialkosten, die bei der Entwicklung, Evaluation und Verbesserung des Auswahlverfahrens anfallen, oft genauer quantifiziert werden. Die Qualität von Nutzenanalysen muß auch daran bemessen werden, inwieweit die Kosten der Interventionsmaßnahmen vollständig berücksichtigt wurden.

4.4 Der Skalierungsfaktor SD_y als Achillesferse

Schon Cronbach und Gleser (1965, S. 121) erkennen die Schätzung der Standardabweichung der in Geldeinheiten bewerteten Leistung als die Achillesferse der NA. Diese Vorhersage hat sich leider bewährt. Nachdem es um die NA nach der Monographie von Cronbach und Gleser vergleichsweise ruhig geworden war, führen Schmidt, Hunter, McKenzie und Muldrow (1979) dies wesentlich darauf zurück, daß Informationen für die Schätzung des Parameters SD_y nur schwierig zu bekommen sind. Roche (1965, basierend auf seiner Dissertation 1961) schätzte den Leistungsstandardabweichungsparameter SD_y sehr aufwendig mit Hilfe der

innerbetrieblichen Kostenrechnung. Nicht nur dieser Aufwand, sondern auch sein Bekenntnis, daß trotz seiner Bemühungen viele Schätzungen und willkürliche Wertzuweisungen in die Kostenrechnung eingingen, haben potentielle Anwender der NA möglicherweise abgeschreckt. Obwohl Roche (1961, 1965) große Anstrengungen unternahm, die nötigen Informationen von Kostenrechnungsspezialisten zu erhalten, bemerkt Cronbach (1965, S. 266) in einem Kommentar zur Arbeit von Roche: "... make it clear that the dollar payoff from an employee is an elusive concept, and that our theory is monstrously oversimplified."

Der Aufsatz von Schmidt et al. (1979) war es schließlich, der die Nutzenanalyседiskussion wieder in Gang brachte. Diese Wirkung konnte er durch den Vorschlag einer vergleichsweise einfachen Schätzmethode des Parameters SD_y erzielen. Obwohl ich der Methode von Schmidt et al. (1979) sehr kritisch gegenüberstehe, kann den Autoren sicherlich der Verdienst zugesprochen werden, wieder das Forschungsinteresse an der NA erweckt zu haben. Gleichzeitig aber haben Schmidt et al., wie noch gezeigt werden wird, auch den Weg für viele Mißverständnisse geebnet.

Kennzeichnend für das Forschungsfeld der NA in den achtziger Jahren und Anfang der neunziger Jahre war eine ausgiebige Diskussion über geeignete Methoden, SD_y zu schätzen. Selbst Arbeiten welche sich um die praktische Anwendung der NA im Feld bemühten, kamen an dieser Diskussion nicht vorbei. Dies ist insofern nachvollziehbar, als daß eine stärkere Änderung der SD_y -Größe entscheidend dafür sein kann, ob ein Interventionsprogramm als durchführungswürdig angesehen wird oder nicht. Die anderen Parameter der Nutzenmodelle sind dagegen vergleichsweise leicht zu schätzen.

Beim Bearbeiten der Literaturfülle zu diesem Thema können zwei Hauptprobleme ausfindig gemacht werden. Das erste davon ist theoretischer Natur, das zweite betrifft die praktischen Parameterschätzmethode.

Die theoretischen Probleme betreffen die Definition des bewerteten Arbeitsergebnisses y . So einfach die Formulierungen 'bewertetes Arbeitsergebnis' bzw. 'Bewertung der Handlungsergebnisse durch die Institution' (Abb. 2) klingen, so verschieden wird die Variable y in der Literatur verstanden. Was mich dabei besonders überraschte, war die Einsicht, daß dieses Problem dabei nur selten und wenn überhaupt, dann nicht ausdrücklich genug diskutiert wird. Diese theoretischen Aspekte werden in diesem Abschnitt besprochen.

Das zweite Problem besteht in der Vielfalt der Parameterschätzmethode. Anstatt die Anstrengung zu unternehmen, die vorhandenen, meist subjektiven Urteilsfindungsmethode genauer zu untersuchen (wie dies von Psychologen auch zu erwarten wäre), werden immer neue Schätzmethode entwickelt. Die gängige Art, eine NA im Feld für praktische Zwecke

durchzuführen, besteht darin, mehrere Methoden zur Schätzung von SD_y heranzuziehen. Dies führt aber unweigerlich zu verschiedenen Ergebnissen. Damit wird der Vorteil der Evaluation durch die NA aufgehoben, daß ihre Ergebnisse transparent und leicht kommunizierbar sind. Auf die Probleme der Parameterschätzmethoden gehe ich in Abschnitt 4.5 ein.

4.4.1 Definitionsprobleme der Leistung y

Die (nicht bewerteten) Ergebnisse einer Behandlung definieren Cronbach und Gleser (1965, S. 22) allgemein als "... all the consequences of a given decision that concern the person making the decision (or the Institution he represents)." Bis zu diesem Punkt herrscht auch innerhalb der neueren Literatur noch Einigkeit. Das Prozeßmodell der Personalauswahl (Abb. 2, Stufe 7, 8) verdeutlicht die Unterscheidung zwischen den Handlungsergebnissen, die einer Behandlung folgen, und der Bewertung dieser Ergebnisse. Innerhalb des BCG-Nutzenmodells erfolgt diese Bewertung in Geldeinheiten. Der Einfachheit halber nenne ich sie die 'Leistung y '. Auch sie trägt die Einheit DM.

Bei den bewerteten Handlungsergebnissen, also der Leistung y und ihrer Standardabweichung SD_y , verliert sich die Einstimmigkeit der jüngeren Autoren. Nachfolgend findet sich eine Zusammenstellung der Auffassungen verschiedener Autoren von y bzw. SD_y . Zunächst ist die Unterschiedlichkeit der Auffassungen vom Begriff der Leistung y bzw. dessen Standardabweichung SD_y augenfällig. Bei der genaueren Analyse lassen sich zwei grundlegende Richtungen erkennen. Die Protagonisten der ersten Richtung sind um die Autoren Schmidt und Hunter versammelt. Sie definieren y ziemlich vage als den Verkaufswert der Leistung, die ein Mitarbeiter erbringt. Eine häufig verwendete Hilfe zur Schätzung dieses Verkaufswerts ist der Vergleich mit den Kosten, den eine externe Organisation für die gleiche Leistung verursachen würde, indem sie beauftragt würde, die Aufgaben des Mitarbeiters zu erfüllen. Im Rahmen dieser Sichtweise sind die Lohnkosten des Mitarbeiters ein Indiz für seine Leistung. Die Anhänger dieser Denkweise definieren y als

- "the yearly value to your agency ... consider the quality and quantity of output ... consider what the cost would be of having an outside firm provide these products and services" (Schmidt et al., 1979, S. 621; sinngemäß: Bobko, Karren & Parkington, 1983, S. 172; Burke & Frederick, 1984, S. 484; Burke & Frederick, 1984, S. 336; DeSimone, Alexander & Cronshaw, 1986, S. 95)
- "dollar value as a increase in output ... dollar value as sold" (Hunter, Schmidt & Coggin, 1988, S. 526)

- "Overall Index of an employee's value to the hiring Organisation" (Brogden & Taylor, 1950, S. 134)
- " SD_y = the yearly Standard deviation of output in dollars" (Schmidt & Hunter, 1983, S. 413).

Diese Sichtweise wird von anderen Autoren bestritten. Schon die frühen Nutzenanalytiker hatten die Vorstellung, daß y der Beitrag eines Mitarbeiters zu den Organisationszielen in Geldeinheiten darstellt. Hierbei ist es egal, ob das Organisationsziel im Erreichen von Gewinnen liegt (profit organizations), oder nicht (non profit organizations). Auch bei letzteren läßt sich der Beitrag der Mitarbeiter in DM ausdrücken. Dies ist z.B. der Fall, wenn ein besonders qualifizierter und motivierter Mitarbeiter Lohnkosteneinsparungen ermöglicht, da er mit seiner Leistung dazu beiträgt, daß Neueinstellungen unnötig sind. Diese Sichtweise trennt die Konzepte von festen und variablen Kosten der Leistung. Der Wert eines Mitarbeiters liegt darin, daß er zu den festen Kosten und dem Gewinn der Organisation seine Beiträge leistet. Insbesondere die Arbeiten von Boudreau (1983a, 1983b) waren für die Sichtweise *Leistung als Beitrag zur Organisation* richtungsweisend. Einige Definitionsversuche von y sind

- "The criterion should measure the contribution of the individual to the Overall efficiency of the organisation." (Brogden & Taylor, 1950, S. 139; Roche, 1965, S.255)
- "... the dollar profit which accrues to the company as a result of an individual's work ..." (Roche, 1965, S.257)
- "... distinguishing the sales value and cost of productivity gains due to treatments requires redefining y ..." (Boudreau, 1983a, S. 554)
- "...the cost to the company of having his or her Services and the dollar benefits of the Output produced. The difference between benefits and costs would yield a net dollar benefit (or loss) to the firm." (Cascio & Ramos, 1986, S.20)
- "... the total amount (in dollars and Cents) contributed toward the coverage of fixed costs, and then profit ..." (Greer & Cascio, 1987, S. 590).

Gemäß meiner Klassifikation in die Begriffsdefinitionen von y als 'Verkaufswert' und 'Beitragswert' können die Äußerungen von Brogden und Taylor (1950) einmal der einen, dann wieder der anderen Kategorie zugeordnet werden. Die unklaren Begriffsdefinitionen der frühen Nutzenanalytiker haben sicherlich erheblich dazu beigetragen, daß bis heute der akademische Disput zwischen den beiden Richtungen anhält. Will man einen Eindruck von der erfrischenden Schärfe gewinnen, mit der dieser geführt wird, so sei der Aufsatz von

Hunter et al. (1988) empfohlen. Unter anderem schreiben die Autoren (S. 526, Hervorhebung durch den Verf.): "... Brogden (1946¹⁰), 1949) and Brogden and Taylor (1950) conceptualized utility as the dollar value of the increase in Output resulting from a personnel program, that is, the increase in dollar value as sold." Boudreau (1991, S. 646, Hervorhebung durch den Verf.) behauptet dagegen: "This view¹¹ is difficult to reconcile with the general payoff definition originally proposed by Brogden and Taylor (1950) and Cronbach and Gleser (1965), both of whom included the notion of revenue minus costs Beide Autoren beanspruchen also die "Vaterschaft" des BCG-Ansatzes für ihre Denkweise. Nach langer Überlegung komme ich nicht umhin zuzugeben, daß die frühen Autoren auch in beide Richtungen interpretiert werden können.

Wichtig für das Verständnis der folgenden Abschnitte ist die Feststellung, daß es verschiedene Auffassungen von y und damit auch von SD_y gibt. Im folgenden Abschnitt kommt zu den beschriebenen zwei Begriffsauffassungen noch eine hinzu, so daß drei Ansätze ausführlich von mir dargestellt werden. Ich halte es für sehr schwierig, Ordnung in die verschiedenen Auffassungen von y und SD_y zu bringen. Deshalb möchte ich in Abschnitt 4.4.3 ein Modell vorstellen, welches zum Ziel hat, die Gültigkeitsbereiche der verschiedenen Sichtweisen abzustecken.

4.4.2 Wie beeinflußt die Leistung y den Ertrag?

In diesem Abschnitt werden drei Auffassungen beschrieben, wie die Leistung y zum Ertrag führt. Diese Unterscheidung ist wichtig, wenn in Abschnitt 4.5 die Verfahren zur Schätzung von SD_y besprochen werden. Da die Standardabweichung der Leistung nicht nur schwierig zu schätzen ist, sondern darüber hinaus auch einen entscheidenden Einfluß auf das Ergebnis der Nutzenberechnung hat, ist es wichtig, die hier beschriebenen Hintergründe zu kennen.

4.4.2.1 Ertrag der Leistung durch Kostenreduktion

Der Ansatz, der Leistung als Kostenreduktion auffaßt, hat bisher in Forschung und Praxis die geringste Bedeutung gewonnen. Die Darstellung orientiert sich weitgehend an Boudreau (1991).

Van Naerssen (1965) berichtet vom Einsatz eines Fragebogens bei der Auswahl von Fahrschülern beim holländischen Militär. Da Fahrschüler keine Wertschöpfungen

¹⁰ gemeint ist Brogden (1946a)

¹¹ gemeint sind Hunter und Schmidt (1982)

produzieren, stehen den Kosten des PAP in diesem Fall keine direkten Erträge, sondern indirekte Erträge gegenüber. Sie werden im Gegensatz zu den bisher besprochenen Fällen nicht direkt erwirtschaftet, sondern entstehen durch Einsparungen. Solche Einsparungen resultieren z.B. aus der Reduktion von Unfällen, der Verringerung von benötigten Fahrstunden und niedrigeren Durchfallquoten.

Eaton, Wing und Mitchell (1985) untersuchten die Leistung von Panzerkommandanten in der U.S. Armee. Da durch diese Tätigkeit ebenfalls keine Erträge direkt erwirtschaftet werden, entstehen diese indirekt durch Einsparungen bzw. vermiedenen Kosten aufgrund der Leistungssteigerungen, welche den Interventionsmaßnahmen zugerechnet werden. Eaton et al. (1985) nennen hier als indirekte Erträge die Anzahl von Panzern, die, bei gleichbleibendem militärischem Auftrag, nicht angeschafft werden müssen.

Schmidt und Hunter (1983) besprechen in ihrem Aufsatz hauptsächlich den Nutzen, der einem Unternehmen aufgrund erhöhter Produktivität durch ein geeignetes PAP entsteht. Dennoch erwähnen sie auch kurz die Konsequenzen eines PAP für ein Unternehmen, welches nicht zum Ziel hat, seine Produktivität zu steigern. Plausiblerweise kann dieses Unternehmen mit einem leistungsfähigeren Personal Lohnkosten dadurch einsparen, indem es Stellen abbaut.

Allen diesen Ansätzen ist gemeinsam, daß Erträge aus Interventionsmaßnahmen indirekt durch die Vermeidung von Kosten entstehen. Die Logik des Nutzenmodells ist prinzipiell demjenigen gleich, welches direkte Erträge berücksichtigt. Werden von den (indirekten) Erträgen die Kosten der Maßnahme subtrahiert, so ergibt sich der Nutzen der Interventionsmaßnahme. Auch in diesen Modellen kommt der Skalierungsfaktor SD_y vor. Die Definitionsprobleme der Leistung y bleiben bestehen.

Eine etwas andere Herangehensweise propagiert Sands (1973). Sein CAPER Modell (Cost of Attaining Personnel Requirements) befaßt sich mit Prozessen, die vor, während und nach der Personalauswahl ablaufen. Es erlaubt eine Optimierung der Aufwendungen für Anwerbung, Auswahl, Einführung und Training, so daß minimale geschätzte Kosten entstehen. Das Modell kann als eine Erweiterung des Taylor-Russel Modells aufgefaßt werden. Es hat zum Ziel, einen Nutzen von PAP dadurch zu erreichen, daß die Kosten der Personalauswahlprozesse reduziert werden. Es hat den Vorteil, daß mehrere Einzelbereiche von PAP explizit berücksichtigt werden. Auf diese Weise kann eine falsche Überbetonung von Einzelbereichen vermieden werden. Da es das Taylor-Russel Modell durch die Kosten verschiedener Maßnahmen erweitert, vermeidet es auch die (falsche) Empfehlung, die Selektionsquote unendlich niedrig zu wählen. Die Vorteile des CAPER Modells werden

teilweise dadurch kompensiert, daß, wie bei seinem Vorgänger, das Kriterium dichotom behandelt wird und damit keine Unterschiede zwischen den Geeigneten bzgl. des Kriteriums gemacht werden.

4.4.2.2 Der Verkaufswert der Leistung erbringt den Ertrag

Die Hauptvertreter der Sichtweise, daß die Leistung y als der Verkaufswert der Leistung eines Mitarbeiters verstanden werden sollte, sind die Forscher Hunter und Schmidt. Sie definieren y als "the yearly value [...] of the products and Services", "the cost [...] of having an outside firm provide these products and Services" (Schmidt et al. 1979, S. 621), "output as Sold" oder womit der Arbeitgeber "charges the customer" (Hunter & Schmidt, 1982, 268-269). Andere Autoren sprechen in Anlehnung; von "value of sales" (Cascio & Silbey, 1979, S. 110; Bobko & Karren, 1982, zit. nach Boudreau, 1983a). Gemeint ist mit allen diesen Umschreibungen der Verkaufswert der Leistung, die ein Mitarbeiter erbringt. Wie aus den Zitaten deutlich wird, ist eine häufig verwendete Hilfe zur Schätzung dieses Verkaufswertes der Vergleich mit den Kosten, den eine externe Organisation für die gleiche Leistung verursachen würde. Mit anderen Worten kann der Lohn eines Mitarbeiters als ein Indikator für seine Leistung angesehen werden.

Wie diese Auffassung von y entstanden ist, läßt sich nur sehr schwierig erschließen. Nach meinem Dafürhalten ist sie jedoch ein Nebenprodukt der GLOBAL-Schätzmethode von SD_y (Schmidt et al., 1979), die in Abschnitt 4.5 besprochen wird. Diese Methode beruht kurz gesagt darauf, daß Experten den Dollarwert der Leistung von Mitarbeitern verschiedener Leistungsstufen schätzen müssen. Um diese Aufgabe zu bewältigen, erhalten die Experten eine ausführliche Instruktion. Diese ist bei Schmidt et al. (1979) wörtlich abgedruckt und enthält u.a. die oben beschriebenen Definitionen von y . Außerhalb dieser Instruktion befassen sich die Autoren bedauerlicherweise kaum mit der genauen Beschreibung des Konstruktes 'Leistung y '. Auch nachfolgende Autoren, die sich der Sichtweise von Schmidt et al. (1979) angeschlossen haben, leisten diese Arbeit nicht. Vielmehr erscheint es, als ob sie dankbar wären, daß man ihnen einen einfachen Weg der Schätzung von SD_y geebnet hat. Viele Forscher begehen diesen Weg weitgehend unkritisch, so lange dies noch möglich ist. Als Indiz für die Richtigkeit dieser Aussage möchte ich anführen, daß viele dieser Autoren statt einer Besprechung des SD_y -Konstruktes die wörtliche Instruktion an die beurteilenden Experten in ihren Aufsätzen wiedergeben. Nur daraus ist oft ihre Auffassung von y bzw. SD_y überhaupt erschließbar.

Die GLOBAL-Schätzmethode von Schmidt et al. (1979) hat sehr viele Anhänger gefunden. Durch sie ist eine vergleichsweise einfache Schätzung des

'Achillesfersen'-Parameters SD im BCG-Modell möglich geworden. Der Aufsatz y von Schmidt et al. (1979) hat zweifelsohne eine Vielzahl von Forschungsarbeiten zur NA stimuliert. Dennoch vermochten viele der nachfolgenden Arbeiten nicht die Unzulänglichkeit des Schmidt et al. (1979) Aufsatzes auszugleichen, ein genaues und plausibles Konzept von y zu vermitteln. Ich möchte in dieser Arbeit das Konzept von y als 'Verkaufswert der Leistung' deutlicher kritisieren.

Ich möchte meine Argumentation mit zwei fiktiven Beispielen einleiten. Man stelle sich einen Computerprogrammierer vor. Er hat die Aufgabe, vorgegebene Flußdiagramme in eine Programmiersprache zu übersetzen und die so entstandenen Programme auf ihre Fehlerfreiheit zu testen. Ein durchschnittlicher Programmierer erledigt zwei Programme pro Monat. Bedenkt man, daß er sechs Wochen Urlaub hat, dann wäre sein jährliches Handlungsergebnis (Abb. 2, Stufe 7) für das Unternehmen 21 Programmumsetzungen. Anstatt den Programmierer die Aufgabe erledigen zu lassen, können wir uns auch vorstellen, daß ein externes Unternehmen den Auftrag erhält. Die Auftrags erledigung von 21 Programmen berechnet das externe Unternehmen mit 80000 DM. Diese Zahl kann dafür dienlich sein, das Handlungsergebnis des Programmierers für das Unternehmen in Geldeinheiten zu bewerten (Abb. 2, Stufe 8), oder anders ausgedrückt, seine jährliche Leistung y zu schätzen. Auch das Gehalt des Programmierers könnte auf ähnliche Weise festgelegt worden sein. Innerhalb des BCG-Nutzenmodells wird jedoch nicht der Parameter y , sondern SDy benötigt. Um die Standardabweichung der Leistung SDy zu erhalten, genügt in unserem fiktiven Beispiel, mehrere Programmierer zu untersuchen. Die Anzahl ihrer jährlich erledigten Programme wird dabei eine gewisse Variabilität aufweisen. Sagen wir, daß unsere Programmiererstichprobe durchschnittlich 21 Programme mit einer Standardabweichung von 5 Programmen jährlich zum Abschluß bringen. Gemäß der oben beschriebenen Logik können nun diese Handlungsergebnisse in DM bewertet werden. Unsere Programmierer erbringen demnach für das Unternehmen die durchschnittliche Leistung $M_y=80000$ DM mit einer Standardabweichung von $SD_y = 19048$ DM. Die Einfachheit dieser Logik ist verblüffend. In diesem fiktiven Beispiel wurde ein vergleichsweise einfacher Weg beschrieben, den Achillesfersenparameter SD_y zu schätzen. Wenn es möglich ist, durch ein Interventionsprogramm (Personalauswahl, Training, etc.) die Leistungsfähigkeit der Programmierer um eine halbe Standardabweichung anzuheben, dann erbringt dieses dem Unternehmen einen jährlichen Ertrag von $E=9524$ DM pro Mitarbeiter. Werden davon die

Kosten des Interventionsprogramms je Mitarbeiter abgezogen, so erhalten wir dessen Nutzen. Besteht das Unternehmen aus 30 Programmierern, und kostet die Interventionsmaßnahme insgesamt 100000 DM, dann berechnet sich der Nutzen der Maßnahme im ersten Jahr nach (12)

$$(12) \quad \Delta U = 30 * 9524 \text{ DM} - 100000 \text{ DM} = 185720 \text{ DM}$$

So fiktiv, wie bisher behauptet, ist dieses Beispiel allerdings gar nicht. Es ist in Anlehnung an die empirische Nutzenanalyse entstanden, die bei Schmidt et al. (1979) beschrieben ist. In Zusammenhang mit diesem empirischen Sachverhalt haben Schmidt et al. (1979) ihre GLOBAL-Schätzmethode entwickelt und das Verständnis von SD_y als Standardabweichung des Verkaufswertes der Leistung geschaffen.

Das zweite Beispiel ist nun allerdings wirklich fiktiv. Man stelle sich vor, jemand vertreibt ein beliebiges Produkt. Da dies als Nebentätigkeit geschieht und der Vertreiber kaum Zeit übrig hat, beschafft er sich für diese Aufgabe Zwischenhändler, so daß er selbst nur mit Bestellungen, Abrechnungen und sonstigen kleinen Bürotätigkeiten belastet ist. Das Produkt kostet im Einkauf 30 DM und es wird für 100 DM weiterverkauft. Die Zwischenhändler erhalten keinen Grundlohn, aber dafür eine Provision von 60 DM je verkauftem Stück. Analog zum Programmiererbeispiel betrachte ich zunächst einen durchschnittlichen Verkäufer. Dieser schafft es, pro Jahr 1000 Stück des Produktes zu verkaufen. Um dieses Ergebnis zu bewerten, braucht man nicht lange nach den Angeboten von externen Firmen Ausschau halten, sondern man betrachtet die Verkäufer selbst als solche. Sie kosten bei der abgesetzten Stückzahl im Jahr $y=60000$ DM. Dies ist gemäß der Auffassung von Schmidt et al. (1979) die Leistung y , die sie erbringen.

Betrachtet man auch die anderen Verkäufer, so stellt man fest, daß eine Standardabweichung von 300 verkauften Stück pro Jahr besteht, was $SD_y = 18000$ DM bedeutet. Ich will wieder die Nutzenrechnung einer Interventionsmaßnahme wie im letzten Beispiel anstellen. Die Intervention (z.B. ein Verkaufstraining) erbringt eine Leistungssteigerung von 0,5 Standardabweichungen und kostet 100000 DM. Bei 30 Verkäufern ergibt sich für das erste Jahr ein Nutzen von

$$(12) \quad \Delta U = 30 * 9000 \text{ DM} - 100000 \text{ DM} = 170000 \text{ DM}$$

Dieses Ergebnis ist für den Vertreiber erfreulich. Da die Intervention einen positiven Nutzen erbringt, entscheidet er sich natürlich dazu, in einigem zeitlichen Abstand weitere Interventionsmaßnahmen zur Steigerung des Umsatzes durchzuführen. Nach einiger Zeit stellt er aber fest, daß er immer weniger Geld besitzt die Verkäufer aber in Saus und Braus leben.

Er legt die nutzenanalytischen Modelle beiseite und rechnet mit seinem gesunden Menschenverstand noch einmal nach:

Ein durchschnittlicher Verkäufer setzt im Jahr 1000 Produkte um. Dies entspricht einem Umsatz von $1000 * 100 \text{ DM} = 100000 \text{ DM}$. Dadurch entstehen Einkaufskosten von $1000 * 30 \text{ DM} = 30000 \text{ DM}$ und Provisionskosten von $1000 * 60 \text{ DM} = 60000 \text{ DM}$. Ein Verkäufer erwirtschaftet für den Vertreiber damit durchschnittlich 10000 DM/ Jahr . Die Standardabweichung des erwirtschafteten Betrags liegt bei 3000 DM pro Jahr. Unsere Maßnahme kostet uns 100000 DM , wodurch alle 30 Verkäufer eine halbe Standardabweichung mehr verkaufen können. Dadurch sollten alle Verkäufer durchschnittlich 1500 DM mehr für den Vertreiber erwirtschaften als vor der Maßnahme. Bei 30 Verkäufern macht dies 45000 DM jährlich. Unter Beachtung der Kosten für die Intervention ist dieser Betrag zu gering, um einen positiven Nutzen zu erzeugen. Die Schmälerung des Geldbeutels des Vertreibers ist jetzt erklärbar. Er hat die falsche Entscheidung auf der Grundlage des Nutzenmodells getroffen. Was aber ist an der Nutzenberechnung falsch, die doch so viel Geld versprochen hat?

Falsch ist die Definition des Skalierungsfaktors SD_y . Die weitverbreitete Auffassung von y als Verkaufswert von Produkten und Dienstleistungen wurde erstmalig von Boudreau (1983a) hart kritisiert. Unter Bezugnahme auf mehrere Definitionen der Leistung y , die auf Schmidt et al. (1979) basieren und zu Anfang dieses Abschnittes und in Abschnitt 4.4.1 dargestellt sind, schreiben die Autoren (S. 553): "For several reasons, this payoff function is a deficient expression of the 'institutional'¹² benefit." Der Autor begründet diese Behauptung, indem er psychologische Interventionsmaßnahmen unter Investitionsgesichtspunkten betrachtet und mehrere ökonomische Konzepte der Investitionsrechnung heranzieht. Worin besteht der Unterschied zwischen dem ersten und zweiten Beispiel? Boudreau (1983a) würde sagen: Im ersten Beispiel war die Herstellung des Produkts oder die Dienstleistung nur mit wenigen variablen Kosten verbunden. In den Worten eines Wirtschaftswissenschaftlers würde dies heißen, daß die Bereitstellung des Produkts mit nur wenigen ausbringungsmengenabhängigen Kosten verbunden ist (zum Konzept der variablen und festen Kosten siehe z.B. Wöhe, 1993). Die Programmierer brauchen zur Erstellung ihrer Produkte ihren Arbeitsplatz und ihre Arbeitsmittel und diese verursachen Kosten. Allerdings ist die Höhe dieser Kosten weitgehend davon unabhängig, ob ein Programmierer pro Monat zwei oder fünf Programme schreibt. Es handelt sich um feste Kosten.

¹² Hier wird auf die Unterscheidung institutioneller vs. individueller Nutzen von Cronbach und Gleser (1965) angespielt. Vgl. auch die Unterscheidung von individueller vs. institutioneller Entscheidung (Abschnitt 2.2.3).

Im zweiten Beispiel dagegen sind die Kosten der Leistung in hohem Maße ausbringungsmengenabhängig. Sowohl die Herstellungskosten der Produkte, als auch die Provision, die der Vertreter bezahlen muß, sind abhängig von den Umsatzzahlen der Verkäufer. Lediglich die Büroarbeit des Vertreibers verursacht feste Kosten. Offensichtlich wirkt sich dieser Unterschied deutlich aus. Dieses Beispiel zeigt, daß die Definition der Leistung y im Sinne von Schmidt et al. (1979) keinesfalls allgemeingültig ist. Es ist festzuhalten, daß bei der Begriffsbestimmung von y bzw. SD_y die Konzepte der variablen und festen Kosten der Leistung wichtig sind. Betrachtet man weiter die beiden Beispiele, so läßt sich fragen, wo der zunächst errechnete Ertrag der Interventionsmaßnahme in Beispiel 2 geblieben ist. Der Vertreter hat ihn offensichtlich nicht geerntet. Die Antwort ist, daß ein Teil des Ertrags der Interventionsmaßnahme der Hersteller des Produkts eingestrichen hat, die eigentlichen Nutznießer sind jedoch die Verkäufer. Durch ihre gesteigerte Provision floß der Großteil des Ertrags des Verkaufstrainings in ihre Taschen.

In der Literatur zur Nutzenanalyse wird die Tatsache, daß eine Steigerung der beruflichen Handlungsergebnisse bei vielen Arbeitssituationen auch mit mehr oder weniger variablen Kosten für die arbeitgebende Organisation verbunden sind, weitgehend nicht erkannt. Wenn sie aber erkannt wird, unterbleibt es meist, daraus die richtigen Schlüsse zu ziehen. Schmidt und Hunter (1983) unterscheiden z.B. auf Vorschlag von Rothe (1978) und Rothe und Nye (1959, beide zit. nach Hunter & Schmidt, 1979) zwischen Arbeiten mit Akkordlohnsystemen und Nichtakkordlohnsystemen. Die wesentliche Konsequenz daraus, daß ein Teil des Ertrags auf die Akkordarbeiter übergeht, und damit dem Unternehmen abgezogen werden muß, wird bei der Schätzung von SD_y nicht berücksichtigt. Ein Paradebeispiel für das (Miß-)Verständnis von 'Leistung als Verkaufswert' liefern Bobko et al. (1983). Sie schätzen nach der GLOBAL-Methode die Standardabweichung der Leistung SD_y von Versicherungsvertretern. Dabei verwenden sie einmal in Anlehnung an das Original von Schmidt et al. (1979) die Instruktion "yearly dollar sales", was den Verkaufswert der Leistung darstellt. Darüber hinaus nehmen sie eine zweite Schätzung mit der Instruktion "value to the company" vor. Beide Schätzungen werden tabellarisch dargestellt. Die Schätzwerte von SD_y stehen zueinander im Verhältnis von ungefähr 1 zu 10. Spätestens dieses Ergebnis hätte zu denken geben müssen. Seltsamerweise gehen die Autoren auf die grundlegende Sachfrage, welche SD_y -Schätzung die bessere sei, nicht weiter ein, sondern diskutieren akademische Nebenfragen, welche sich im Zusammenhang mit der GLOBAL-Methode stellen.

Zusammenfassend kann festgestellt werden, daß das Verständnis der Leistung y als der Verkaufswert der Leistung nicht allgemeingültig ist. Die Schätzung der

Standardabweichung der Leistung SD_y , welche auf einem derart definierten y basiert, führt zu Nutzenüberschätzungen, wenn das Wertesystem der Organisation im Mittelpunkt stehen soll.

Neben diesen praktischen Konsequenzen ist auch das theoretische Konzept der Leistung y als Verkaufswert der Leistung fragwürdig. Aus der Sicht der Organisation stellt die Entlohnung einen Kostenfaktor dar. Dieser entsteht bei der 'Herstellung' der Produkte oder der Dienstleistungen genauso wie andere Kosten. Von der Kostenrechnungsperspektive aus betrachtet gibt es keinen Grund, warum die Konzepte der Leistung und der Kosten miteinander vermischt werden sollten. Das heißt, die Entlohnung würde im ersten Fall als Lohnkosten für die Organisation angesehen, im zweiten Fall aber als Leistungs- und damit Ertragsindikator. Wie das fiktive Verkäuferbeispiel aber deutlich macht, ist die Höhe der Entlohnung der Mitarbeiter bei der Leistungsberechnung zweitrangig. Was zählt ist der Umsatz abzüglich der variablen Kosten.

4.4.2.3 Der Deckungsbeitrag der Leistung erbringt den Ertrag

Das größte Problem der Sichtweise 'Leistung als Verkaufswert' besteht in dem Umstand, daß SD_y wohl meist überschätzt wird. Boudreau (1983a) zeigt eindrucksvoll, daß sich der Nutzen von Interventionsmaßnahmen bei publizierten Nutzenanalysen erheblich verringert, wenn ökonomische Konzepte, wie z.B. das der variablen Kosten der Leistung, in die Überlegungen mit einbezogen werden. Er trennt die Leistung y auf in den Verkaufswert w und die 'Servicekosten' s , die dadurch entstehen, daß w erbracht werden kann (Materialkosten, Leistungslohn, ...). Die Servicekosten sind ausbringungsmengenabhängige variable Kosten. Die Nettogleistung (net benefits) y setzt sich aus dem Verkaufswert der Leistung w abzüglich der Servicekosten zusammen.

$$y = w - s \quad (14)$$

Damit läßt sich y als Deckungsbeitrag ansehen. Der Deckungsbeitrag ist definiert als der Umsatz abzüglich der variablen Kosten. Boudreau führt den Parameter V ein, der den Anteil der variablen Servicekosten am Verkaufswert angibt. V berechnet sich durch

$$V = s/w. \quad (15)$$

Das allgemeine BCG-Nutzenmodell läßt sich folgendermaßen darstellen

$$\Delta U = Na T r_{xy} \bar{z}_x * SD_y - C. \quad (16)$$

C sind dabei die Gesamtkosten der Interventionsmaßnahme, Na die Anzahl der durch die Intervention behandelten bzw. die akzeptierten Bewerber bei der Personalauswahl und die T Wirkdauer der Maßnahme. y ist die Leistung der Mitarbeiter, also die bewerteten

Handlungsergebnisse in Geldeinheiten aus Sicht der Organisation. Die Validität r_{xy} ist die Korrelation zwischen Prädiktor x und y , SD_y die Standardabweichung der Leistung und \bar{z}_x^* der durchschnittliche standardisierte Prädiktorwert der selektierten Stichprobe. Wird im allgemeinen BCG-Nutzenmodell statt y der Parameter w (Verkaufswert der Leistung) verwendet; so muß der Anteil der variablen Kosten am Verkaufswert im Modell berücksichtigt werden. Dies geschieht durch Aufnahme des Parameters V ins Modell.

$$\Delta U = Na T r_{xw} \bar{z}_x^* SD_w (1-V) - C \quad (17)$$

Es gilt also

$$SD_y = SD_w (1-V). \quad (18)$$

Diese Gleichung stammt von Boudreau (1983a). Der Verkaufswert der Leistung w wird als das aufgefaßt, was Schmidt et al. (1979) als "value as sold" bezeichnen. Beträgt der Anteil der variablen Kosten am Verkaufswert der Leistung z.B. $V=30\%$ (bzw. $V=0,3$), dann muß die Standardabweichung des Verkaufswertes der Leistung SD_w um den Faktor 0,7 vermindert werden, um die Standardabweichung der Leistung SD_y zu erhalten. Es läßt sich also festhalten, daß unter bestimmten Umständen die Parameterschätzung SD_y nach unten korrigiert werden muß. Dies ist dann der Fall, wenn der Anteil der variablen Servicekosten an den Verkaufskosten bedeutsam wird, V damit einen Wert annimmt, der von Null abweicht.

4.4.3 Die Leistung y - Ordnungsversuche

Wenn man sich mit der Nutzenanalyse beschäftigt, besteht die Hauptschwierigkeit darin, das Konzept der Standardabweichung der Leistung SD_y zu verstehen. Dies liegt weniger daran, daß es schwer verständlich, und damit nur eingeschränkt brauchbar ist. Vielmehr besteht die Schwierigkeit darin, daß es in der Literatur in mehreren Bedeutungen gebraucht wird. Die Definitionsversuche der Forscher, die den Grundstein zur NA legten, haben sich als mehrdeutig erwiesen. Ich möchte mit eigenen Systematisierungsversuchen Ordnung in das Konzept SD_y bringen.

Auf Stufe 7 des Prozeßmodells der Personalauswahl (Abb. 2) sind die Handlungsergebnisse der behandelten Individuen angeführt. Es ist angedeutet, daß Ergebnisse für verschiedene Zielobjekte entstehen können. Als Beispiel ist das Individuum, die Institution und andere Organisationen genannt. Stufe 8 beschreibt die bewerteten Handlungsergebnisse. Hierbei wird explizit zwischen verschiedenen Wertsystemen unterschieden. Im Zusammenhang der Nutzenmodelle sollte in erster Linie das institutionale Wertsystem im Vordergrund stehen. Um der Tatsache Rechnung zu tragen, daß die

Aufwendungen der Institution für die Personalauswahl in Geldeinheiten bewertet werden, werden auch die institutionellen Bewertungen der Handlungsergebnisse in DM vorgenommen. Die Definition des bewerteten Kriteriums als 'Verkaufswert' mißachtet die Unterscheidung und Trennung zwischen mehreren möglichen Wertsystemen. In einem Nutzenmodell, welches die Standardabweichung des bewerteten Kriteriums auf der Grundlage des 'Verkaufswertes' definiert, sind die Bewertungen durch die Institution mit denen anderer Organisationen und Individuen vermischt. Ein solchermaßen definiertes Kriterium bezeichne ich mit w (Verkaufswert), seine Standardabweichung mit SD_w . Es läßt sich also festhalten:

Bei der Nutzenanalyse muß deutlich gemacht werden, welches Wertsystem herangezogen wird, um die Ergebnisse einer Intervention zu bewerten. Im Regelfall handelt es sich um das Wertsystem der Organisation, welches die Intervention herbeiführt. Die Bewertung ist normalerweise mit der Einheit DM versehen.

Die in DM bewerteten Handlungsergebnisse für die Organisation müssen variable Kosten der Handlungen mitberücksichtigen. Variable Kosten s sind solche Kosten, die mit dem Verkaufswert der Handlungsergebnisse w positiv korreliert sind. Sind solche variable Kosten vorhanden, so trägt eine Steigerung der Handlungsergebnisse nach einer Intervention nicht in vollem Umfang zum Ertrag für die Organisation bei. Die Leistung y eines Mitarbeiters ist nach meiner Sichtweise die Gesamtheit der bewerteten Handlungsergebnisse in DM für die Organisation (w), abzüglich der variablen Kosten (s), welche durch die Leistung für die Organisation entstehen. Ein ökonomisches Konzept kann hier fruchtbar gemacht werden. Der Deckungsbeitrag ist definiert als der Umsatz weniger den variablen Kosten. Der erhöhte Deckungsbeitrag, den ein Mitarbeiter aufgrund einer Intervention erwirtschaftet, trägt damit nicht nur zur Finanzierung der Intervention bei, sondern auch zur Steigerung des Gewinns. Letztendlich ist der zusätzliche Deckungsbeitrag, der durch eine Intervention entsteht, das, was die Organisation interessiert. Werden vom zusätzlichen Deckungsbeitrag die Kosten der Intervention abgezogen, so erhält man den Nutzen ΔU der Maßnahme.

