Evaluation der Eignungsdiagnostik bei der Personalauswahl von Unternehmensberatern

Validität und Nutzen am Beispiel eines Beratungsunternehmens

Tobias Plate

April 2006

Evaluation der Eignungsdiagnostik bei der Personalauswahl von Unternehmensberatern

Validität und Nutzen am Beispiel eines Beratungsunternehmens

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Sozialwissenschaften der Universität Mannheim

vorgelegt von
Dipl.-Betriebswirt Dipl.-Psychologe Tobias Plate
aus Heidelberg

Fakultät für Sozialwissenschaften
Dekan: Prof. Dr. Josef Brüderl
Prodekan: Prof. Dr. Jan van Deth
Referent: Prof. Dr. Werner W. Wittmann
Koreferent: Prof. Dr. André Beauducel

Tag der mündlichen Prüfung: 28. April 2006

Meinen lieben Eltern

Inhaltsverzeichnis

1 Einführung ... 1

1.1 Historischer Abriss der Eignungsdiagnostik ... 1

1.2 Humankapital bei Unternehmensberatungen ... 4

1.3 Die ökonomische Perspektive der Personalauswahl ... 5

1.4 Zielsetzung, Inhalt und Struktur der vorliegenden Arbeit ... 7

1.5 Einordnung der Arbeit in die Evaluationsforschung ... 9

2 Theoretischer Teil ... 11

2.1 Grundlagen ... 11

2.1.1 Evaluation ... 11

2.1.2 Die Evaluation der Personalauswahl als Intervention ... 15

2.1.3 Psychologische Diagnostik und Berufseignungsdiagnostik ... 18

2.1.4 Anforderungen an die Berufseignungsdiagnostik ... 21

2.1.5 Zusammenfassung ... 22

2.2 Der eignungsdiagnostische Prozess ... 23

2.2.1 Arbeits- und Anforderungsanalyse als Ausgangspunkt ... 24

2.2.2 Generierung und Operationalisierung der Kriterien ... 29

2.2.2.1 Taxonomie der Leistungskriterien ... 30

2.2.2.2 Zur Problematik der Erfassung und Bewertung der Kriterien ... 32

2.2.2.3 Das Vorgesetztenurteil im Kreuzfeuer der Kritik ... 34

2.2.3 Auswahl und Operationalisierung der Prädiktoren ... 37

2.2.4 Validitätsanalyse ... 40

2.2.5 Kombination der Prädiktordaten ... 48

2.2.5.1 Klinische vs. statistische Urteilsforschung ... 48

2.2.5.2 Statistische Prognoseverfahren und methodische Perspektiven ... 52

2.2.6 Entscheidungsmodelle und Ergebnisse der Auswahlentscheidung ... 58

2.2.7 Zusammenfassung ... 67

2.3 Nutzenmodelle bei der Personalauswahl ... 69

2.3.1 Übersicht historischer Ansätze ... 69

2.3.2 Zur Problematik der Parameteroperationalisierung ... 74

2.3.3 Neuere Entwicklungen bei Nutzenmodellen ... 78

2.3.4 Empirische Arbeiten zu Nutzenberechnungen im Auswahlkontext ... 82

2.3.5 Implikationen für die Nutzenberechnungen bei Unternehmensberatungen ... 83

2.3.6 Zusammenfassung ... 88

2.4 Berufserfolg in Unternehmensberatungen ... 89

2.4.1 Unternehmensberatung – Begriff, Wesen und Funktionen 90
2.4.2 Zur Problematik der Bewertung von Beratungsleistungen 92
2.4.3 Ableitung von Kriterien erfolgreicher Beratung 93
2.4.4 Implikationen für die vorliegende Studie 97
2.4.5 Zusammenfassung 98