Die Leistung y ist die Gesamtheit der bewerteten Handlungsergebnisse in DM für die Organisation abzüglich der variablen Kosten, welche durch die Leistung für die Organisation entstehen. Die Leistung y kann somit als der Deckungsbeitrag definiert werden, den ein Mitarbeiter für die Organisation erwirtschaftet. $SD y$ ist die Standardabweichung dieser Leistung y zwischen den Mitarbeitern einer Berufsgruppe.

Der Nutzen ΔU einer Intervention ist der Ertrag, der aufgrund der Intervention entsteht, abzüglich der variablen Kosten und der festen Kosten der Intervention.

Nicht bei allen Arbeitstätigkeiten fallen in gleichem Maße variable Kosten durch Leistung an. Wie bei den Beispielen in Abschnitt 4.4.2.2 (Programmierer und Produktverkäufer) deutlich wurde, gibt es Berufsfelder, in welchen Leistungssteigerung nahezu ohne Kostensteigerung möglich sind. Typische Beispiele hierfür sind Programmierer, Verwaltungsberufe, Militär und Polizei, etc. Andere Berufsfelder hingegen weisen sehr hohe variable Kosten auf. Dies ist z.B. immer dann der Fall, wenn reine Leistungslohnsysteme bestehen. Auf diesen Umstand weise ich deswegen hin, weil variable Kosten der Leistung bei der Durchführung von Nutzenanalysen oft nicht beachtet wurden. Dies sollte weniger schlimm sein, wenn Arbeitstätigkeiten untersucht werden, bei welchen der Anteil der variablen Kosten im Vergleich zum Verkaufswert der Leistung ohnehin gering ist.

Abbildung 8 soll diese Unterscheidung veranschaulichen. Der Koeffizient V bezeichnet das Verhältnis der variablen Kosten (in DM) zum Verkaufswert eines Handlungselementes (in DM).

$$(15) \quad V = s/w$$

Auf der Abszisse sind die Werte von V abgetragen. Der Wertebereich geht von 0 bis ∞ . Auf der Ordinate ist der Anteil des Deckungsbeitrags am Verkaufswert der Leistung w (d.h. y/w) abgetragen. Wie der Graph veranschaulicht, sinkt der anteilige Deckungsbeitrag je Handlungseinheit, je größer V wird. In wirtschaftlich arbeitenden Unternehmen sollten nur Arbeitstätigkeiten zu finden sein, aus welchen ein positiver Deckungsbeitrag resultiert. Theoretisch denkbar sind jedoch auch Tätigkeiten mit negativem Deckungsbeitrag. Dies könnte z.B. bei Arbeitsbeschaffungsmaßnahmen der Fall sein. Interventionsmaßnahmen, welche die Handlungsergebnisse steigern, führen in diesem Fall zu negativen Erträgen.

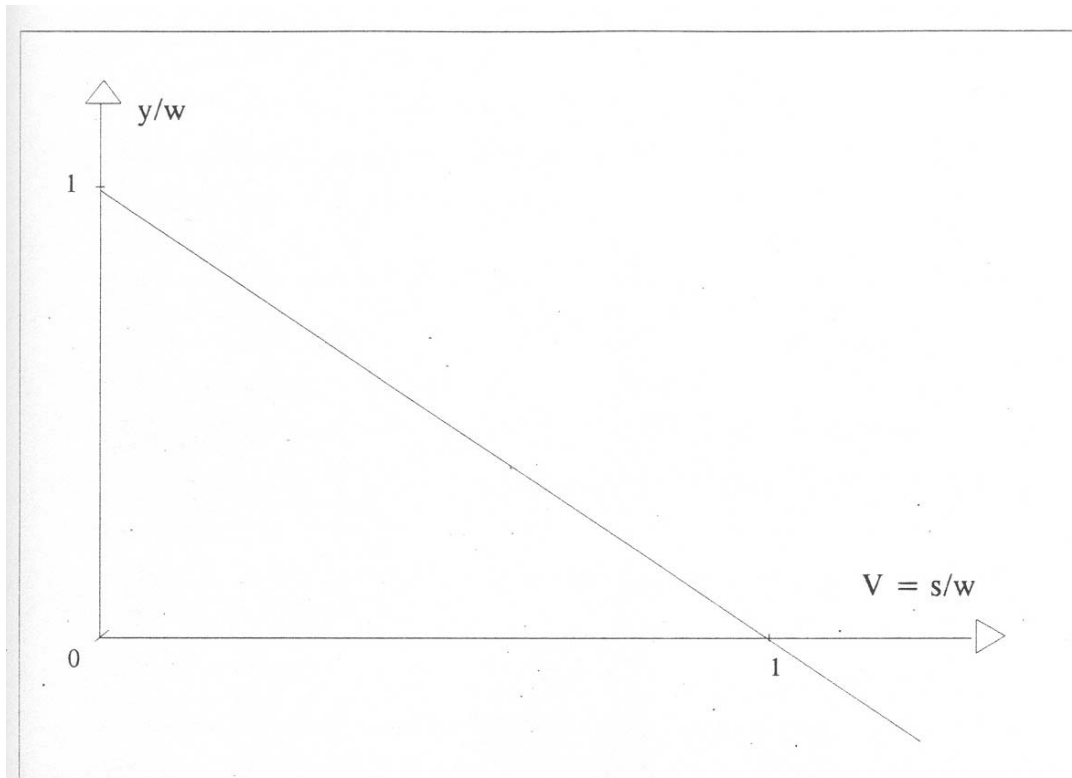


Abbildung 8: Der Zusammenhang zwischen dem Anteil des Deckungsbeitrages y am Verkaufswert der Leistung w und dem Parameter V (Anteil der variablen Kosten s am Verkaufswert der Leistung w).

Ob bei der Schätzung von SD_y das Verständnis von y als der Verkaufswert der Leistung zu gravierenden Verzerrungen führt, kann durch Abbildung 8 veranschaulicht werden. Liegt der Anteil der variablen Kosten am Verkaufswert niedrig, d.h. ist V klein, dann kann die Leistung y mit dem Verkaufswert der Leistung w annähernd gleichgesetzt werden. Die Schätzungen von SD_y , die auf einer solchen Definition basieren, sind nur unwesentlich verzerrt. Dies ist bei früheren nutzen-analytischen Untersuchungen oft der Fall, welche der SD_y -Schätzung zwar das Verkaufswertkonzept zugrundelegen, dieses aber trotz theoretischer Unzulänglichkeiten mit dem Deckungsbeitragkonzept gleichgesetzt werden kann.

Sobald jedoch ein Arbeitsfeld vorliegt, bei welchem die Annahme eines geringen Anteils der variablen Kosten am Verkaufswert nicht mehr gemacht werden kann, sind erhebliche Zweifel geboten. Schätzungen von SD_y können dann nicht mehr akzeptiert werden, wenn sie direkt oder indirekt auf der Entlohnungshöhe basieren. Nur solche Nutzenmodelle, die den Parameter V oder ein ähnliches Konzept beinhalten, spiegeln die Wirklichkeit einigermaßen richtig wieder.'

Im Abschnitt 4.4 habe ich versucht, wichtige theoretische Überlegungen zur Definition von y und seiner Standardabweichung SD_y festzuhalten. Hierbei sollte insbesondere deutlich geworden sein, daß der Parameter y begrifflich verschieden aufgefaßt wird. Ich verstehe ihn

als die bewerteten Handlungsergebnisse in DM durch die Organisation, welche die Intervention trägt. Diese nenne ich die Leistung y des Mitarbeiters. Sie stellt formal den Deckungsbeitrag des Mitarbeiters zum Unternehmen dar.

Die Unterscheidung zwischen der Leistung als Deckungsbeitrag (y) und dem Verkaufswert der Leistung (w) ist sehr wichtig. Im weiteren Verlauf der Arbeit wird erstere durch y und letztere durch w bezeichnet.

4.5 Methoden zur Ermittlung von SD_y

Da die Schätzung der Standardabweichung der Leistung SD_y verglichen mit den anderen Parametern des BCG-Nutzenmodells schwierig ist, und weil der Parameter aufgrund seiner Nichtstandardisierung überproportional das Ergebnis der NA beeinflusst, wurden verschiedene Methoden zur Ermittlung von SD_y entwickelt. Die meisten NA verwenden mehrere Schätzmethoden. Ähnliche Schätzergebnisse werden dann oft als Indikator dafür interpretiert, daß die Schätzungen valide sind.

Die Schätzmethoden lassen sich in vier Kategorien ordnen (vgl. auch Boudreau, 1991). Die erste Klasse umfaßt dabei Ansätze, die sich Kostenrechnungsmethoden bedienen (z.B. Roche, 1961, 1965). Prinzipien der Kostenrechnung werden dabei verwandt, um individuellen Handlungen oder Handlungsergebnissen Geldwerte zuzuordnen. Die Standardabweichung dieser Geldwerte zwischen den Individuen ist die Schätzung für SD_y . Die zweite Klasse besteht aus Schätzmethoden, bei welchen Experten für die in Frage kommenden Tätigkeiten (in der Regel Vorgesetzte) globale Schätzungen abgeben, wieviel die beruflichen Handlungen der Ausführenden wert sind. Diese Schätzung wird für Vertreter verschiedener Leistungsklassen ausgeführt, so daß auf diesem Wege die Standardabweichung ermittelt wird. Oft wird diese Methode 'GLOBAL-Schätzmethode' genannt. Sie findet ihren Ursprung bei Schmidt et al. (1979). Die dritte Klasse der Schätzverfahren sind sogenannte proportionale Verfahren. Diese Art der Verfahren gehen auf Hunter und Schmidt (1982) zurück. Sie sind sehr leicht zu handhaben, dafür aber auch umstritten. Dieser Ansatz hat die Beobachtung zur Grundlage, daß SD_y in den meisten vergangenen Studien einen bestimmten Prozentsatz der mittleren Lohnhöhe erreicht hat. Da dieser vergleichsweise stabil bleibt, kann, so die Vertreter dieses Verfahrens, eine Schätzung damit erfolgen. Daß dieser empiristische Ansatz viele Kritiker aufs Feld gerufen hat, ist verständlich, wenn man bedenkt, daß seine

Vertreter eine Abhängigkeit der Standardabweichung vom Mittelwert behaupten. Bei der letzten Klasse der Schätzmethoden werden Verhaltensunterschiede auf Individuumsebene betrachtet (z.B. Cascio & Ramos, 1986). Das Grundprinzip der Methode besteht darin, daß einer Stichprobe von beobachtbaren Verhaltenseinheiten Geldwerte zugeordnet werden. Auf dieser Grundlage kann für jedes Individuum die Leistung in DM berechnet werden. Die Standardabweichung zwischen den Individuen stellt die Schätzung für SD_y dar. Diese vier Schätzverfahren werden nachfolgend vertieft besprochen.

4.5.1 Der Kostenrechnungsansatz

Die betriebliche Kostenrechnung verfolgt das Ziel, die Wirtschaftlichkeit des Betriebsprozesses zu kontrollieren und die Kalkulation des Angebotspreises zu ermöglichen. Diese Ziele sollen durch die Erfassung, Verteilung und Zurechnung der Kosten erreicht werden, die bei der betrieblichen Leistungserstellung und -verwertung entstehen (Wöbe, 1993).

Kostenrechnungsprozeduren innerhalb der NA gehen zurück auf Brogden und Taylor (1950), die sich in ihrer Arbeit um die Konstruktion eines Leistungskriteriums in Geldwerteinheiten ("dollar criterion") bemühen. Hierzu machen sie Methoden der Kostenrechnung fruchtbar. Die Bewertung der Leistung eines Mitarbeiters verlangt die Beachtung verschiedener Sachverhalte. Einen Eindruck von den wichtigsten Sachverhalten, die in die Bewertung der Leistung eingehen sollten, vermitteln Brogden und Taylor (1950, S. 146, Übersetzung durch den Verf.).

- Durchschnittlicher Wert einer Produktions- oder Dienstleistungseinheit.
- Qualität der produzierten Objekte oder der ausgeführten Dienstleistungen.
- Gemeinkosten - beinhalten Miete, Licht, Heizung, Abschreibung, Pacht von Maschinen und Ausrüstung.
- Fehler, Unfälle, Verderb, Verschwendung, Beschädigung von Maschinen oder Ausrüstung durch nicht sachgemäße Handhabung.
- Faktoren wie Erscheinung, Freundlichkeit, Ausgeglichenheit und allgemeine soziale Fertigkeiten, wenn Außenbeziehungen betroffen sind. (...)
- Die Kosten der Zeit, welche von anderem Personal in Anspruch genommen werden. Diese beinhaltet nicht nur die Zeit des Führungspersonals, sondern auch die anderer Mitarbeiter.

Alle diese Punkte, für die Brogden und Taylor keine Vollständigkeit beanspruchen, müssen in Geldwerteinheiten übersetzt werden. Es ist klar, daß dies eine sehr komplexe Aufgabe ist. Trotzdem ist der Kostenrechnungsansatz zur Schätzung von SD_y theoretisch plausibel und damit auch leicht und überzeugend kommunizierbar.

Roche (1961, 1965) hat versucht, den Kostenrechnungsansatz bei seiner Nutzenanalyse eines PAP bei Bohrmaschinenbedienern konsequent umzusetzen. Er kommt zu der Einschätzung, daß das Leistungskriterium in Geldeinheiten, wie es von Brogden und Taylor vorgeschlagen wurde, mit Hilfe der Kostenrechnungsspezialisten eines Unternehmens konstruiert werden kann. Dessen Standardabweichung über die verschiedenen Mitarbeiter hinweg bildet die Schätzung für SD_y . Diese kann dann innerhalb des BCG-Nutzenmodells zur Anwendung kommen. Trotz dieser klaren Aussagen steht auch Roche der Kostenrechnungsmethode kritisch gegenüber. Durch die Komplexität des Verfahrens waren bei ihm viele Schätzungen und willkürliche Wertzuweisungen bei der Kostenrechnung notwendig.

Cascio & Ramos (1986) berichten im Rahmen ihrer nutzenanalytischen Studie, daß zunächst beabsichtigt war, auf der Grundlage des Kostenrechnungssystems, welches im Unternehmen vorhanden war, den Nettobeitrag eines jeden Mitarbeiters zum Unternehmen in Dollar zu schätzen. Die Aufgabe sollte in Zusammenarbeit mit einer Gruppe von Wirtschaftsprüfern vor sich gehen. Allerdings fanden sich die Wissenschaftler bald in der Situation wieder, daß eine Reihe von schwer vertretbaren Schätzungen bei dieser Vorgehensweise unumgänglich wurde. Sie verweisen auf die gleichen Schwierigkeiten, welche verschiedene Forscher mit dem Ansatz in der Vergangenheit hatten und verwarfen das Verfahren, um stattdessen eine nutzerfreundlichere Technik zu entwickeln. In Abschnitt 4.5.3 wird diese beschrieben werden.

Greer und Cascio (1987) sehen den Kostenrechnungsansatz der Schätzung von SD_y als den konzeptuellen Vergleichsstandard für andere Schätzverfahren, da Kostenrechnungsdaten im allgemeinen objektiv und überprüfbar sind. Allerdings erkennen auch sie, daß die Nutzung von Kostenrechnungsdaten sowohl kostenintensiv als auch zeitkonsumierend sind. In diesen Tatsachen sehen die Autoren den Umstand begründet, daß sich verhaltensorientierte Schätzmethoden (GLOBAL, CREPID, ...) auf diesem Sektor weitgehend durchgesetzt haben. Auch Boudreau (1991) kommt zu dem Schluß, daß einfachere Methoden zur Schätzung des SD_y -Parameters nicht nur häufig gefordert wurden, sondern sich auch in der nutzenanalytischen Literatur durchgesetzt haben.

4.5.2 Die GLOBAL-Methode

Dieses Schätzverfahren für SD_y wurde von Schmidt et al. (1979) vorgeschlagen. Ich nenne es in Anlehnung an Boudreau (1991) die 'GLOBAL'-Methode. Schmidt et al. (1979) begründen die Notwendigkeit der Entwicklung eines alternativen Schätzverfahrens damit, daß der Aufwand für die Kostenrechnungsmethode sehr hoch ist und diese trotzdem keine verlässlichen Ergebnisse sicherstellt.

Beim GLOBAL-Verfahren werden Expertenurteile über eine bestimmte Arbeitstätigkeit eingeholt. Im Regelfall handelt es sich bei den Beurteilern um die Vorgesetzten der Mitarbeiter, welche die Zieltätigkeit ausführen. Die Beurteiler sollen zunächst den Wert schätzen, den ein durchschnittlicher Stelleninhaber jährlich zum Unternehmen beiträgt. Dieses Werturteil wird in Geldeinheiten verlangt und sollte theoretisch dem Mittelwert von y entsprechen. Wie in Abschnitt 4.4 ausführlich besprochen wurde, wird den Beurteilern meist als Hilfestellung angeboten, sich ein externes Unternehmen vorzustellen, welches die Aufgaben des Stelleninhabers zu erledigen hätte. Die Kosten, welche durch die externe Auftragsvergabe entstehen würden, entsprechen damit konzeptionell dem Verkaufswert der Mitarbeiterleistung. Ich bezeichne dieses Urteil deshalb nicht mit y sondern mit w .

Um die Standardabweichung der Leistung SD_w schätzen zu können, wird den Experten danach die Aufgabe gestellt, den Beitragswert eines fiktiven Mitarbeiters einzuschätzen, welcher den Prozentrang 85 (PR 85) einnimmt. Der Prozentrang 85 bedeutet, daß 85% der Kollegen des fiktiven Mitarbeiters eine schlechtere Leistung erbringen als er, während 15% besser sind. Es handelt sich also um einen Mitarbeiter mit überdurchschnittlicher Leistung. Ebenso wird nach dem beurteilten Beitragswert eines unterdurchschnittlichen PR 15 Mitarbeiters gefragt (15% schlechter, 85% besser).

Unter der Annahme, daß die Leistung in der Population der Tätigkeitsträger normal verteilt ist, müßten die Abstände zwischen der Leistung des (fiktiven) durchschnittlichen PR 50 Mitarbeiters und des PR 85 Mitarbeiters genauso groß sein wie die, zwischen dem PR 50 und dem PR 15 Mitarbeiter. Kann die Annahme der Normalverteilung aufrechterhalten werden, dann sollte die mittlere Differenz zwischen PR 50 und PR 85 bzw. PR 15 einen Schätzwert der Standardabweichung des Verkaufswertes der Leistung SD_w darstellen.

Die Logik des Verfahrens ist vergleichsweise einfach. Das Verfahren wurde sowohl häufig angewandt, als auch häufig kritisiert. Bobko et al. (1983) beklagen insbesondere die hohe Varianz der SD_w -Schätzwerte zwischen den Beurteilern. Diese wurde bereits von Schmidt et al. (1979) gefunden. Da sie auf eine niedrige Reliabilität der Kriteriumsmessung

schließen läßt, schlagen Bobko et al. (1983) eine zweistufige Vorgehensweise bei der GLOBAL-Schätzmethode vor. Zunächst sollten die Experten die Leistung der PR 50 Mitarbeiter schätzen. Danach erhält jeder der Beurteiler eine Rückmeldung darüber, wie die anderen Beurteiler die PR 50 Leistung eingeschätzt haben. Durch diesen informationalen Einfluß sollten die Urteile näher zusammenrücken, d.h. die Varianz zwischen den Beurteilern sollte sich verringern. Dies ist psychologisch begründbar, da solche Beurteiler, die weit von der Norm abweichen, sich wahrscheinlich noch einmal Gedanken über die Richtigkeit ihrer Schätzung machen. Bei bestehender Unsicherheit kann der Durchschnitt der Urteile als Information über das richtige Urteil gewertet werden (zum informationalen Einfluß von Gruppen: z.B. Sherif, 1966).

Im zweiten Schritt erst bekommen die Beurteiler die Aufgabe, die Leistung der PR 15 und PR 85 Mitarbeiter zu schätzen. Nach Überprüfung der Normalverteilungsannahme wird, wie bei Schmidt et al., deren gemittelte Differenz zur Leistung der PR 50 Mitarbeiter als Schätzmaß für SD_w herangezogen. Darüber hinaus nehmen Bobko et al. auch noch die Schätzung der Leistung eines fiktiven PR 97 Mitarbeiters dazu. Diese Leistung sollte 2 Standardabweichungen vom fiktiven PR 50 Mitarbeiter abweichen. Allerdings kommen die Autoren zu dem Schluß, daß die Aufnahme der PR 97-Schätzung keine Verbesserung der Schätzung erbringt. Vielmehr komprimieren die Beurteiler die Verteilung der Leistung im oberen Bereich in Relation zur Normalverteilung, so daß die SD_w auf der Grundlage von PR 97 unterschätzt wird. Die Normalitätsannahme wurde von mehreren Autoren überprüft und tendenziell als haltbar befunden (Bobko et al., 1983; Burke und Frederick, 1984).

Die Forschung zur GLOBAL-Methode und deren Derivate kann wesentlich aus zwei Perspektiven kritisiert werden. Zum ersten kümmert sie sich in keiner Weise um die kognitiven Prozesse, die bei den Beurteilern ablaufen. Da die wenigsten Beurteiler ein ausgeprägtes Konzept zur Normalverteilung haben werden, wäre interessant, welche Verteilungsform sie der Beurteilung implizit unterlegen. Möglicherweise wäre eine Einweisung in die Grundlagen des Verteilungsgedankens der Statistik und ihre graphische Verdeutlichung bei der Beurteilung hilfreich, so daß genauere Urteile abgegeben werden können. Statt dessen werden die Beurteiler im Unklaren darüber gelassen, was es mit den Prozenträgen auf sich hat, um danach die Normalverteilungsannahme anhand der Urteile zu überprüfen. Wenn man aber bedenkt, daß DeSimone et al. (1986) eine Retestrelabilität von GLOBAL SD_w -Schätzungen nach 6 Monaten von $r=0,38$ erhalten haben, dann wird deutlich, daß noch einiger Forschungsaufwand nötig ist, die Urteilsprozesse zu optimieren, die bei der GLOBAL-Schätzung ablaufen.

Darüber hinaus ist bei der Forderung nach der Erforschung der kognitiven Prozesse bei der GLOBAL-Schätzung das Verständnis der Beurteilerinstruktion ein zentraler Punkt. Wie in Abschnitt 4.4 ausführlich angesprochen, beziehen sich Instruktionen zur GLOBAL-Prozedur auf den Verkaufswert von Leistungen. Wenn die Standardabweichung der Schätzungen des Verkaufswertes der Leistung eines PR 50 Mitarbeiters in Geldeinheiten zwischen den Beurteilern betrachtet wird, so fällt auf, daß sie in manchen Studien größer als der mittlere Verkaufswert ist. Diese hohe Interratervarianz kann meines Erachtens nur dadurch zustande kommen, daß es sich bei den Befragten entweder nicht um Experten handelt oder daß die Instruktion mehrdeutig ist. Die Instruktionen erfragen klar den Verkaufswert der Leistung. Dennoch verstehen aber möglicherweise einige Befragte, die mit Kostenrechnungskonzepten vertraut sind, die Instruktion als die Frage nach den Deckungsbeiträgen der Mitarbeiter oder ähnlichen Konzepten.

Der zweite Hauptkritikpunkt bezieht sich auf den Begriff des Verkaufswertes der Leistung. Er wurde in Abschnitt 4.4 ausführlich besprochen. Den Ergebnissen des Abschnitts wurde hier Rechnung getragen, indem der Verkaufswert nicht mit y , sondern mit w bezeichnet wurde. Die frühen Anhänger der GLOBAL-Methode setzten diese Begriffe noch gleich. Bereits bei Burke und Frederick (1986) findet sich aber das Verständnis der Leistung y als Deckungsbeitrag eines Mitarbeiters. Die Instruktion der GLOBAL-Prozedur wird dabei beibehalten. Es wird also der Verkaufswert w explizit erfragt. Das Nutzenmodell von Burke und Frederick (1986) beinhaltet korrigiert die Standardabweichung des Verkaufswertes der Leistung durch den Parameter V (Boudreau, 1983b). Dadurch wird die Standardabweichung des Verkaufswertes SD_w in die des Deckungsbeitrags SD_y umgerechnet.

Ein der GLOBAL-Methode ähnliches Verfahren zur Bestimmung von SD_w , stellen Eaton et al. (1985) vor. Auch bei diesem Verfahren fungieren Experten als Beurteiler. Auch sie haben die Aufgabe, den Verkaufswert der Leistung eines durchschnittlichen Mitarbeiters (PR 50) in Geldwerteinheiten zu schätzen. Im Unterschied zur Prozedur von Schmidt et al. (1979), wird im weiteren Verlauf keine SD_w -Schätzung von PR 85 bzw. PR 15 Mitarbeitern verlangt. Vielmehr gilt die Leistung einer festgelegten Anzahl von PR 50 Mitarbeitern als Vergleichswert für die Einschätzung, wie viele überdurchschnittliche Mitarbeiter (PR 85) notwendig wären, die gleiche Leistung zu erbringen, wie die festgelegte Anzahl der durchschnittlichen Mitarbeiter. Auf dieser Grundlage wird nach Gleichung (19) die Standardabweichung des Verkaufswertes der Leistung SD_w errechnet.

$$SD_w = W_{(PR50)} ((N_{PR50} / N_{PR85}) - 1) \quad (19)$$

Dabei ist $W_{(PR50)}$ der Verkaufswert der Leistung eines PR 50 Mitarbeiters, N_{50} die festgelegte Anzahl der PR 50 Mitarbeiter und N_{85} die geschätzte Anzahl der PR 85 Mitarbeiter, die notwendig wären, um die gleiche Leistung zu erbringen, wie die durchschnittlichen PR 50 Mitarbeiter.

4.5.3 Proportionale Schätzung: Die 40-70 Regel

Diese Schätzmethode für SD_y ist die einfachste und gleichzeitig die am meisten umstrittene. Sie behauptet einen Zusammenhang zwischen der Höhe der Entlohnung für eine bestimmte Tätigkeit und der Standardabweichung der Leistung zwischen den Personen, welche diese ausführen. Nachdem Hunter und Schmidt (1982) mehrere Studien zur Bestimmung von SD_y in Augenschein genommen hatten, und deren SD_y -Schätzwerte mit dem durchschnittlichen Einkommen verglichen hatten, kommen die Autoren zu dem Schluß, daß die meisten geschätzten SD_y im Intervall zwischen 40% und 70% der durchschnittlichen Entlohnung liegen.

Diese Behauptung ist jedoch vergleichsweise leicht empirisch zu widerlegen und wurde häufig kritisiert. In einer Zusammenstellung von nahezu allen international veröffentlichten empirischen Arbeiten zur NA zählt Boudreau (1991), daß von 66 Studien, welche SD_y und die Lohnhöhe berichten, SD_y von 24 Studien unter der 40% Marke liegt, bei 18 liegt sie zwischen 40% und 70%, und bei 22 über 70%.¹³ Hierdurch wird deutlich, wie sehr die Standardabweichung der Leistung über verschiedene Berufe hinweg variiert. Die Regel von Hunter und Schmidt (1982) kann jedoch als Faustregel nützlich sein, wenn andere SD_y -Schätzungen kritisch überprüft werden. Die 40% Marke des durchschnittlichen Lohnes stellt für Schmidt und Hunter (1983) eine konservative Schätzung von SD_y dar.

Alles in allem halte ich diese Regeln nur für heuristische Zwecke tauglich. Es gibt sicherlich einige Arbeitstätigkeiten, bei welchen kaum eine Variation des Kriteriums zwischen verschiedenen Mitarbeitern vorkommt, während bei anderen die Standardabweichung der Leistung ein Vielfaches der Entlohnung darstellt. Zeigt sich jedoch bei einer bestimmten Tätigkeit, bei welcher die SD_y geschätzt werden soll, kein Hinweis auf eine Spezifität, dann können die beschriebenen Proportionalregeln als Faustregeln dienlich sein.

¹³ Wie hoch die SD_y -Werte der verbleibenden zwei Studien liegen, verschweigt Boudreau.

4.5.4 Individuelle Schätzmethoden

Alle individuellen Schätzmethoden von SD_y verfahren nach demselben Grundprinzip. Sie stellen das Individuum in den Fokus der Aufmerksamkeit. Zunächst wird eine Stichprobe der Tätigkeiten eines Individuums gebildet. Im zweiten Schritt wird jeder Teiltätigkeit ein Geldwert für die Organisation zugeordnet. Danach wird der Wert des Individuums für die Organisation durch Addition der Werte der Teiltätigkeiten bestimmt. Auf diese Weise wird für jedes untersuchte Individuum seine Leistung y geschätzt. Die Standardabweichung der Leistungen der Individuenstichprobe ergibt den Schätzwert für SD_y . Damit sind die individuellen Schätzverfahren mit dem Kostenrechnungsansatz verwandt. Allerdings verwenden sie keine Daten der innerbetrieblichen Kostenrechnung, um y zu bestimmen, und sind somit wesentlich einfacher zu handhaben.

4.5.3.1 Die CREPID-Methode

Ein bekanntes individuelles Schätzverfahren für SD_y wurde von Cascio (1982, 1987) und Cascio und Ramos (1986) entwickelt. Den Anstoß für ihre Bemühungen ergab die Erkenntnis, daß Kostenrechnungsprozeduren meist sehr aufwendig und kostspielig sind, aber trotzdem keine so verlässlichen Ergebnisse erbringen, daß sie zu rechtfertigen wären. Die CREPID-Methode (Cascio-Ramos estimate of Performance in dollars) ist ein verhaltensorientiertes und vergleichsweise nutzerfreundliches Verfahren, um SD_y zu schätzen.

Cascio und Ramos (1986) unterteilen das Verfahren in zwei Phasen. Zuerst wird eine Arbeitsanalyse durchgeführt, danach folgt die Leistungsbeurteilungsphase. Die Arbeitsanalyse hat zum Ziel, die Haupttätigkeiten eines Mitarbeiters zu identifizieren. Das sind solche Tätigkeiten, aus welchen die Arbeit innerhalb eines Jahres zu mindestens 10% besteht. Die Autoren schreiben, daß eine Tätigkeit in der Regel aus 4-5 solcher Haupttätigkeiten besteht. Nachdem diese feststehen, wird jede Haupttätigkeit auf vier Dimensionen eingestuft. Die erste Dimension ist die Zeit, welche für die Tätigkeit aufgewendet wird, bzw. die Häufigkeit, mit welcher sie auftritt. Diese Einstufung erfolgt auf einer Skala von 0% bis 100%. Die anderen drei Dimensionen werden auf einer verbal verankerten Ratingskala (0 bis 7) eingeschätzt. Es handelt sich um den Schwierigkeitsgrad und die Wichtigkeit der Tätigkeit und darum, wie schwerwiegend die Konsequenzen von Fehlern sind. Die vier Werte werden multiplikativ verknüpft. Der so ermittelte Gesamtwert stellt einen Index für die Wichtigkeit der jeweiligen

Haupttätigkeit dar. Ziel dieser Prozedur ist also, die Haupttätigkeiten ausfindig zu machen und diesen relative Gewichte zuzuteilen.

Nachdem die jeweiligen Haupttätigkeiten und deren Wichtigkeit ermittelt worden sind, folgt die Leistungsbeurteilungsphase. Sie zeichnet sich dadurch aus, daß Vorgesetzte ihre Mitarbeiter einzeln danach beurteilen, wie gut sie die Haupttätigkeiten ausführen. Die Beurteilung erfolgt auf einer Skala mit den Werten von 0 bis 200. Die Skala ist so verankert, daß ein Wert von 100 einen durchschnittlichen Mitarbeiter kennzeichnen sollte, einen Wert von 200 sollte ein Mitarbeiter bekommen, der die Aufgabe besser als 99% seiner Kollegen löst. Null kennzeichnet den schlechtesten Mitarbeiter, der jemals bei einer solchen Tätigkeit erlebt wurde, und 50 einen solchen, der den Prozenrang 25 erreicht. Cascio und Ramos (1986) weisen darauf hin, daß die Ratings damit eine rechteckige Verteilung aufweisen. Die Ratingwerte werden anschließend durch 100 geteilt. Auf diese Weise wird jedem beurteilten Mitarbeiter ein Wert zwischen 0 und 2 für jede Haupttätigkeit zugeordnet.

Um die Bewertung der einzelnen Mitarbeiter in Geldeinheiten vorzunehmen, wird einem durchschnittlichen Mitarbeiter (Summe der mit den Tätigkeitsgewichten multiplizierten Ratings gleich 1) der durchschnittliche Verdienst aller Mitarbeiter zugeordnet. Die Definition der bewerteten Leistung eines Mitarbeiters durch die CREPID-Methode entspricht also dem Konzept des Verkaufswertes der Leistung w (siehe Abschnitt 4.4). Diese Vorgehensweise illustriert Tabelle 4 am Beispiel eines fiktiven Mitarbeiters.

Tabelle 4: Die Zuordnung des Verkaufswertes der Leistung (w) bei der CREPID-Methode. Dargestellt ist das Beispiel eines fiktiven Mitarbeiters (in Anlehnung an Cascio & Ramos, 1986, S. 22).

Haupttätigkeit	relatives Gewicht	durchschnittlicher jährlicher Verdienst (in \$)	Dollarwert für das Unternehmen	Leistungsrating	Verkaufswert w
1	0,29	30000	8700	1,0	8700
2	0,32	30000	9600	2,0	19200
3	0,03	30000	900	0,5	450
4	0,36	30000	10800	1,0	10800
Total					39150

Im letzten Schritt wird SD_w durch die Standardabweichung der Leistung zwischen den Mitarbeitern berechnet.

Kritisch zu beurteilen ist diese Methode deshalb, weil sie das Grundproblem der Definition der Leistung y zudeckt. Gemäß meinen Ausführungen in Abschnitt 4.4 schätzen

Cascio und Ramos den Verkaufswert w der Handlungsergebnisse eines Mitarbeiters und nicht dessen Deckungsbeitrag y . Das Verhältnis der variablen Kosten der Leistung zum Verkaufswert (V) muß also auch nach der Schätzung von SD_w beachtet werden. Die Schätzung von SD_y ist also gemäß Gleichung (18) vorzunehmen.

Edwards, Frederick und Burke (1988) beschreiben ein modifiziertes CREPID-Verfahren. Bei diesem werden statt der Leistungsbeurteilung Archivdaten herangezogen, um die individuellen Unterschiede zwischen den Beurteilten zu ermitteln. Diese archivgestützte Schätzung erbringt ähnliche Werte wie die originale CREPID-Prozedur. Wenn in einem Unternehmen ein funktionierendes Leistungsbeurteilungssystem besteht, können diese Daten gemäß dieser Studie herangezogen werden, um SD_y zu schätzen, ohne daß gegenüber der originalen CREPID-Vorgehensweise große Abweichungen zu erwarten sind.

4.5.3.2 Individuelle Leistungsschätzung bei Verkaufspersonal

Ein anderer Weg der individuellen Schätzung der Leistung y bzw. des Verkaufswertes der Leistung w wird bei Berufen möglich, deren Leistung objektiv und relativ einfach meßbar ist. Dies ist z.B. bei Verkaufspersonal der Fall. Da die Entlohnung innerhalb dieser Berufe meist durch den Umsatz mitbestimmt wird, werden die Daten zur Schätzung der Leistung zur Entlohnungsberechnung ständig miterhoben und sind somit oft leicht verfügbar.

Diesen Vorteil nutzen einige Forscher aus. Bobko et al. (1983) untersuchten Schätzmethoden von SD_y bei Versicherungsvertretern auf Grundlage der erzielten individuellen Umsatzzahlen. Da das Ziel der Forschungsarbeit von Bobko et al. (1983) aber darin besteht, Schätzungen der Standardabweichung des Verkaufsvolumens durch die GLOBAL-Methode mit den objektiven Umsatzzahlen zu vergleichen, strengen die Autoren leider keine SD_y -Schätzung durch die objektiven Daten an. Greer und Cascio (1987) untersuchten mehrere Schätzmethoden von SD_y bei Handlungsreisenden. Als Prüfstein der verschiedenen SD_y -Schätzmethoden benutzen sie ein Verfahren, das auf den objektiven Umsatzzahlen basiert. Ihre Definition der Leistung als Deckungsbeitrag zur Organisation ist konsistent mit dem Konzept des Parameters y . Sie wird dadurch ermittelt, daß vom Verkaufsumsatz die variablen Kosten, die durch den Verkauf entstehen, abgezogen werden. Aus den derart geschätzten Leistungen der Kaufleute wird die Standardabweichung zwischen ihnen ermittelt. Diese stellt einen vergleichsweise verlässlichen Schätzwert für SD_y dar. Greer und Cascio (1987) nennen ihren Ansatz selbst 'Kostenrechnungsmethode'. Nach meiner Ansicht nimmt er tatsächlich eine Mittelstellung zwischen den Kostenrechnungsverfahren und

den individuellen Verfahren ein, da die variablen Kosten ohne Unterstützung der betrieblichen Kostenrechnung abgeschätzt werden. Ein strikter Kostenrechnungsansatz würde wieder die Probleme mit sich bringen, daß viele Parameter schwer zu ermitteln sind und deshalb nur geschätzt werden können.

Das Problem der Ermittlung von SD_y durch die Methode von Greer und Cascio (1987) liegt weniger darin, den Umsatz oder den Verkaufswert der Leistung der Mitarbeiter zu ermitteln. Vielmehr verlagert es sich auf die Schätzung des Anteils der variablen Kosten an den Gesamtkosten der Leistung. Ein Ansatz zur Lösung des Problems stammt von Barthel (1989; siehe auch Barthel, 1988; bzw. der Zeitschriftenartikel von Barthel & Schuler, 1989). Er überprüfte den Nutzen eines PAP für Außendienstmitarbeiter (AM) bei einer Versicherungsgesellschaft. Zur Ermittlung der individuellen Leistung y standen ihm die Umsatzzahlen der einzelnen AM zur Verfügung. Diese waren in sogenannten 'Wertungspunkten' (WP) codiert. Dabei bedeutet ein WP eine Beitragseinnahme von 12.- DM/Jahr für die Organisation. Die Entlohnung der Mitarbeiter wird ebenfalls durch die Anzahl der WP bestimmt. Wie hoch diese im Vergleich zu den WP liegt, bleibt bei Barthel (1989) allerdings etwas unklar. Ich nehme an, daß Barthel (1989, S. 204) mit "... beträgt die Provision etwa das 6-7fache der Wertungspunkte" meint, daß pro Wertungspunkt 6-7 DM einmalige Provision an den AM ausbezahlt werden.

Sehr klar drückt sich der Autor jedoch aus, wenn er eine Leistungsschätzung auf der Basis der Entlohnungshöhe ablehnt. Diese Konsequenz ist bislang in der Literatur zur NA kaum zu finden. Dies möchte ich damit unterstreichen, daß sowohl die GLOBAL-Methode (Schmidt et al., 1979), die Prozentregeln (Hunter & Schmidt, 1982), als auch das CREPID-Verfahren (Cascio & Ramos, 1986) ihre Schätzungen an die Höhe der Entlohnung anlehnen. Die Auffassung von Barthel stellt also eine wirkliche Neuerung dar.

Barthel (1989) verfolgt eine andere Logik, als die bisher beschriebenen Verfahren. Auch er beklagt das Problem, daß der Beitrag eines Mitarbeiters zur Organisation¹⁴ nicht eindeutig zu berechnen ist. Stattdessen legt er einen Teil der Beitragseinnahmen als Beitrag eines AM zur Organisation fest. Dieser Teil ist 1,-DM je WP und Jahr, also 1/12 der Beitragseinnahmen, die das Unternehmen pro WP und Jahr erhält. Wie der Autor zu dieser willkürlichen Setzung kommt, umschreibt er leider sehr vage: "Verschiedene Betrachtungen führten letztlich zur Festlegung." (Barthel, 1989, S. 204-205). Eine der 'Betrachtungen' war

¹⁴ Einer der Schwachpunkte der Arbeit von Barthel (1989) liegt meines Erachtens in der begrifflichen Unschärfe. So wird z.B. der Beitrag eines Mitarbeiters zum Unternehmen als 'Nutzen' bezeichnet. Der selbe Begriff findet auch Verwendung für den gesteigerten Ertrag durch die Intervention abzüglich ihrer Kosten. Aus diesem Grunde werden an dieser Stelle die vermuteten intendierten Bedeutungen bei Barthel in meiner Terminologie ausgedrückt.

die Überlegung, daß der Anteil des Beitrags eines Mitarbeiters für das Unternehmen an seinem Umsatz grob durch die Umsatzrendite geschätzt werden kann. Barthel (1989) kommt zu dem Schluß, daß die erwartete Umsatzrendite 8% beträgt. Der Anteil der Leistung 1.- DM am Umsatz von 12.- DM/ WP entspricht ungefähr diesen 8 %. Prinzipiell ist auch diese Schätzmethode nicht ganz befriedigend. Die Aufschlüsselung der variablen Kosten, die durch den Verkauf eines Versicherungsvertrags entstehen, und die Schätzung ihres prozentualen Anteils am Umsatz, wären hier sicher eleganter. Leider können solche wichtigen Forschungsvorhaben aber oft nicht in die Tat umgesetzt werden, da sie in konflikthaftem Zusammenhang zu den Unternehmensinteressen stehen. Der Ansatz von Barthel wird auch bei meiner Nutzenschätzung der Personalauswahl in einem Versicherungsunternehmen eine Rolle spielen.

4.6 Fazit

Im vierten Kapitel habe ich die Anstrengungen beschrieben, die unternommen wurden, um die Parameter des BCG-Nutzenmodells zu schätzen. Das Modell ist im Vergleich zur Realität bestimmt sehr einfach und berücksichtigt deshalb nur wenige Parameter, die auf den Nutzen von Interventionsmaßnahmen eine Wirkung haben. Trotz dieser relativen Einfachheit bereitet die Schätzung der Parameter große Schwierigkeiten. Insbesondere die Schätzung von SD_y ist problematisch.

Ich habe mich bemüht, nicht nur die praktischen Maßnahmen zur Parameterschätzung, die bisher vorgeschlagen wurden, darzustellen, sondern auch die Konzepte theoretisch zu durchdringen. Mein besonderes Augenmerk lag dabei auf der Leistung y , die ich als Beitrag eines Individuums zur Organisation definiert habe. Wird die Leistung y in Geldwerteinheiten ausgedrückt, dann kann dieser Beitrag mit dem ökonomischen Konzept des Deckungsbeitrages der Leistung gleichgesetzt werden. In der Literatur bestehen bezüglich des Parameters y sehr unterschiedliche Auffassungen. Oftmals habe ich den Eindruck, daß die Verschiedenheit der Begriffe gar nicht bemerkt wird. Problematisiert wurde dies jedenfalls selten.

5 Weiterentwicklung des BCG-Nutzenmodells

Bei der kritischen Begutachtung des BCG-Nutzenmodells kann man sich fragen, ob das Modell bei der Bestimmung des Nutzens von Interventionsmaßnahmen nicht auch weitere Faktoren berücksichtigen sollte, um der Realität gerecht zu werden. Da die Frage berechtigt ist, wurden solche Modellerweiterungen von einigen Forschern vorgeschlagen. Ich werde diese Vorschläge in diesem fünften Kapitel besprechen.

Wenn der Nutzen von Interventionsprogrammen in Geldwerteinheiten bestimmt werden soll, so ist es natürlich wünschenswert, wenn diese Erweiterungen direkt in das formale BCG-Modell in Form weiterer Parameter aufgenommen werden. Diese Möglichkeit wurde insbesondere von Boudreau (1983a, 1983b) ausgenutzt. Andere Vorschläge gehen jedoch nicht diesen Weg. Vielmehr geben sie zu bedenken, daß wichtige Sachverhalte bei der Nutzerschätzung bedacht werden müssen, ohne daß sie formal in das Modell aufgenommen werden. Dadurch wird die Quantifizierung des Nutzens der Intervention natürlich erschwert. Trotzdem haben auch diese Erweiterungen ihre Bedeutung. Sie finden deshalb in diesem fünften Kapitel ebenso ihren Platz, wie formalisierte Erweiterungsvorschläge.

So wünschenswert solche Erweiterungen des Modells im Sinne einer Annäherung an die Wirklichkeit sind, so sehr steht ihnen eine Überlegung entgegen: Wenn es schon schwierig ist, ein einfaches Nutzenmodell so zu operationalisieren, daß seine Parameterschätzungen und damit sein Ergebnis glaubhaft gemacht werden können, dann sollte man das Modell nicht noch komplizierter und damit sein Ergebnis noch fragwürdiger machen. Im Sinne von Cronbach und Gleser könnte man auch argumentieren, daß es bei einem Nutzenmodell nicht darauf ankommt, daß es die Realität mit aller seiner Vielfalt möglichst detailgetreu nachzeichnet, sondern daß zwischen möglichen Alternativen richtige Entscheidungen getroffen werden. In diesem Sinne ist sicherlich Zurückhaltung geboten. In ein Nutzenmodell, das auf eine Frage der Praxis eine Antwort geben soll, sollten nur diejenigen Parameter aufgenommen werden, die bei der Antwortfindung unumgänglich sind.

5.1 Formalisierte Erweiterungen des BCG-Modells durch ökonomische Konzepte

In diesem Abschnitt werde ich formalisierte Erweiterungen des BCG-Nutzenmodells vorstellen und diskutieren. Diese basieren insbesondere auf ökonomischen Überlegungen. Konzepte der Investitionsrechnung haben hierbei in besonderem Maße Eingang in bestehende Nutzermodelle gefunden. Eine der Hauptfunktionen der NA besteht darin, die Bewertung psychologischer Interventionsmaßnahmen den Entscheidungsträgern in Organisationen verständlich darzustellen. Da diese Entscheidungsträger meist wirtschaftswissenschaftlich ausgebildet sind, ist die Entwicklung, ökonomische Konzepte in Nutzenmodelle aufzunehmen, im allgemeinen sehr zu begrüßen. Darüber hinaus werden oft auch andere, nichtpsychologische Interventionsmaßnahmen auf der Grundlage ökonomischer Konzepte bewertet. Wenn psychologische Interventionsmaßnahmen durch nutzenanalytische Modelle eine ähnliche oder gar gleiche Bewertung erfahren sollen, so kann die Vergleichbarkeit mit konkurrierenden Vorschlägen anderer Disziplinen einfacher hergestellt werden. Bevor ich jedoch zur Diskussion dieser Ansätze komme, möchte ich kurz wiederholen, wie weit ich bei der Darstellung der Entwicklung der Nutzenmodelle gekommen war.

Gleichung (16) stellt das grundlegende BCG-Nutzenmodell dar, wie es nach der Erweiterung durch Schmidt et al. (1979, Wirkdauer T) von vielen Autoren bei der Analyse des Nutzens von PAP Verwendung fand.

$$(16) \quad \Delta U = Na T r_{xy} \bar{z}_x^* SD_y - C$$

ΔU bezeichnet hierbei den inkrementellen Nutzen in Geldwerteinheiten, der durch die Behandlung 'Aufnahme in die Organisation' einer Kohorte der Bewerber entsteht. Na ist deren Anzahl, die Größe T die Wirkdauer der Interventionsmaßnahme. Diese ist bei der Personalauswahl normalerweise der durchschnittliche Verbleib der Ausgewählten in der Organisation. Die Validität des Prädiktors x bzgl. der Leistung y ist mit r_{xy} , bezeichnet, SD_y ist die Standardabweichung des in Geldeinheiten bewerteten Organisationsbeitrages der Individuen der Bewerberpopulation und C sind die Gesamtkosten des Programms. Mit \bar{z}_x^* ist der durchschnittliche standardisierte Prädiktorwert der ausgewählten Bewerber bezeichnet. Unter Annahme der normalen Verteilung der Prädiktorwerte und der Selektion der Bewerber 'von oben nach unten', läßt sich \bar{z}_x^* mittels Gleichung (3) schätzen.

$$(3) \quad \bar{z}_x^* = \lambda_{(Q)}/Q$$

Q ist die Selektionsquote und $\lambda_{(Q)}$ der Funktionswert der Normalverteilungsfunktion an der Stelle Q. Die Definitionsprobleme der Leistung veranlaßten mich zu der Unterscheidung zwischen dem 'Deckungsbeitrag durch die Leistung' (y) und dem 'Verkaufswert der Leistung'

(w). Da einige Schätzmethoden der Standardabweichung der Leistung diese Verkaufswertdefinition zugrundelegen, schlägt Boudreau (1983a) vor, den Anteil der variablen Kosten am Verkaufswert der Leistung zu ermitteln und den Anteil der Standardabweichung, der durch die variablen Kosten verursacht wird, vom Verkaufswert der Leistung abzuziehen.

$$(15) \quad V = s/w$$

$$(18) \quad SD_y = SD_w (1-V)$$

Wird die Gleichung (18) in (16) eingesetzt, so ergibt sich

$$(17) \quad \Delta U = N_a T r_{xw} \bar{z}_x * SD_w (1-V) - C$$

Dieses Modell stellt die Ausgangssituation für die Erweiterungen dar, die im weiteren Verlauf beschrieben und diskutiert werden. Ich werde mich bei der Darstellung dieser Entwicklung nicht an den chronologischen Verlauf der Vorschläge halten. Vielmehr will ich die Ordnung der Darstellung so aufbauen, wie sie mir am einfachsten verständlich erscheint und didaktisch sinnvoll ist.

5.1.1 Intervention als Investition

Obwohl die NA schon eine längere Geschichte aufweisen kann, wurde ihr sowohl in der akademischen Ausbildung als auch von Praktikern vergleichsweise wenig Beachtung geschenkt. Mitverantwortlich für diesen Umstand ist sicherlich, daß die Sprache der Nutzenmodelle weder zum 'mainstream' der Psychologie noch zu dem der Wirtschaftswissenschaften paßt. Daß die schlechte Passung zu den Wirtschaftswissenschaften aber eher eine Folge der Ausdrucksweise und weniger der konzeptionellen Denkweise ist, stellen Cronshaw und Alexander (1985) fest. Ihr Aufsatz verfolgt das Ziel, die strukturelle Ähnlichkeit von Nutzenmodellen der Personalauswahl zu vorhandenen wirtschaftswissenschaftlichen Modellen der Investitionsrechnung aufzuzeigen und die Nutzenmodelle diesen weiter anzupassen. Diese Anstrengung begründen die Autoren unter anderem damit, daß die Vergleichbarkeit der Bewertung von psychologischen Interventionsmaßnahmen mit anderen Interventionsmaßnahmen gegeben sein muß, um rationale Entscheidungen zu rechtfertigen. Cronshaw und Alexander (1985) sehen die Entscheidungen über Interventionen aus der Investitionsperspektive. Sie beklagen die Sichtweise vieler Personalabteilungen, die ihre Kosten als Gemeinkosten betrachten. Statt dessen vertreten sie vehement den Standpunkt, die Ausgaben der Personalabteilung als Investitionen zu betrachten. Ein Gesinnungswandel in diese Richtung hätte verschiedene

Konsequenzen. So wird verständlicherweise in konjunkturell schlechten Zeiten Kostensenkung gefordert. Die Ausgaben der Personalabteilung werden in einem solchen Fall Zielpunkt der Kostensenkungsforderung sein, sofern sie als Gemeinkosten angesehen werden. Betrachtet man diese Ausgaben jedoch als normale Investitionen, d.h. als solche Ausgaben, die den Marktwert der Organisation und damit ihre Überlebensfähigkeit erhöhen sollen, verändert sich die Sachlage vollkommen. Die Investitionsvorschläge der Personalabteilung müssen in einem solchen Fall mit denen anderer Abteilungen konkurrieren. Dies ist nur möglich, wenn eine sinnvolle Vergleichbarkeit der Wertmaßstäbe aller Investitionsabsichten gegeben sind.

Eine Methode der Investitionsrechnung stellt die Kapitalplanungsrechnung dar (Wöbe, 1993). Sowohl die Ansätze der Kapitalplanungsrechnung als auch die Nutzenmodelle haben folgende Struktur:

$$(12) \quad \Delta U = E - C$$

In Worten ausgedrückt bedeutet dieser Zusammenhang, daß der Nutzen einer Investition ihr Ertrag gemindert um ihre Kosten ist. Trotz dieser strukturellen Ähnlichkeit vernachlässigen Nutzenmodelle im Gegensatz zu den Methoden der Investitionsrechnung wichtige ökonomische Konzepte. Die Trennung der festen von den variablen Kosten der Intervention wurde bereits in Kapitel 4 vorgestellt, als die Operationalisierung des Parameters C besprochen wurde (Abschnitt 4.3.1).

$$\Delta U = E - C_1 - C_0 \quad (20)$$

Damit wird dem Umstand Rechnung getragen, daß PAP z.B. meist einmalige Entwicklungskosten haben. Diese Kosten sind ausbringungsmengenunabhängig, und somit feste Kosten (C_0). Darüber hinaus verursacht aber die Anwendung des PAP auch variable Kosten (C_1), die mit seiner Anwendungsintensität in Zusammenhang stehen. Diese Kosten können z.B. in DM je Bewerber ausgedrückt werden.

Weitere ökonomische Konzepte, die bei der Investitionsrechnung regelmäßig Anwendung finden, sind z.B. die Kapitalverzinsung, das Projektrisiko, etc. Um verschiedene Interventionsmaßnahmen, die in der Ökonomie oft auch 'Projekte' genannt werden, zu vergleichen, haben sich innerhalb der Investitionsrechnung verschiedene Indizes etabliert. Diese dienen der Unterstützung der Entscheidungsfindung, welche Projekte durchgeführt werden sollten und welche nicht. Cronshaw und Alexander nennen einige dieser Indizes, von welchen ich die wichtigsten hier vorstelle.

Die Amortisationsdauer (AD, engl.: pay-back period) ist die Zeitspanne, in der investiertes Kapital durch den Ertrag der Maßnahme zurückgeflossen ist. E_i ist dabei der Ertrag der Maßnahme pro Jahr, $(C_1 + C_0)$ sind die Gesamtkosten der Maßnahme.

$$AD = (C_1 + C_0) / E_j \quad (21)$$

Bei der Betrachtung dieser Denkweise wird unmittelbar deutlich, daß die Zeitdimension bei nutzenanalytischen Modellen weitgehend nicht beachtet wurde. Dies zeigt sich noch klarer, wenn bedacht wird, daß gebundenes Kapital verzinst werden muß. Dieser Verzinsung trägt der Index AD allerdings auch keine Rechnung.

Der Barwert (net present value) ist der auf den Investitionszeitpunkt bezogene Wert des investierten Kapitals und der Rendite, also dem Kapital, welches später in die Organisation zurückfließt und der Investition zuzuschreiben ist. Unter der Annahme, daß das Kapital einer Verzinsung unterliegt, ist der Barwert von zeitlich später zurückfließendem Kapital und zeitlich früher investiertem Kapital höher, als die fließende Geldmenge. Unter den gegenteiligen Bedingungen ist der Barwert niedriger als die fließende Geldmenge. Gleichung (22) zeigt, wie sich der Barwert des Ertrages unter Beachtung der Verzinsung ändert. Der betrachtete Zeitraum wird hierzu in mehrere diskrete Wirkperioden aufgeteilt.

$$\Delta U = \sum_{t=1}^T (E_t / (1+i)^t) - C_1 - C_0 \quad (22)$$

t bezeichnet dabei eine Wirkperiode der Behandlung. Im Regelfall wird eine solche Wirkperiode aus Vereinfachungsgründen auf ein Jahr festgesetzt. Die Gesamtzeit der Behandlungswirkung ist mit T gekennzeichnet. E_t ist der Ertrag in der Wirkperiode t , der Kalkulationszinsfuß heißt i . Durch die Verzinsung der zukünftigen Erträge in t verringert sich deren Barwert. Nicht verzinst sind in diesem Modell die variablen und festen Kosten der Maßnahme, d.h. es wird angenommen, daß diese zum Investitionszeitpunkt fällig sind. Wären z.B. die variablen Kosten C_1 erst in späteren Wirkperioden t fällig, so würde sich der Barwert der Kosten ebenfalls verringern. An dieser Stelle soll auf die Verzinsung nicht weiter eingegangen werden, da sie in den nächsten Abschnitten noch näher beschrieben wird.

Hunter et al. (1988) kritisieren scharf die Auffassung von Personalauswahl als Investition. Meines Erachtens sind ihre Argumente jedoch nicht stichhaltig genug, um auf die großen Vorteile zu verzichten, welche die Investitionsperspektive mit sich bringt. In einer Antwort auf den Aufsatz von Hunter et al. (1988) verteidigen Cascio und Morris (1990) sowie Cronshaw und Alexander (1991) den Investitionsansatz. Sie begründen ihren Standpunkt hauptsächlich damit, daß durch die Beachtung der Prinzipien der Investitionsrechnung

Investitionen im Personalbereich grundsätzlich mit Investitionen in anderen Bereichen vergleichbar sind.

5.1.2 Variable Kosten der Leistung, Zinsen und Steuern

Die Aufnahme der ökonomischen Konzepte 'variable Kosten der Leistung', 'Verzinsung' und 'Besteuerung' schlägt Boudreau (1983a) vor. Die Gründe für die Beachtung der variablen Kosten der Leistung wurde bereits in Abschnitt 4.4.2 ausführlich diskutiert, als die Definitionsprobleme der Leistung y zur Debatte standen. Boudreau löst das Problem dadurch, daß er die Standardabweichung des Verkaufswertes der Leistung (SD_w) mittels seines Parameters V korrigiert. V ist in Gleichung (15) definiert und ist der Anteil der variablen Kosten der Leistung an ihrem Verkaufswert. Dadurch ergibt sich das Nutzenmodell aus Gleichung (17).

$$(17) \quad \Delta U = N_a T r_{xw} \bar{z}_x * SD_w (1-V) - C.$$

Um eine weitere Übereinstimmung des Nutzenmodells mit der Realität zu erreichen, nimmt Boudreau das Konzept des Barwertes in sein Modell auf. Er berücksichtigt durch diese Maßnahme, daß, wie bei allen Investitionen, eingesetztes Kapital verzinst werden muß. Das Konzept der Gesamtzeit der Behandlungswirkung T , wie es von Schmidt et al. (1979) verwendet wurde, wird in einzelne Wirkperioden t aufgespalten. Aus Einfachheitsgründen wird die Dauer der Wirkperioden meist auf ein Jahr festgelegt. Das Nutzenmodell erweitert sich mit Aufnahme der Verzinsung zu Gleichung (23).

$$\Delta U = \sum_{t=1}^T N_a [r_{xwt} \bar{z}_x * SD_{wt} (1-V_t) (1 / (1+i)^t)] - C \quad (23)$$

t bezeichnet die Wirkperiode der Behandlung. Die Gesamtzeit der Behandlungswirkung ist T , der Kalkulationszinsfuß heißt i . C bezeichnet die Gesamtkosten der Interventionsmaßnahme. Wie aus der Gleichung (23) erkennbar ist, bietet sich die Möglichkeit, für verschiedene Wirkperioden t unterschiedliche Parameter r_{xwt} , SD_{wt} oder V_t anzugeben. Davon Gebrauch zu machen, wird sicherlich selten möglich sein, da die Parameter ohnehin schon mit einiger Unsicherheit geschätzt werden müssen. Es gibt allerdings auch sinnvolle Einsatzgebiete dieser Möglichkeit. Man denke z.B. an den Fall, bei welchem ein diagnostisches Verfahren nur langfristige zuverlässige Leistungsvorhersagen machen kann. Es ist z.B. denkbar, daß ein biographischer Fragebogen bzgl. der Leistung erst ab zwei Jahre nach Berufseintritt valide ist.

In diesem Fall könnte r_{xw1} , und r_{xw2} auf 0 geschätzt werden und r_{xw3} bis r_{xwT} auf einen anderen Wert.

Bei diesem Modell zeigt sich das Grunddilemma der Ausdifferenzierung der Nutzenmodelle zum erstenmal deutlich: So lange das Problem der sehr ungenauen Parameterschätzung gegeben ist, muß genau bedacht werden, unter welchen Bedingungen es sinnvoll ist, Modellerweiterungen vorzunehmen.

Eine weitere Annäherung des Nutzenmodells an die Realität verspricht sich Boudreau von der Aufnahme des Steuersatzes in das Modell. Steuern vermindern die Erträge der Intervention, andererseits nehmen durch Steuereinsparungen auch die Kosten der Maßnahme ab. Boudreau berücksichtigt dies, indem sowohl die Erträge, als auch die Kosten proportional zum Steuersatz verringert werden. Inwieweit eine solche Behandlung des Problems angemessen ist, soll dem Steuerfachmann überlassen und hier nicht weiter hinterfragt werden. Gleichung (23) erweitert sich zu (24).

$$\Delta U = N_a \sum_{t=1}^T r_{xwt} \bar{z}_x * SD_{wt} (1-V_t) (1/(1+i))^t (1-TAX_t) - C (1-TAX_t) \quad (24)$$

Die Variable TAX_t bezeichnet hierbei, in Anlehnung an den englischen Ausdruck, den Steuersatz in der jeweiligen Wirkperiode t .

5.1.3 Das Mitarbeiterflußmodell

Im Regelfall wird ein PAP nicht nur bei einer Bewerberkohorte angewandt, wie dies z.B. die Modelle der Gleichungen (16), (17), (23) und (24) annehmen. Diese Modelle 'tun so', als ob aus den Bewerbern einmal Mitarbeiter ausgewählt würden, und diese Kohorte dann eine bestimmte Zeitdauer der Organisation angehören würden. Diese Zeitdauer wird durch die Wirkdauer T berücksichtigt. Die Realität ist jedoch komplexer. Boudreau (1983b) weist richtigerweise darauf hin, daß ein PAP über einen bestimmten Zeitraum zur Anwendung kommt und mehrere Bewerberkohorten behandelt.

Ich möchte dies an einem Beispiel deutlich machen. Ein PAP eines Unternehmens - z.B. eine bestimmte Kombination aus Würdigung der Bewerbungsunterlagen, eines Interviews und eines biographischen Fragebogens- wird konstruiert. Nach der Konstruktion kommt es ohne große Veränderungen 10 Jahre zur Anwendung. Es wird zunächst angenommen, daß die Mitarbeiter exakt 5 Jahre im Unternehmen verbleiben. Die Wirkung des PAP beginnt mit der Auswahl des ersten Mitarbeiters, der durch das PAP behandelt (akzeptiert oder abgelehnt) wird. Im Laufe der Zeit ändert sich jedoch die Anzahl der

Mitarbeiter im Unternehmen, welche durch das PAP behandelt wurden. Sie nimmt in den ersten 5 Jahren ständig zu, da immer mehr Mitarbeiter ausgewählt werden. Nach 5 Jahren jedoch verlassen die ersten Mitarbeiter, die 5 Jahre zuvor eingestellt wurden, das Unternehmen. Da trotz der Fluktuation neue Mitarbeiter in das Unternehmen aufgenommen werden, bleibt die Anzahl der Mitarbeiter im Unternehmen, welche durch das PAP behandelt wurden, etwa gleich. Nach 10 Jahren wird das PAP abgesetzt. Dennoch wirkt es weiter, denn viele Mitarbeiter sind noch im Unternehmen, die früher mit dem Programm ausgewählt wurden. Diese Anzahl sinkt allerdings von da an stetig ab. Nach 15 Jahren endet die Programmwirkung, da alle PAP-behandelten Mitarbeiter das Unternehmen verlassen haben.

Diese Programmwirkungsdauer bezeichne ich mit F . Um den Mitarbeiterfluß zu erfassen, wird F in mehrere gleich große Zeiträume unterteilt. Diese Zeiträume werden aus Vereinfachungsgründen oft auf die Länge von einem Jahr festgelegt. Jede andere Länge ist natürlich theoretisch denkbar. Ich nenne diese Zeiträume der Programmwirkungsdauer die Behandlungsperioden k . Im obigen Beispiel gibt es die Behandlungsperioden $k = 1, 2, \dots, 15$.

Um die Anzahl der Mitarbeiter zu berechnen, die in einer Behandlungsperiode k im Unternehmen sind und durch das PAP behandelt wurden, können die Anzahl der akzeptierten Bewerber einer jeden Wirkperiode t bis zur Behandlungsperiode k , abzüglich der ausgeschiedenen Mitarbeiter der jeweiligen Wirkperioden, aufsummiert werden.

$$N_k = \sum_{t=1}^k (N_{at} - N_{zt}) \quad (25)$$

N_k ist die Anzahl der durch das PAP behandelten Mitarbeiter der Organisation in der Behandlungsperiode k . Die Anzahl der akzeptierten Bewerber in der Wirkperiode t ist mit N_{at} bezeichnet, die Anzahl der behandelten ausgeschiedenen Mitarbeiter in t mit N_{zt} . Indem die Maßzahl N_k mit $k = 1, 2, \dots, F$ in die Gleichung des BCG-Nutzenmodells aufgenommen wird, beinhaltet diese implizit die Wirkdauer T der Intervention durch das PAP.

Die Bestimmung von N_k will ich kurz illustrieren. Angenommen man will die Anzahl der Mitarbeiter im Unternehmen in der Behandlungsperiode 3, also im 3. Jahr nach Beginn des PAP, berechnen. Dann ist dies (aufgrund der groben Auflösung von einem Jahr natürlich nur näherungsweise) möglich, indem wir für jedes Jahr bis zu dieser Behandlungsperiode (entspricht den Wirkperioden $t=1, 2, 3$) die Differenz der durch das PAP akzeptierten Bewerber und der zuvor durch das PAP behandelten ausgeschiedenen Mitarbeiter berechnen. Im ersten Jahr war $N_{a1}=50$ und $N_{z1}=0$. Im zweiten Jahr wurden weitere $N_{a2}=70$ Mitarbeiter neu ausgewählt, während $N_{z2}=20$ aus dem ersten Jahr wieder ausschieden. In der 3. Wirkperiode, der aktuellen Behandlungsperiode k , war $N_{a3}=60$ und $N_{z3}=50$. N_k berechnet sich nach

Gleichung 25 dann zu 110 Mitarbeitern, die zum fraglichen Zeitpunkt im Unternehmen sind und welche mittels des PAP ausgewählt worden sind. Durch diese Mitarbeiter, und nur durch diese, kann das PAP einen Ertrag für das Unternehmen erbringen.

Boudreau (1983b) integriert die Logik der Behandlungsperioden in sein Nutzenmodell (24), in welchem die ökonomischen Konzepte 'variable Leistungskosten', 'Verzinsung' und 'Steuern' (Boudreau, 1983a) berücksichtigt sind. Dadurch ergibt sich Gleichung (26).

$$\Delta U = \sum_{k=1}^F [N_k r_{xwt} \bar{z}_x^* SD_w (1-V) (1/(1+i)^k (1-TAX))] - \sum_{k=1}^F [C_k (1/(1+i)^{k-1} (1-TAX))] \quad (26)$$

Die Struktur der Gleichung (26) ist vergleichsweise einfach, wenn man sich klar macht, daß in der oberen Zeile der Gleichung die Erträge des PAP getrennt für die Behandlungsperioden k der Maßnahme stehen und in der unteren Zeile die Gesamtkosten der Maßnahme je Behandlungsperiode k . Diese Kosten werden wie gewohnt von den Erträgen subtrahiert.

Sowohl die Erträge als auch die Kosten werden verzinst und versteuert. Die Verzinsung erfolgt mit dem Faktor $(1/(1+i)^{k-1})$, da angenommen wird, daß die Kosten jeweils am Anfang der Periode k anfallen, während sich der Ertrag zeitverzögert erst am Ende von k einstellt. N_k ist die Anzahl der Mitarbeiter einer jeden Behandlungsperiode k , die durch das PAP ausgewählt wurden und während der Behandlungsperiode in der Organisation sind. Im Gegensatz zum früheren Modell von Boudreau (1983a, Gleichung 24) sieht dieses Modell keine verschiedenen Werte von r_{xw} , SD_w , V und TAX für die verschiedenen Behandlungsperioden k vor. Ich halte dies auch für sinnvoll, wie bei meiner Kritik des Modells (24) bereits deutlich wurde.

Ich möchte dieses Modell durch die Übernahme eines Beispiels aus Boudreau (1983b) verdeutlichen. Das Beispiel entstammt ursprünglich der Untersuchung von Schmidt et al. (1979) und wurde so modifiziert, daß sowohl die ökonomischen Konzepte V , i und TAX , als auch das Mitarbeiterflußmodell darin integriert sind. Schmidt et al. untersuchten den Nutzen der Personalauswahl von Programmierern durch den Programmer Aptitude Test (PAT). Jedes Jahr wurden dabei durchschnittlich 618 Programmierer ausgewählt, welche im Schnitt ca. 10 Jahre im Unternehmen blieben. Die Validität wird mit $r_{xy} = 0,76$, SD_y mit \$ 10413 und die Kosten der Auswahl je Bewerber mit \$ 10 angegeben. Bei einer Selektionsquote von $Q=50\%$ und unter Anwendung von Gleichung (3) und ihrer Annahmen, ergibt sich ein $\bar{z}_x^* = 0,8$. Die

alternative Auswahlprozedur ist die Zufallsauswahl, so daß der inkrementelle Nutzen ΔU der Maßnahme gegenüber dieser Zufallsauswahl angegeben wird. Inwieweit die Höhe der Maßzahlen von Schmidt et al. realistisch sind, soll hier nicht diskutiert werden. Dennoch stehe ich ihnen z.T. äußerst kritisch gegenüber.

Boudreau (1983a, 1983b) schätzt den Anteil der variablen Kosten an den Verkaufskosten auf $V=0,05$ und den Steuersatz auf $TAX=0,45$. Der Kalkulationszinsfuß wird mit $i=0,1$ angenommen. Der Gesamtnutzen des PAP wird bei einem Einsatz von 15 Jahren geschätzt. In Tabelle 5 sind die Zwischenergebnisse und der Nutzen des Programms für jede Behandlungsperiode aufgeführt.

In der ersten Spalte ist dabei die Behandlungsperiode k aufgeführt. Die zweite Spalte zeigt die Anzahl der Ausgewählten im Unternehmen N_k . Die dritte Spalte enthält den nichtverzinsten Ertrag der Maßnahme in der jeweiligen Behandlungsperiode k (E'_k , in Mio Dollar).

$$E'_k = N_k r_{xw} \bar{z}_x * SD_w (1-V) (1-TAX) \quad (27)$$

In der vierten Spalte ist der sog. Diskontierungsfaktor des Ertrages in k angegeben. Dies ist der Faktor, um den der Ertrag aufgrund der Kapitalverzinsung gemindert wird. Die fünfte Spalte zeigt die verzinsten Erträge. Dies sind die unverzinsten Erträge multipliziert mit dem Diskontierungsfaktor. In ähnlicher Weise stehen in Spalte sechs bis acht die unverzinsten Kosten, der Diskontierungsfaktor für die Kosten und die verzinsten Kosten. Die letzte Spalte gibt den Nutzen in der jeweiligen Behandlungsperiode k an. Der Nutzen ist definiert als der Ertrag weniger die Kosten. Beachtenswert sind insbesondere die enormen Wirkungen, welche durch die Verzinsung verursacht werden.

Tabelle 5: Illustration des Mitarbeiterflußmodells. Erklärungen der Abkürzungen im Text (in Anlehnung an Boudreau, 1983b). Gekennzeichnete Angaben in Mio \$ (10^6 \$).

k	N_k	E'_k	$1/(1+i)^k$	$D(E'_k)$	$C_k(1-Tax)$	$1/(1+i)^{(k-1)}$	$D(C_k)$	ΔU_k
1	618	10^6 \$ 2,04	0,91	10^6 \$ 1,86	\$ 6798	1,00	\$ 6798	10^6 \$ 1,85
2	1236	4,09	0,83	3,39	6798	0,91	6186	3,38
3	1854	6,13	0,75	4,60	6798	0,83	5642	4,59
4	2472	8,18	0,68	5,56	6798	0,75	5099	5,55
5	3090	10,22	0,62	6,34	6798	0,68	4623	6,33
6	3706	12,27	0,56	6,87	6798	0,62	4215	6,87
7	4326	14,31	0,51	7,30	6798	0,56	3807	7,30
8	4944	16,35	0,47	7,68	6798	0,51	3467	7,68
9	5562	18,40	0,42	7,73	6798	0,47	3795	7,73
10	6180	20,44	0,39	7,97	6798	0,42	2855	7,97
11	6180	20,44	0,35	7,15	6798	0,39	2651	7,15
12	6180	20,44	0,32	6,54	6798	0,35	2379	6,54
13	6180	20,44	0,29	5,93	6798	0,32	2175	5,93
14	6180	20,44	0,26	5,31	6798	0,29	1971	5,31
15	6180	20,44	0,24	4,91	6798	0,26	1767	4,91
16	5562	18,40	0,22	4,05	0	0,24	0	4,05
17	4944	16,35	0,20	3,27	0	0,22	0	3,27
18	4326	14,31	0,18	5,58	0	0,20	0	2,58
19	3708	12,27	0,16	1,96	0	0,18	0	1,96
20	3090	10,22	0,15	1,53	0	0,16	0	1,53
21	2472	8,18	0,14	1,14	0	0,15	0	1,14
22	1854	6,13	0,12	0,74	0	0,14	0	0,74
23	1236	4,09	0,11	0,45	0	0,12	0	0,45
24	618	2,04	0,10	0,20	0	0,11	0	0,20
25	0	0	0,09	0	0	0,10	0	0

Der Gesamtnutzen des Programmes ist \$ 105,01 * 10^6

Nach diesem ausführlichen Anwendungsbeispiel des erweiterten Nutzenmodells von Boudreau (1983a, 1983b) möchte ich zu dessen Kritik kommen. Insgesamt gesehen überzeugt das Modell durch seine Klarheit und innere Logik. Nachdem allerdings in Abschnitt 5.1.1 der Ansatz von Cronshaw und Alexander (1985) beschrieben wurde, innerhalb dessen Interventionsmaßnahmen als Investitionsentscheidungen behandelt werden, erscheint es als nachteilig, daß die Programmkosten bei Boudreau nicht in fixe und variable Kosten getrennt

werden. Insbesondere das obige Beispiel in Tabelle 5 zeigt deutlich die enormen Auswirkungen, welche die Verzinsung des investierten Kapitals hat. Gerade aber diagnostische Instrumente, wie z.B. AC, biographische Fragebögen und computergestützte Verfahren, welche innerhalb moderner PAP immer häufiger zum Einsatz kommen, haben oft sehr hohe einmalige Entwicklungskosten. Da dieses Kapital noch vor der ersten Behandlungsperiode investiert werden muß, wirkt sich die Verzinsung nachteilig aus.

Durch die Aufnahme der ökonomischen Konzepte in das BCG-Nutzenmodell und der Berücksichtigung des Mitarbeiterflusses, ist dieses zu einem unangenehm langen algebraischen Ausdruck angewachsen. Im nächsten Abschnitt (5.1.4) erfolgt noch eine kleine Erweiterung des Modells. Diese verkürzt zwar den Ausdruck nicht, hat aber den Vorteil, daß ich auf ihrer Grundlage versuchen werde, die Sichtweisen von Boudreau (1983a, 1983b) und Cronshaw und Alexander (1985) zu vereinen. Dadurch entsteht ein formales Nutzenmodell, in welchem alle bisher beschriebenen Erweiterungsvorschläge des BCG-Modells vereint sind. Ich werde dann dieses allgemeine Modell in seiner Struktur noch einmal verdeutlichen. Die Struktur ist bei weitem nicht so kompliziert, wie das Modell zunächst aussehen mag.

5.1.4 Integration von festen und variablen Investitionskosten

Die Behandlung der Kosten von Interventionsmaßnahmen im Personalbereich als Investitionskosten propagieren Cronshaw und Alexander (1985). Wie die Autoren behaupten, sind Nutzenmodelle strukturell mit Modellen der Investitionsrechnung gleich. Das Modell (20) erfüllt diese Forderung.

$$(20) \quad \Delta U = E - C_1 - C_0$$

Der inkrementelle Nutzen der Interventionsmaßnahme gegenüber einer alternativen Maßnahme ΔU berechnet sich nach dieser Gleichung als der Ertrag der Maßnahme in DM (E) abzüglich der variablen (ausbringungsmengenabhängigen) Kosten der Maßnahme (C_1) und der (ausbringungsmengenunabhängigen) festen Kosten der Maßnahme (C_0). Da ich mich hier schwerpunktmäßig mit PAP beschäftige, soll angenommen werden, daß die festen Kosten C_0 zu Beginn der ersten Behandlungsperiode anfallen. Diese Annahme dürfte im Zusammenhang mit PAP meist ungefähr der Realität entsprechen, da der weitaus größte Teil der festen Kosten durch die Entwicklung der diagnostischen Instrumente und der Implementation des PAP entstehen dürfte. Natürlich entspricht das Modell nicht vollständig der Realität, da z.B. Evaluationsmaßnahmen (Validierungstudie, Nutzenanalyse, ...) erst nach einiger Zeit durchgeführt werden können. Da die Kosten für diese Maßnahmen als feste Kosten behandelt

werden müssen, sind diese unter C_0 mit ihrem Barwert zu verrechnen. Die Festkostenverzinsung kann in das Modell leicht dadurch integriert werden, indem diese Festkostenanteile mit ihrem jeweiligen Diskontierungsfaktor multipliziert in die Nutzenberechnung unter C_0 eingehen. Ich nehme diese Überlegung nicht explizit in das Modell auf, - um es nicht unnötig kompliziert zu machen. Mein Modell führt die Auffassungen von Boudreau (1983a, 1983b) und Cronshaw und Alexander (1985) zusammen und hat die folgende Struktur:

$$\Delta U = \sum_{k=1}^F E_k - \sum_{k=1}^F C_{1k} - C_0 \quad (28)$$

ΔU gibt dabei den inkrementellen Gesamtnutzen der Interventionsmaßnahme an, welcher durch die Erträge der Investition von Entwicklungs-, Implementations- und Evaluationskosten (C_0) und der Investition der variablen Kosten (C_1) entstehen. Die Erträge und variablen Kosten werden für die Behandlungsperioden k des PAP (zwecks Verzinsung) gesondert berechnet und danach aufsummiert. Der Nutzen heißt deshalb 'inkrementell', weil berücksichtigt werden soll, daß eine Interventionsmaßnahme immer vor dem Hintergrund von Alternativen bewertet werden muß. Im Zusammenhang von PAP könnte eine Alternative die Zufallsauswahl oder ein anderes PAP sein.