2.5 Theoretische Perspektiven und empirische Befunde 99
2.5.1 Theoretische Grundlagen zu ausgewählten Prädiktoren des Berufserfolges . 99
2.5.1.1 Kognitive Fähigkeiten 99
2.5.1.2 Persönlichkeitskonstrukte 102
2.5.1.3 Interessen 103
2.5.1.4 Motivation 105
2.5.1.5 Wissen 106
2.5.2 Ausgewählte empirische Befunde zu Prädiktoren des Berufserfolges 107
2.5.2.1 Intelligenz 107
2.5.2.2 Persönlichkeit 110
2.5.2.3 Motivation 113
2.5.2.4 Biographie 114
2.5.2.5 Wissen 116
2.5.3 Überblick ausgewählter eignungsdiagnostischer Verfahren 117
2.5.3.1 Analyse der Bewerbungsunterlagen 121
2.5.3.2 Interviews 122
2.5.3.3 Assessment Center 127
2.5.4 Zusammenfassung 130
3 Empirischer Teil 133
3.1 Zentrale Fragestellungen der Untersuchung 133
3.2 Untersuchungsvorgehen 137
3.2.1 Untersuchungsgegenstand 137
3.2.2 Beschreibung des Auswahlverfahrens 138
3.2.3 Beschreibung der Stichprobe 141
3.2.4 Operationalisierung der Prädiktoren 142
3.2.5 Operationalisierung der Kriterien 146
3.2.6 Methode 150
3.2.6.1 Kohortenübergreifende Prädiktorenanalyse 151
3.2.6.2 Kohortenspezifische Analysen 155
3.2.6.3 Erweiterte Fragestellungen 165
3.3 Ergebnisse 166
3.3.1 Kohortenübergreifende Prädiktorenanalyse 166

3.3.1.1 Methodische Restriktionen ... 167
3.3.1.2 Deskriptive Analyse ... 168
3.3.1.3 Reliabilitäts- und Beurteileranalyse ... 170
3.3.1.4 Profilanalyse ... 175
3.3.1.5 Faktorenanalyse ... 180
3.3.1.6 Logistische Regressionsanalyse ... 183
3.3.2 Analyse der Kohorte Junior-Consultants ... 186
3.3.2.1 Methodische Restriktionen ... 187
3.3.2.2 Deskriptive Analyse ... 188
3.3.2.3 Mittelwert- und Profilvergleiche ... 189
3.3.2.4 Korrelationsanalyse ... 193
3.3.2.5 Regressionsanalyse ... 200
3.3.2.6 Faktorenanalyse ... 203
3.3.3 Analyse der Kohorte Consultants ... 207
3.3.3.1 Methodische Restriktionen ... 208
3.3.3.2 Deskriptive Analyse ... 208
3.3.3.3 Mittelwert- und Profilvergleiche ... 210
3.3.3.4 Korrelationsanalyse ... 215
3.3.3.5 Regressionsanalyse ... 220
3.3.3.6 Faktoren- und Pfadanalyse ... 223
3.3.4 Analyse der Kohorte Senior-Consultants ... 227
3.3.4.1 Deskriptive Analyse ... 229
3.3.4.2 Mittelwert- und Profilvergleiche ... 230
3.3.4.3 Korrelationsanalyse ... 234
3.3.4.4 Regressionsanalyse ... 237
3.3.4.5 Faktorenanalyse ... 238
3.4 Erweiterte Fragestellungen ... 240
3.4.1 Zur Auswahl und Operationalisierung von Prädiktor und Kriterium ... 241
3.4.2 Urteilsbildung und Verteilung diagnostischer Entscheidungen ... 243
3.5 Nutzenanalyse ... 244
3.5.1 Operationalisierung der Parameter ... 245
3.5.2 Nutzenberechnung ... 247
3.5.3 Sensitivitäts- und Break-Even-Analyse ausgewählter Parameter ... 248
3.5.4 Validitätserhöhung und Nutzenvorteil ... 251
3.5.5 Fazit ... 253
3.6 Implikationen für die Praxis ... 255
4 Zusammenfassung und Ausblick ... 265
Anhang ... 279
Literaturverzeichnis ... 327