Ein allgemeines BCG-Nutzenmodell, welches durch die ökonomischen Konzepte und das Mitarbeiterflußmodell von Boudreau (1983a, 1983b) und die Konzepte von Cronshaw und Alexander (1985) erweitert wird, läßt sich damit durch Gleichung (29) ausdrücken.

Dieses Modell beinhaltet alle mir bekannten Vorschläge, welche bisher gemacht wurden, um das Nutzenmodell auf der Basis von Brogden (1946a, 1949) und Cronbach und Gleser (1957, 1965) formal zu erweitern. Es wird als Grundlage für alle weiteren Betrachtungen und die empirische Untersuchung in Kapitel 6 herangezogen. Darüber hinaus liegt es als Computerprogramm vor. Das Turbo Pascal-Listing des Programms ist in Anhang A abgedruckt.

$$\Delta U = \sum_{k=1}^F [N_k r_{xw} \bar{z}_x^* SD_w (1-V) (1/(1+i)^k) (1-TAX)] - \sum_{k=1}^F [C_{1k} (1/(1+i)^{k-1}) (1-TAX)] - [C_0 (1-TAX)] \quad (29)$$

Dabei gilt:

$$(25) \quad N_k = \sum_{t=1}^k (N_{at} - N_{zt})$$

$$(3) \quad \bar{z}_x^* = \lambda(Q)/Q \quad \text{und (15)} \quad V = s/w.$$

Die Parameter bedeuten dabei:

ΔU	Inkrementeller Nutzen. Dies ist der Betrag, um den sich der Nutzen verändert, wenn das untersuchte PAP relativ zu einem alternativen PAP bewertet wird.
k	Behandlungsperiode. Sie hat die Länge eines Jahres und bezeichnet das jeweilige k-te Jahr nach Implementation des PAP.
F	Programmwirkungsdauer. Diese besteht solange, wie $N_k > 0$ ist.
N_k	Anzahl der Mitarbeiter in der Organisation, welche mit der Intervention behandelt wurden, bzw. welche durch das PAP ausgewählt worden sind. N_k berechnet sich nach Gleichung (25).
N_{at}	Anzahl der akzeptierten Bewerber in der Wirkperiode t .
N_{zt}	Anzahl der ausgeschiedenen Mitarbeiter in der Wirkperiode t .
r_{xw}	Validität des Verfahrens. Dies ist der Produkt-Moment Korrelationskoeffizient zwischen dem/n Prädiktor/en x und dem Verkaufswert der Leistung w , bzw. dem Beitrag zur Organisation y . r_{xw} bezieht sich auf die Bewerberpopulation und bezeichnet die Differenz der Validität zu einem alternativen PAP.
\bar{z}_x^*	Durchschnittlicher standardisierter Prädiktorwert der akzeptierten Bewerber. Unter der Annahme einer normalen Verteilung von x und einer Selektionsstrategie 'von oben nach unten', kann diese berechnet werden durch $\bar{z}_x^* = \lambda(Q)/Q$. Q ist dabei die Selektionsquote und $\lambda(Q)$ der Funktionswert der Normalverteilungsfunktion an der Stelle Q .
SD_w	Standardabweichung des Verkaufswertes der Leistung. Dies ist die Standardabweichung des Verkaufswertes der bewerteten Handlungsergebnisse zwischen den Mitarbeitern. SD_w bezieht sich auf die Bewerberpopulation.
V	Anteil der variablen Kosten der Leistung für die Organisation am Verkaufswert der Leistung. V berechnet sich nach Gleichung (15). Der Faktor $(1-V)$ stellt die Proportion zwischen SD_y und SD_w dar.
i	Kalkulationszinsfuß.
TAX	Steuersatz.
C_{1k}	Inkrementelle variable Kosten des PAP in der Behandlungsperiode k (für alle Bewerber in k).
C_0	Inkrementelle feste Entwicklungs-, Implementations-, Evaluations- und andere Kosten des PAP.

5.2 Nutzen der Personalauswahl jenseits der Leistungssteigerung

Bei allen Nutzenmodellen, die auf dem BCG-Ansatz beruhen, entsteht der Nutzen der Personalauswahl direkt oder indirekt durch die gesteigerten Erträge, welche durch die ausgewählten Mitarbeiter entstehen. Diese Ertragssteigerung beruht auf der höheren durchschnittlichen Leistung der ausgewählten Bewerber gegenüber der Bewerberpopulation. Die durchschnittliche standardisierte Leistungssteigerung läßt sich durch Gleichung (2) ermitteln

$$(2) \quad \bar{z}_y^* = r_{xw} \bar{z}_x^*$$

Durch Multiplikation der standardisierten Leistungssteigerung mit dem Faktor SD_y , welcher die Einheit DM trägt, läßt sich der Ertrag der Maßnahme berechnen.

$$(5) \quad E = r_{xw} \bar{z}_x^* SD_y$$

Diese Ertragssteigerung führt gemäß der Logik der BCG-Nutzenmodelle direkt zum Nutzen des PAP, wenn die Kosten des Programms davon subtrahiert werden. Ein indirekter Nutzen kann entstehen, wenn nicht der Ertrag direkt (z.B. durch erhöhten Umsatz) gesteigert wird, sondern wenn durch die gesteigerte Leistung des Personals Lohn- und Sachkosten eingespart werden können (z.B. durch Personaleinsparung oder Kosteneinsparung bei Fehlerbehebung).

Dies habe ich in den vorausgehenden Kapiteln ausführlich dargestellt. In diesem Abschnitt werde ich nun besprechen, welche Dinge darüber hinaus zum Nutzen von PAP beitragen können. Wie ich zeigen werde, erbringen psychologische Interventionsprogramme oft einen Nutzen, der nicht von den BCG-Nutzenmodellen berücksichtigt wird. Dies werde ich wieder am besonderen Beispiel der PAP illustrieren. Da dieser erweiterte Nutzen von den BCG-Modellen nicht erfaßt wird, muß im allgemeinen davon ausgegangen werden, daß BCG-Nutzenschätzungen den tatsächlichen Nutzen von PAP unterschätzen, sofern einer oder mehrere der nachfolgenden Punkte zutreffen. Die folgende Sammlung der erweiterten Nutzenquellen von PAP versteht sich weder als erschöpfend, noch behaupte ich, daß bei jedem PAP alle Punkte zutreffend sein müssen. Vielmehr stellt sie eine Auswahl der möglichen zusätzlichen Nutzenquellen dar. Auf ihrer Grundlage können bestimmte PAP daraufhin untersucht werden, ob es lohnenswert ist, bestimmte Punkte vertieft zu bedenken.

Alle diese Vorschläge lassen sich schwer in quantitative Erträge übersetzen. Bei der Analyse des Nutzens eines konkreten PAP sollte deswegen in zwei Schritten vorgegangen werden. Zum ersten sollte eine Nutzenschätzung auf der Grundlage eines erweiterten BCG-Modells, z.B. des Modells (29), erfolgen. Danach kann eine Untersuchung des PAP

hinsichtlich der nachfolgenden Nutzenquellen angestrebt werden. Kommt der Untersucher zu dem Schluß, daß ein PAP einen wesentlichen Nutzen aus einer dieser Quellen zieht, dann kann er versuchen, diesen überschlägig in DM zu quantifizieren. Wie dabei vorgegangen werden kann, behandle ich hier nicht.

Ich möchte diese zusätzlichen Nutzenquellen in zwei Kategorien aufteilen. Die erste Kategorie beinhaltet Nutzenquellen, die innerhalb einer Organisation durch die Personalauswahl entstehen. Sie werden im nächsten Abschnitt (5.2.1) behandelt. Im Abschnitt 5.2.2 dagegen werde ich auf den Nutzen von PAP eingehen, der einer Organisation erst aus dem Zusammenspiel mit der Umwelt und anderen Organisationen erwächst. Auf diesen systemischen Betrachtungen soll der Schwerpunkt liegen.

5.2.1 Nutzen innerhalb der Organisation

Die Annahme, daß der Nutzen der Personalauswahl in einer Organisation nur durch die höheren Erträge der ausgewählten Mitarbeiter aufgrund ihrer höheren Leistung entsteht, kann wichtige indirekte Nutzenquellen vernachlässigen. Wenn höher qualifizierte Mitarbeiter ausgewählt werden, können bei manchen Berufsfeldern teilweise die Kosten für bestimmte Ausbildungs-, Weiterbildungs- und Trainingsmaßnahmen entfallen. Diesen Punkt nenne ich *Nutzen durch verminderte Ausbildungskosten*. Er bedeutet, daß die Mitarbeiter, die durch ein gutes PAP ausgewählt wurden, bei gleichem Leistungsziel insbesondere in der Berufseingangsphase weniger Ausbildungskosten verursachen können. Neben diesen Einsparungen, die durch verminderte Inanspruchnahme von Ausbildungsprogrammen entstehen, benötigen diese höherqualifizierten Mitarbeiter auch *weniger zeitliche Ressourcen ihrer Mitarbeiter*, insbesondere ihrer Vorgesetzten. Brogden und Taylor (1950) versuchen diesen Punkt bei ihrem Kostenrechnungsansatz zur Ermittlung von SD_y zu berücksichtigen. Bei allen anderen Schätzmethoden von SD_y fehlt er aber.

Unter der Annahme, daß bei vielen Berufsfeldern gute Leistungen dem Leistungsträger rückgemeldet werden, gehe ich davon aus, daß die verbesserten Leistungen der ausgewählten Bewerber tendenziell auch zu einer Steigerung der Arbeitszufriedenheit führen. Darüber hinaus werden leistungsfähigere Mitarbeiter durchschnittlich auch mehr verdienen, mehr Anerkennung finden und einen höheren sozialen Status innerhalb ihrer Arbeitsumwelt haben. Diese Auswirkungen der Personalauswahl auf individueller Ebene (vgl. Abb. 2) können langfristig dazu beitragen, daß die hohen Kosten, welche durch Krankenstand und unerwünschte Fluktuation entstehen, gesenkt werden können. Diese Überlegung wird

auch bei der Besprechung des Nutzens der Personalauswahl durch das AC von Versicherungsaußendienstmitarbeitern (AM) bei der empirischen Analyse in Kapitel 6 wichtig, wenn ich mich mit der Operationalisierung des Konstruktes 'Berufserfolg' beschäftige. Die Fluktuationsrate von AM liegt sehr hoch. Man kann davon ausgehen, daß weit mehr als die Hälfte der AM zwei Jahre nach der Einstellung ihr Unternehmen wieder verlassen haben. In diesem besonderen Berufsfeld, so behaupte ich, ist eine wesentliche Determinante beruflicher Zufriedenheit der Verdienst. Man kann dies damit begründen, daß die wenigsten AM ihre primäre Berufsausbildung auf diese Tätigkeit ausgerichtet haben. Vielmehr kommen die AM aus den verschiedensten Berufen, so daß man sich fragen kann, was sie zu dieser beruflichen Veränderung motiviert hat. Da eines der Stereotype der Versicherungsaußendiensttätigkeit darin besteht, daß damit vergleichsweise viel Geld zu verdienen ist, gehe ich davon aus, daß diese Erwartung eine der Hauptmotivatoren der Bewerbung für diese Tätigkeit darstellt. Ich behaupte nicht, daß bei Eintreten der Erwartung auch automatisch die Zufriedenheit mit dem Beruf folgt. Meine Hypothese besteht vielmehr darin, daß die Enttäuschung der Verdiensterwartung eine massive Arbeitsunzufriedenheit nach sich zieht und damit der Wunsch nach beruflicher Veränderung entsteht. Da jedoch bei AM die Entlohnung weitgehend an die Leistung gekoppelt ist (Provision), sollte die Auswahl von weniger leistungsfähigeren Mitarbeitern in einer erhöhten Fluktuationsrate resultieren, was wiederum hohe Personalwerbungs-, Personalauswahl- und Ausbildungskosten verursacht. Darüber hinaus bedeutet eine hohe Fluktuationsrate für das Unternehmen auch, daß das Erfahrungspotential und das 'know-how' des Personalbestands ständig und systematisch verloren geht. Die Folgen daraus bestehen allerdings nicht nur darin, daß diese Erfahrung dem Unternehmen nicht mehr zur Verfügung steht. Man tut gut daran, noch weitreichendere Konsequenzen zu bedenken. Der nächste Abschnitt soll dies verdeutlichen.

5.2.2 Der Arbeitsmarkt als limitiertes offenes System

Alle hier besprochenen Nutzenmodelle, die auf dem BCG-Ansatz beruhen, beziehen die durchschnittliche Steigerung der standardisierten Kriteriumswerte der ausgewählten Bewerber \bar{z}_y^* auf die Bewerberpopulation. Der Durchschnitt der standardisierten Kriteriumswerte in der Bewerberpopulation wird dabei per definitionem auf Null gesetzt ($\bar{z}_y^*=0$). Gegen diese Herangehensweise an das Problem wäre dann nichts einzuwenden, wenn die Bewerberpopulation unendlich groß wäre. Dann würde eine valide Personalauswahl lediglich eine Änderung der durchschnittlichen Kriteriumswerte der ausgewählten Bewerber bewirken,

diejenigen der abgelehnten Bewerber wären weiterhin Null. Die Annahme der unendlichen Größe der Bewerberpopulation ist allerdings oftmals alles andere als gerechtfertigt. Die Anzahl der Bewerber für eine bestimmte Berufsgruppe auf dem Arbeitsmarkt in bestimmten Regionen ist meist nur begrenzt. Eine valide Personalauswahl ändert damit nicht nur die Leistung der ausgewählten Gruppe gegenüber der Population, sie ändert auch die potentielle durchschnittliche Leistung der Bewerberpopulation nach der Personalauswahl gegenüber der Bewerberpopulation vor der Maßnahme. Die potentielle Leistung der nicht ausgewählten Bewerber sinkt ab, und zwar um so stärker, je mehr Bewerber ausgewählt worden sind, je höher die Validität des Personalauswahlverfahren ist und je weniger potentielle Bewerber auf dem Arbeitsmarkt sind. Man kann sich dies so vorstellen, daß ein gutes PAP die leistungsfähigeren Bewerber vom Arbeitsmarkt 'abzieht', während die weniger leistungsfähigen darin verbleiben. Die BCG-Nutzenmodelle berücksichtigen diesen Sachverhalt nicht.

Ich betrachte hier den Arbeitsmarkt als offenes System. Die Elemente des Arbeitsmarktes sind die Leistungspotentiale aller potentiellen Bewerber. Die Elemente des Systems sind also nicht die Bewerber. Diese sind lediglich die Träger verschiedener Leistungspotentiale für bestimmte Berufstätigkeiten. Das Arbeitsmarktsystem ist offen, das heißt es steht in Wechselwirkung mit anderen Systemen. Dies sind z.B. das Bildungssystem und das Berufsausbildungssystem, welches dem Arbeitsmarkt Leistungsressourcen zuführt, wirtschaftliche Unternehmen und andere Organisationen, die sowohl die Leistungsressourcen aus dem Arbeitsmarkt entnehmen (Einstellung) als auch zuführen (Personalentwicklung, Berufserfahrung, ...). Neben der Eigenschaft der Offenheit hat das Arbeitsmarktsystem einen endlichen Umfang. Es ist limitiert. Das bedeutet, daß aus dem System nur begrenzt Leistungsressourcen entnommen werden können.

Personalauswahl ist eine der Interventionsmaßnahmen, die den Fluß der Leistungsressourcen vom Arbeitsmarkt in eine Organisation steuern können. Nun ist aber noch ein weiterer Sachverhalt zu berücksichtigen. In den seltensten Fällen wird nur eine Organisation in Wechselwirkung mit einem Arbeitsmarkt stehen, sondern viele Organisationen und wirtschaftliche Unternehmen, die miteinander in Konkurrenz stehen und versuchen, den Systeminput und -output zu kontrollieren. Diese systemische Sichtweise illustriert Abbildung 9.

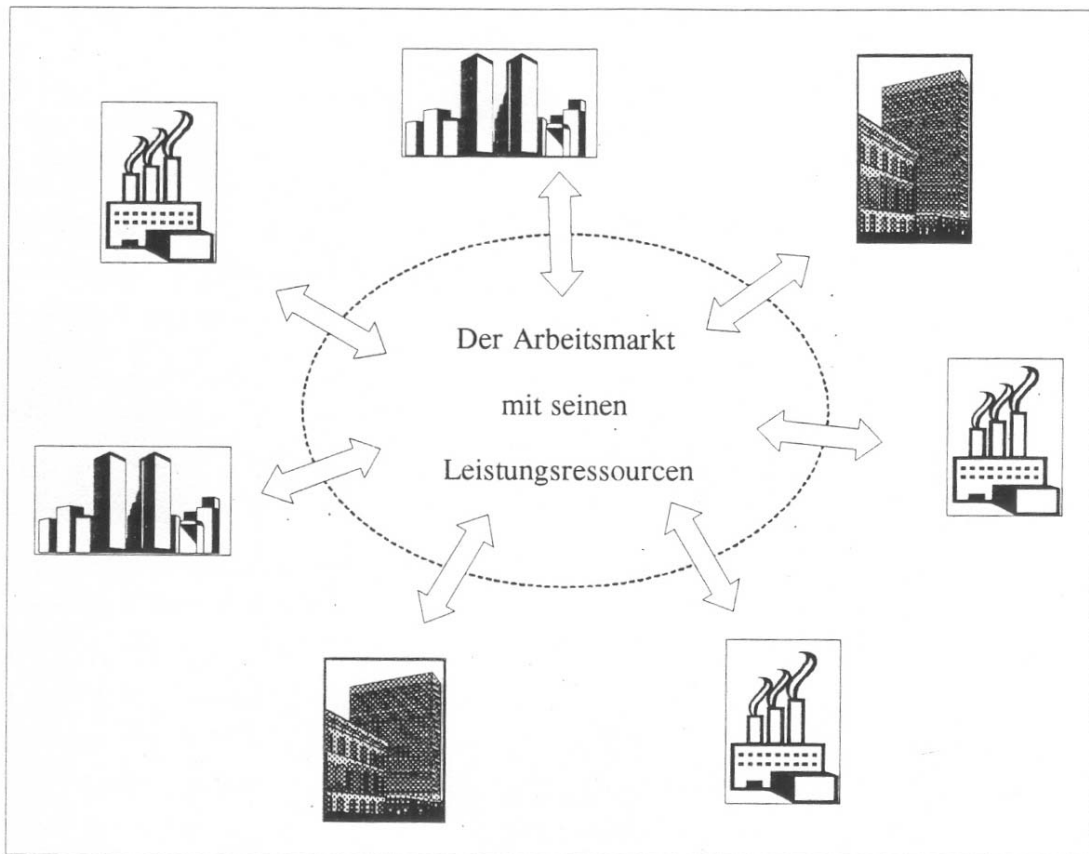


Abbildung 9: Der Arbeitsmarkt als limitiertes offenes System. Verschiedene Organisationen geben und entnehmen ihm Leistungsressourcen für bestimmte Berufstätigkeiten.

Wird der Arbeitsmarkt als limitiertes offenes System angesehen und bedenkt man, daß mehrere Organisationen um die begrenzten Leistungsressourcen auf diesem Arbeitsmarkt in Konkurrenz treten, dann werden neue Aspekte des Nutzens der Personalauswahl sichtbar. Betreibt eine Organisation nämlich ihre Personalauswahl besonders effektiv, so bedeutet dies nicht nur, daß sie besonders qualifizierte Mitarbeiter erhält, welche mit ihrer Leistung vergleichsweise hohe Erträge für die Organisation erbringen, wie die Nutzenmodelle des BCG-Ansatzes dies feststellen. Über diesen Vorteil hinaus zieht eine solche Organisation auch die begrenzten Ressourcen vom Arbeitsmarkt ab, so daß sie nicht mehr für eventuelle Konkurrenzorganisationen zur Verfügung stehen. Eine Entscheidung für ein PAP sollte deshalb nicht nur berücksichtigen, welche alternativen PAP zur Verfügung stehen, sondern ebenso, welche PAP die Konkurrenzorganisationen besitzen und damit auf demselben Arbeitsmarkt intervenieren. Diese Überlegungen müssen natürlich nicht bei jeder Art von Arbeitsmarkt vorgenommen werden. Wichtig können sie aber werden, wenn bestimmte Leistungsressourcen auf dem Arbeitsmarkt besonders knapp sind (vgl. z.B. Facharbeitermangel in bestimmten Regionen der BRD). Eine konkurrierende Organisation, die versäumt hat, ihr Personal in geeigneter Weise zu behandeln und deshalb Wettbewerbsnachteile hat oder befürchtet, kann versuchen, die Defizite im Nachhinein

auszugleichen. Dies kann zum Beispiel durch Interventionsprogramme zur Anhebung der beruflichen Qualifikation des bestehenden Mitarbeiterstammes geschehen (Personalentwicklung) oder durch Personalabwerbungsversuche von anderen Organisationen. In jedem Fall verursachen solche Aktivitäten Kosten. Diese Interventionskosten sind Folgekosten, die entstehen, weil frühere PAP nicht die qualifizierten Mitarbeiter erbracht haben, wie die PAP der Konkurrenzorganisationen.

Schafft es eine Organisation, die derart in Schwierigkeiten gekommen ist, nicht rechtzeitig, ihr Personal auf das Niveau der erfolgreichen Konkurrenz anzuheben, dann sind weitere Konsequenzen zu erwarten. Handelt es sich z.B. um ein wirtschaftliches Unternehmen, dann wird es aufgrund der verminderten Umsätze gezwungen sein, besondere Leistungen für seine Mitarbeiter zu streichen (freiwillige Sozialleistungen, übertarifliche Bezahlung, Weiterbildung, ...). Dadurch könnten die höher qualifizierten Mitarbeiter sich stärker von Konkurrenzunternehmen angezogen fühlen, was zu einer weiteren Verschlechterung der Lage führen kann. Spätestens dann beginnt sich der Teufelskreis zu schließen.

Der Gedanke, daß der Nutzen von PAP durch die Eigenschaft der Arbeitsmärkte als offene limitierte Systeme beeinflusst wird, ist schwer in einem formalen mathematischen Modell auszudrücken. Dies liegt nicht nur daran, daß solche Modelle dynamische Systemabläufe abbilden müßten. Darüber hinaus dürften sie nicht vernachlässigen, daß der Fluß von Leistungsressourcen vom Arbeitsmarkt in die Organisationen und umgekehrt natürlich nicht nur von PAP allein beeinflusst wird. Andere wichtige Determinanten sind z.B. die Verdienstmöglichkeiten, der Standort, Weiterbildungs- und Aufstiegsmöglichkeiten, aber auch schwierig operationalisierbare Dinge wie das Organisationsklima und das Image der Organisation nach außen.

Andererseits sollte ein Forscher oder Praktiker, der den Nutzen von Interventionsmaßnahmen zu beurteilen hat, diese Überlegungen nicht deswegen ignorieren, weil sie nicht formal ausgedrückt werden können. PAP sind im Verbund mit anderen Maßnahmen in der Lage, die Leistungsressourcen des Mitarbeiterbestandes von Organisationen entscheidend zu beeinflussen. Da die Leistungsressourcen des Personals vielleicht das wertvollste sind, was eine Organisation besitzt, sollte diese ganzheitliche Betrachtungsweise des Arbeitsmarktes als limitiertes offenes System bei der langfristigen Nutzenbetrachtung von PAP Berücksichtigung finden.

5.3 Nutzen allgemeiner psychologischer Interventionsmaßnahmen

Ich habe insbesondere im zweiten Kapitel zu zeigen versucht, daß es vorteilhaft sein kann, Personalauswahl als psychologische Interventionsmaßnahme aufzufassen. Ich habe im weiteren Verlauf dieser Arbeit anstatt des Begriffs 'Personalauswahl' an vielen Stellen einfach den Begriff 'Interventionsmaßnahme' verwendet. Diese Begriffswahl ist allerdings nur dann gerechtfertigt, wenn die Personalauswahl im Kontext der Nutzenanalyse wirklich nur einen Spezialfall der Intervention darstellt, die Personalauswahl sozusagen dazu benutzt wird, die Anwendung der Nutzenmodelle auf psychologische Interventionsmaßnahmen am speziellen Beispiel der Personalauswahl darzustellen.

Nun wäre es übertrieben zu behaupten, daß die Personalauswahl ein beliebiges Illustrationsbeispiel wäre. Wie schon mehrfach angemerkt, hat sich die Logik der BCG-Nutzenmodelle in enger Anlehnung an die Personalauswahl entwickelt. Darüber hinaus findet sich bis heute die überwiegende Mehrzahl der Anwendungen von NA im Zusammenhang mit PAP. Dennoch bin ich der Auffassung, daß das Potential der BCG-Nutzenmodelle nicht genutzt wird, wenn sie nicht auch außerhalb der Personalauswahl Beachtung finden. Meines Erachtens kann nahezu jedes Interventionsprogramm, welches zum Ziel hat, die Leistungsfähigkeit von Menschen in Organisationen zu beeinflussen und mit psychologischen oder anderen Methoden arbeitet, mittels der BCG-Modelle nutzenanalytisch untersucht werden.

Diese Überlegungen sind Gegenstand des vorliegenden Abschnittes. Um Ordnung in die Menge solcher Interventionsprogramme zu bringen, soll zunächst eine Klassifikationsmöglichkeit von Interventionsprogrammen vorgestellt werden, die zum Ziel haben, die Leistungsfähigkeit von Menschen in Organisationen zu verändern. Die Personalauswahl wird sich innerhalb dieser Klassifikationsmöglichkeit als Spezialfall erweisen. Um die besprochenen Nutzenmodelle auch für die anderen Interventionsmöglichkeiten fruchtbar zu machen, soll im zweiten Teil des Abschnittes eine Anpassungsmöglichkeit der formalen Nutzenmodelle an die neue Situation besprochen werden.

5.3.1 Klassifikation von Interventionsprogrammen im Personalbereich

Boudreau (1991) klassifiziert Interventionsprogramme, welche die Leistung von Mitarbeitern in Organisationen beeinflussen sollen und auf welche deshalb Nutzenmodelle prinzipiell

anwendbar sein sollten, in zwei Bereiche. Die erste Interventionsklasse umfaßt solche Programme, die eine Wirkung auf den *Fluß der Mitarbeiter* innerhalb der Organisation, in sie hinein oder aus ihr heraus, ausüben. Das Ziel dieser Programme besteht darin, die Leistungsfähigkeit des Personals positiv zu beeinflussen, indem sie steuern, welche Mitarbeiter der Organisation angehören, und in welcher Position diese Mitarbeiter innerhalb der Organisation sind. Das Grundprinzip der Behandlung dieser Programme besteht also im Verschieben von Leistungsressourcen. Unser Paradebeispiel, die Personalauswahl, gehört dieser Interventionsklasse an. Natürlich gibt es auch weitere Beispiele, die dieses Interventionsprinzip anwenden. Outplacementprogramme betreiben Personalauswahl unter umgekehrten Vorzeichen. Sie versuchen, direkt oder indirekt, den Fluß solcher Mitarbeiter aus der Organisation hinaus zu steuern, die vergleichsweise geringe Leistungsressourcen für die speziellen Aufgaben besitzen, für die sie vorgesehen waren. Auch wenn Outplacement in wirtschaftlichen Unternehmen wegen der nachteiligen Folgen für die Betroffenen oft mit Vorsicht zu genießen ist, so sollte man die vielen Organisationen nicht vergessen, in welchen diese Behandlungsart zur Tagesordnung gehört. Dies sind z.B. Universitäten, die sich regelmäßig von ihren leistungsschwächsten Angehörigen trennen oder Leistungssportgruppen. Eine weitere Programmart in dieser Kategorie nenne ich 'organisationsinterne Personalpolitik'. Sie bestimmt den Fluß von Organisationsmitgliedern sowohl in horizontaler als auch in vertikaler Richtung. Solche Programme könnten z.B. in einem Arbeitsplatzrotationssystem oder einer systematischen internen Führungskräfteerkrutierung bestehen. Allen Interventionsprogrammen der ersten Klasse ist gemeinsam, daß sie Mitarbeiter für eine bestimmte Behandlung auswählen müssen. Diese Auswahl kann mehr oder weniger systematisch erfolgen. Die Grundlage einer systematischen Auswahl können hochwertige diagnostische Instrumente sein oder nicht. Oft lassen sich bei solchen Programmen Validierungsstudien durchführen und die BCG-Nutzenmodelle ohne Veränderungen anwenden.

Die zweite Interventionsklasse umfaßt Programme, die nicht den Fluß des Personals, sondern das *bestehende Personal in seinen Positionen* beeinflussen wollen. Diese Programme behandeln das bestehende Personal von Organisationen mit Maßnahmen wie z.B. Training, Entlohnungssystem, Zielvereinbarung und Ergebnisrückmeldung, Leistungsbeurteilung, etc. Das Ziel dieser Maßnahmen besteht in der Veränderung von Wissen, Fähigkeiten, Fertigkeiten, Einstellungen, Werten, Motivation und anderer Charakteristika der Mitarbeiter der Organisation, um ihre Leistungen positiv zu beeinflussen. Sie sind also prinzipiell nicht anders in ihrer Struktur und ihren Zielen als die Interventionsprogramme, die den Fluß der

Mitarbeiter ändern wollen. Die BCG-Nutzenmodelle sollten deshalb auch auf sie anwendbar sein. Betrachten wir das allgemeine Nutzenmodell, das in Gleichung (29) festgelegt ist und ein Trainingsprogramm als Interventionsmaßnahme zur Leistungssteigerung von bestimmten Mitarbeitern, dann lassen sich einige Parameter mit nicht größeren Schwierigkeiten als bei PAP operationalisieren. Die festen Entwicklungs-, Durchführungs- und Evaluationskosten und die variablen Trainingskosten bereiten genausowenig Probleme wie die Diskontierung und die Versteuerung von eingesetztem Kapital. Die Probleme mit dem Parameter SD_w bleiben ebenfalls ähnlich gelagert.

Unterschiede bestehen allerdings im Zusammenhang mit dem Validitätskoeffizienten rx_w und dem Parameter \bar{z}_x^* da aufgrund des Fehlens der Prädiktorwerte x bei Trainingsprogrammen natürlich weder die durchschnittlichen standardisierten x -Werte der ausgewählten Gruppe, noch die Korrelation zwischen x und w angegeben werden kann. Weitere Probleme bereitet der Parameter N_k . Im Gegensatz zur Personalauswahl, bei welcher z.B. Persönlichkeitseigenschaften, Fähigkeiten und Fertigkeiten als zeitlich stabil angenommen werden, hat man bei Trainings festgestellt, daß die Maßnahmen oft einen erstaunlichen Anfangserfolg verbuchen können, die positiven Effekte aber meist sehr bald zurückgehen.

Die spezielle Form der Evaluation, die Validierungsstudie (s. Abschnitt 2.3), welche bei der Überprüfung von Interventionsmaßnahmen der ersten Kategorie oft anwendbar ist, ist nicht bei den Interventionsmaßnahmen brauchbar, die das bestehende Personal beeinflussen wollen. Werden solche Interventionsmaßnahmen überprüft, so geschieht dies meist durch experimentelle oder quasiexperimentelle Untersuchungen. Diese drücken ihre Ergebnisse allerdings meist nicht in Validitätskoeffizienten aus, sondern in Effektstärken und t - bzw. F -Werten. Studien mit mehreren Meßzeitpunkten liefern darüber hinaus noch Hinweise über die Wirkdauer der Interventionsmaßnahme.

5.3.2 BCG-Nutzermodelle bei der Veränderung des bestehenden Personals

Erste Anwendungen des BCG-Nutzenmodells auf die Bewertung von Interventionsmaßnahmen des bestehenden Personals stammen von Landy, Farr und Jacobs (1982) und Schmidt, Hunter und Pearlman (1982). Landy et al. (1982) untersuchten den Nutzen eines Leistungsfeedbackprogrammes, bei Schmidt et al. (1982) bestand die Intervention in Trainingsprogrammen. Einige der folgenden Überlegungen finden sich auch bei diesen Autoren wieder.

Wie in Gleichung (2) ausgedrückt ist, läßt sich bei der Personalauswahl der durchschnittliche standardisierte Kriteriumswert der ausgewählten Gruppe bei bekanntem Validitätskoeffizient und bekanntem durchschnittlichem standardisierten Prädiktorwert der ausgewählten Gruppe berechnen.

$$(2) \quad \bar{z}_y^* = r_{xy} \bar{z}_x^*$$

Wie Abbildung 2 (Stufe 5) verdeutlicht, besteht die Behandlung bei der Personalauswahl darin, daß ein Bewerber entweder akzeptiert oder abgelehnt wird. Es bestehen also zwei Gruppen, die verschieden behandelt werden, die Gruppe der aufgenommenen und die der abgelehnten Bewerber. Der Parameter \bar{z}_y^* kann auch als der Unterschied zwischen einer behandelten Gruppe (aufgenommene Bewerber) und der Gesamtgruppe (aufgenommene und abgelehnte Bewerber) bzgl. der durchschnittlichen Kriteriumswerte aufgefaßt werden. Der Unterschied wird dabei in Standardabweichungseinheiten angegeben. Ein durchschnittlicher standardisierter Kriteriumswert von $\bar{z}_y^*=0,5$ bedeutet nichts anderes, als daß eine in bestimmter Form behandelte Gruppe (Aufnahme in die Organisation) durchschnittlich 0,5 Standardabweichungen höhere Kriteriumswerte (z.B. Arbeitsleistungen) erbringt als die Population der Behandlungsfähigen.

Im allgemeinen Fall der Interventionsmaßnahmen kann die Behandlung in vielfältiger Form bestehen. Dies habe ich anhand der oben beschriebenen Klassifikation von Boudreau (1991) gezeigt. Bei Interventionsmaßnahmen, die das bestehende Personal behandeln, kann man die Kriteriumswerte der behandelten Gruppe mit einer Stichprobe aus der unbehandelten Population vergleichen. Wird die Differenz der mittleren Kriteriumswerte der behandelten Gruppe (\bar{y}^*) und der Gesamtpopulation (\bar{y}) gebildet, und diese durch die Standardabweichung der Kriteriumswerte der Gesamtpopulation (SD_y) geteilt, dann erhält man die standardisierte Effektgröße d_y der Behandlung der Gruppe bzgl. des Kriteriums y . Diesen Sachverhalt verdeutlicht Gleichung (30).

$$d_y = (\bar{y}^* - \bar{y})/SD_y \quad (30)$$

Die an der Standardabweichung relativierte Effektgröße der Behandlung nenne ich in Anlehnung an gängige Bezeichnungen d_y . Wenn das Zustandekommen von d_y auf die Behandlung zurückgeführt werden kann, was durch entsprechende experimentelle Techniken bzw. Versuchsplanung sichergestellt werden sollte, dann hat diese Größe eine hohe konzeptionelle Ähnlichkeit mit \bar{z}_y^* . Man erhält durch diese Vorgehensweise eine direkte Schätzung für \bar{z}_y^* , ohne vorher den Umweg über r_{xy} und \bar{z}_x^* gehen zu müssen.

Das einfache BCG-Nutzenmodell läßt sich auf diese Weise auch für den allgemeinen Fall der Behandlung darstellen. In diesem Fall wird die \bar{z}_y^* -Schätzung ($r_{xy} \bar{z}_x^*$) durch d_y ersetzt. Das Nutzenmodell z. B. der Gleichung (16) ändert sich in Gleichung (31).

$$\Delta U = N_a T d_y S D_y - C \quad (31)$$

Natürlich läßt sich d_y in jedes Nutzenmodell auf Grundlage des BCG-Ansatzes einbauen. Werden Interventionsmaßnahmen als Investitionsentscheidungen gesehen, und sollen ökonomische Konzepte und die Zeitdimension Berücksichtigung finden, so läßt sich das erweiterte Nutzenmodell (29) auf den allgemeinen Fall einfach übertragen, indem ($r_{xy} \bar{z}_x^*$) durch d_y ersetzt wird.

Ich habe gezeigt, daß die NA eine sinnvolle Erweiterung quantitativer Evaluation aller Interventionsmaßnahmen ist, welche die Leistung der Mitglieder von Organisationen beeinflussen sollen. In Relation zum Aufwand, den eine Evaluation erfordert, die 'lediglich' Effektgrößen und Validitätskoeffizienten ermittelt, verlangt die nutzenanalytische Bewertung der Interventionsmaßnahmen einen geringen Mehraufwand. Die Ergebnisse der Evaluation werden durch die NA jedoch stark aufgewertet. Durch nutzenanalytische Betrachtungen können Interventionsprogramme optimal gestaltet und Entscheidungen über Interventionsprogramme rationaler gefällt werden.

5.4 Nebennutzen psychologischer Intervention

In diesem Abschnitt möchte ich eine weitere Art des Nutzens von Interventionsmaßnahmen vorstellen und diskutieren.

5.4.1 Rückschau der bisher betrachteten Nutzenformen

Ich habe bisher zwei Nutzenarten unterschieden, die im Zusammenhang mit psychologischen Interventionsmaßnahmen beachtet werden sollten. Die erste Nutzenart ist diejenige, die durch die BCG-Nutzenmodelle beschrieben werden. Dieser Nutzen basiert auf den gesteigerten Erträgen, die der Personalbestand aufgrund der Intervention erbringt. Die Erträge sind deshalb gesteigert, weil die Mitarbeiter der Organisation durch die Behandlung leistungsfähiger geworden sind. Diese Leistungsfähigkeitssteigerung resultiert z.B. aus der Personalauswahl, aus einem Training oder aus einem veränderten Entlohnungssystem. Die Erträge können auf der Grundlage dieser Leistungsfähigkeit entweder direkt erbracht werden,

indem z.B. die Produktivität über die Quantität oder die Qualität des Arbeitsergebnisses gesteigert wird. Diese direkten Ertragssteigerungen werden insbesondere in wirtschaftlich arbeitenden Unternehmen interessant sein. Die gesteigerte Leistungsfähigkeit kann aber auch indirekte Erträge erbringen, indem durch die Leistungsträger zwar beispielsweise mehr produziert wird, dieser vermehrte Output aber nicht durch einen erhöhten Umsatz genutzt wird, sondern indem Stellen abgebaut werden und dadurch Lohnkosten eingespart werden. Der Nutzen der Interventionsmaßnahme besteht in beiden Fällen im Ertrag der Maßnahme abzüglich ihrer Kosten.