Abbildungsverzeichnis

Abbildung 1: Gesamtvorgehen der Evaluation des Personalauswahlverfahrens 8
Abbildung 2: Modell der fünf Datenboxen 13
Abbildung 3: Schema der Interventionsstrategien 15
Abbildung 4: Arten der Personalentscheidungen 16
Abbildung 5: Beispiel für Evaluationsansätze bei der Personalauswahl 17
Abbildung 6: Eignungsdiagnostischer Prozess 23
Abbildung 7: Beschreibungsebenen eignungsdiagnostisch relevanter Instrumente 26
Abbildung 8: Beschreibungsdimensionen des O*Net 29
Abbildung 9: Kriterienwürfel 33
Abbildung 10: Drei-Ebenen-Modell der sozialen Urteilsbildung 35
Abbildung 11: Mögliche personelle Determinanten des beruflichen Erfolges 38
Abbildung 12: Ablauf einer Testkonstruktion nach der KTT 39
Abbildung 13: Beispiel für ein hierarchisches Brunswik'sches Linsenmodell 43
Abbildung 14: Darstellung des Gesamtzusammenhangs der Varianzen 46
Abbildung 15: Urteilsfehler 51
Abbildung 16: Beispiele bivariater Zusammenhänge von Prädiktor und Kriterium 52
Abbildung 17: Beispiel für ein lineares Strukturgleichungsmodell 57
Abbildung 18: Konjunktive und kombinatorisch-kompensatorische Strategie 59
Abbildung 19: Sequentielle Strategien 60
Abbildung 20: Klassenzuordnungen bei der Personalauswahl 62
Abbildung 21: Erweiterte Vierfeldertafel 63
Abbildung 22: Einfluss von Basisrate und Selektionsrate bei gegebener Validität 64
Abbildung 23: Wert eines Testverfahrens in Abhängigkeit situativer Parameter 65
Abbildung 24: Verteilung von 2 Gruppen unterschiedlicher Eignung und ROC-Kurve 67
Abbildung 25: Mögliche weitere ökonomische Nutzenaspekte bei Auswahlverfahren 80
Abbildung 26: Funktionen von Beratungen 91
Abbildung 27: Modell zur Herleitung der Kriterien erfolgreicher Beratung 95
Abbildung 28: Kriterienmodell der fünf Säulen erfolgreicher Beratung 96
Abbildung 29: Implikationen für die Evaluationsstudie 98
Abbildung 30: Das Fünf-Faktoren-Modell der Persönlichkeit 102
Abbildung 31: RIASEC-Taxonomie von Holland 104
Abbildung 32: Zusammensetzung der Beratungsteams nach *Competence Center* 138
Abbildung 33: Ablauf des Auswahlverfahrens der ISB 139
Abbildung 34: Häufigkeiten der Beobachtungen und Interviews je Bewerber 143

Abbildung 35: Häufigkeiten der Urteile aus Interviews und Beobachtung 145
Abbildung 36: Zusammensetzung Projektbeurteilungen nach Funktionsstufe..... 148
Abbildung 37: Personenanzahl nach Menge aggregierter Urteile je Funktionsstufe..... 149
Abbildung 38: Datenstruktur der Studie 150
Abbildung 39: Vorgehen der kohortenübergreifenden Prädiktorenanalyse 154
Abbildung 40: Urteilsstrenge nach Funktionsstufe und Competence Center 174
Abbildung 41: Profilvergleiche am Beispiel *Forced Exit*..... 177
Abbildung 42: Profilvergleiche am Beispiel *MBA-Abschluss* und *Dr.-Titel* 178
Abbildung 43: 25%-Randquartile kommunikativer und logisch-analytischer Fähigkeiten 179
Abbildung 44: Profilvergleich von zwei Bewerbern bei identischem Recruiting-Gesamtwert 180
Abbildung 45: Beispielhafte Profilvergleiche der Projektbeurteilungen..... 192
Abbildung 46: Korrelationen Abiturnote und Recruiting-Gesamtwert mit Gesamtwert aus Projektbeurteilungen (JC) 197
Abbildung 47: Profilvergleich Projektbeurteilung Randquartile Level- und Profilfaktor 207
Abbildung 48: Boxplotbeispiele für Recruiting-, CV- und Projektbeurteilungsvariablen..... 210
Abbildung 49: Beispielhafte Profilvergleiche der Projektbeurteilungen bei Cs..... 213
Abbildung 50: Korrelationen Abiturnote und Recruiting-Gesamtwert mit Gesamtwert aus Projektbeurteilungen (C)..... 217
Abbildung 51: Profilvergleich Projektbeurteilung Randquartile Level- und Profilfaktor.. 226
Abbildung 52: Symmetrieneuordnung und Pfadmodell 228
Abbildung 53: Verteilung der diagnostischen Entscheidungen bei der C-Kohorte..... 244
Abbildung 54: Sensitivitätsanalyse mit Validität und Selektionsquote..... 250
Abbildung 55: Sensitivitätsanalyse mit SDy und Selektionsquote SQ..... 251
Abbildung 56: Inkrementeller Ergebniseffekt bei Änderung des Auswahlinstrumentes ... 253
Abbildung 57: Praktische Ansatzpunkte zur Verbesserung des Auswahlverfahrens 256

Tabellenverzeichnis

Tabelle 1: Vier Alternativdimensionen psychodiagnostischer Zielsetzungen 18
Tabelle 2: Leistungskriterien und ihre Vor- und Nachteile 31
Tabelle 3: Urteilstendenzen 36
Tabelle 4: Ergebnisse der Sawyer'schen Reanalyse und Beispiele 49
Tabelle 5: Ausgewählte kompensatorische Prognoseverfahren 54
Tabelle 6: Ausgewählte nichtkompensatorische Prognoseverfahren 56
Tabelle 7: Entscheidungsarten nach Cronbach & Gleser 58
Tabelle 8: Gütekriterien diagnostischer Entscheidungsstrategien 63
Tabelle 9: Entscheidungsregeln zur Minimierung des Zuordnungsfehlers 66
Tabelle 10: Methoden zur Schätzung von SDy 77
Tabelle 11: Mögliche Alternativen zur Nutzenanalyse 81
Tabelle 12: Charakteristika der Dienstleistung Unternehmensberatung 91
Tabelle 13: Evaluierung des Beratungserfolges 93
Tabelle 14: Durchschnittliche Validitätskoeffizienten ausgewählter Auswahlverfahren zur Vorhersage von Berufserfolg (Berufsleistung) und inkrementelle Validitäten zu kognitiven Leistungstest 118
Tabelle 15: Validität, Praktikabilität und Akzeptanz eignungsdiagnostischer Verfahren für externe Personalauswahl in der Einschätzung der Verwender 120
Tabelle 16: Einstiegsstufe nach Einstiegsjahr der Untersuchungsstichprobe 142
Tabelle 17: Beurteilte Dimensionen des AC 144
Tabelle 18: Erfasste CV-Variablen der Bewerberstichprobe 146
Tabelle 19: Kriteriendimensionen der Projektbeurteilungen 147
Tabelle 20: Gruppierungssystematik explorativer t-Tests nach Kohorten 156
Tabelle 21: Gruppierungssystematik für t-Tests nach Kohorten 157
Tabelle 22: Set-Bildung der Recruitingvariablen bei der JC-Kohorte 161
Tabelle 23: Inhaltlich sachlogisch zugeordnete Variablen je Kriteriendimension (Cs) 162
Tabelle 24: Deskriptive Kennwerte der Recruitingvariablen 169
Tabelle 25: Deskriptive Kennwerte nicht-dichotomer CV-Variablen 170
Tabelle 26: ICC-Koeffizienten mit Einweg-Zufallseffekten nach Recruitingvariablen 171
Tabelle 27: Variablen der Profilvergleiche und ihre Häufigkeiten 176
Tabelle 28: t-Test für abhängige Stichproben zum Vergleich der oberen Quartilwerte auf den Variablen Kommunikationsfähigkeit und logisch-analytisches Denken 179
Tabelle 29: Rotierte und unrotierte Komponentenmatrizen der Recruitingvariablen 181