Die zweite Art des Nutzens von Interventionsmaßnahmen habe ich in Abschnitt 5.2 dargestellt. Es handelt sich um den Nutzen, der nicht direkt oder indirekt durch den Ertrag der Maßnahme entsteht, sondern erst durch das Wechselspiel mehrerer konkurrierender Organisationen mit einem limitierten Arbeitsmarkt. Das Prinzip dieses *Nutzens zweiter Art* besteht darin, daß ein offenes Arbeitsmarktsystem Leistungsressourcen besitzt, und diese an verschiedene Organisationen abgeben oder von diesen aufnehmen kann. Die Leistungsressourcen des Arbeitsmarktsystems sind nicht unendlich vorhanden, sie sind limitiert. Einmal abgegebene Leistungsressourcen fehlen dem Arbeitsmarkt so lange, bis neue hinzugeführt werden. Im Zusammenhang mit der Personalauswahl führt die Betrachtung des Zusammenspiels mehrerer konkurrierender Organisationen und des limitierten offenen Arbeitsmarktsystems unweigerlich zu der Erkenntnis, daß den Organisationen, die durch gute Personalauswahl dem Arbeitsmarkt überproportional viele Ressourcen entnehmen, -nicht nur ein direkter Nutzen erster Art erwächst. Über diesen hinaus erlangen sie auch einen Nutzen dadurch, daß konkurrierende Organisationen die fehlenden Leistungsressourcen auf dem Arbeitsmarkt nicht mehr verwenden können. Dieser Nutzen wird durch die BCG-Modelle nicht abgebildet.

Man könnte sich fragen, ob dieser 'Nutzen zweiter Art' nur im Zusammenhang mit der Personalauswahl interessant ist oder auch bei Interventionsmaßnahmen, die bestehendes Personal verändern wollen (vgl. Abschnitt 5.3). Was passiert mit dem Arbeitsmarkt, wenn z.B. ein Unternehmen intensive Aus- und Weiterbildungsprogramme für sein Personal durchführt? Zunächst wird es dadurch hoffentlich einen Nutzen erster Art davontragen. Darüber hinaus gibt es dadurch über kurz oder lang auch Leistungsressourcen an den Arbeitsmarkt ab, die andere Unternehmen für sich einsetzen können. In diesem Fall erbringen die Aus- und Weiterbildungsmaßnahmen der Konkurrenz einen Nutzen zweiter Art, was den Gesamtnutzen der Maßnahmen für das durchführende Unternehmen mindern sollte.

Ich möchte diesen Sachverhalt mit einem Beispiel verdeutlichen. Ein Versicherungsunternehmen A sieht für seine neuen Außendienstmitarbeiter (AM) im ersten Jahr nach dem Berufseintritt ein ausführliches Ausbildungsprogramm vor. Dem Unternehmen entstehen dadurch sehr hohe Kosten, die insbesondere durch die Kosten der Ausbildungszentren und dessen Personal verursacht werden, aber auch durch die Lohn-, Unterbringungs- und Fahrtkosten der Auszubildenden. Nach diesem ersten Jahr erwirtschaften die AM Erträge, welche u.a. ihre hohen Ausbildungskosten decken müssen. Ein kleines Versicherungsunternehmen B hat sich aufgrund dieser Lage darauf spezialisiert, diese gut ausgebildeten Mitarbeiter von A verdeckt abzuwerben. Die Abwerbemaßnahme kann z.B. lediglich darin bestehen, daß Unternehmen B deutlich höhere Löhne als A bezahlt. B kann sich diese erhöhten Lohnkosten deshalb leisten, weil es keine Ausbildungskosten zu decken hat. In unseren nutzenanalytischen Termini könnte man auch sagen, daß Unternehmen B ein Abwerbeprogramm installiert hat. Dieses verursacht wie alle Interventionsprogramme Kosten, aber auch Erträge. Ob die Durchführung des Programms sinnvoll ist, kann nach der Analyse des Programmnutzens und des Nutzens alternativer Programme entschieden werden.

5.4.2 Nutzen der dritten Art

Über die beschriebenen zwei Nutzenformen hinaus unterscheide ich eine dritte Art des Nutzens von Interventionsprogrammen. Ich nenne diese Art auch *Nebennutzen*. Er ist nicht immer sicher von den anderen beiden Nutzenarten abzugrenzen. Allgemein gesagt, zeichnet den Nebennutzen die Tatsache aus, daß er im Zusammenhang mit einem Interventionsprogramm entsteht, ohne daß er das primäre Ziel der Maßnahme darstellt. Das Ausbildungsprogramm des Versicherungsunternehmens A aus dem letzten Beispiel hat z.B. das Hauptziel, AM hervorzubringen, die den Kunden qualifiziert beraten können, und das Unternehmen dadurch eine qualitativ hochwertigere Leistung erbringen kann als die Konkurrenz. Allerdings hat die Maßnahme auch einen Nebennutzen, der möglicherweise zum Teil bei der Entwicklung des Interventionsprogrammes im Sinne von Nebenzielen bedacht wurde, der aber zum anderen Teil unbeabsichtigt entstand, sozusagen als Nebenwirkung der Maßnahme. Ein solcher Nebennutzen besteht z.B. darin, daß mehrjährige AM, die sich aus dem Außendienst zurückziehen wollen, für die aber keine anderen Stellen zur Verfügung stehen, dem Unternehmen erhalten bleiben können, indem sie in der Ausbildung tätig werden können. Dadurch wird nicht nur deren Wissen und Erfahrung für das Unternehmen erhalten.

Das 'Berufsaufstiegsprogramm' bringt sicherlich auch eine Fluktuationsminderung bei den jüngeren AM mit sich, da diesen eine langfristige Perspektive geboten wird.

Das Konzept des Nebennutzens von Interventionsmaßnahmen halte ich für sehr wichtig. Es wird in den BCG-Nutzenmodellen nicht berücksichtigt. So wie beim Nutzen zweiter Art halte ich eine Quantifizierung des Nebennutzens in Geldwerteinheiten für sehr schwierig. Ich möchte mich deshalb auf eine qualitative Betrachtung beschränken. Anstatt das Konzept theoretisch zu durchdringen, möchte ich den Nebennutzen von Interventionsmaßnahmen exemplarisch am Beispiel der Personalauswahl mittels Assessment Centern darstellen.

5.4.3 Nebennutzen am Beispiel eines Auswahl-AC

Das Hauptziel von Auswahl-AC besteht einfach formuliert darin, aus allen Bewerbern diejenigen auszuwählen, welche die höchsten bewerteten Beiträge zur Organisation erbringen werden. Neben dieser Hauptfunktion erfüllen Auswahl-AC auch Nebenfunktionen, die z.T. nicht unwichtiger als die Hauptfunktion sind. Diese Nebenfunktionen erbringen für die Organisation einen Nebennutzen, der ebenso wie der Nutzen erster und zweiter Art Beachtung finden sollte.

Bei der Konstruktion von Auswahl-AC beginnt man im allgemeinen zunächst mit der Analyse der Tätigkeiten, die bei den zu besetzenden Stellen anfallen. In der Regel werden die Vorgesetzten und Kollegen der späteren Stelleninhaber in diesen Analyseprozeß mit einbezogen. Oft erleben diese dann die sonst seltene Situation, daß die Ziele und Aufgaben einer bestimmten Position explizit erarbeitet werden. Auf diese Weise werden oft Zielkonflikte, Rollenprobleme und andere Schwachpunkte im Unternehmen erkannt. Durch die Einbeziehung der genannten Personen in den Analyseprozeß kann eine Diskussion über die Schwachpunkte der Organisation ausgelöst werden, was wiederum zur Lösung bestehender Probleme führen kann. Darüber hinaus werden sich die Mitarbeiter, die an der Definition der Ziele, Aufgaben und Abhängigkeiten der neuen Position teilhaben, auch für die erfolgreiche Zielerreichung des oder der 'Neuen' verantwortlich fühlen und bei Schwierigkeiten helfend eingreifen. Mindestens aber ist eine bewußte Torpedierung der neuen Stelle nicht von den Mitarbeitern zu erwarten, die an ihrer Konstruktion teilgenommen haben. Dies ist nicht immer selbstverständlich, wenn einem eine neue Stelle 'vor die Nase gesetzt wird', welche möglicherweise eigene Kompetenzen oder Freiheiten einschränkt.

Die 'Meßfühler' bei AC-Verfahren sind im Regelfall spätere Linienvorgesetzte der neuen Mitarbeiter.¹⁵ Damit diese Ihre Aufgabe als Beobachter erfüllen können, werden regelmäßig Beobachtertrainings durchgeführt. Diese sollen durch die Weitergabe psychologischer Erkenntnisse dazu führen, daß der Beobachtungsprozeß optimiert wird, insbesondere daß Beobachtungsfehler vermieden werden. Neben der Funktion der Verbesserung des AC-Verfahrens erfüllen diese Trainings wichtige Personalentwicklungsfunktionen für die Vorgesetzten, da ein Teil ihrer beruflichen Aufgabe in der Beobachtung und Beurteilung von Menschen besteht. Darüber hinaus lernen sie beim Training auch, daß soziale Handlungen oft mehrdeutig sind und von ihren Kollegen teilweise völlig anders beurteilt werden.

Viele Führungsaufgaben bestehen aus Beobachtung, Beurteilung des Beobachteten und Kommunikation des Urteils. Dies ist z.B. der Fall bei der Leistungsbeurteilung untergebener Mitarbeiter, bei Zielvereinbarungsgesprächen und der Zielerreichungsrückmeldung, bei der Einschätzung von Absichten und zukünftigen Handlungen von Kunden, Untergebenen, Kollegen und eigenen Vorgesetzten. Durch das AC-Auswahlssystem werden diese Fähigkeiten ständig geschult und von den Kollegen und evtl. einem Psychologen kritisch überprüft. Insbesondere der Abgleich der Urteile verschiedener Beobachter in Gruppendiskussionen, wie er in vielen AC-Verfahren vorgesehen ist, fördert die Kommunikationsfähigkeit der Beurteiler auch in sensiblen Bereichen.

Ein Vorgesetzter kann beim AC seine zukünftigen Mitarbeiter im Prinzip selbst auswählen. Ihm wird durch das AC lediglich eine standardisierte Entscheidungshilfe angeboten. Durch diese Mitbestimmung wird er den zukünftigen Mitarbeiter sicherlich wesentlich einfacher akzeptieren bzw. sogar an seiner Förderung interessiert sein, als wenn die Personalabteilung ihm den Mitarbeiter auswählt und zuweist. Man muß kein großer Psychologe sein, um zu begründen, warum dies so sein wird: Würde der Vorgesetzte den Mitarbeiter, den er letztendlich selbst ausgewählt hat, nicht akzeptieren, so käme das einem Eingeständnis der eigenen Unzulänglichkeit gleich. Dies kann in besonderen Ausnahmen sicherlich einmal vorkommen, der Regelfall wird es aber nicht sein. Viel lieber beweist man aber der Personalabteilung, welche Fehlentscheidung sie wieder einmal getroffen hat, wenn der neue Mitarbeiter ohne Zutun des Vorgesetzten von ihr zugewiesen wurde.

Dieser Punkt wird in der Literatur unter verschiedensten Bezeichnungen (sich selbst erfüllende Prophezeiungen, Kriterienkontamination, Reaktivitätseffekte, ...) immer wieder dafür verantwortlich gemacht, daß bei der empirischen Validitätsermittlung von AC meist

¹⁵ Eine genauere Beschreibung von AC-Verfahren wird in Kapitel 6 vorgenommen.

vergleichsweise hohe Koeffizienten gefunden werden. Obwohl dieses Thema hier nicht weiter verfolgt werden soll, möchte ich auf einen interessanten Punkt hinweisen: Das Problem der z.B. durch sich selbst erfüllende Prophezeiungen verfälschten Validität der diagnostischen Information existiert nur, wenn das diagnostische Verfahren 'AC' unter dem Meß- und Entscheidungsgesichtspunkt betrachtet wird (vgl. Kapitel 2). Betrachtet man das Personalauswahlprogramm 'Assessment Center unter Einbeziehung der Vorgesetzten in Entscheidungen zur Förderung ihrer Akzeptanz' als Interventionsmaßnahme, verschwindet das Problem. In diesem Fall ist beabsichtigt, daß die Erträge der Maßnahme nicht nur durch die hohe Validität der diagnostischen Information erzeugt werden, sondern ebenso oder vielleicht sogar maßgeblich durch die Veränderung der Akzeptanz der Vorgesetzten von neuen Mitarbeitern und der Motivation, diese Mitarbeiter zu fördern. Werden Vorgesetzte als Beurteiler bei AC eingesetzt, so haben diese die Möglichkeit, sich ein genaues Bild vom Arbeitsmarkt zu verschaffen, da sie die Leistung vieler Bewerber in der AC-Situation beobachten können. Der Nebennutzen des Verfahrens könnte hierbei z.B. darin bestehen, daß diese Vorgesetzten bei verschiedenen Entscheidungen die Lage auf dem Arbeitsmarkt besser beurteilen können, oder darin, daß die Leistung langjähriger Untergebener, die schon zur Gewohnheit geworden ist, vor dem Hintergrund des Mitarbeiterangebotes auf dem Arbeitsmarkt neu bewertet wird.

AC-Verfahren erbringen aber nicht nur durch die Beurteiler einen Nebennutzen für die Organisation. Die Tatsache, daß eine Organisation ein modernes und teures AC-Verfahren bei der Personalauswahl benutzt, wird zumindest bei seinen potentiellen Bewerbern die Wahrnehmung der Organisation beeinflussen. AC erfüllen also die Nebenfunktion der positiven Darstellung des Unternehmens nach außen. Doch nicht nur auf potentielle Bewerber, sondern auch auf potentielle und tatsächliche Kunden, Kooperationspartner und Konkurrenten könnte dieser Teil der Unternehmensidentität Auswirkungen haben.

Nicht nur eine Wirkung nach außen, sondern auch eine Wirkung nach innen kann durch ein gutes AC-System erzielt werden. So könnten z.B. Auswahl-AC, aber auch AC mit anderem Zweck, entscheidend zu einer Unternehmenskultur beitragen, welche durch Transparenz und Fairneß anstatt von Vetternwirtschaft und Vereinsmeierei bei Personalentscheidungen geprägt ist.

Nicht zuletzt können diese Kulturelemente auch auf die Teilnehmer eines Auswahl-AC handlungssteuernd wirken. Durch die Selbstdarstellung der Organisation im AC und die stichprobenartige Darstellung der Inhalte der angebotenen Arbeitstätigkeit können die Teilnehmer selbst beurteilen, ob sie für die Tätigkeit geeignet sind und Interesse haben. Die

darauf wahrscheinlich folgenden Selbstselektionsprozesse tragen mit zum Nutzen von Auswahl-AC bei.

6 Nutzen der Personalauswahl durch ein Assessment Center: Eine empirische Untersuchung

In diesem Kapitel soll zunächst allgemein darauf eingegangen werden, was ein AC ist. Danach konzentriere ich mich auf die Darstellung der empirischen Untersuchung. Hierzu wird zunächst das untersuchte Auswahl-AC, das bei einem Versicherungsunternehmen zur Anwendung kommt, dargestellt. Danach werden die vorliegenden empirischen Daten kritisch diskutiert. Der Schwerpunkt liegt bei der Evaluation des AC-Verfahrens. Da eine umfassende Darstellung der Evaluationsmaßnahmen im Rahmen dieser Arbeit nicht möglich ist, möchte ich lediglich kurz auf einige 'Highlights' eingehen, um mich dann ausführlich der nutzenanalytischen Bewertung des Auswahlprogramms zuzuwenden.

6.1 Was ist ein Assessment Center?

Personalauswahlprogramme können verschiedene diagnostische Verfahren beinhalten. Dies sind z.B. Auswertung der Bewerbungsunterlagen, Einstellungsinterviews, psychologische Testverfahren, Arbeitsproben, biographische Fragebögen, computergestützte Simulationen, Assessment Center, etc. Das AC nimmt meines Erachtens eine Sonderstellung innerhalb dieser Verfahren ein, da es so gut wie alle anderen Verfahren beinhalten kann. Darüber hinaus werden die Ergebnisse von AC nicht nur von Fachleuten der Eignungsdiagnostik interpretiert, wie dies oft bei den anderen Verfahren der Fall ist (also z.B. Psychologen, Personalfachleute), sondern hauptsächlich von den Vorgesetzten der zukünftigen Mitarbeiter selbst. Eine pragmatische Definition von Assessment Centern gibt Obermann (1992, S. 11):

Ein Assessment Center ist

- ein ein- bis dreitägiges Seminar
- mit acht bis zwölf Mitarbeitern oder Bewerbern,
- die von Führungskräften und Personalfachleuten
- in Rollenübungen und Fallstudien
- beobachtet und beurteilt werden.
- Diese Rollenübungen und Fallstudien
- sind charakteristisch für
- bestehende oder zukünftige

- Arbeitssituationen und Aufgabenfelder.

Ich würde diese Definition noch dahingehend erweitern, daß ein AC nicht nur Rollenübungen und Fallstudien umfassen kann, sondern alle denkbaren eignungs diagnostischen Instrumente. Nicht ungewöhnlich in AC sind psychometrische Leistungstests, Persönlichkeitstests, biographische Fragebögen und Interviews. Zu unterstützen ist die Betonung Obermanns, daß die AC-Übungen charakteristisch für die Arbeitstätigkeit sein sollen. Diese Forderung nach der inhaltlichen Validität von AC wird in der Praxis seit dem Erscheinen leichtverdaulicher halbwissenschaftlicher Literatur zu Assessment Centern leider oft genug ignoriert.

AC-Verfahren werden neben der Personalauswahl zu sehr unterschiedlichen Zwecken eingesetzt. In Anlehnung an Schuler (1992) sollen die wichtigsten genannt sein:

- interne Bewerberauswahl
- Auswahl externer Bewerber
- Laufbahnplanung
- Ausbildungsberatung, Berufsberatung
- Beurteilung, insbesondere Potentialberatung
- Trainingsbedarfsanalyse, Bildungsbedarfsanalyse
- Personalentwicklung, Teamentwicklung
- berufliche Rehabilitation
- Arbeitsplatzgestaltung
- Evaluation von Personalentwicklungsmaßnahmen
- Forschung

Im Rahmen der Personalauswahl nehmen AC-Verfahren seit einiger Zeit einen bedeutenden Raum ein. Ihnen wird nicht nur die Erfüllung notwendiger Testgütekriterien wie Objektivität, Reliabilität oder prognostische Validität in einer Vielzahl von Studien bescheinigt. Sie erfüllen gleichermaßen andere berechnete Forderungen an die Methoden der Eignungsdiagnostik. Dies wären beispielsweise die Aspekte der sozialen Validität, wie sie von Schuler und Stehle (1983) vorgeschlagen wurden: Die Verfahren stellen dem Bewerber Informationen über Charakteristika des Arbeitsplatzes und der Organisation zur Verfügung, die Führungskräfte der Organisation werden an der Entwicklung und Durchführung der eignungsdiagnostischen Instrumente beteiligt, die Verfahren sind bzgl. der Durchführung und der daraus gezogenen Schlüsse für Beurteiler und Beurteilte transparent und daher vergleichsweise leicht kommunizierbar. Diese Eigenschaften des AC-Verfahrens haben zu einer hohen Akzeptanz bei Stellenbewerbern als auch bei den Organisationsmitgliedern geführt, welche für die Personalauswahl verantwortlich sind.

Eine Vielzahl von Einzelstudien, Übersichtsartikeln und Metaanalysen (z.B. Hunter und Hunter, 1984; Maukisch, 1986; Schmitz, Gooding, Noe & Kirsch, 1984; Thornton, Gaugler, Rosenthal & Bentson, 1992) beschäftigen sich mit der Frage der prognostischen Validität von AC-Verfahren. Die Validitätskoeffizienten, welche von diesen Arbeiten berichtet werden, liegen im Bereich von $r_{xy} = 0,3$ bis $0,4$.

Diese kurze Einführung soll hier genügen. Seit einiger Zeit gibt es eine Vielzahl einführender, deutschsprachiger Bücher über das AC, wie z.B. Jeserich (1981), Lattmann (1989), Obermann (1992) oder Schuler und Stehle (1983).

6.2 Beschreibung des AC

In diesem Abschnitt möchte ich den Anlauf des AC-Verfahrens und die Datenbasis der Untersuchung beschreiben. Darüber hinaus werden Merkmale der Stichprobe beschrieben und die Art und Weise der Entscheidungsfindung diskutiert.

6.2.1 Der Ablauf des AC und die vorliegenden Daten

6.2.1.1 Ablauf und Prädiktoren

Das AC-Verfahren, welches von dieser Studie untersucht werden soll, wird bei einer mittelgroßen deutschen Versicherungsgesellschaft durchgeführt. Es kommt im Rahmen der Personalauswahl von Versicherungsaußendienstmitarbeitern (AM) zum Einsatz. Das Verfahren wurde von einer Unternehmensberatung konzipiert, welche seit längerer Zeit auf Personalentwicklungsmaßnahmen mittels der AC-Methode spezialisiert ist. Die AC werden unter Leitung eines versicherungsinternen Moderators durchgeführt.

Die AC dauern insgesamt vier Tage. Der erste Tag ist dabei für die Vorbereitung und das Beobachterauffrischungstraining vorgesehen. Am zweiten und dritten Tag finden die eigentlichen AC-Übungen statt. Neben den Übungen werden z.T. den Teilnehmern (Tn) ein Persönlichkeitstest und ein selbstkonstruierter biographischer Fragebogen vorgelegt. Auf diese beiden Elemente des AC werde ich aufgrund mangelnder Datenbasis nicht weiter eingehen. Dies sollte jedoch die Ergebnisse der Studie nicht erheblich beeinflussen, da die Verfahren nicht regelmäßig appliziert wurden, und deren Ergebnisse wenig zur Entscheidung über Annahme oder Ablehnung der Bewerber beigetragen haben.

Der Kern der AC besteht aus 6 *Übungen*. Die Leistung der Tn in den Übungen werden von den Beobachtern auf jeweils verschiedenen Dimensionen eingeschätzt. Leider liegen von den einzelnen Übungen differenziert für die verschiedenen Dimensionen keine Daten vor, sondern lediglich der Mittelwert einer jeden Übung. Auf der Grundlage der protokollierten Übungsleistungen geben die Beobachter im Anschluß an die Übungen auf der Beobachterkonferenz per Gruppenentscheidung ein *Eigenschaftsurteil* über jeden Teilnehmer auf 10 Eigenschaftsdimensionen ab. Diese 10 Eigenschaftsdimensionen entstammen der Tätigkeitsanalyse in der Konstruktionsphase des AC. Sie sollten nach Meinung derjenigen, die an der AC-Konstruktion beteiligt waren, wichtige Determinanten des beruflichen Erfolgs von AM sein. Zur Messung dieser zuvor festgelegten Dimensionen wurden die 6 Übungen konzipiert. Die Namen der 10 Eigenschaftsdimensionen, ihre kurze Beschreibung sowie ihre Mittelwerte und Standardabweichungen sind in Tabelle 6 aufgeführt. In Tabelle 7 finden sich die Bezeichnungen der 6 Übungen, eine kurze Beschreibung ihrer Inhalte sowie deren Gesamtmittelwerte und Standardabweichungen.

Bei jeder Übung schätzen die Beobachter die Tn auf verschiedenen Dimensionen ein. Dieses Beobachterrating wird bei den Übungen auf einer vierstufigen Skala durchgeführt (1 =selten/ kaum beobachtbar, 2=gelegentlich/ ansatzweise beobachtbar, 3=regelmäßig/ ausgeprägt beobachtbar, 4=sehr stark/ überragend beobachtbar). Da wie bereits bemerkt keine differenzierteren Daten verfügbar sind, ist das Übungsergebnis eines Tn das arithmetische Mittel der Einschätzungen über alle Dimensionen und Beobachter innerhalb einer Übung.

Die Urteile über die Ausprägung der Tn auf den Eigenschaftsdimensionen auf der Beobachterkonferenz werden ebenfalls auf der oben beschriebenen vierstufigen Skala abgegeben, wobei hier Zwischeneinstufungen in Form einer Nachkomma-dezimalen möglich sind. Das Eigenschaftsurteil über einen Tn ist das Gruppenurteil der Beobachter auf der Beobachterkonferenz.

Auf der Grundlage dieser Daten entscheiden die Beobachter auf der Beobachterkonferenz über die Annahme oder Ablehnung eines jeden Bewerbers. Die Entscheidung wird dabei nach einer Diskussion per Gruppenkonsens auf der Grundlage impliziter Regeln getroffen. Die Selektionsquote des AC liegt bei etwa $Q \approx 50\%$. Aus datenschutzrechtlichen Gründen liegen für abgelehnte Bewerber keine Daten vor.

Tabelle 6: Die Eigenschaftsdimensionen des AC. In Klammern finden sich die Mittelwerte (M) und Standardabweichungen (s) der Eigenschaftsurteile.

Eigenschaftsdimension	Beschreibung des Eigenschaftskonstruktes
1 Beharrlichkeit	Ziele beharrlich verfolgen; läßt sich nicht vom Wesentlichen abbringen; nicht gemeint: Starrsinnigkeit (M=2,77; s=0,42)
2 Belastbarkeit	emotionalen, kognitiven oder psychischen Anforderungen standhalten; keine Anzeichen von Nervosität, Verkrampftheit zeigen; zeigt normale emotionale Regungen (M=2,91; s=0,42)
3 Initiative	eigene oder fremde Ideen oder Handlungen aktiv vorantreiben; aktives Verhalten; kommt zu Ergebnissen; nicht gemeint: ziel-loser Aktivismus (M=2,70; s=0,42)
4 Kontaktfähigkeit	zwischenmenschliche Beziehungen einleiten und aufbauen; Kontaktangebote geben und annehmen; Blickkontakt (M=2,88; s=0,42)
5 Leistungsmotivation	einen hohen Leistungsanspruch erheben und verfolgen; Ehrgeiz; Zielstrebigkeit; Konfrontationsbereitschaft (M=2,75; s=0,39)
6 Lern-/ Umstellverhalten	eigene Erkenntnisse und Meinungen in Frage stellen, ggf. korrigieren und Neues aufnehmen und anwenden; intellektuelle Fähigkeiten (M=2,72; s=0,45)
7 Persönliches Auftreten	gewandt und verbindlich auftreten und damit die emotionale Grundhaltung des Gesprächspartners positiv beeinflussen (M=2,88; s=0,54)
8 Selbständigkeit	unabhängig von anderen urteilen, entscheiden und handeln; Äußern von konkreten Stellungnahmen; Treffen von Entscheidungen auch bei unvollständiger Information (M=2,85; s=0,47)
9 Selbstvertrauen	Vertrauen in die eigene Person im Handeln zum Ausdruck bringen; Personen und Situationen realistisch einschätzen (M=2,84; s=0,40)
10 Verhandlungsgeschick	Gespräche und Diskussionen in die von einem selbst gewünschte Richtung lenken, diese dort halten und sich damit durchsetzen; nicht: 'jemanden über den Tisch ziehen' (M=2,69; s=0,49)

Tabelle 7: Die Übungen des AC. In Klammern finden sich die Mittelwerte (M) und Standardabweichungen (s) der Übungsergebnisse.

Assessment Center-Übung	Beschreibung des Übungsinhalts
1 Selbstpräsentation	Jeder Teilnehmer stellt sich selbst im Stehen den anderen Tn vor. Hierzu hat er einige Minuten Vorbereitungszeit. (M=2,57; s=0,31)
2 Gruppendiskussion	Zu einem problemhaften Thema soll jeder Teilnehmer sich ein Urteil bilden und dies der Gruppe mitteilen. Anschließende führerlose Gruppendiskussion mit dem Ziel, eine einstimmige Gruppenmeinung zu bilden. (M=2,45; s=0,42)
3 Präsentation	Jeder Teilnehmer soll in einer 15 bis 30-minütigen Vorbereitungszeit Gestaltungsvorschläge für einen gemeinsamen Betriebsausflug erarbeiten. Diese sollen in einer 5-minütigen Präsentation den anderen Tn möglichst überzeugend mitgeteilt werden. (M=2,58; s=0,37)
4 Erkundungsgespräch	Jeder Tn soll nach einer kurzen Vorbereitungszeit ein Gespräch mit einem anderen Tn führen, bei welchem eine bestimmte vorgegebene Meinung des Gesprächspartners erkundet werden soll. Der Gesprächspartner darf dabei nicht ausdrücklich über das Gesprächsziel informiert werden. (M=2,62; s=0,44)
5 Einwandbehandlung	Jeder Tn erhält einen konflikthaften Kurzfall. Er soll dazu persönlich Stellung beziehen und seine Entscheidung begründen. Die Beobachter diskutieren mit dem Tn und konfrontieren ihn mit gezielten Einwänden. (M=2,75; s=0,44)
6 Planung	Jeder Tn erhält eine Liste von fiktiven Aufgaben, die in einer vorgegebenen Zeit und unter bestimmten Randbedingungen zu erledigen sind. Die Zweckmäßigkeit der Reihenfolge der Aufgabenerledigung ist dabei nicht beliebig. Der Tn hat die Aufgabe, die Reihenfolge der Aufgabenausführung zu planen. (M=2,38; s=0,68)

Der vierte Tag des AC ist für Feedback und Einzelgespräche mit den Teilnehmern vorgesehen. Die Anzahl der Teilnehmer je AC liegt zwischen 3 und 8 (M=5,3 s=1,57). Das Verhältnis der Anzahl der Beobachter zu den Tn liegt bei 1:2 oder besser.

6.2.1.2 Umsatzzahlen als Berufserfolgsindikatoren

Als Grundlage für die Einschätzung des Berufserfolgs dient der Umsatz der AM. Eine unternehmensinterne Codierung (UP=Umsatzpunkte) läßt aus der Art und dem Umfang abgeschlossener bzw. gekündigter Versicherungsverträge einen direkten Rückschluß auf den individuellen monatlichen Umsatz und Verdienst eines jeden AM zu. Die Anzahl der Umsatzpunkte eines jeden AM im ersten Jahr nach Eintritt ins Unternehmen dient in dieser Studie als Indikator des Berufserfolgs. Ziel der Indikatorbildung war dabei, daß er monatliche

und saisonale Schwankungen nicht abbilden, aber dennoch die Information des Umsatzes in einem standardisierten Zeitraum beinhalten sollte. Dadurch daß der Indikator V12 durch Summierung der Umsätze eines vollen Kalenderjahres gebildet wird, sollten saisonale Schwankungen zu keinen Umsatzunterschieden zwischen den AM führen. Der Berufserfolgsindikator V12 errechnet sich für jeden AM aus den erreichten UP in den ersten zwölf Monaten nach Tätigkeitsbeginn im Versicherungsunternehmen.

V12 stellt ein objektives Berufserfolgs- und Leistungskriterium dar. Natürlich wird das weitreichende Konstrukt des Berufserfolgs durch V12 nicht umfassend operationalisiert. Andere relevante Teilbereiche des Konstrukts könnten neben dem erzielten Umsatz z.B. die langfristige Sicherung der Kundschaft, die subjektiv erlebte berufliche Zufriedenheit oder der langfristige Verbleib eines AM im Berufsfeld sein. Dennoch kann der Indikator für die beabsichtigten Zwecke als befriedigend angesehen werden, da er einerseits den größten Teil des Berufserfolgs der AM aus der Sicht der Organisation beinhaltet, nämlich die bewertete beruflichen Leistung in DM. Dies ist im Zusammenhang mit der NA wichtig. Andererseits bin ich der Überzeugung, daß durch die Umsatzzahlen weitere wichtige Teile des Berufserfolgskonstrukts operationalisiert werden. Ich habe z.B. im Abschnitt 5.2 (Nutzen 'zweiter Art') darauf hingewiesen, daß beim spezifischen Berufsfeld des Versicherungsaußendienstes eine wesentliche Determinante beruflicher Zufriedenheit der Verdienst ist. Bei der vorliegenden Form der provisionsgebundenen Entlohnung ist der Verdienst aber mit dem Umsatz gleichbedeutend.

Die Studie wurde als Längsschnittstudie angelegt. Sie umfaßt eine Zeitspanne von insgesamt 3 Jahren und AC-Ergebnisse von insgesamt 316 Probanden. Die Umsatzdaten wurden allerdings erst ein Jahr nach Beginn der Studie erfaßt, so daß eine kriteriumsorientierte Validierung des AC für die Probanden, die im ersten Jahr des AC ausgewählt worden sind, nicht möglich ist. Da das Kriterium V12 natürlich erst ein Jahr nach Tätigkeitsbeginn erfaßbar ist, entfallen bei der Studie die Probanden, welche weniger als ein Jahr im Unternehmen sind. Darüber hinaus wird die Stichprobengröße noch einmal durch die hohe Fluktuation der AM gemindert, so daß von insgesamt n=114 AM die erzielten UP im ersten Jahr vorliegen und V12 berechnet werden kann. Sofern nicht ausdrücklich angegeben, beziehen sich die Betrachtungen in diesem Kapitel auf diese Probandengruppe.

6.2.2 Beschreibung der Stichprobe

Die Teilnehmer der AC werden nach ihrer Bewerbung einer Vorauswahl unterzogen. Hierzu dienen Ausschlußkriterien. Diese beziehen sich auf das Alter, nicht vorhandene Berufsausbildung, erhebliche Diskontinuitäten im beruflichen Werdegang, Häufung von Wechseln des Arbeitsplatzes bzw. der Versicherungsgesellschaften, Nichtbesitz der Fahrerlaubnis und auf bestimmte Konstellationen der familiären Verhältnisse. Allgemein kann gesagt werden, daß der Ausschluß durch diese Kriterien nicht zu streng gehandhabt wird, so daß die Stichprobe dadurch nicht erheblich vorselektiert ist.

Wie bereits bemerkt, besteht die Stichprobe (n=114) nur aus positiv beurteilten Probanden. Sie entstammen sehr unterschiedlichen früheren Berufssparten. Sie haben eine unterschiedliche schulische bzw. universitäre Vorbildung und sind im Durchschnitt 36,2 Jahre alt ($s=7,9$). Die Untersuchung wurde in der Zeit nach der 'Maueröffnung' in Deutschland durchgeführt. 21 % der Probanden sind ehemalige DDR-Bürger. Hinsichtlich diesen und weiteren Variablen bestehen keine Unterschiede bzgl. der AC-Leistung und des Umsatzes, so daß keine gesonderte Betrachtung von Unterstichproben durchgeführt wird.

6.2.3 Klinische vs. statistische Urteilsbildung

Die Beobachtung von Rollenspielen bei AC hat zum Ziel, selektiv bestimmte Sachverhalte zu quantifizieren. Die Information, welche die Übungen des AC über jeden Bewerber bereitstellt, ist beim untersuchten AC sowohl in den Übungsergebnissen als auch in den Eigenschaftsurteilen enthalten. Die Ergebnisse der sechs Übungen entstehen beobachtungsorientiert während und kurz nach den AC-Übungen. Im Gegensatz dazu werden die Einschätzungen auf den Eigenschaftsdimensionen zeitlich abgekoppelt von den Übungen auf der Beobachterkonferenz vorgenommen. Grundlage hierfür sind die protokollierten Übungsergebnisse, die Erinnerung an die Übungen sowie der Gesamteindruck eines jeden Bewerbers. Dieser Gesamteindruck entsteht natürlich nicht nur in den formellen Übungssituationen, sondern auch zwischen den Übungen und bei den Abendprogrammen. Wichtig ist, daß diese Eigenschaftsurteile Gruppenentscheidungen sind, die auf der Beobachterkonferenz nach der Diskussion der Beobachter per Konsens abgegeben werden. Sie sind also nicht beobachtungsorientiert, sondern eigenschaftsorientierte Globalurteile nach sozialem Abgleich.

Auf der Beobachterkonferenz wird unter Beachtung der Übungsergebnisse und der geschätzten Eigenschaften auch die Entscheidung über Annahme oder Ablehnung des Bewerbers getroffen. Diese Entscheidungsfindung folgt dabei keinen expliziten Regeln, wie dies z.B. bei der statistischen Urteilsbildung der Fall wäre. Statt dessen findet in der Gruppe eine klinische Urteilsbildung statt.¹⁶ In Abschnitt 2.3.2.1 (Kriterienorientierte Validitätsbestimmung eines eignungsdiagnostischen Verfahrens bei der Personalauswahl) habe ich ausführlich die Konsequenzen abgehandelt, die für die Validierung eines eignungsdiagnostischen Verfahrens entstehen, wenn die Prädiktorinformation mehrdimensional vorliegt, die Entscheidungsregeln für die Behandlung nicht explizit sind und abgelehnte Bewerber nicht in die Organisation aufgenommen werden. In diesem Abschnitt wurde gezeigt, daß eine empirische Validitätsbestimmung in diesem Fall streng genommen nicht möglich ist. Das untersuchte AC hat diesen Nachteil.

Obwohl eine strenge Überprüfung der Validität der tatsächlichen Auswahlentscheidungen nicht möglich ist, kann man trotzdem den Informationsgehalt der Prädiktoren, die das AC liefert, untersuchen. Auf diese Weise kann die prognostische Validität des AC unter Beachtung einer bestimmten Datenkombinationsregel ermittelt werden. Gelingt der Nachweis einer positiven Validität, dann ist allerdings noch lange nicht gesagt, daß die Beobachter bei ihrer Entscheidung gerade diese oder eine ähnliche Datenkombinationsregel angewandt haben. Im schlimmsten Falle haben die Entscheidungen, die sie getroffen haben, keine oder sogar eine negative Validität, obwohl die vorliegende Information eine bedeutsame prognostische Validität besitzt. Dies kann an einem extremen Beispiel leicht verdeutlicht werden. Angenommen die Eigenschaft 'Beharrlichkeit' (1) hätte eine sehr hohe Validität von 0,8 bzgl. des Umsatzes, die Beobachter würden aber ihre Auswahlentscheidung im wesentlichen auf die Eigenschaftsdimensionen 4,5,8 und 9, die keine Validität aufweisen, gründen, dann wäre die Entscheidung nicht valide, obwohl sie es sein könnte. Dies wäre dann natürlich der Fall, wenn die Entscheidungsregel lauten würde: Bringe die Bewerber durch die Eigenschaftswerte der 'Beharrlichkeit' in eine Rangreihe, und wähle von oben nach unten die Bewerber aus, bis die gewünschte Selektionsquote erreicht ist.

Bei der Bestimmung der Validität wird die tatsächliche Auswahlentscheidung in den AC nicht bewertet. Stattdessen wird die Validität der einzelnen Prädiktoren bzgl. des

¹⁶ Die Begriffe 'klinische Urteilsbildung' und 'statistische Urteilsbildung' beziehen sich auf den Unterschied in der Explizitheit der Regeln, nach welchen die Datenerhebung und die Datenkombination erfolgt. Liegen die Daten quantifiziert vor, und erfolgt ihre Kombination durch einen Algorithmus, so spricht man von statistischer Urteilsbildung. Liegen qualitative Daten vor, oder erfolgt ihre Kombination nicht explizit, sondern auf Grundlage von Erfahrung, Fachwissen und Intuition des Diagnostikers, so spricht man von klinischer Urteilsbildung (vgl. z.B. Fisseni, 1990).

Berufserfolgsindikator bestimmt. Mittels der multiplen Regressionsanalyse soll versucht werden, die optimale Kombinationsregel der multiplen Prädiktoren zu schätzen. Diese Kombinationsregel sollte die Auswahlentscheidungen ermöglichen, welche die höchsten Umsätze innerhalb des ersten Jahres erwarten lassen. Darüber hinaus wird hier angenommen, daß das Umsatzergebnis innerhalb des ersten Jahres im Unternehmen ein hochvalider Prädiktor für den weiteren Berufserfolg darstellt.