Tabelle 30: Korrelation der Faktoren mit *Angestelltenverhältnis* und *Forced Exit* 183
Tabelle 31: Logistische Regression der abhängigen Variable *Angestelltenverhältnis* 183
Tabelle 32: Logistische Regression der abhängigen Variable *Forced Exit* 185
Tabelle 33: Deskriptive Kennwerte der Kriterien (Projektbeurteilungen) bei JCs 189
Tabelle 34: Explorative t-Tests ausgewählter CV-Kategorien für JCs 190
Tabelle 35: Gruppenvergleich (t-Test) biographischer Kategorien für JCs 191
Tabelle 36: Übersicht Mittelwertunterschiede im Zeitverlauf 193
Tabelle 37: Korrelationen Projektbeurteilungsdimensionen mit ausgewählten CV-Variablen (JCs) 194
Tabelle 38: Gesamtübersicht der Regressionsergebnisse der JC-Kohorte 201
Tabelle 39: Korrelationsmatrix der Faktorwerte der Recruitingfaktoren mit Projektbeurteilungsdimensionen 203
Tabelle 40: Komponentenmatrizen der Projektbeurteilungsvariablen 204
Tabelle 41: Korrelationsmatrix Projektbeurteilungsfaktoren mit Recruitingvariablen 205
Tabelle 42: Korrelationen ausgewählter Recruitingvariablen mit verschiedenen Kriterien 207
Tabelle 43: Deskriptive Kennwerte der Kriterien (Projektbeurteilungen) bei *Cs* 209
Tabelle 44: Explorative t-Tests ausgewählter CV-Kategorien für Cs 211
Tabelle 45: Gruppenvergleich (t-Test) biographischer Kategorien für Cs 212
Tabelle 46: Übersicht Mittelwertunterschiede im Zeitverlauf 214
Tabelle 47: Korrelationen Projektbeurteilungsdimensionen mit ausgewählten CV-Variablen (Cs) 215
Tabelle 48: Gesamtübersicht der Regressionsergebnisse der C-Kohorte 220
Tabelle 49: Korrelation Faktorwerte der Recruitingfaktoren der C-Kohorte mit Projektbeurteilungsdimensionen 224
Tabelle 50: Rotierte und unrotierte Komponentenmatrizen der Projektbeurteilungsvariablen 225
Tabelle 51: Korrelationen von Recruitingvariablen mit verschiedenen Kriterien 227
Tabelle 52: Deskriptive Kennwerte der Kriterien (Projektbeurteilungen) bei SCs 229
Tabelle 53: Explorative t-Tests ausgewählter CV-Kategorien für SCs 231
Tabelle 54: Gruppenvergleich (t-Test) biographischer Kategorien für SCs 232
Tabelle 55: Mittelwertunterschiede im Zeitverlauf 233
Tabelle 56: Korrelationen Projektbeurteilungsdimensionen mit ausgewählten CV-Variablen (SCs) 235
Tabelle 57: Korrelationen Gesamtwerte aus Projektbeurteilung und Recruiting 236

Tabelle 58: Gesamtübersicht der Regressionsergebnisse Kohorte SC 237
Tabelle 59: Komponentenmatrizen der Projektbeurteilungsvariablen 239
Tabelle 60: Operationalisierung der Parameter zur Nutzenberechnung 246
Tabelle 61: Kennwerte der einzelnen Behandlungsperioden 246
Tabelle 62: Mitarbeiterfluss bei der Nutzenberechnung 247
Tabelle 63: Kostenübersicht nach Perioden 248
Tabelle 64: Nutzenübersicht und Ergebnisbeitrag nach Perioden 248
Tabelle 65: Prognostische Nutzenübersicht nach Perioden 249
Tabelle 66: Angenommene differierende variablen Kosten 252
Tabelle 67: Angenommene Entwicklungskosten 252
Tabelle 68: Komponenten strukturierter Interviews 259

Abkürzungsverzeichnis

AA	Arbeits- und Anforderungsanalyse
AP	Associate Partner
bzw.	beziehungsweise
C	Consultant
etc.	et cetera
f.	folgende
ff.	fortfolgende
FN	Falsch Negative (zu Unrecht abgelehnte Bewerber)
FP	Falsch Positive (zu Unrecht eingestellte Bewerber)
i.H.v.	in Höhe von
i.d.R.	in der Regel
JC	Junior Consultant
KTT	Klassische Testtheorie
MBA	Master of Business Administration
o.Ä	oder Ähnliches
o.g.	oben genannte
P	Partner
PM	Project Manager
PTT	Probabilistische Testtheorie
RN	Richtig Negative (zu Recht abgelehnte Bewerber)
RP	Richtig Positive (zu Recht eingestellte Bewerber)
SC	Senior Consultant
s.o.	siehe oben
u.a.	unter anderem
u.Ä.	und Ähnliches
usf.	und so fort
usw.	und so weiter
z.B.	zum Beispiel
z.T.	zum Teil

Danksagung

Die Erstellung der vorliegenden Dissertation habe ich einer Vielzahl von Personen zu verdanken, die mich entweder durch Zuarbeit, als „Sparringspartner“ bei zahlreichen Diskussionen oder einfach als „seelischer Beistand“, wenn ich im Laufe der Fertigstellung an der einen oder anderen Stelle der Verzweiflung nahe war, unterstützt haben und für mich da waren. In der Hoffnung, keinen zu vergessen, möchte ich meinen Dank in drei Bereiche unterteilen: einen akademischen, einen privaten und einen beruflichen Teil.