Mit anderen Worten wird den Beobachtern nicht unterstellt, daß ihre Entscheidung dieses Optimum erreicht hat. Vielmehr stellt der so errechnete Validitätskoeffizient einen Wert dar, der zukünftig innerhalb des PAP erreicht werden kann. Auch die nachfolgenden nutzenanalytischen Schätzungen beziehen sich nicht auf den Nutzen, der mit dem PAP bisher erreicht wurde, sondern auf den Nutzen, der mit ihm erreicht worden wäre, wenn man die optimale Vorhersagekraft der Prädiktorinformation des AC genutzt hätte. Diese Angaben sind gleichzeitig eine Schätzung darüber, welcher Nutzen aus dem Programm unter Ausnutzung der optimalen Datenkombinationsregel zukünftig zu erwarten ist.

Einen Vorteil allerdings haben auch die vermutlich suboptimalen tatsächlichen Auswahlentscheidungen. Paradoxerweise wird man bei der empirischen Validitätsbestimmung auf der Grundlage der Daten der angenommenen Bewerber bei mehrdimensionaler Prädiktorinformation und impliziten Auswahlentscheidungsregeln eine um so höhere Validität der Information finden, je weniger valide die tatsächliche Bewerberauswahl war. War sie hingegen hoch valide, dann wird eine Validierungsstudie lediglich eine niedrige Validität der Prädiktorinformation feststellen. Dieser Effekt stellt sich ein, weil eine valide Auswahl die Gruppe der angenommenen Bewerber bzgl. der validen Prädiktoren homogenisiert hat. Die untersuchte Stichprobe ist dadurch stark selektiert und die Verteilung der validen Prädiktoren in der Stichprobe varianzeingeschränkt. Je weniger valide die tatsächliche Auswahl hingegen ist, desto weniger gravierend sollte die Stichprobenselektion gegenüber der Bewerberpopulation sein.

6.3 Evaluation der Personalauswahl durch das AC

In Abschnitt 2.3.1 wurden in Anlehnung an Wottawa und Thierau (1990) mehrere Kennzeichen wissenschaftlicher Evaluation genannt. Evaluation umfaßt die Bewertung von Handlungsalternativen mit dem Ziel, Planungs- und Entscheidungshilfen zu erhalten. Sie dient dazu, Interventionen zu überprüfen, zu verbessern oder über ihre Durchführung zu entscheiden und bedient sich dazu wissenschaftlicher Methoden und Techniken. Eine

Evaluation benötigt als Grundlage Zielkriterien. Sie hat die Aufgabe, den Grad der Erreichung dieser Ziele zu bewerten, die mit der Intervention erreicht werden sollten. In Abschnitt 2.3 wird die Ansicht vertreten, daß die Validierungsstudie und die Nutzenanalyse spezielle quantitative Formen der Evaluation von Interventionsmaßnahmen sind. Diese Gliederungsmöglichkeit bei der Darstellung der Evaluation des AC-PAP wird beibehalten. Der Schwerpunkt der Darstellung liegt natürlich auf der NA, so daß die Validierungsbemühungen hier nur kurz dargestellt werden. Eine Validierungsstudie des vorliegenden AC wurde ausführlich an anderer Stelle vorgenommen (Weiß, in Vorbereitung).

6.3.1 Validierungsstudie

Um die Validität der Prädiktoren des AC zu beurteilen, werden zunächst die univariaten Produkt-Moment Korrelationen der 6 Übungsergebnisse und der 10 Eigenschaftsurteile mit dem Kriterium V12 berechnet. Die Normalverteilungsannahme der meisten dieser Variablen ist nur schwer haltbar (Kolmogorov-Smirnov goodness-of-fit-Test: $p < 0,05$ für V12, Übung 6 und Eigenschaft 1 bis 10). Aus diesem Grunde wird die univariate Validität der Prädiktoren zusätzlich mittels der Spearman'schen Rangkorrelation berechnet. Die Ergebnisse veranschaulicht Tabelle 8.

Tabelle 8: Validitätskoeffizienten der durch Gruppendiskussion ermittelten Eigenschaftsurteile und der beobachtungsorientierten AC-Übungsergebnisse bezüglich des im Text beschriebenen Außenkriteriums V12. Angegeben ist die Produkt-Moment Korrelation (Pearson), die Rangkorrelation (Spearman) und die Stichprobengröße (n).

Prädiktor	V12 (Pearson)	V12 (Spearman)	n
1 Beharrlichkeit	-.001	.043	105
2 Belastbarkeit	-.053	.022	105
3 Initiative	.177	.154	105
4 Kontaktfähigkeit	.127	.260*	105
5 Leistungsmotivation	.218*	.116	105
6 Lern-/ Umstellverhalten	-.009	.142	105
7 Persönliches Auftreten	.116	.162	105
8 Selbständigkeit	.165	.127	105
9 Selbstvertrauen	.213*	.136	105
10 Verhandlungsgeschick	.184	.180	105
<hr/>			
1 Selbstpräsentation	.030	.153	95
2 Gruppendiskussion	.047	.069	95
3 Präsentation	.028	.172	95
4 Erkundungsgespräch	-.034	.081	94
5 Einwandbehandlung	.157	.040	94
6 Planung	.400**	.337**	56

* p < 0,05 (zweiseitiger Test)
 ** p < 0,01 (zweiseitiger Test)

Die Stichprobengrößen liegen deshalb jeweils unter 114, da z.T. bei den AC-Ergebnissen fehlende Werte auftreten. Dies gilt insbesondere für Übung 6 (Planungsaufgabe), da diese Übung nicht bei jedem AC durchgeführt wird.

Zunächst fällt auf, daß die Prädiktoren, zumindest wenn sie univariat jeweils für sich betrachtet werden, überraschend niedrig valide sind. Bei den Eigenschaftswerten ergeben sich nur bei zweien statistisch signifikante Korrelationen (Leistungsmotivation und Selbstvertrauen). Diese liegen allerdings nur knapp über 0,2 und halten dem nichtparametrischen Prüfverfahren der Spearman'schen Rangkorrelation nicht stand. Das Spearman-Verfahren hingegen weist die Eigenschaft 'Kontaktfähigkeit' als niedrig valide aus, was bei der Pearson'schen Produkt-Moment Korrelation nicht der Fall ist. Insgesamt gesehen

ergeben diese Eigenschaftsprädiktoren das Bild, daß aus ihnen eine Umsatzvorhersage kaum möglich ist. Mindestens ein Teil der Eigenschaftsprädiktoren ist bzgl. des gewählten Kriteriums völlig unvalid.

Die Übungsergebnisse zeigen ein ähnliches Bild. Die Ergebnisse Übungen 1 bis 5 scheinen in keinem Zusammenhang zum Umsatzindikator zu stehen. Allerdings läßt die Planungsübung (6) eine statistisch hochsignifikante und praktisch bedeutsame Validität von deutlich über 0,3 erkennen. Dieser Befund bleibt auch bei der nichtparametrischen Rangkorrelation nach Spearman stabil.

Um die Vorhersagefähigkeit der simultan verwendeten Prädiktoren zu überprüfen und 'um eventuelle Suppressoreffekte zu erkennen, wurden mehrere multiple Regressionsanalysen mit V12 als Kriteriumsvariable durchgeführt. Diese sollten darüber hinaus Hinweise auf die optimale Datenkombinationsregel im Sinne einer statistischen Urteilsbildung liefern. Bei den Ergebnissen der multiplen Regressionsanalyse ist festzuhalten:

- 1) Werden die sechs Übungsergebnisse in das Regressionsmodell aufgenommen, dann trägt nur Übung 6 statistisch signifikant ($p < 0,01$) zur Erklärung des Kriteriums bei.
- 2) Wird das Modell mit den Eigenschaftsurteilen gebildet, so leistet nur die Eigenschaftsdimension 5 einen statistisch signifikanten Erklärungsbeitrag ($p < 0,05$).
- 3) Werden alle Prädiktoren zur Modellbildung herangezogen, so erbringt nur die Übung 6 einen statistisch signifikanten Erklärungsbeitrag ($p < 0,01$).

Werden alle nichtsignifikanten Prädiktoren aus dem Regressionsmodell entfernt, dann verbleibt lediglich die Planungsübung (6) im Modell. Dieses Ergebnis ist mindestens hinsichtlich zweier Punkte interessant.

Zum ersten bestätigt das Ergebnis die häufig geäußerte Hypothese, daß die 'mechanisch' erstellten AC-Prädiktoren, die direkt auf den AC-Übungen basieren und beobachtungsorientiert sind, eine mindestens gleich hohe oder höhere kriteriumsorientierte Validität erbringen, als die eigenschaftsorientierten Ergebnisse, welche durch den sozialen Urteilsbildungsprozeß in der Beobachterkonferenz entstehen (Feltham, 1988; Pynes, Bernardin, Benton & McEvoy, 1988; Pynes & Bernardin, 1992). Durch die Beobachterkonferenz werden die vorher z.T. guten Ergebnisse der AC-Übungen sozusagen 'verwässert' und die Validität der Prädiktorinformation zerstört. Allerdings kann daraus nicht direkt der Schluß abgeleitet werden, daß die Beobachterkonferenz überflüssig ist und deshalb nicht stattfinden sollte. Sie erfüllt andere wichtige Funktionen. So wird durch sie die Grundlage für ein verständlich kommunizierbares Feedback an die Tn geschaffen. Die Beobachter haben außerdem bei der Konferenz die Möglichkeit des sozialen Vergleichs,

wodurch eine kritische Überprüfung und Verbesserung ihrer Beobachtungsleistung möglich wird. Allerdings legen die Befunde nahe, daß die Auswahlentscheidung nicht durch Gruppenentscheidung auf der Beobachterkonferenz gefällt werden sollte, sondern mittels statistischer Urteilsbildung auf der Basis der beobachtungsorientierten Übungsergebnisse.

Der zweite interessante Befund ist überraschend. Offensichtlich erbringt nur eine der sechs AC-Übungen valide Ergebnisse bzgl. des Berufserfolgsindikators V12. Kann daraus geschlossen werden, daß das AC stark verkürzt werden könnte, ohne in seinen positiven Funktionen beeinträchtigt zu werden?

Um diese Frage zu beantworten, wurde eine Faktorenanalyse mit den Übungen 1 bis 6 durchgeführt. Per Hauptkomponentenanalyse wurden zwei Faktoren extrahiert, was sowohl durch den Scree-Test (Gatten, 1966) als auch durch das Kaiser-Guttman Kriterium (Guttman, 1954, Kaiser & Dickmann, 1959, zit. nach Bortz, 1989) indiziert ist (vgl. Abbildung 10).

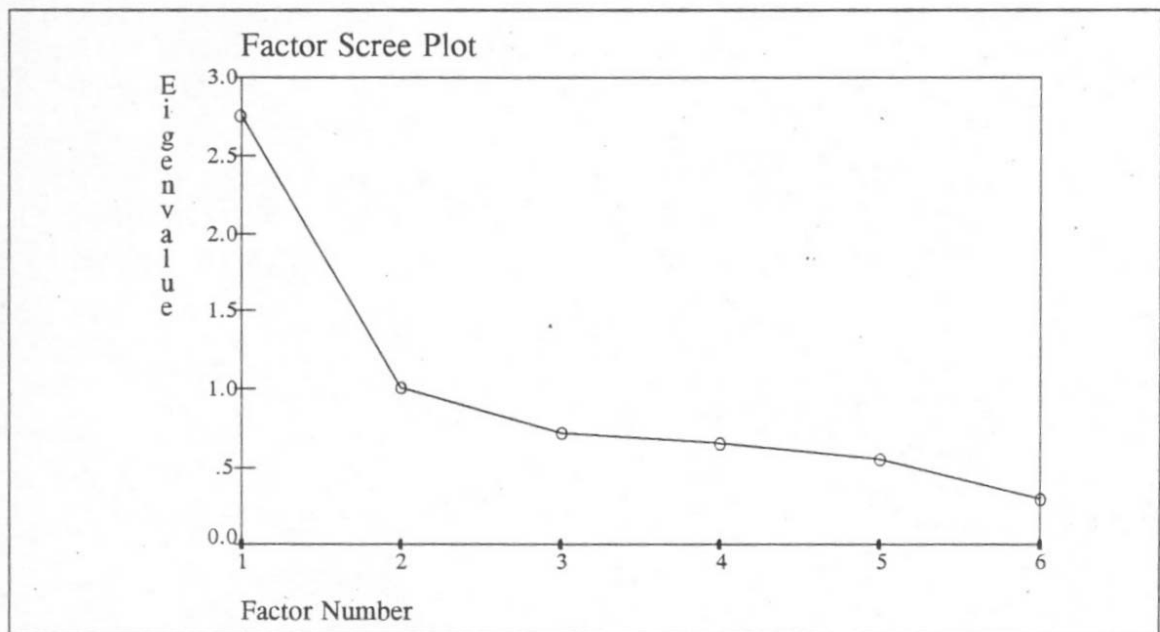


Abbildung 10: Der Scree-Test und das Kaiser-Guttman Kriterium schlagen vor, zwei Faktoren aus den sechs Übungsergebnissen zu extrahieren. Abgetragen sind auf der Ordinate die Eigenwerte der Faktoren und Einzelrestfaktoren nach der Hauptkomponentenanalyse.

Die beiden Faktoren erklären insgesamt 62,9 % der Varianz der Übungsergebnisse. Interessant ist die rotierte Faktorenlösung (Varimax). Faktor 1 wird weitgehend durch die Übungen 1 bis 5 repräsentiert, während die Übung 6 ein nahezu perfekter Indikator des orthogonalen zweiten Faktors darstellt. Abbildung 11 zeigt das Schaubild der Faktorenladungen aller Übungen. Die Kreise stellen dabei die Übungen dar. Übung 6 lädt auf Faktor 2 sehr hoch. Die Punktwolke mit den hohen Ladungen auf Faktor 1 sind die Ergebnisse der Übungen 1 bis 5. Sie haben alle niedrige Ladungen auf dem zweiten Faktor.

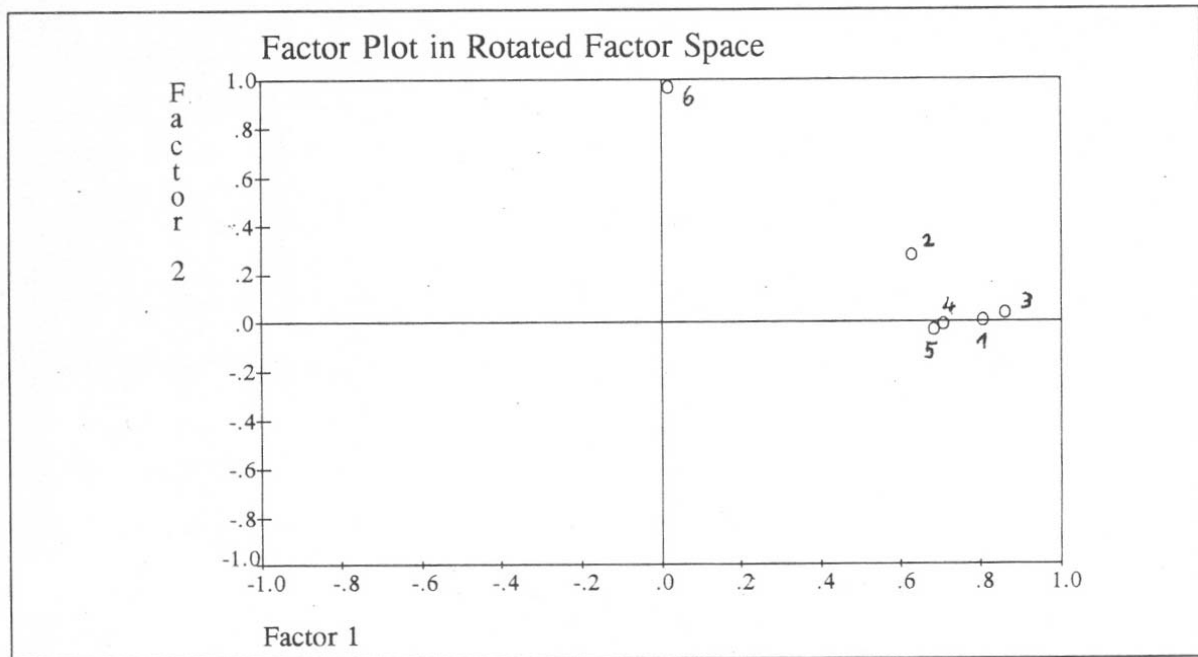


Abbildung 11: Die rotierten Faktorenloadungen der sechs AC-Übungen. Faktor 2 wird durch die Planungsübung (Übung 6) repräsentiert, die anderen Übungen laden hoch auf Faktor 1 ($n=130$).

Offensichtlich sind die Übungsergebnisse 1 bis 5 hoch miteinander korreliert, es fällt also durch die Übungen 1 bis 5 sehr viel redundante Information an. Im Gegensatz dazu wird augenscheinlich bei Übung 6 etwas anderes gemessen als in den anderen Übungen. Darüber hinaus kann auf der Grundlage der Information des Faktors 2 der Umsatz vergleichsweise gut vorhergesagt werden, während dies Faktor 1 nicht gestattet, wie schon die Validitätskoeffizienten in Tabelle 8 belegen. Wird eine multiple Regression von V12 auf die Faktorwerte berechnet, so ergibt sich ein multipler Korrelationskoeffizient von $R=0,403$. Das Betagewicht des Faktors 1 ist niedrig und statistisch nicht signifikant ($\beta=0,096$ $p > 0,45$), während der Faktor 2 zur Umsatzvorhersage geeignet ist ($\beta=0,403$ $p < 0,01$).

6.3.2 Diskussion der Validierungsstudie

Es kann m.E. aus diesen Befunden der Schluß gezogen werden, daß die Auswahlentscheidung nur auf der Planungsübung basieren sollte, da ein solcher Prädiktor die höchste Validität bzgl. des Umsatzkriteriums hat. Des weiteren wird deutlich, daß es falsch war, die Planungsübung nicht bei jedem AC einzusetzen. Sie erbringt nicht nur die validesten Prädiktoren, sondern auch neue Information, während die Werte der anderen fünf Übungen hoch miteinander korreliert und weitgehend unvalide sind. Bevor weitreichende Änderungen des AC-Verfahrens vorgenommen werden, sollten diese Behauptungen zur Vorsicht an einer neuen

Stichprobe noch einmal überprüft werden, da die Stichprobengrößen an der unteren Grenze für einigermaßen zuverlässige Aussagen liegen. Auf dieser Grundlage können dann konkrete Handlungsempfehlungen erarbeitet werden, wie das AC umstrukturiert werden sollte.

Vorsicht ist dabei in jedem Fall aus zwei Gründen geboten. Zum ersten sollte durch eine Änderung des AC in keinem Fall eine Verringerung der sozialen Validität in Kauf genommen werden. So erfüllen die AC-Übungen noch andere Funktionen als die Gewinnung möglichst hoch valider Information bzgl. der Prognose des Kriteriums. Beispielsweise hat die Selbstpräsentationsübung eine wichtige 'Eisbrecherfunktion'. In der Gruppendiskussion haben die Tn die Möglichkeit, sich gegenseitig im direkten Wettstreit zu vergleichen, so daß nicht nur für die ausgewählten Bewerber wichtige Erkenntnisse daraus resultieren können. Zweitens muß bedacht werden, daß die ermittelte Validität sich ausschließlich auf den Berufserfolgsindikator V12 bezieht. Dieser deckt aber nur einen Teil des Konstruktes 'Berufserfolg' ab, nämlich den erzielten Umsatz im ersten Jahr. Wichtige weitere Inhalte werden von ihm nicht berücksichtigt, wie z.B. Kollegialität, Schaffung von langfristigen Vertrauensverhältnissen zu den Kunden, Führungs- und Innovationspotential, etc. Darüber hinaus konnte in V12 nur der Umsatz im ersten Tätigkeitsjahr Eingang finden, da die Dauer der Studie noch zu kurz ist, um auch langfristige Erfolge untersuchen zu können. Da die Datensammlung noch anhält, muß in einer späteren Studie untersucht werden, ob einige andere Übungen bzw. Eigenschaftsprädiktoren bei der langfristigen Umsatzvorhersage möglicherweise doch prognostisch valide sind.

Eine Bemerkung zur Validierungsstudie sollte noch gemacht werden. Wenn beachtet wird, daß Thornton et al. (1992) nach einer Metaanalyse die durchschnittliche Validität von AC mit $r_{xy} = 0,37$ beziffern, so fällt auf, daß die Validität des untersuchten AC mit der Planungsübung als alleinigem Prädiktor nicht niedriger liegt als andere AC-Verfahren. Dies ist insofern interessant, da die metaanalytisch untersuchten Validierungsstudien 'weiche' Berufserfolgskriterien benutzen. Dies sind z.B. Vorgesetzten-, Gleichgestellten- und Eigenbeurteilung der Arbeitsleistung auf der Grundlage der Erinnerung oder von Verhaltensbeobachtungen, Einschätzungen des Potentials von Führungskräften, Leistungen bei Trainings- oder Fortbildungsseminaren, Karriereentwicklungsvariablen, wie die Anzahl von Beförderungen, Gehaltshöhe oder Gehaltsentwicklung im Laufe der Zeit, Stellenwechsel, etc. Bei allen diesen Validierungsstudien kann eine Alternativinterpretation die empirischen ermittelten Validitätskoeffizienten erklären: Da die Führungskräfte einer Organisation sowohl bei der Entwicklung des AC-Verfahrens als auch bei der Beurteilung potentieller Stelleninhaber beteiligt sind, fließt ihr Normen- und Wertesystem und ihre Vorstellung über

die Eigenschaften erfolgreicher Mitarbeiter massiv in das diagnostische Urteil über einen Bewerber ein. Auf der anderen Seite werden nun Leistungskriterien zur prognostischen Validierung der Urteile aus AC herangezogen, die ebenso durch die Normen, Werte und Vorstellungen des Führungspersonals beeinflusst werden. Ausgehend von der Annahme, daß organisationale Sozialisationsprozesse zu einer Homogenisierung der Normen, Werte und Vorstellungen der (noch verbliebenen) Führungskräfte führen, wird immer ein Zusammenhang zwischen Prädiktor und Kriterium zu finden sein. Dieser als Kriterienkontamination bekannte Effekt führt zu einer Überschätzung der tatsächlichen Validität (z.B. Klimoski & Strickland, 1977; Maukisch, 1986).

Diese Alternativinterpretation des Befundes ist bei den objektiven Umsatzdaten, die diese Studie als Berufserfolgskriterium verwendet, nur schwer zu halten.

6.4 Nutzenanalyse

In diesem Abschnitt soll nun exemplarisch eine Nutzenanalyse dargestellt werden. Die Vorgehensweise kann in drei Stufen eingeteilt werden. Auf der ersten Stufe hat sich der Nutzenanalytiker für ein Nutzenmodell zu entscheiden. Hierbei gibt es z.B. die Wahlalternativen zwischen dem Taylor-Russel Modell oder dem BCGModell. Fällt die Entscheidung, wie in diesem Falle, auf das BCG-Modell, so ist weiter zu überlegen, welche Erweiterungskonzepte mit in die Nutzenbetrachtungen einfließen sollen. Stufe 2 erfordert die Schätzung der Parameter des Nutzenmodells. Bisherige nutzenanalytische Untersuchungen legen nahe, daß insbesondere die Schätzung des Parameters SD_y schwierig ist. Auch bei dieser Studie wird dieses Problem auftreten.

Man kann davon ausgehen, daß im Regelfall nicht alle Parameter innerhalb enger Vertrauensbereiche geschätzt werden können. Aus diesem Grunde verlangt jede gute Nutzenanalyse auch eine kritische Diskussion ihrer Ergebnisse. Diese wird auf der dritten Stufe vorgenommen, indem für kritische Parameter beispielsweise Break-even-points angegeben werden. Diese Punkte, auch Kostendeckungspunkte genannt, geben an, bei welchem Wert eines bestimmten Parameters die Erträge gleich den Kosten sind. Die Vorgehensweise hat den Vorteil, daß den Kritikern der Parameteroperationalisierung bei einem Nutzenmodell eine untere Schranke eines Parameterwertes geboten wird, ab welcher der Nutzen eines PAP negativ wird.

6.4.1 Das verwendete Nutzenmodell

Bei dieser Nutzenanalyse findet das Modell Verwendung, welches in Gleichung (29) ausgedrückt ist. Es basiert auf dem ursprünglichen BCG-Nutzenmodell und wurde von mir so formuliert, daß es die Überlegungen von Cronshaw und Alexander (1985). und Boudreau (1983a, 1983b) beinhaltet. Der Nutzen des PAP, welcher über die Überlegungen des Modells (29) hinausgeht, und wie er in den Abschnitten 5.2 und 5.4 beschrieben wurde ('Nutzen zweiter und dritter Art'), soll hier nicht weiter betrachtet werden. Dies liegt weniger daran, daß er nicht wichtig ist, sondern daß er kaum quantifiziert werden kann. Da aber ein wesentliches Ziel dieses sechsten Kapitels darin besteht, die Vorgehensweise einer quantitativen NA zu illustrieren, steht die Schätzung der Parameter und deren Anwendung im Nutzenmodell im Vordergrund. Eine weitere Nutzenabschätzung, die den Nutzen 'zweiter und dritter Art' berücksichtigt, ist sicherlich sinnvoll. Allerdings sind die Modelle, wie die Abschnitte 5.2 und 5.4 deutlich machen, hierzu noch zu wenig entwickelt, so daß deren Berücksichtigung und Diskussion den Rahmen dieser Arbeit sprengen würden.

6.4.2 Operationalisierung der Parameter

In diesem Abschnitt diskutiere ich die Operationalisierung der 'einfachen' Parameter: Die Selektionsquote Q und λ , die Validität r_{xw} , die festen und variablen Kosten des PAP C_0 und C_1 , die Dauer und Anzahl (F) der Behandlungsperioden k , die Anzahl der AM in jeder Behandlungsperiode N_k , den Steuersatz TAX und den Kalkulationszinsfuß i . Im nächsten Abschnitt soll es dann um den 'Achillesfersen'-Parameter SD_y bzw. SD_w und V gehen.

6.4.2.1 Die Selektionsquote und Lambda

Die Selektionsquote wird auf ca. 50% geschätzt. Aus datenschutzrechtlichen Gründen liegen keine Daten über abgelehnte Bewerber vor, so daß eine genauere Bestimmung dieser Quote nicht möglich ist. In Zeiten der Personalknappheit wird diese Quote gelegentlich angehoben. Daß dies sehr negative Folgen für den Nutzen des PAP haben kann, wurde schon mehrfach angesprochen, insbesondere im Zusammenhang mit dem Taylor-Russel Modell (Abschnitt 3.2).

Die Selektionsquote wird für die folgenden Betrachtungen auf $Q=0,5$ geschätzt. Der Funktionswert der Normalverteilungsfunktion an der Stelle $Q=50\%$ ist $\lambda_{(Q)}=0,399$. Gemäß

Gleichung (3) berechnet sich damit der durchschnittliche standardisierte Prädiktorwert der ausgewählten Bewerbergruppe auf

$$(3) \quad \bar{z}_x^* = 0,399 / 0,5 = 0,8$$

6.4.2.2 Behandlungsperioden, Anzahl der Bewerber und behandelten Mitarbeiter

Die Datenerfassung der AC-Ergebnisse beginnen im November 1989 und enden im Oktober 1992. Umsatzdaten und Fluktuationsdaten liegen darüber hinaus ab November 1991 bis Dezember 1992 vor. Die Dauer der Behandlungsperioden k wird willkürlich auf 1 Jahr festgelegt. Die Anzahl der ausgewählten AM und der ausgeschiedenen AM in den ersten drei Behandlungsperioden wurde empirisch ermittelt. Die Anzahl der Bewerber je AM kann unter Zuhilfenahme der durchschnittlichen Selektionsquote geschätzt werden. Da zum Zeitpunkt der Erstellung dieser Arbeit das PAP noch nicht beendet war, gehe ich bei den weiteren Berechnungen davon aus, daß es 5 Jahre in nicht erheblich veränderter Form zum Einsatz kommt. Die Anzahl der AM, die nach der Datenerfassung in das Unternehmen ein- und austreten, wird durch Interpolation geschätzt. Tabelle 9 veranschaulicht die z.T. geschätzte Anzahl der Bewerber, der aufgenommenen AM und der ausgeschiedenen AM je Behandlungsperiode. Dabei wird angenommen, daß das PAP 5 Jahre zum Einsatz kommt. Aus der Tabelle ist ersichtlich, daß das PAP eine Wirkdauer von 6 Jahren hat.

6.4.2.3 Die Kosten

Das Nutzenmodell, welches in Gleichung (29) dargestellt ist, verlangt die Trennung der festen Kosten des PAP (sie entstehen z.B. bei der Entwicklung, Implementation, Evaluation, Optimierung, ...) von den variablen Kosten des PAP, welche in einer bestimmten Behandlungsperiode k entstehen. Die festen Kosten enthalten nicht die anteiligen Kosten, welche der Organisation beispielsweise dadurch entstehen, daß die Personalabteilung das PAP betreuen muß oder daß sie Personalakquisition betreibt. Diese Kosten, so wird angenommen, fallen auch bei jedem anderem PAP an, auch bei einem, in welchem die Bewerber per Zufall ausgewählt würden. Da sich BCG-Nutzenmodelle nicht auf den absoluten Gesamt nutzen, sondern nur den inkrementellen Nutzen von PAP beziehen (auf der Ertragsseite ist der Bezug, wie schon bemerkt, implizit im Validitätskoeffizienten enthalten), ist diese Behandlung des Problems meines Erachtens gerechtfertigt.

Tabelle 9: Die Mitarbeiterkennwerte in den einzelnen Behandlungsperioden. Die mit "*" gekennzeichneten Werte sind Schätzwerte.

Behandlungs- periode k	Zeitraum	Anzahl der Bewerber	Akzeptierte Bewerber (N_{ak})	Ausgeschie- dene AM (N_{zk})	Anzahl der behandelten AM in k (N_k)
1	11/89-10/90	178	89	0	89
2	11/90-10/91	252	126	0	215
3	11/91-10/92	200	100	98	217
4	11/92-10/93	210*	105*	132	190*
5	11/93-10/94	210*	105*	132*	163*
6	11/94-10/95	0*	0*	132*	31*
7	11/95-10/96	0*	0*	132*	0*

Die Schätzungen der Kostenparameter müssen in dieser Studie relativ willkürlich vorgenommen werden. Diese Sachlage entstand aus der nicht befriedigend differenzierten Datenlage. Ich trenne die festen Kosten in die Entwicklungs- und Implementationskosten (geschätzter Wert 100000 DM), die grundlegenden Ausbildungskosten für die 28 Moderatoren des AC und die Grundausbildung für die 116 Beobachter. Die Kosten für die Moderatoren- und Beobachterausbildung werden dabei jedoch nur soweit berechnet, wie sie über denjenigen liegen, die bei einem 'einfachen' PAP (welches z.B. ein Einstellungsinterview als diagnostisches Verfahren vorsieht) entstehen würden. Es werden erweiterte Ausbildungskosten von 1000 DM je Moderator und 500 DM je Beobachter veranschlagt.

Entwicklung und Implementation des AC	100 000 DM
Erweiterte Ausbildungskosten der Moderatoren	28 000 DM
Erweiterte Ausbildungskosten der Beobachter	58 000 DM
Gesamt	<u>186 000 DM</u>

Die geschätzten inkrementellen festen Kosten des PAP belaufen sich damit auf 186000 DM.

Neben den festen Kosten sind die variablen Kosten des PAP je Bewerber zu schätzen. Das Versicherungsunternehmen geht bei seinen Kalkulationen davon aus, daß die AC-Beobachter im Vergleich zu anderen PAP keine zusätzlichen Personalkosten verursachen, da sie gleichzeitig Vorgesetzte sind und ihr Personal selbst auswählen müssen. Im Falle eines PAP, welches z.B. als diagnostisches Verfahren Einstellungsinterviews vorsieht, wären die Vorgesetzten statt der Beobachtungstätigkeit damit beschäftigt, Einstellungsinterviews zu führen. Zusätzliche Personalkosten entstehen lediglich für die Moderatoren. Diese werden je

4-tägigem AC auf 2000 DM geschätzt. Aus der Anzahl der eingestellten AM und der Selektionsquote kann die Anzahl der Bewerber je Behandlungsperiode geschätzt werden. Da durchschnittlich 5,3 Bewerber an einem AC teilnehmen, kann des weiteren die Anzahl der AC aus den Bewerberzahlen geschätzt werden. Daraus wiederum können die gesamten Personalkosten geschätzt werden, welche durch die Moderatoren verursacht werden.

Neben diesen Personalkosten entstehen Unterbringungs- und Verpflegungskosten für Moderatoren, Beobachter und Bewerber. Diese werden vom Versicherungsunternehmen mit 100 DM je Tag und Person angegeben. Dies bedeutet also für die Moderatoren und Beobachter jeweils 400 DM je AC und für die Tn 300 DM je AC. Des weiteren wird eine Materialkostenpauschale je Bewerber von 50 DM angenommen. Die Anzahl der Beobachter wird mit dem durchschnittlich 0,6-fachen der Bewerberanzahl angenommen, da das Verhältnis Teilnehmer-Beobachter 1:2 oder leicht besser ist. Die geschätzten variablen Kosten in 5 angenommenen Anwendungsjahren des PAP sind aus Tabelle 10 ersichtlich.

Tabelle 10: Die variablen Kosten des PAP in den 5 Anwendungsjahren des AC. Angegeben sind geschätzte Mehrkosten gegenüber einem fiktiven einfachen PAP, welches z.B. Auswahlinterviews beinhaltet. Die Berechnungsvorschriften sind im Text genannt.

k	Anzahl der AC	variable Kosten: Moderatoren (in DM)	Anzahl Beobachter	variable Kosten: Beobachter (in DM)	Anzahl Bewerber	variable Kosten: Bewerber (in DM)	gesamte variable Kosten in k (C_{1k}) (in DM)
1	34	81600	107	42800	178	62300	186700
2	48	115200	151	60400	252	88200	263800
3	38	91200	120	48000	200	70000	209200
4	40	96000	126	50400	210	73500	219900
5	40	96000	126	50400	210	73500	219900

Diese geschätzten variablen Kosten lassen sich auch durch ihren Betrag, der auf die Anzahl der Bewerber bezogen ist, ausdrücken. Die variablen Kosten des PAP je Bewerber belaufen sich auf ca. 1050 DM. Wenn man bedenkt, daß diese Zahl die zusätzlichen Kosten der Personalauswahl mittels dem AC-Verfahren gegenüber einem herkömmlichen Verfahren ausdrücken soll und wenn berücksichtigt wird, daß die AC fast jede Woche routinemäßig von unternehmensinternem Personal durchgeführt werden, welche dadurch von anderen Personalauswahlaufgaben entlastet werden, so dürfte die Schätzung realistisch sein.

6.4.2.4 Die Validität

Wie in Abschnitt 6.2.3 gezeigt wurde, ist es streng genommen nicht möglich, die kriteriumsorientierte Validität des AC, wie es bisher durchgeführt wurde, zu bestimmen, da keine expliziten Entscheidungsregeln vorliegen und die AC-Ergebnisse und die Leistung der abgelehnten Bewerber nicht bestimmbar sind. Es wird für die folgenden Betrachtungen angenommen, daß die Bewerberauswahl des AC zukünftig durch die beobachtungsorientierten Ergebnisse der Planungsübung (Übung 6) erfolgt. Wenn der Validierungsstudie Glauben geschenkt werden darf, beruht die optimale Personalauswahlentscheidungsregel auf diesem Prädiktor. Die empirisch ermittelte Validität der Ergebnisse aus Übung 6 liegt zwischen 0,34 und 0,4. Diese Schätzwerte stehen in Einklang mit der generalisierten Validitätsschätzung von AC durch die Metaanalyse von Thornton et al. (1992).

Bei der Bewertung des Validitätskoeffizienten muß bedacht werden, daß dieser nach der Vorauswahl durch Ausschlußkriterien und nach der weiteren Stichprobenselektion durch die tatsächliche Personalauswahl berechnet wurde. Wenn man annimmt, daß die Entscheidungen, welche bei einem einfachen Auswahlprogramm gefällt werden, höchstens die Qualität der Auswahlentscheidungen auf Grundlage der AC-Information, wie sie tatsächlich getroffen wurden, erreichen, dann kann der ermittelte Validitätskoeffizient als eine Näherung für die inkrementelle Validität des optimalen Auswahlverfahrens gegenüber einem einfachen Verfahren angesehen werden. Das optimale Verfahren verwendet als Prädiktor, wie bereits gesagt, die Ergebnisse der Planungsübung (Übung 6). Die tatsächliche Entscheidung hat bei dieser Untersuchung per Stichprobenselektion sozusagen den Anteil der Validität, den ein einfaches Verfahren haben würde, an der Validität des optimalen Verfahrens 'herauspartialisiert'.

Man kann diese Annahme durch ein Gedankenexperiment veranschaulichen: Wenn z.B. Einstellungsinterviews innerhalb des AC durchgeführt worden wären, und deren Ergebnisse von einem Interviewer quantifiziert worden wären, der nur Informationen über die T_n aus dem Interview hätte, dann wäre eine empirisch ermittelte Validität von kaum über 0 zu erwarten, sofern die Validität durch eine Validierungsstudie ermittelt würde, welche auch auf der Gruppe der angenommenen Bewerber basiert. Dies ist so, weil die Stichprobe der Validierungsstudie durch die tatsächliche Auswahl selektiert ist und die tatsächliche Auswahl durch das AC als ähnlich valide angenommen wird, wie die Auswahl auf der Basis des Interviews.

Ich gehe deshalb im weiteren Verlauf davon aus, daß das AC der vorliegenden Studie eine inkrementelle Validität von $r_{xy}=0,3$ hat. Der Vergleichsmaßstab ist dabei, wie bei den Kostenbetrachtungen, ein einfaches PAP, bei welchem z.B. die diagnostische Information durch Auswahlinterviews gesammelt wird.

6.4.2.5 Steuersatz und Kalkulationszinsfuß

Bei der vorliegenden NA wird davon ausgegangen, daß das Kapital mit durchschnittlich $i=0,1$ bzw. 10% verzinst wird. Ich werde die Versteuerung von Kosten und Erträgen zunächst mit einem Steuersatz von $TAX=0,4$ vornehmen. Es werden auch die Ergebnisse einer NA unter Vernachlässigung der Besteuerung angegeben ($TAX = 0$).

6.4.3 Operationalisierung von SD_y

Ich gehe davon aus, daß der Anteil der variablen Kosten am Verkaufswert der Leistung (V) in diesem Berufsfeld besonders hoch ist. Dies kann durch folgende Überlegungen illustriert werden: Die Umsätze, welche die AM machen, beziehen sich auf die Beitragseinnahmen, die das Versicherungsunternehmen in einem bestimmten Zeitraum macht. Die Höhe der Beitragseinnahmen, die ein AM in einem Jahr erwirtschaftet, kann natürlich keinesfalls als dessen Leistung y gewertet werden. Die Leistung y wurde in Abschnitt 4.4 als der Deckungsbeitrag der Leistung, welche ein AM erbringt, definiert. Die erwirtschafteten Beitragseinnahmen entsprechen dem Verkaufswert der Leistung w . Von diesen müssen im Versicherungsbereich erhebliche variable Kosten der Leistung (s) bestritten werden. Dies sind z.B. die Versicherungsleistungen, die Provision der AM, die Superprovision der Vorgesetzten der AM und die zusätzlichen Verwaltungskosten, die aus den abgeschlossenen Versicherungsverträgen entstehen. Alle diese Kosten der Leistung sind variable Kosten, d.h. es besteht zwischen ihnen und der Menge der Beitragseinnahmen ein positiver Zusammenhang, die Kosten sind ausbringungsmengenabhängig. Der Deckungsbeitrag der Leistung y eines AM berechnet sich aus den erwirtschafteten Beitragseinnahmen abzüglich der variablen Kosten dieser Leistungen. Um die Standardabweichung der Leistung SD_y zu ermitteln, sind zwei Probleme zu lösen. Zum einen muß die Standardabweichung der erwirtschafteten Beitragseinnahmen zwischen den AM pro Behandlungsperiode bekannt sein (SD_w). Zum anderen verlangt die Schätzung von SD_y den Anteil der variablen Kosten an diesen Beitragseinnahmen (V). Beide Parameter sind schwierig zu schätzen.

6.4.3.1 Standardabweichung der Beitragseinnahmen zwischen den AM (SD_w)

Wie bereits im Zusammenhang mit der Validierungsstudie des AC bemerkt wurde, sind die sog. Umsatzpunkte (UP) bekannt, die jeder Mitarbeiter im ersten Jahr nach Tätigkeitsbeginn im Versicherungsunternehmen erwirtschaftet hat. Diese UP werden einem AM bei jedem Versicherungsvertragsabschluß gutgeschrieben. Ihre Anzahl ist abhängig von der Versicherungssparte (z.B. Kfz-Haftpflichtversicherung, Lebensversicherung, Krankenversicherung, ...) und der Höhe des Versicherungsvertrages. Die UP bilden die Grundlage für die Berechnung der Provision der AM.

Die UP stehen in indirektem Zusammenhang mit den Beitragseinnahmen, die das Versicherungsunternehmen durch den Vertragsabschluß erhält. Es läßt sich jedoch keine allgemeingültige und direkte Proportion zwischen den UP und den Beitragseinnahmen angeben, da diese für die Sparten unterschiedlich ist. Wünschenswert wären empirische Daten, die darüber Aussagen machen, wieviele UP jeder AM innerhalb jeder Sparte erwirtschaftet hat. Würden auch noch Daten über die Laufzeit der Verträge vorliegen, ließe sich ohne Schwierigkeit die jährlichen Beitragseinnahmen berechnen, die auf den Vertragsabschlüssen eines jeden AM beruhen. Leider sind diese differenzierten Daten in dieser Untersuchung nicht zugänglich. Statt dessen muß geschätzt werden, in welchem durchschnittlichen Verhältnis die UP zu den jährlichen Beitragseinnahmen stehen. Auf der Grundlage dieser Schätzung und der Standardabweichung der UP über die AM kann dann die Standardabweichung des Verkaufswertes der Leistung der AM für das Unternehmen geschätzt werden.

Zum Zwecke der Bestimmung des durchschnittlichen Verhältnisses der UP zu den jährlichen Beitragseinnahmen wurden mehrere Interviews mit Kalkulationsexperten im Versicherungswesen durchgeführt. Um eine eventuell vorliegende Spezifität der untersuchten Versicherungsgesellschaft zu berücksichtigen, wurden auch andere Versicherungsgesellschaften in diese Befragung miteinbezogen. Um die Vergleichbarkeit der Ergebnisse zu erreichen, liegt den folgenden Ausführungen die Annahme zugrunde, daß es sich bei den AM um selbständige Handelsvertreter nach § 84 HGB (Handelsgesetzbuch) handelt. Diese werden aus schließlich über Provisionszahlungen entlohnt. Da die Versicherungsunternehmen verschiedene interne Verrechnungssysteme zur Bestimmung der Umsatzleistung haben, wurde versucht, das durchschnittliche Verhältnis der Provision der AM zu den jährlichen Beitragseinnahmen zu bestimmen. Da die Provision der AM in direktem linearen Zusammenhang zu den UP unseres Versicherungsunternehmens stehen und

die Umrechnungsvorschrift bekannt ist, kann daraus dann das Verhältnis der UP zu den jährlichen Beitragseinnahmen ermittelt werden.

Insgesamt kann festgehalten werden, daß das Verhältnis der Provision zu den jährlichen Beitragseinnahmen von der Versicherungsgesellschaft, der Sparte und dem jährlichen Umsatz des AM (Bonuspunktesystem) abhängig ist. Darüber hinaus besteht bei den selbständigen Handelsvertretern auch ein gewisser Verhandlungsspielraum, um dieses Verhältnis zu beeinflussen. Die niedrigste Expertenschätzung des durchschnittlichen Anteils der einmaligen Provision an der jährlichen Beitragsleistung lag bei 0,5. Die höchste wurde mit 0,9 angegeben, wobei die Schätzwerte über verschiedene Versicherungsunternehmen im Mittel bei 0,8 lagen. Die Schätzungen im Unternehmen unserer Studie lag bei 0,8. Ich nehme deshalb für die weiteren Berechnungen an, daß das durchschnittliche Verhältnis der Provision zu den jährlichen Beitragseinnahmen bei 0,8 liegt. Da diese Schätzung sehr vage ist, wird im weiteren Verlauf der Nutzenschätzung jeweils die untere Grenze von 0,5 und die obere Grenze von 0,9 mitberücksichtigt.

Die Provision im untersuchten Versicherungsunternehmen beträgt das 20- bis 25-fache der Anzahl der erzielten UP. Diese Zahl schwankt, weil das Verhältnis durch Bonussysteme und durch Verhandlungen zwischen dem AM und dem Unternehmen beeinflusst wird. Bei einer angenommenen durchschnittlichen Provision von 22,5 DM je UP betragen damit die durchschnittlichen jährlichen Beitragseinnahmen ca. das 28-fache [in DM] ($22,5 \text{ DM}/0,8$) der Umsatzpunkte. Die obere Grenze der Beitragseinnahmenschätzung ist das 45-fache ($22,5 \text{ DM}/0,5$) und die untere Grenze das 25-fache ($22,5 \text{ DM}/0,9$) der Umsatzpunkte.

Wenn der Verkaufswert der Leistung w eines AM so geschätzt werden soll, dann ist weiterhin zu berücksichtigen, wie lange durchschnittlich ein Versicherungsvertrag besteht. Dies ist deshalb erforderlich, da die Gesamthöhe der Beitragszahlungen je abgeschlossenem Vertrag nicht nur von der Sparte und der Vertragshöhe abhängt, sondern auch von dessen Laufzeit. Spätestens an dieser Stelle wird deutlich, wie detailliert die Datenlage sein müßte, um SD_w einigermaßen genau zu ermitteln. Wünschenswert wäre nicht nur die Information, wieviele UP jeder AM in jeder Sparte erreicht, sondern auch die Laufzeit der einzelnen Verträge. Auch die durchschnittliche Laufzeit in den verschiedenen Sparten würde nicht genügen, da durchaus vorstellbar ist, daß die AM unterschiedlich lange Laufzeiten 'produzieren'. Diese Laufzeitunterschiede müssen theoretisch natürlich auch den Leistungsunterschieden der AM zugerechnet werden.

In Ermangelung genauerer Daten nehme ich eine durchschnittliche Versicherungsvertragslaufzeit von 4 Jahren an. Ich halte diese Schätzung für äußerst niedrig,

wenn man bedenkt, daß Kranken- und Lebensversicherungsverträge oft während eines großen Teils des Lebens des Versicherungsnehmers aufrechterhalten werden. Ich gebe zu bedenken, daß sowohl diese kurze Laufzeitschätzung, als auch die Nichtberücksichtigung der durchschnittlichen Laufzeitunterschiede zwischen den AM, zu einer konservativen Schätzung von SDI, führen. Der tatsächliche Wert könnte erheblich über demjenigen liegen, den ich schätze.

Die Standardabweichung der erzielten UP im ersten Tätigkeitsjahr (V12) zwischen den AM beträgt $SD_{V12}=941$. In Abbildung 12 ist die Verteilung der erzielten UP dargestellt. Daraus wird u.a. deutlich, daß die Höhe der Standardabweichung nicht unerheblich durch die vier AM zustandekommt, die über 4000 UP/Jahr liegen. Als Psychologe wird man durch ein solches Verteilungsbild normalerweise mißtrauisch, da man an mäßig reliable Messungen gewöhnt ist, bei welchen ab und zu sog. 'Ausreißer' auftreten. Ich betrachte diese vier Fälle jedoch keinesfalls als Artefakte, die möglichst zu eliminieren sind. Da die Messungen der Verkaufsleistung hochreliabel sind, würde man geradezu einen großen Fehler begehen, wenn diese Fälle von der Datenanalyse ausgeschlossen würden. Sie spiegeln die Realität, daß enorme Leistungsunterschiede zwischen den AM bestehen, wider. Dies schlägt sich richtigerweise in der Maßzahl SD_{V12} nieder. Es ist eher zu vermuten, daß in der Stichprobe einige AM mit V12 über 3500 in der Stichprobe fehlen, obwohl sie in der Population vorhanden sind. In diesem Fall würde sich SD_{V12} noch erhöhen.

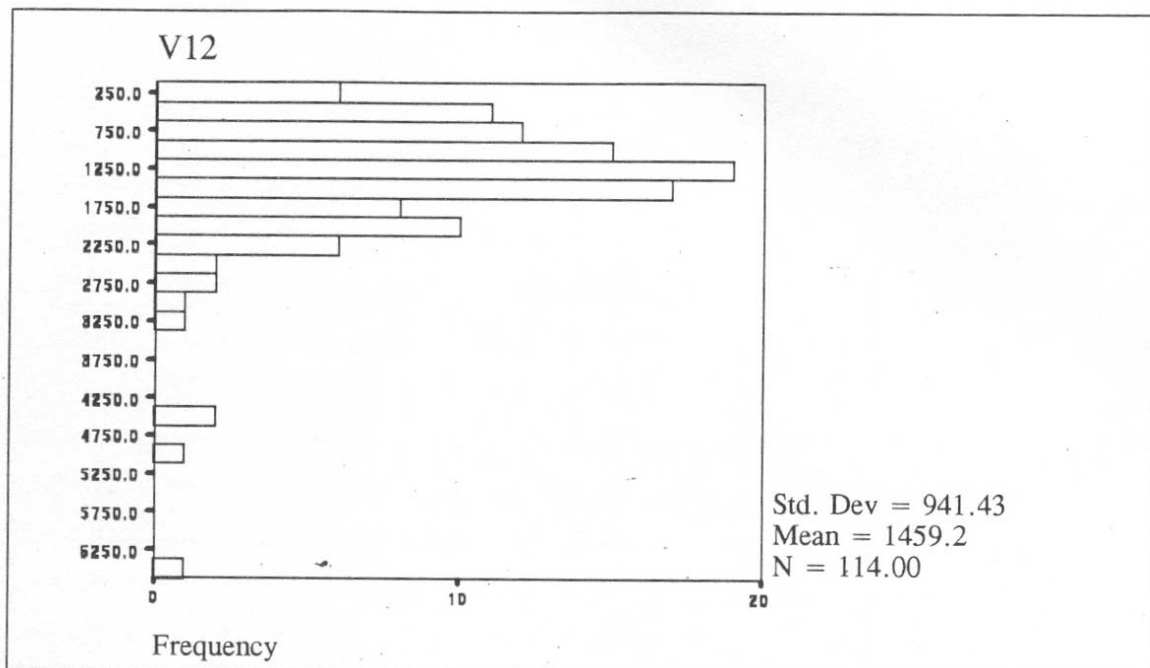


Abbildung 12: Verteilung der erzielten Umsatzpunkte im ersten Jahr nach Eintritt in das Versicherungsunternehmen (V12).

Die Standardabweichung der erzielten UP im ersten Berufsjahr wird, wie oben beschrieben, in die Standardabweichung der erwirtschafteten jährlichen Beitragsleistungen umgerechnet. Diese erhält man durch Multiplikation mit dem geschätzten Faktor 28 DM (bzw. 25 und 45 DM als untere und obere Grenze). Man erhält eine Standardabweichung der erwirtschafteten jährlichen Beitragseinnahmen zwischen den AM von $SD_{w/1 \text{ Jahr}}=26360$ DM (mit den Grenzen 23536 DM und 42364 DM). Dieser Wert ist die Standardabweichung des Verkaufswertes der Leistung bei einem Jahr Versicherungsvertragslaufzeit. Bei den sehr vorsichtig angenommenen 4 Jahren durchschnittlicher Laufzeit ergibt sich ein Wert von

$$SD_w = 105440 \text{ DM.}$$

Dies ist die Standardabweichung des Verkaufswertes der Leistung zwischen den AM in einem Arbeitsjahr. Die geschätzte untere und obere Grenze ist $SD_{w/u}=94143$ DM und $SD_{w/o}=169457$ DM.

Wie realistisch sind diese Werte? Auf die Frage an einen Mitarbeiter des untersuchten Unternehmens mit Führungsfunktion, wie hoch er die erwirtschafteten Beitragseinnahmen eines durchschnittlichen AM pro Jahr einschätze, erhielt ich zur Antwort, daß dies etwa 50000 bis 70000 DM (!) für einen durchschnittlichen AM wären, die Frage aber damit eigentlich nicht beantwortet sei, da es zwischen den AM enorme Unterschiede gäbe. Ich halte diese Aussage mit meiner unabhängig davon geschätzten Zahl $SD_{w/1 \text{ Jahr}}=26360$ DM vereinbar.

Interessant in diesem Zusammenhang ist auch der Vergleich zu den Zahlen, die Barthel (1988, 1989) in seiner Studie im Versicherungsbereich berichtet (vgl. auch Abschnitt 4.5.3.2 'Individuelle Leistungsschätzung bei Verkaufspersonal'). Auch bei der Versicherungsgesellschaft der Barthel-Studie werden die Vertragsabschlüsse mittels einem internen Codierungssystem quantifiziert. Diese Wertungspunkte (WP) erreichen ein jährliches Mittel von 6400 mit einer Standardabweichung von 4750. Die Standardabweichung der WP ist also etwa 74% des Mittelwertes. Bei der vorliegenden Studie beträgt dieses Verhältnis 64,5%. Barthel Probanden sind allerdings bereits verschieden lange im Unternehmen und verfügen deshalb über einen unterschiedlichen Erfahrungsschatz. Bei der vorliegenden Studie befinden sich alle AM im ersten Tätigkeitsjahr im Unternehmen. Die Stichprobe der untersuchten AM ist dadurch homogener, was zu einer niedrigeren Varianz der erzielten Umsätze führen sollte als bei der Studie von Barthel. Barthel gibt an, daß 1 WP einer jährlichen Beitragseinnahmen von 12 DM entspricht.¹⁷ Die Standardabweichung der

¹⁷ Woher diese Zahl stammt, erklärt Barthel (1988, 1989) allerdings nicht. Nach den Schwierigkeiten, die ich hatte, eine solche Zahl zu schätzen, scheint mir die Aussage "Ein Wertungspunkt entspricht der Beitragseinnahme von DM 1,--/Monat bzw. DM 12,--/Jahr für die Organisation" erklärungsbedürftig. Da Barthel keine genauen Angaben macht, ist diese Schätzung sicherlich mit Vorsicht zu genießen.

Beitragseinnahmen pro Jahr liegt damit bei 57000 DM, einem Wert, der weit über meiner (allerdings sehr konservativen) Schätzung von $SD_w/1 \text{ Jahr} = 26360 \text{ DM}$ liegt. Ich werte dies als weiteren Hinweis, daß der Parameter $SD_w/1 \text{ Jahr}$ keinesfalls überschätzt wurde.

6.4.3.2 Anteil der variablen Kosten an den Beitragseinnahmen (V)

Die bisherigen Überlegungen bezogen sich alle auf den Verkaufswert der Leistung w . Letztendlich interessiert allerdings nicht w , sondern der Deckungsbeitrag der Leistung y , der sich dadurch berechnet, daß vom Verkaufswert die variablen Kosten der Leistung abgezogen werden. Der Anteil der variablen Kosten der Leistung an ihrem Verkaufswert wurde V genannt. Ist dieser Parameter bekannt und die Standardabweichung des Verkaufswertes der Leistung geschätzt, dann berechnet sich SD_y durch

$$(18) \quad SD_y = SD_w (1-V).$$

Die variablen Kosten, die von den eingenommenen Versicherungsbeiträgen abgezogen werden müssen, entstehen hauptsächlich durch die Provision des AM., die Superprovision an dessen Vorgesetzte, die Verwaltungskosten, die durch den Vertrag entstehen, insbesondere aber durch die Versicherungsleistungen, die an andere Kunden bezahlt werden. Diese stellen auch die Kalkulationsgrundlage für die Beitragsbemessung dar. Der AM-Beruf zählt zu den Tätigkeiten, die in Abbildung 8 auf der Abszisse sehr weit rechts eingeordnet werden müßte. Dies bedeutet, daß der Deckungsbeitrag der Leistung nur einen kleinen Teil des Verkaufswertes der Leistung darstellt. Der Parameter V liegt also sehr hoch. $(1-V)$ kann als der Anteil der festen Kosten am Verkaufswert der Leistung interpretiert werden. Diese festen Kosten entstehen z.B. durch feste Angestelltenlöhne, Gebäude, Verwaltungsgemeinkosten, Personalprogramme (incl. PAP), etc. Außer den festen Kosten sind in $(1-V)$ auch die Gewinne der Gesellschaft enthalten.

Eine direkte Berechnung von V mittels der innerbetrieblichen Kostenrechnung halte ich für sehr aufwendig. Trotz dieses Aufwandes würde das Ergebnis möglicherweise weiterhin fraglich sein. Barthel (1988, 1989) nutzt bei einer (bei ihm allerdings nicht expliziten) Schätzung von V die Tatsache, daß in den Wert $(1V)$ auch der Anteil der Gewinne am Umsatz eingeht. Er beziffert die Umsatzrendite mit 8% und benutzt diesen Wert als Schätzung für $(1-V)$. Der Parameter SD_y ist also 8% der Standardabweichung der Beitragseinnahmen des Unternehmens.

Diese Logik entspricht nicht vollkommen meiner Definition von y als Deckungsbeitrag der Leistung. Da der Deckungsbeitrag als der Umsatz weniger die variablen

Kosten definiert ist, enthält er neben dem Gewinnanteil auch die festen Kosten. Diese werden von Barthel (1988, 1989) nicht berücksichtigt. Ich schätze $(1-V)$ auf 0,2. Dabei gehe ich von 5% Umsatzrentabilität und 15% Deckung der festen Kosten durch die Beitragseinnahmen aus. Die restlichen 80% der Beitragseinnahmen sind zur Deckung der variablen Kosten notwendig. Der Parameter SD_y berechnet sich nach Gleichung (18) auf

$$SD_y = 105440 \text{ DM} * (1-0,8) = 21088 \text{ DM.}$$

Die geschätzte untere und obere Grenze ist $SD_{y/u}=18829 \text{ DM}$ und $SD_{y/o}=33891 \text{ DM}$.

In Abschnitt 4.5.3 habe ich die proportionalen Regeln (40-70% Regel) von Hunter und Schmidt (1982) zur Schätzung von SD_y besprochen. Im vorliegenden Fall beträgt der mittlere jährliche Bruttoverdienst der untersuchten AM ca. 32850 DM. Der Schätzwert für SD_y liegt bei 64,2 % dieser mittleren Entlohnung ($SD_{y/u}$: 57% und $SD_{y/o}$: 103%). Der Befund von Hunter und Schmidt (1982), daß die meisten Schätzwerte für SD_y in den Bereich von 40% bis 70% des mittleren Verdienstes fallen, wird in unserem Fall also weitgehend bestätigt.

6.4.4 Ergebnisse der Nutzenanalyse

Die theoretischen Hintergründe der Nutzenmodelle und ihrer Parameter sowie die Schätzmethoden und -werte der Parameter wurden ausführlich dargestellt. Die Nutzenberechnungen erfolgen nach dem Nutzenmodell, welches in Gleichung (29) spezifiziert ist. Tabelle 11 faßt noch einmal alle Parameterschätzwerte zusammen. Die Ergebnisse der Nutzenanalyse zeigt Tabelle 12.

Tabelle 11: Die Parameterschätzwerte der Nutzenanalyse.

Bezeichnung des Parameters	Symbol	Schätzwert	Einheit
durchschnittlicher standardisierter Prädiktorwert der ausgewählten Gruppe	\bar{z}_x^*	0,8	
inkrementelle Validität	r_{xy}	0,3	
mittlere Standardabweichung des Verkaufswertes der Leistung	SD_w	105440	DM
Anteil der variablen Kosten der Leistung	V	0,8	
Standardabweichung des Deckungsbeitrages der Leistung			
mittlere Schätzung	SD_y	21088	DM
minimale Schätzung	$SD_{y/u}$	18829	DM
maximale Schätzung	$SD_{y/o}$	33891	DM
inkrementelle Kosten des PAP			
feste Kosten (Entwicklung, Implementation, ...)	C_0	186000	DM
variable Kosten je Bewerber	C_1 / N_a	1050	DM
Kalkulationszinsfuß	i	0,1	
Steuersatz	TAX	0,4	

Wie aus Tabelle 12 ersichtlich ist, hat das untersuchte Personalauswahlprogramm, welches ein Assessment Center beinhaltet, unter den gegebenen Randbedingungen einen hohen positiven Nutzen. Die Erträge sind in jeder Behandlungsperiode höher als die Kosten. Die investierten festen Kosten zur Entwicklung und Implementation des PAP haben sich schon nach einem Jahr amortisiert. Auch wenn der Nutzen des PAP mit der unteren Grenze des $SD_{y/u}$ -Schätzwertes ermittelt wird, ergibt sich noch ein hoher positiver Nutzen von $\Delta U = 1154964$ DM. Wird die obere Grenze des Schätzwertes dazu herangezogen ($SD_{y/o}$), dann beläuft sich der Nutzen auf über 2,6 Mio DM. Findet die Besteuerung keine Berücksichtigung ($TAX=0$), so beläuft sich der Nutzen mit dem mittleren SD_y -Schätzwert auf $\Delta U = 2288248$ DM. Auch diese Nutzenanalyse bestätigt damit die Befunde vieler anderer Untersuchungen, welche den ökonomischen Nutzen von psychologischen Interventionsprogrammen z.T. eindrucksvoll bestätigen (eine nahezu vollständige Auflistung publizierter NA auf der Grundlage des BCG-Modells findet sich bei Boudreau, 1991).

Tabelle 12: Nutzenanalyse des PAP. In der letzten Zeile ist der Gesamtnutzen des PAP zu finden. Die Angaben sind ohne Einheit oder in DM.

k	Ertragsvariablen								Kostenvariablen				Kosten	Nutzen
	N_k	r_{xy}	\bar{z}_x^*	SD_w	(1-V)	(1-Tax)	$1/(1+i)^k$	E	C_0	C_{1k}	(1-Tax)	$1/(1+i)^{k-1}$	gesamt	U
	DM								DM	DM	DM	DM	DM	DM
1	89	0,3	0,8	105440	0,2	0,6	0,91	245694	186000	186900	0,6	1,00	223740	21954
2	215	0,3	0,8	105440	0,2	0,6	0,83	539574	0	264600	0,6	0,91	144327	395247
3	217	0,3	0,8	105440	0,2	0,6	0,75	495085	0	210000	0,6	0,83	104132	390953
4	190	0,3	0,8	105440	0,2	0,6	0,68	394077	0	220500	0,6	0,75	99399	294678
5	163	0,3	0,8	105440	0,2	0,6	0,62	307342	0	220500	0,6	0,68	90363	216979
6	31	0,3	0,8	105440	0,2	0,6	0,56	53138	0	0	0,6	0,62	0	53138
7	0	0,3	0,8	105440	0,2	0,6	0,51	0	0	0	0,6	0,56	0	0
								2034910					661961	1372949

Bei der Bewertung des DM-Betrags des Nutzens ist zu bedenken, daß die Berechnung z.T. auf äußerst konservativen Annahmen beruhen. Insbesondere die durchschnittliche Versicherungsvertragslaufzeit ist mit 4 Jahren möglicherweise weit unterschätzt. Der tatsächliche Nutzen in DM könnte ein Vielfaches über dem berichteten Wert liegen.

Wie ausführlich in den Abschnitten 5.2 und 5.4 besprochen wurde, berücksichtigt das verwendete Nutzermodell nur den Nutzen, welcher direkt aus den Erträgen der Leistungssteigerung der behandelten Mitarbeiter resultiert. Der Nutzen zweiter Art, der aufgrund der begrenzten Leistungsressourcen des Arbeitsmarktsystems entsteht, und der Nutzen 'dritter Art', der Nebennutzen des AC, welcher z.B. der positiven Auswirkung auf die Führungskompetenz der Beobachter oder der Auswirkungen auf die Unternehmenskultur entspringt, ist in dieser Nutzenkalkulation nicht berücksichtigt. Ich möchte eine derartige Kalkulation auch nicht versuchen, da solche Nutzermodelle bisher nicht existieren. Es muß deshalb an dieser Stelle der Hinweis auf die Vermutung genügen, daß die Nutzen zweiter und dritter Art unter bestimmten Randbedingungen sehr bedeutsam sein können.

6.5 Diskussion der Nutzenanalyse

Nutzenanalytischen Ergebnissen wird immer wieder mit viel Skepsis begegnet. Dies liegt vielleicht an den beeindruckenden Befunden von Nutzenanalysen, welche psychologischen Personalprogrammen immer wieder einen hohen ökonomischen Nutzen zuschreiben wollen,

vielleicht auch nur an menschlichen Vorbehalten gegenüber der Bewertung 'weicher' psychologischer Maßnahmen in Geldeinheiten. Dieser kritischen Haltung will ich in diesem Abschnitt dadurch Rechnung tragen, daß ich die Information, welche in das Nutzenmodell eingegangen ist, kritisch überprüfe. Dies soll im wesentlichen in der Form quantitativer Verfahren geschehen: Durch die Suche nach sog. Break-even-points und durch die Überprüfung der Sensitivität der Parameter.

6.5.1 Break-even-points

Der Break-even-point, der in der deutschen Literatur gelegentlich auch als Kostendeckungspunkt bezeichnet wird, ist jeweils der Wert eines Parameters, bei dessen Durchgang sich ein positiver Nutzen eines Interventionsprogrammes in einen negativen Nutzen wandelt. Um diesen Punkt zu ermitteln, wird jeweils ein Parameter so lange variiert, bis sich ein Nullnutzen der Maßnahme einstellt, wobei die ursprünglichen Schätzwerte der anderen Parameter konstant gehalten werden. Auf diese Weise soll ermittelt werden, mit welchem Risiko eine Investitionsentscheidung gefällt wird. Liegen die Parameter deutlich über ihrem Break-evenpoint, so kann davon ausgegangen werden, daß das Programm auch bei einer unsicheren Schätzung der Parameter oder beim Auftreten vorher nicht erwarteter ungünstiger Bedingungen noch einen positiven Nutzen erbringt.

Diese Art der Risikoabschätzung hat jedoch einen großen Nachteil. Da immer nur ein Parameter variiert wird und die anderen mit ihren ursprünglichen Werten verrechnet werden, hängt der Break-even-point eines jeden Parameters im wesentlichen von den Schätzwerten der anderen Parameter ab. Wurde ein Parameterwert beispielsweise grob überschätzt, so wirkt sich das darin aus, daß die Break-evenpoints aller Parameter zu niedrig eingeschätzt werden. Die Ergebnisse des Verfahrens sollten also erst dann interpretiert werden, nachdem der Schätzwert eines jeden einzelnen Parameters kritisch hinterfragt worden ist und an dessen einigermaßen realistischer Höhe nicht mehr gezweifelt wird. Die Break-even-points der vorliegenden Parameter zeigt Tabelle 13. Der Berechnung wurden die Parameterschätzwerte der Tabelle 11 zugrundegelegt.

Tabelle 13: Die Break-even-points der Parameter. Es sind diejenigen Parameterwerte angegeben, bei welchen das PAP gerade noch einen positiven Nutzen erbringt, während jeweils alle anderen Parameter mit ihrem ursprünglichen Schätzwert (s. Tabelle 11) verrechnet werden.

Parameter	Break-even-point	Bemerkung
r_{xy}	0,10	
\bar{z}_x^*	0,27	Dies entspricht einer Selektionsquote $Q \approx 85\%$.
V	0,935	
SD_w	34300 DM	Dies entspricht $SD_y = 6860$ DM.
C_0	2475000 DM	
C_1 /Bewerber	3665 DM	Bei durchschnittlich 5,3 Tn je AC sind das variable inkrementelle Kosten je AC von 19425 DM.

Die Werte in Tabelle 13 machen deutlich, daß das PAP auch unter sehr ungünstigen Bedingungen noch einen positiven Nutzen erbringt. Die Investitionsentscheidung unter den gegebenen Bedingungen ist nicht nur aufgrund des hohen Nutzens der Intervention zu rechtfertigen, sondern auch durch das geringe Risiko, welches mit dem PAP verbunden ist.

6.5.2 Sensitivitätsanalysen

Das Nutzenmodell, welches durch Gleichung (29) dargestellt ist, verknüpft die Parameter in einer nicht linearen Weise, sondern durch verschiedene arithmetische Operationen. Wenn der Wert eines Parameters zu einem bestimmten prozentualen Anteil geändert wird, so ergibt sich eine parameterspezifische prozentuale Änderung des Nutzens. Das Verhältnis der prozentualen Änderung des Nutzenwertes zur prozentualen Änderung des Parameterwertes nennt man die Sensitivität des Parameters (z.B. Cascio & Silbey, 1979). Diese ist abhängig von der Art der arithmetischen Operationen, mit welchen die Parameter in das Nutzenmodell eingebunden sind und ihren Schätzwerten. Eine Sensitivität des Parameters r_{xy} von 1,48 bedeutet z.B., daß sich der Gesamtnutzen des Programmes um den Faktor 1,48 seines ursprünglichen Wertes nach oben bzw. unten verändert, wenn der Parameter r_{xy} um 100% seines ursprünglichen Wertes vergrößert bzw. vermindert wird. Eine Vergrößerung bzw. Verminderung von r_{xy} um 20% seines Wertes, hat entsprechend eine Veränderung des $(0,2 \cdot 1,48 =) 0,296$ -fachen des ursprünglichen ΔU -Wertes nach oben bzw. unten zur Folge. Eine negative Sensitivität bedeutet, daß der Nutzen sinkt, wenn der Wert des Parameters zunimmt und umgekehrt. Die Sensitivität der Parameter des untersuchten PAP soll auch hier

angegeben werden. Die Sensitivitätswerte sind in Tabelle 14 abgebildet. Die Parameter sind dabei in der Reihenfolge des Betrages ihrer Sensitivität geordnet.

Tabelle 14: Die Sensitivität der Parameter des untersuchten Personalauswahlprogramms.

Parameter	Sensitivität (prozentuale Nutzenänderung je prozentualer Parameteränderung)
V	- 5,92
r_{xy}	1,48
\bar{z}_x^*	1,48
SD_w	1,48
TAX	- 0,67
C_1 /Bewerber	- 0,40
i	- 0,28
C_0	- 0,08

Zunächst fällt der große Unterschied zwischen den Sensitivitätswerten auf. In den seltensten Fällen werden die Werte in dieser Höhe und Konstellation erwartet. Eine enorme Sensitivität erreicht der Parameter V. Da dieser nur sehr schwer zu schätzen ist, ist dies ein bedenkliches Ergebnis. Dank der konservativen Schätzung von $V = 0,8$ ist jedoch kaum eine Bedrohung des Nutzens durch die Annahme eines zu niedrigen Wertes zu erwarten. Wäre der wahre Wert von V um 10% größer, also 0,88, so würde noch ein Gesamtnutzen des PAP von

$$1372949 \text{ DM} + (1372949 \text{ DM} * 0,1 * (-5,92)) = 560163 \text{ DM}$$

entstehen. Erstaunlich niedrige Sensitivitäten erreichen die Kostenkomponenten des PAP. Bei der Entscheidung über die Neuentwicklung eines PAP tut man gut daran, dieses Ergebnis zu berücksichtigen. Offensichtlich spart man am falschen Ende, wenn man sich auf Kosten anderer Parameterwerte für ein billig zu entwickelndes und durchzuführendes PAP entscheidet. Diese Ergebnisse sind konsistent mit Cascio und Silbey (1979), die ein ähnliches Verfahren bei der NA eines AC-PAP angewandt haben.

6.5.3 Zusammenfassung der Ergebnisse

Es kann zusammenfassend festgehalten werden, daß das untersuchte PAP unter den gegebenen Umständen einen positiven Nutzen erbringt. Die mittlere Nutzenschätzung ist bei dieser Untersuchung etwa 1,4 Millionen DM. Die absolute Höhe dieser Zahl soll jedoch nicht interpretiert werden, weil die Schätzung der Parameter des verwendeten Nutzenmodells z.T.

mit erheblicher Unsicherheit erfolgt ist und deshalb sehr konservative Schätzwerte angenommen wurden. Wichtiger ist, daß die NA trotz konservativer Parameterschätzungen einen positiven Nutzen des PAP erbringt. Da ein wesentliches Ziel der NA darin besteht, rationale und richtige Entscheidungen zu fördern, genügt diese vorsichtige Interpretation, um das AC-PAP zu rechtfertigen und aufrechtzuerhalten. Das Ergebnis der nutzenanalytischen Evaluation des PAP ist aufgrund der Tatsache, daß die Aussagen in DM erfolgen, vergleichsweise einfach auch Angehörigen nichtpsychologischer Fachdisziplinen kommunizierbar.

Der positive Nutzen des PAP, der in dieser Untersuchung gefunden wurde, steht im Einklang mit vielen früheren nutzenanalytischen Befunden. Bei Boudreau (1991) findet sich eine Zusammenstellung nutzenanalytischer Studien. Diese belegen eindrucksvoll den durchgehend positiven, meist sehr hohen Nutzen von Personalauswahl- und anderen Personalinterventionsprogrammen. Der, in Relation zu diesen Ergebnissen, niedrige absolute Nutzen bei dieser Studie ist wesentlich darauf zurückzuführen, daß sehr konservative Parameterschätzwerte Verwendung finden. Aber auch die Besonderheiten des untersuchten PAP sind in diesem Zusammenhang zu beachten. Insbesondere die extrem hohe Personalfluktuation vermindert sehr stark die Wirkdauer des Personalauswahlprogramms und damit den Betrag des Nutzens.

Wie deutlich wurde, bereitet die Operationalisierung einiger Nutzenmodellparameter erhebliche Schwierigkeiten. Dies gilt insbesondere für r_{xy} und SD_y , bzw. dessen Komponenten SD_w und V , aber auch für die Kostenparameter C_1 und C_0 . Kommt diese Schätzunsicherheit mit einer hohen Sensitivität des Parameters zusammen, dann ist ein Parameterschätzwert besonders kritisch. Dies trifft bei der vorliegenden Untersuchung auf den Parameter V zu. Dieser wiederum bestimmt neben SD_w den Wert von SD_y , so daß auch bei dieser Untersuchung die Standardabweichung der Leistung in DM (SD_y) der am schwierigsten operationalisierbare Parameter ist.

Die Kostenparameter C_1 und C_0 dagegen erreichen sehr niedrige Sensitivitätswerte. Dies gilt insbesondere für die festen Kosten C_0 des PAP, die z.B. bei der Entwicklung, Implementation und Evaluation des Programms entstehen. Hieraus wird deutlich, daß es nicht sinnvoll ist, diese Kosten eines PAP zu reduzieren, wenn sich dadurch andere Charakteristika des PAP, insbesondere die Validität, verschlechtern.

7 Fazit und Ausblick

Die Nutzenanalyse ist ein Verfahren, welches bei der Evaluation von psychologischen Interventionsprogrammen eingesetzt werden kann. Es ist vorzüglich dazu geeignet, ökonomische Auswirkungen von Interventionsprogrammen für die Organisation zu bewerten und dadurch Entscheidungen zu unterstützen. Nutzenmodelle vermitteln Einsicht in relevante Größen, welche den Nutzen psychologischer Interventionsprogramme beeinflussen. Sie können als normative Theorien die Rationalität von Entscheidungen beträchtlich erhöhen. Weil sie im Grenzgebiet zwischen der Psychologie und wirtschaftswissenschaftlichen Disziplinen angesiedelt sind, haben sie den Vorteil, von Vertretern dieser Disziplinen relativ leicht verstanden zu werden. Sie können dadurch an der Schnittstelle dieser Disziplinen wichtige Aufgaben übernehmen, um so Kommunikationsschwierigkeiten abzubauen.

Bisher veröffentlichte Nutzenanalysen zeigen nahezu einheitlich einen großen positiven Nutzen von Interventionsprogrammen im Personalbereich. Auch die empirische Untersuchung, die in dieser Arbeit beschrieben ist, bestätigt diesen Befund. Bei der Interpretation des Nutzens ist aber Vorsicht geboten. Bei vielen empirischen NA werden die Parameter sehr konservativ geschätzt. Dies geschieht mit dem Ziel, das Risiko einer fälschlichen Entscheidung zur Durchführung eines negativ nützlichen Interventionsprogrammes zu minimieren. Aus diesem Grund wird der Nutzen von Interventionsprogrammen im Personalbereich wahrscheinlich tendenziell unterschätzt. Eine Interpretation des absoluten Betrages des Nutzens könnte deshalb zu falschen Schlüssen führen.

Durch die strukturelle Ähnlichkeit der Nutzenmodelle mit Modellen der Investitionsrechnung kann die Vergleichbarkeit von Investitionen in psychologische Interventionsprogramme untereinander und mit anderen Investitionen hergestellt werden. Dadurch können die Nutzenbeträge konkurrierender Investitionen relativ zueinander begutachtet und auf diese Weise richtige Entscheidungen abgeleitet werden. Wenn die konkurrierenden Investitionen im selben Umfeld geplant sind, dann können verschiedene Parameterschätzungen möglicherweise gemeinsam vorgenommen werden. Auch wenn diese Parameter in ihrer absoluten Höhe von der Realität abweichen, so können Entscheidungen zwischen den Programmen aus dem relativen Nutzen der Programme zueinander abgeleitet werden. Auf diesem Wege können Nutzenmodelle einen wichtigen Beitrag zu rationalen Entscheidungen in Organisationen leisten, obwohl die absolute Nutzenhöhe kritisierbar bleibt.

Eine der Voraussetzungen der Vergleichbarkeit verschiedener Investitionen besteht darin, daß ihr Nutzen mit ähnlichen Modellen bewertet wird. Einen wichtigen Schritt zur Anpassung der psychologischen Nutzenmodelle an wirtschaftswissenschaftliche Modelle der Investitionsrechnung wurde durch die Integration ökonomischer Konzepte in nutzenanalytische Modelle vollzogen (z.B. Boudreau, 1983a, 1983b, Cronshaw & Alexander, 1985). Eine weitere Anpassung der Modelle unter Beachtung der Erkenntnisse der wirtschaftswissenschaftlichen Nachbardisziplinen ist sicherlich möglich. Da die Verbreitung der Nutzenanalyse in Deutschland erst vor wenigen Jahren begonnen hat, könnten solche angepaßten Modelle leichter akzeptiert werden, als dies in den USA mit dem ursprünglichen BCG-Modell der Fall war. Der beschleunigte Verlauf der Verbreitung der NA in Deutschland spricht für diese Entwicklung.

Neben dieser Anpassung sind natürlich auch weitere Anstrengungen notwendig, um verfeinerte Methoden zur Parameteroperationalisierung zu entwickeln. Insbesondere der Parameter SD_y ist nach wie vor nicht einfach zu schätzen. Mit zunehmender Anzahl von empirischen Nutzenanalysen werden jedoch auch mehr Parameterschätzungen in verschiedenen Berufsbereichen vorliegen. Insbesondere für den Parameter SD_y könnten sich auf metaanalytischem Wege Anhaltspunkte dafür ergeben, wie hoch er im Durchschnitt anzusetzen ist und bei welchen Berufsfeldern mit Abweichungen zu rechnen ist.

In Kapitel 5 habe ich mich mit dem Nutzen von psychologischen Interventionsprogrammen, welcher nicht durch die BCG-Nutzenmodelle abgebildet wird, beschäftigt. Diese und weitere Nutzenarten könnten in Zukunft eine größere Beachtung finden. Möglicherweise könnten solche Überlegungen auch in einem formalen Modell festgehalten werden oder sogar ihre Integration in BCG-Modelle möglich sein.

Nicht zuletzt ist im Zusammenhang mit der Nutzenanalyse zu bedenken, daß sie bisher nur auf sehr spezifische Fragestellungen Anwendung fand. Durch historische Gründe bedingt ist das Hauptanwendungsgebiet der NA immer noch die Personalauswahl. Dadurch daß die Weiterentwicklung der Nutzenmodelle hauptsächlich innerhalb der US-amerikanischen Organisationspsychologie stattfand, wurden sie auch auf weitere organisationspsychologische Probleme angewandt. Hier ist insbesondere die Evaluation von Personalentwicklungsmaßnahmen zu nennen. Die Anwendung der NA auf Fragestellungen außerhalb der Organisationspsychologie blieb bisher weitgehend aus. Ich bin der Überzeugung, daß die nutzen-analytischen Grundgedanken sowohl bei weiteren Gebieten innerhalb als auch außerhalb dieses Faches auf fruchtbaren Boden fallen und wichtige Anregungen und Problemlösungen mit sich bringen können. Man kann hierbei z.B. an den

Nutzen umfassender Organisationsentwicklungsmaßnahmen oder den volkswirtschaftlichen Nutzen pädagogisch-psychologischer und psychotherapeutischer Intervention denken. So wäre, neben anderen Überlegungen, z.B. die volkswirtschaftliche Nutzenbestimmung von bundesweiten psychotherapeutischen Interventionsprogrammen im Zusammenhang mit der aktuellen Diskussion um das Psychotherapeutengesetz interessant. Solche volkswirtschaftliche Nutzenmodelle sind jedoch kaum direkt aus den BCG-Modellen ableitbar, da sie komplexe Systemzusammenhänge und -wechselwirkungen zu berücksichtigen hätten.

Die Nutzenanalyse und ihre Modelle haben bereits einige Aufmerksamkeit erregt und Anerkennung gewonnen. Dennoch besitzt dieser Forschungsbereich noch ein großes Entwicklungspotential. Wenn es den Vertretern der nutzen-analytischen Forschung gelingt, ihre Denkweise weiterhin erfolgreich bekanntzumachen, dann könnte die NA eines Tages ein wichtiger Baustein quantitativer Evaluation darstellen.

Anhang A: Das Turbo-Pascal-Listing eines Programms zur Nutzenanalyse auf Grundlage des Nutzenmodells der Gleichung (29)

```

PROGRAM Nutzenanalyse;

{***** (c) Wolfram Reiners (1993) *****}

USES Dos, Crt;
{$M 65520,0,655360}

VAR
  Einsatz, t, k, F, l           : Integer;
  N, Na, Nz                    : Array [0..30] of Integer;
  r, Zx, V, i, TAX             : Real;
  C1                            : Array [1..30] of LongInt;
  Diskontfaktor                : Array [0..30] of Real;
  varKosten, Ertrag, Nutzen    : Array [1..30] of Real;
  C0, C1a, SDw                 : LongInt;
  versteuerteFestkosten, GesamtNutzen : Real;
  ch, Auswahl                  : char;
  runten, roben, rinc          : Real;
  zxunten, zxoben, zxinc       : Real;
  Vunten, Voben, Vinc         : Real;
  iunten, ioben, iinc         : Real;
  TAXunten, TAXoben, TAXinc    : Real;
  SDwunten, SDwoben, SDwinc    : LongInt;
  C0unten, C0oben, C0inc       : LongInt;
  Claunten, Claoben, Clainc    : LongInt;
  Sensitivity                  : Byte;

Procedure N_von_k_ermitteln;

{**** N(k) sind die Mitarbeiter im Unternehmen, die in der
Behandlungsperiode k jeweils wirksam sind. ****}

Begin
  TextBackGround(1);
  TextColor(12);
  ClrScr;
  WriteLn(^j, '-----');
--');
  WriteLn('      Programm zur Analyse des Nutzens von Personalauswahlprogrammen');
  WriteLn(^j, '      auf Grundlage des Nutzenmodells von Reiners (1994)');
  WriteLn(^j, '-----');
--');
  TextColor(14);
  Write(^j^j^j, 'Wie lange soll das PAP eingesetzt werden [Jahre]? ');
  ReadLn(Einsatz);

  ClrScr;
  WriteLn(^j, 'Geben Sie bitte die Anzahl der Organisationsmitglieder an, die im jeweiligen');
  WriteLn('Wirkjahr nach Beginn des PAP eingetreten bzw. ausgetreten sind. ');
  WriteLn;
  k:=0;
  N[0]:=0;

```

```

Repeat
  Inc(k);
  If k<=Einsatz then
    Begin
      Write(k, '. Jahr:   Einstellungen   '); Read(Na[k]);
    End
    else
      Begin
        WriteLn(k, '. Jahr:   Einstellungen   0');
        Na[k]:=0;
      End;
  GotoXY(40, WhereY-1);
  Write('   Austritte   '); Read(Nz[k]);
  N[k]:=N[k-1]+Na[k]-Nz[k];
Until N[k]<=0;
F:=k;
If N[F]<0 then F:=F-1;

TextColor(12);
WriteLn('^j^j'                                     'Weiter bitte mit beliebiger Taste ...');
ch:=Readkey;
TextColor(14);

ClrScr;
WriteLn('^j, 'Die Einsatzdauer des PAP beträgt           ', 'Einsatz, ' Jahre. ');
WriteLn('^j, 'Die Wirkdauer des PAP beträgt             ', 'F, ' Jahre.. ');
WriteLn('^j, 'In den jeweiligen Behandlungsperioden sind an behandelten Mitarbeitern');
WriteLn('im Unternehmen tätig:');
WriteLn;
For k:=1 to F do
  Begin
    WriteLn(k, '.-te Behandlungsperiode:   ', N[k], ' Mitarbeiter');
  End;

TextColor(12);
WriteLn('^j^j'                                     'Weiter bitte mit beliebiger Taste ...');
ch:=Readkey;
TextColor(14);
End;

Procedure Parametererfassung;

Begin
  ClrScr;
  WriteLn('^j, 'Bitte geben Sie die Werte für die jeweiligen Parameter ein:');
  WriteLn;
  Write('^j, 'Validitätskoeffizient r(xy) bzw. r(xw)           '); Read(r);
  Write('^j, 'Durchschnittl. Prädiktorwert der ausgewählten Gruppe Z(x) '); Read(Zx);
  Write('^j, 'Anteil der variablen Kosten am Verkaufswert der Leistung V '); Read(V);
  Write('^j, 'Kalkulationszinsfuß i                             '); Read(i);
  Write('^j, 'Steuersatz TAX                                     '); Read(TAX);
  Write('^j, 'Standardabweichung der Verkaufsleistung SD(w)   [in DM] '); Read(SDw);
  Write('^j, 'Feste Kosten des PAP zu Beginn der Maßnahme CO   [in DM] '); Read(CO);
  Write('^j, 'Variable Kosten des PAP je ausgewähltem MA Cla   [in DM] '); Read(Cla);

  TextColor(12);
  WriteLn('^j^j'                                     'Weiter bitte mit beliebiger Taste ...');
  ch:=Readkey;
  TextColor(14);
End;

```

```

Procedure Standardparameter;

Begin
  {*****
  {***** hier Parameter bei Sensitivitätsanalysen definieren *****
  {*****}

  r := 0.3;
  Zx := 0.8;
  V := 0.8;
  i := 0.1;
  TAX:= 0.4;
  SDw:= 105440;
  CO := 186000;
  Cla:= 2100;

  {*****}
End;

Procedure Standardanzeigen;

Begin
  ClrScr;
  WriteLn;
  TextColor(12);
  WriteLn('Die Parameter werden mit folgenden Werten vorläufig festgelegt:');
  TextColor(14);
  WriteLn('^j^j','Validitätskoeffizient r(xy) bzw. r(xw) ',r:9:2);
  WriteLn('^j','Durchschnittl. Prädiktorwert der ausgewählten Gruppe Z(x) ',Zx:9:2);
  WriteLn('^j','Anteil der variablen Kosten am Verkaufswert der Leistung V ',V:9:2);
  WriteLn('^j','Kalkulationszinsfuß i ',i:9:2);
  WriteLn('^j','Steuersatz TAX ',TAX:9:2);
  WriteLn('^j','Standardabweichung der Verkaufsleistung SD(w) [in DM] ',SDw);
  WriteLn('^j','Feste Kosten des PAP zu Beginn der Maßnahme CO [in DM] ',CO);
  WriteLn('^j','Variable Kosten des PAP je aufgenommenem Bewerber Cla ',Cla);

  TextColor(12);
  Write('^j^j' Weiter bitte mit beliebiger Taste ...');
  ch:=Readkey;
  TextColor(14);
End;

Procedure Sensitivitaetserfassung;

Begin
  ClrScr;
  WriteLn('^j','Welchen Parameter wollen sie variieren?');
  WriteLn; WriteLn;
  WriteLn(' 0.....keinen');
  WriteLn(' 1.....r');
  WriteLn(' 2.....Zx');
  WriteLn(' 3.....V');
  WriteLn(' 4.....i');
  WriteLn(' 5.....TAX');
  WriteLn(' 6.....SDw');
  WriteLn(' 7.....Cla');
  WriteLn(' 8.....CO');
  Write('^j^j','Geben sie die Zahl ein ');
  Auswahl:=Readkey;
  WriteLn;WriteLn; WriteLn;
  Repeat
    If Auswahl=#49 then
      Begin
        Write('Untere Grenze von r ');ReadLn(runten);
        Write('Obere Grenze von r ');ReadLn(roben);
        Write('Inkrement ');ReadLn(rinc);
      End;
  End;

```

```

If Auswahl=#50 then
Begin
  Write('Untere Grenze von Zx ');ReadLn(Zxunten);
  Write('Obere Grenze von Zx ');ReadLn(Zxoben);
  Write(' Inkrement ');ReadLn(Zxinc);
End;
If Auswahl=#51 then
Begin
  Write('Untere Grenze von V ');ReadLn(Vunten);
  Write('Obere Grenze von V ');ReadLn(Voben);
  Write(' Inkrement ');ReadLn(Vinc);
End;
If Auswahl=#52 then
Begin
  Write('Untere Grenze von i ');ReadLn(iunten);
  Write('Obere Grenze von i ');ReadLn(ioben);
  Write(' Inkrement ');ReadLn(iinc);
End;
If Auswahl=#53 then
Begin
  Write('Untere Grenze von TAX ');ReadLn(TAXunten);
  Write('Obere Grenze von TAX ');ReadLn(TAXoben);
  Write(' Inkrement ');ReadLn(TAXinc);
End;
If Auswahl=#54 then
Begin
  Write('Untere Grenze von SDw ');ReadLn(SDwunten);
  Write('Obere Grenze von SDw ');ReadLn(SDwoben);
  Write(' Inkrement ');ReadLn(SDwinc);
End;
If Auswahl=#55 then
Begin
  Write('Untere Grenze von Cla ');ReadLn(Claunten);
  Write('Obere Grenze von Cla ');ReadLn(Claoben);
  Write(' Inkrement ');ReadLn(Clainc);
End;
If Auswahl=#56 then
Begin
  Write('Untere Grenze von C0 ');ReadLn(C0unten);
  Write('Obere Grenze von C0 ');ReadLn(C0oben);
  Write(' Inkrement ');ReadLn(C0inc);
End;
ch:=Auswahl;
Until (ch=#48) or (ch=#49) or (ch=#50) or (ch=#51) or (ch=#52) or (ch=#53) or (ch=#54) or
(ch=#55) or (ch=#56)
End;

```

Procedure Nutzenrechnung;

```

Begin
  Clrscr;
  Diskontfaktor[0]:=1;
  For k:=1 to F do
  Begin
    Diskontfaktor[k]:=1/(1+i)*Diskontfaktor[k-1];
  End;

  For k:=1 to F do
  Begin
    Cl[k]:=Na[k]*Cla;
  End;

  For k:=1 to F do
  Begin
    Ertrag[k]:=N[k]*r*Zx*SDw*(1-V)*Diskontfaktor[k]*(1-TAX);
    varKosten[k]:=Cl[k]*Diskontfaktor[k-1]*(1-TAX);
  End;

  versteuerteFestkosten:=C0*(1-TAX);

```

```

For k:=1 to F do
Begin
  Nutzen[k]:=Ertrag[k]-varKosten[k];
End;

GesamtNutzen:=0-versteuerteFestkosten;
For k:=1 to F do
Begin
  GesamtNutzen:=GesamtNutzen+Nutzen[k];
End;

TextColor(12);
WriteLn('Behandlungs-           Ertrag           variable           Nutzen ohne CO ');
WriteLn('periode k           in k           Kosten in k           in Periode k');
WriteLn('-----');
TextColor(14);
For k:=1 to F do
Begin
  Write('           ',k);
  GotoXY(20,WhereY);
  Write(Ertrag[k]:10:2);
  GotoXY(40,WhereY);
  Write(varKosten[k]:10:2);
  GotoXY(60,WhereY);
  WriteLn(Nutzen[k]:10:2);
End;

TextColor(12);
Write(^j'                               Weiter bitte mit beliebiger Taste ...');
ch:=Readkey;
TextColor(14);

WriteLn;
WriteLn(^j^j,'versteuerte feste Kosten: CO (1-TAX) = ',versteuerteFestkosten:10:2);
WriteLn('Gesamtnutzen der Maßnahme über ',F,' Wirkperioden: U = ',Gesamtnutzen:12:2);

TextColor(12);
Write(^j'                               Weiter bitte mit beliebiger Taste ...');
ch:=Readkey;
TextColor(14);
ClrScr;
End;

{***** Hauptprogramm *****}

BEGIN
Sensitivity:=0;
N_von_k_ermitteln;
Repeat
  ClrScr;
  WriteLn(^j,'Wollen sie eine Nutzenanalyse durchführen oder die Sensitivität');
  WriteLn('verschiedener Parameter überprüfen?');
  Write(^j,'Geben Sie N oder S ein! ');
  Repeat
    ch:=Readkey;
    If (ch=#115) or (ch=#83) then
      Begin
        Standardparameter;
        Standardanzeigen;
        Sensitivitaetserfassung;
        Sensitivity:=1;
        ch:=#115;
      end;
    If (ch=#110) or (ch=#78) then
      Begin
        Parametererfassung;
        ch:=#110;
      end;
  Until (ch=#115) or (ch=#83) or (ch=#110) or (ch=#78);

```

```

{***** Normale NA *****)
If Sensitivity=0 then
Begin
  Nutzenrechnung;
End;

{***** Sensitivitätsanalyse bzw. NA mit vorgegebenen Daten *****)

If Sensitivity=1 then
Begin
  If Auswahl=#48 then
  Begin
    Nutzenrechnung;
  End;

  If Auswahl=#49 then
  Begin
    r:=runten;
    Repeat
      WriteLn(^j^j^j, 'Wert des variierten Parameters: ', r:10:2);
      TextColor(12);
      WriteLn(^j'                               Weiter bitte mit beliebiger Taste ...');
      ch:=Readkey;
      TextColor(14);
      Nutzenrechnung;
      r:=r+rinc;
    Until r>=robent+rinc;
  end;
  If Auswahl=#50 then
  Begin
    Zx:=Zxunten;
    Repeat
      WriteLn(^j^j^j, 'Wert des variierten Parameters: ', Zx:10:2);
      TextColor(12);
      WriteLn(^j'                               Weiter bitte mit beliebiger Taste ...');
      ch:=Readkey;
      TextColor(14);
      Nutzenrechnung;
      Zx:=Zx+Zxinc;
    Until Zx>=Zxoben+Zxinc;
  end;
  If Auswahl=#51 then
  Begin
    V:=Vunten;
    Repeat
      WriteLn(^j^j^j, 'Wert des variierten Parameters: ', V:10:2);
      TextColor(12);
      Write(^j'                               Weiter bitte mit beliebiger Taste ...');
      ch:=Readkey;
      TextColor(14);
      Nutzenrechnung;
      V:=V+Vinc;
    Until V>=Voben+Vinc;
  end;
  If Auswahl=#52 then
  Begin
    i:=iunten;
    Repeat
      WriteLn(^j^j^j, 'Wert des variierten Parameters: ', i:10:2);
      TextColor(12);
      Write(^j'                               Weiter bitte mit beliebiger Taste ...');
      ch:=Readkey;
      TextColor(14);
      Nutzenrechnung;
      i:=i+iinc;
    Until i>=ioben+iinc;
  end;

```

```

If Auswahl=#53 then
Begin
  tax:=taxunten;
  Repeat
    WriteLn(^j^j^j, 'Wert des variierten Parameters: ', tax:10:2);
    TextColor(12);
    WriteLn(^j'                               Weiter bitte mit beliebiger Taste ...');
    ch:=Readkey;
    TextColor(14);
    Nutzenrechnung;
    tax:=tax+taxinc;
  Until tax>=taxoben+taxinc;
end;
If Auswahl=#54 then
Begin
  sdw:=sdwunten;
  Repeat
    WriteLn(^j^j^j, 'Wert des variierten Parameters: ', SDw);
    TextColor(12);
    Write(^j'                               Weiter bitte mit beliebiger Taste ...');
    ch:=Readkey;
    TextColor(14);
    Nutzenrechnung;
    sdw:=sdw+sdwinc;
  Until sdw>=sdwoben+sdwinc;
end;
If Auswahl=#55 then
Begin
  cla:=claunten;
  Repeat
    WriteLn(^j^j^j, 'Wert des variierten Parameters: ', cla);
    TextColor(12);
    WriteLn(^j'                               Weiter bitte mit beliebiger Taste ...');
    ch:=Readkey;
    TextColor(14);
    Nutzenrechnung;
    cla:=cla+clainc;
  Until cla>=claoben+clainc;
end;
If Auswahl=#56 then
Begin
  c0:=c0unten;
  Repeat
    WriteLn(^j^j^j, 'Wert des variierten Parameters: ', c0);
    TextColor(12);
    Write(^j'                               Weiter bitte mit beliebiger Taste ...');
    ch:=Readkey;
    TextColor(14);
    Nutzenrechnung;
    c0:=c0+c0inc;
  Until c0>=c0oben+c0inc;
end;
End;
ClrScr;
Write(^j^j^j^j'Noch eine NA mit denselben Mitarbeiterzahlen N(k)? (j/n) ');
Repeat
  Read(ch);
  Until (ch=#74) or (ch=#106) or (ch=#110) or (ch=#78);
Until (ch=#110) or (ch=#78);

ClrScr;
TextColor(14);
ClrScr;
WriteLn(^j^j'                               Programmende');
TextColor(12);
Write(^j^j^j^j^j^j'                               Weiter bitte mit beliebiger Taste ...');
ch:=Readkey;
TextColor(14);
End.

```

Literatur

- ABRAHAMSON, N.A., ALF, E.F., & WOLFF, J.H. (1971). Taylor-Russel tables for dichotomous criterion variables. *Journal of Applied Psychology*, 55, 449-457.
- BARTHEL, E. (1988). *Nutzen eignungsdiagnostischer Verfahren bei der Bewerberauswahl*. Dissertation an der Universität Hohenheim.
- BARTHEL, E. (1989). *Nutzen eignungsdiagnostischer Verfahren bei der Bewerberauswahl*. Frankfurt am Main: Lang.
- BARTHEL, E. & SCHULER, H. (1989). Nutzenkalkulation eignungsdiagnostischer Verfahren am Beispiel eines biographischen Fragebogens. *Zeitschrift für Arbeits- und Organisationspsychologie*, 33, 73-83.
- BOBKO, P. & KARREN, R. (1982). The estimation of standard deviations in utility analysis. In K.H. Chung (Ed.), *Proceedings of the forty-second annual meeting of the Academy of Management* (p. 272-276). New York: Academy of Management.
- BOBKÖ, P., KARREN, R., & PARKINGTON, J.J. (1983). Estimation of standard deviations in utility analyses: An empirical test. *Journal of Applied Psychology*, 68, 170-176.
- BORTZ, J. (1989). *Statistik für Sozialwissenschaftler*. 3. Aufl. Berlin: Springer.
- BOUDREAU, J.W. (1983a). Economic considerations in estimating the utility of human resource productivity improvement programs. *Personnel Psychology*, 36, 551-576.
- BOUDREAU, J.W. (1983b). Effects of employee flows on utility analysis of human resource productivity improvement programs. *Journal of Applied Psychology*, 68, 396-406.
- BOUDREAU, J.W. (1991). Utility analysis for decisions in human resource management. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of Industrial and Organisational Psychology* (621-745). Vol. 2. 2nd ed. Palo Alto: Consulting Psychologists Press.
- BROGDEN, H.E. (1946a). On the Interpretation of the correlation coefficient as a measure of predictive efficiency. *The Journal of Educational Psychology*, 37, 65-76.

- BROGDEN, H.E. (1946b). An approach to the problem of differential prediction. *Psychometrika*, 11 (3), 139-154.
- BROGDEN, H.E. (1949). When testing pays off. *Personnel Psychology*, 2, 171-183.
- BROGDEN, H.E., & TAYLOR, E.K. (1950). The dollar criterion - Applying the cost accounting concept to criterion construction. *Personnel Psychology*, 3, 133-154.
- BURKE, M.J., & FREDERICK, J.T. (1984). Two modified procedures for estimating Standard deviations in utility analyses. *Journal of Applied Psychology*, 69, 482-489.
- BURKE, M.J., & FREDERICK, J.T. (1986). A comparison of economic utility estimates for alternative SD_y estimation procedures. *Journal of Applied Psychology*, 71, 334-339.
- BURKE, M.J., RAJU, N.S., & PEARLMAN, K. (1986). An empirical comparison of the results of five validity generalisation procedures. *Journal of Applied Psychology*, 71, 349-353.
- LALLENDER, J.C., & OSBURN, H.G. (1981). Testing the constancy of validity with Computer generated sampling distributions for the multiplicative model variance estimate: Results for petroleum industry validation research. *Journal of Applied Psychology*, 66, 274-281.
- CAMPBELL, D.T., & FISKE D.W. (1959). Convergent and discriminant validation by the - multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- CAMPBELL, D.T., & STANLEY, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (ed.), *Handbook of research on teaching*. Chicago: Rand McNally.
- CAMPBELL, D.T., & STANLEY, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- CASCIO, W.F. (1980). Responding to the demand for accountability: A critical analysis of three utility models. *Organizational Behavior and Human Performance*, 25, 32-45.
- CASCIO, W.F. (1982). *Costing human resources: The financial Impact of behavior in organizations*. Boston (Mass.): Kent.
- CASCIO, W.F., & SILBEY, V. (1979). Utility of the assessment Center as a selection device. *Journal of Applied Psychology*, 64, 107-118.

- CASCIO, W.F., & RAMOS, R.A. (1986). Development and application of a new method for assessing job performance in behavioral/ economic terms. *Journal of Applied Psychology*, 71, 20-28.
- CASCIO, W.F., & MORRIS, J.K. (1990). A critical reanalysis of Hunter, Schmidt, and Loggins' 1988 "Problems and pitfalls in using capital budgeting and financial accounting techniques in assessing utility of personnel programs. *Journal of Applied Psychology*, 75, 410-417.
- CATTELL, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- LOOK, T.D., & CAMPBELL, D.T. (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Chicago: Rand McNally.
- COWARD, W.M., & SACKETT, P.R. (1990). Linearity of ability-performance relationships: A re-confirmation. *Journal of Applied Psychology*, 75, 297-300.
- CRONBACH, L.J. (1965). Comments by Lee J. Cronbach. In L.J. Cronbach & G.C. Gleser, *Psychological tests and personnel decisions* (S. 266-267). 2nd ed. Urbana: University of Illinois Press.
- CRONBACH, L.J., & GLESER, G.C. (1957). *Psychological tests and personnel decisions*. Urbana: University of Illinois Press.
- CRONBACH, L.J., & GLESER, G.C. (1965). *Psychological tests and personnel decisions*. 2nd ed. Urbana: University of Illinois Press.
- CRONSHAW, S.F., & ALEXANDER, R.A. (1985). One answer to the demand for accountability: Selection utility as an Investment decision. *Organizational Behavior and Human Decision Processes*, 35, 102-118.
- CRONSHAW, S.F., & ALEXANDER, R.A. (1991). Why capital budgeting techniques are suited for assessing the utility of personnel programs: A reply to Hunter, Schmidt, Loggin. *Journal of Applied Psychology*, 76, 454-457.
- DESIMONE, R.L., ALEXANDER, R.A., & CRONSHAW, S.F. (1986). Accuracy and reliability of SD_y estimates in utility analysis. *Journal of Occupational Psychology*, 59, 93-102.
- DÖRNER, D. (1992). Die Logik des Mißlingens. Strategisches Denken in komplexen Situationen. Hamburg: Rowohlt.
- EATON, N.K., WIND, H., & MITCHELL, K.J. (1985). Alternative methods of estimating the dollar value of performance. *Personnel Psychology*, 38, 27-40.

- ECKARDT, H.H. & SCHULER H. (1992). Berufseignungsdiagnostik. In R.S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (S. 533-551). 2. Aufl. Weinheim: Psychologie Verlags Union.
- EDWARDS, J.E., FREDERICK, L.T., & BURKE, M.J. (1988). Efficacy of modified CREPID SD_y^s an the basis of archival organizational data. *Journal of Applied Psychology*, 73, 529-535.
- FELTHAM, R. (1988). Assessment Center decision making: judgmental vs. mechanical. *Journal of Occupational Psychology*, 61, 237-241.
- FISSENI, H.-J. (1990). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- FUNKE, U., SCHULER, H. & MOSER, K. (in Vorbereitung). Nutzenanalyse zur ökonomischen Evaluation eines Personalauswahlprojekts für Industrieforscher. In T.J. Gerpott & S.H. Siemers (Hrsg.), *Evaluation von Personalprogrammen*.
- GADENNE, V. (1976). *Die Gültigkeit psychologischer Untersuchungen*. Stuttgart: Kohlhammer.
- GADENNE, V. (1984). *Theorie und Erfahrung in der psychologischen Forschung*. Tübingen: Mohr.
- GLASE, G.V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- GREER, O.L., & CASCIO, W.F. (1987). Is cost accounting the answer? Comparison of two behaviorally based methods for estimating the standard deviation of job performance in dollars with a cost-accounting-based approach. *Journal of Applied Psychology*, 72, 588-595.
- GULLIKSEN, H. (1950). *Theory of mental tests*. New York: John Wiley.
- GUTTMAN, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, 19, 149-161.
- HAMBLETON, R.K., SWAMINATHAN, H., & ROGERS, H.J. (1991). *Fundamentals of item response theory*. Newbury Park: Sage Publications.
- HAMBLIN, A.C. (1974). *Evaluation and control of training*. Maidenhead: McGraw Hill.
- HIRSH, H.R., SCHMIDT, F.L., & HUNTER, J.E. (1986). Estimation of employment validities by less experienced judges. *Personnel Psychology*, 39, 337-344.
- HOLLING, H. & LIEPMANN, D. (1993). Personalentwicklung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 285-316). Bern: Huber.
- HORVÀTH, P. (1990). *Controlling*. 3. Aufl. München.

- HULL, C.L. (1928). *Aptitude testing*. Yonkers, New York: World Book.
- HUNTER, J.E., & HUNTER, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- HUNTER, J.E., & SCHMIDT, F.L. (1982). Fitting people to jobs: The Impact of personnel selection on national productivity. In E.A. Fleishman & M.D. Dunnette (Eds.), *Human performance and productivity: Vol. 1: Human capability assessment* (p. 233-284). Hillsdale, N.J.: Erlbaum.
- HUNTER, J.E., SCHMIDT, F.L., & JACKSON, G.B. (1982). *Meta-analysis. Cumulating research findings across studies*. Beverly Hills: Sage Publications.
- HUNTER, J.E., SCHMIDT, F.L., & COGGIN, T.D. (1988). Problems and pitfalls in using capital budgeting and financial accounting techniques in assessing the utility of personnel programs. *Journal of Applied Psychology*, 73, 522-528.
- HUNTER, J.E., & SCHMIDT, F.L. (1990). *Methods of meta-analysis. Correcting error and bias in research findings*. Newbury Park: Sage Publications.
- JÄGER, R.S. (1986). *Der diagnostische Prozeß. Eine Diskussion psychologischer und methodischer Randbedingungen*. 2. Aufl. Göttingen: Hogrefe.
- JÄGER, R.S. & PETERMANN, F. (1992). Einleitung. In R.S. Jäger & - F. Petermann (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch* (S. 11-13). 2. Aufl. Weinheim: Psychologie Verlags Union.
- JESERICH, W. (1981). *Mitarbeiter auswählen und fördern: Assessment-Center-Verfahren*. München: Hanser.
- JÖRESKOG, K.G., & SÖRBOM, D. (1988). *LISREL 7 - Guide to the program and application*. Chicago: SPSS Inc.
- KAHNEMAN, D., & TVERSKY, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology*, 3, 430-454.
- KAHNEMAN, D., & TVERSKY, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- KAISER, H.F., & DICKMANN, K. (1959). Analytic determination for common factors. *American Psychologist*, 14, 452 ff.
- KELLEY, T.L. (1923). *Statistical method*. New York: Macmillan.
- KIRKPATRICK, D.L. (1960). Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, 14, 13-18, 28-32.
- KLIMOSKI, R.J., & STRICKLAND, W.J. (1977). Assessment Centers: Valid or merely prescient. *Personnel Psychology*, 30, 353-363.

- KOMPA, A. (1984). *Personalbeschaffung und Personalauswahl*. Stuttgart: Entre.
- LANDY, F.J., FARR, J.L., & JACOBS, R.R. (1982). Utility concepts in performance measurement. *Organizational Behavior and Human Performance*, 30, 15-40.
- LATTMANN, C. (Hrsg.) (1989). *Das Assessment-Center-Verfahren der Eignungsbeurteilung. Sein Aufbau, seine Anwendung und sein Aussagegehalt*. Heidelberg: Physica-Verlag.
- LIENERT, G.A. (1989). *Testaufbau und Testanalyse*. 4. Aufl. München: Psychologie Verlags Union.
- LORD, F.M., & NOVICK M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison Wesley.
- MAUKISCH, H. (1986). Erfolgskontrollen von Assessment Center-Systemen: Der Stand der Forschung. *Zeitschrift für Arbeits- und Organisationspsychologie*, 30, 8691
- MICHEL, L. & CONRAD, W. (1982). Theoretische Grundlagen psychometrischer Tests. In K.J. Groffmann & L. Michel (Hrsg.), *Grundlagen psychologischer Diagnostik. Enzyklopädie der Psychologie*, B/ 11/ 1 (S. 1-129), Göttingen: Hogrefe.
- MICHEL, L. & MAI, N. (1968). Entscheidungsprobleme und Probleme der Diagnostik bei Cronbach & Gleser. *Diagnostica*, 14, 99-121.
- MURPHY, K.M. (1986). When your top choice turns you down: The effect of rejected offers an the utility of selection tests. *Psychological Bulletin*, 99, 133-138.
- NAERSSSEN, R.F. VAN (1965). Application of the decision theoretical approach to the selection of drivers. In L.J. Cronbach & G.C. Gleser, *Psychological tests and personnel decisions* (S. 273-290). 2nd ed. Urbang: University of Illinois Press.
- NAYLOR, J.C., & SHINE, L.C. (1965). A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology*, 3, 33-42.
- NEUBERGER, O. (1991). *Personalentwicklung*. Stuttgart: Entre.
- NIESCHLAG, R., DICHTL, E. & HÖRSCHGEN, H. (1991). *Marketing*. 16. Aufl. Berlin: Duncker & Humblot.
- OBERMANN, C. (1992). *Assessment Center. Entwicklung Durchführung Trends*. Wiesbaden: Gabler.
- 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.
- PEARLMAN, K., SCHMIDT, F.L., & HUNTER, J.E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65, 373-406.
- POPPER, K.R. (1984). *Objektive Erkenntnis. Ein evolutionärer Entwurf*. 4. Aufl. Hamburg: Hoffmann & Campe.
- PYNES, J., BERNARDIN, H.J., BENTON, A.L., & McEVOY, G.M. (1988). Should assessment center dimension ratings be mechanically-derived? *Journal of Business and Psychology*, 2, 217-227.
- PYNES, J., & BERNARDIN, H.J. (1992). Mechanical vs consensus-derived assessment center ratings: A comparison of job performance validities. *Public Personnel Management*, 21, 17-28.
- RAJU, N.S., & BURKE, M.J. (1983). Two new procedures for studying validity generalization. *Journal of Applied Psychology*, 68, 382-395.
- RASCH, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut.
- ROCHE, W.J. (1961). *The Cronbach-Gleser utility function in fixed-treatment employee selection*. Unpublished doctoral dissertation, Southern Illinois University, Carbondale.
- ROCHE, W.J. (1965). A dollar criterion in fixed-treatment employee selection. In L.J. Cronbach & G.C. Gleser. *Psychological tests and personnel decisions*. 2nd ed. Urbang: University of Illinois Press.
- ROTHER, H.F. (1978). Output rates among industrial employees. *Journal of Applied Psychology*, 63, 40-46.
- ROTHER, H.F., & NYE, C.T. (1959) Output rates among machine Operators: II. Consistency related to methods of pay. *Journal of Applied Psychology*, 43, 417-420.
- SANDS, W.A. (1973). A method for evaluating alternative recruiting-selection strategies: The CAPER model. *Journal of Applied Psychology*, 57, 222-227.
- SCHILDBACH, T. (1992). Begriff und Grundprobleme des Controlling aus betriebswirtschaftlicher Sicht. In K. Spremann & E. Zur (Hrsg.), *Controlling* (S. 21-36). Wiesbaden: Gabler.

- SCHMIDT, F.L., GAST-ROSENBERG, I., & HUNTER, J.E. (1980). Validity generalization results for Computer programmers. *Journal of Applied Psychology*, 65, 643-661.
- SCHMIDT, F.L., HUNTER, J.E., MCKENZIE, R.C., & MULDROW, T.W. (1979). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology*, 64, 609-626.
- SCHMIDT, F.L., HUNTER, J.E., & PEARLMAN, K. (1982). Assessing the economic Impact of personnel programs on workforce productivity. *Personnel Psychology*, 35, 333-347.
- SCHMIDT, F.L., & HUNTER, J.E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology*, 68, 407-414.
- SCHMIDT, F.L., ONES, D.S., & HUNTER, E.H. (1992). Personnel selection. *Annual Review of Psychology*, 43, 627-670.
- SCHMITT, N., GOODING, R.Z., NOE, R.A., & KIRSCH, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422.
- SCHULER, H. (1992). Assessment Center als Auswahl- und Entwicklungsinstrument: Einleitung und Überblick. In H. Schuler & W. Stehle (Hrsg.), *Assessment Center als Methode der Personalentwicklung*. 2. Aufl. (1-35). Göttingen: Verlag für Angewandte Psychologie.
- SCHULER, H. & FUNKE, U. (1993). Diagnose beruflicher Eignung und Leistung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 235-283). Göttingen: Huber.
- SCHULER, H. & STEHLE, W. (1983). Neuere Entwicklung des Assessment-CenterAnsatzes - beurteilt unter dem Aspekt der sozialen Validität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 27, 33-44.
- SCHULER, H. & STEHLE, W. (Hrsg.) (1992). *Assessment Center als Methode der Personalentwicklung*. 2. Aufl. Göttingen: Verlag für Angewandte Psychologie.
- SHERIF, M. (1966). *In common predicament: Social psychology of intergroup conflict and cooperation*. Boston: Houghton Mifflin.
- SIMON, H.A. (1957). *Administrative behavior: A study of decisions-making processes in administrative organization*. New York: McMillan.

- SIMON, H.A. (1981). *Entscheidungsverhalten in Organisationen: Eine Untersuchung von Entscheidungsprozessen in Management und Verwaltung*. 3. Aufl. Landsberg/ Lech: Verlag Moderne Industrie.
- TACK, W.H. (1976). Diagnostik als Entscheidungshilfe. In K. Pawlik (Hrsg.), *Diagnose der Diagnostik: Beiträge zur Diskussion der psychologischen Diagnostik in der Verhaltensmodifikation* (S. 103-130). Stuttgart: Klett.
- TAYLOR, H.C., & KUSSEC, 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.
- THIERAU, H., STANGEL-MESEKE, M. & WOTTAWA, H. (1992). Evaluation von Personalentwicklungsmaßnahmen. In K. Sonntag (Hrsg.), *Personalentwicklung in Organisationen* (229-249). Göttingen: Hogrefe.
- THORNTON, G.C., GAUGLER, B.B., ROSENTHAL, D.B., & BENTSON, C. (1992). Die prädiktive Validität des Assessment Centers - eine Metaanalyse. In H. Schuler & W. Stehle (Hrsg.), *Assessment Center als Methode der Personalentwicklung*. Aufl. (S. 36-77). Göttingen: Verlag für Angewandte Psychologie.
- WEIß, L. (in Vorbereitung). *Untersuchung zur prognostischen Validität und Konstruktvalidität eines Auswahl-Assessment-Centers anhand von Umsatzdaten*. Diplomarbeit an der Universität Mannheim.
- WESMAN, A.G. (1953). Better than Chance. *Test Service Bulletin*, 45, 1-67.
- WÖHE, G. (1993). *Einführung in die allgemeine Betriebswirtschaftslehre*. 18. Auflage. München: Vahlen.
- WOTTAWA, H. & HOSSIEP, R. (1987). *Grundlagen psychologischer Diagnostik. Eine Einführung*. Göttingen: Hogrefe.
- WOTTAWA, H. & THIERAU, H. (1990). *Evaluation*. Stuttgart: Huber.

Eidesstattliche Erklärung

Ich versichere, daß ich diese Diplomarbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Mannheim, den 17.01.1994

Wolfram Reiners