Auf akademischer Seite gilt mein größter Dank meinem Doktorvater Prof. Dr. Werner W. Wittmann. Seit ich mit der Idee zu dieser Arbeit an ihn herantrat, hat er mich laufend unterstützt, wichtige Impulse gegeben, sowie in schwierigen Phasen ermutigt und motiviert, mit Begeisterung an der Fertigstellung der Dissertation weiter zu arbeiten. Eben genau das, was man sich von einem Doktorvater wünscht. Herzlichen Dank dafür. Außerdem gilt mein Dank meinem Zweitgutachter Prof. Dr. André Beauducel, der diese Arbeit nicht nur zügig und sehr detailliert begutachtet hat, sondern auch in unseren Diskussionen wertvolle, die Qualität der Arbeit verbessernde Anregungen gab. Ganz besonderer Dank gilt außerdem Dr. Nicolas Sander, der mir in der alltäglichen Arbeit als wichtigster Ansprechpartner mit zahlreichen wertvollen Tipps und Literaturempfehlungen zur Seite stand und mit dem ich viele Fragestellungen ausdiskutieren durfte.

Auf privater Seite gilt mein größter Dank meinen Eltern Heide und Alexander Plate. Ihnen verdanke ich meine gesamte Ausbildung und mein Werdegang und sie haben mich stets unterstützt und motiviert, das Promotionsvorhaben umzusetzen. Ihnen ist deshalb diese Arbeit gewidmet. Ferner möchte ich ganz herzlich meinen Schwiegereltern Dr. Hedda und Dr. Josef Ferdinand Seeberger danken, die mir insbesondere durch das zur Verfügung stellen der notwendigen Infrastruktur und Versorgung ein konzentriertes und ruhiges Arbeiten ermöglicht haben. Besonderer Dank gilt meiner geliebten Frau Thessa, die mir während der gesamten Zeit den Rücken freigehalten hat und ohne die meine Promotion deshalb nicht möglich gewesen wäre. Schließlich möchte ich auf privater Seite auch meinem Freund und ehemaligen Studien- sowie Arbeitskollegen Tobias Gelbert danken. Gerade in der schwierigen Anfangszeit der Arbeit hat er mir besonders geholfen, schnell auf die „richtige Spur“ zu kommen und das Thema richtig anzugehen.

Dass ich diese Arbeit berufsbegleitend erstellen durfte, verdanke ich zudem auf beruflicher Seite einer Vielzahl von Personen. Zunächst möchte ich meinem Mentor Dr. Ralf Moldenhauer danken, der mich stets ermutigt hat, die Promotion durchzuführen und durch seine Unterstützung es mir ermöglicht hat, dies berufsbegleitend zu tun. Ferner möchte ich mich bei Thomas Eichelmann und Nikolaus Held bedanken, die mir unbürokratisch die für die Arbeit notwendigen Ressourcen und Daten zur Verfügung gestellt haben, ohne die die Erstellung der Arbeit gar nicht möglich gewesen wäre. Ferner gilt besonderer Dank Dr. Nils Bickhoff und Dr. Christoph Kleppel, die in zahlreichen Doktoranden-Zirkeln interessante Diskussionen angeführt und wertvolle Anregungen zu dieser Arbeit geliefert haben. Für die Aufbereitung und Überlassung der Daten danke ich außerdem Ann-Kathrin Hindelang, Bettina Breloh, Silvia Zöller und Marie Knutsson. Vielen Dank für Euren Einsatz!

Schließlich möchte ich all jenen danken, die mir im Rahmen der Erstellung der Arbeit ebenfalls geholfen haben, die ich aber vergessen habe zu erwähnen.

Heidelberg, 29. April 2006

Tobias Plate

1 Einführung

„Das Wichtigste im Leben ist die Wahl eines Berufes.
Der Zufall entscheidet darüber.“

Blaise Pascal (1623-1662), französischer Religionsphilosoph und Naturwissenschaftler

Die Aussage von Blaise Pascal scheint auf den ersten Blick auf Grund der Unvereinbarkeit von Wahlmöglichkeit und Zufall widersprüchlich. Sie muss jedoch vor dem Hintergrund der zu jener Zeit den Personen angebotenen Berufsmöglichkeiten gesehen werden. Diese waren eher durch nicht beeinflussbare (also zufallsbedingte) Umstände, wie Geburtsort oder gesellschaftlicher Status, geprägt. Demgegenüber sind in der heutigen Zeit die Möglichkeiten der Berufswahl weitaus breiter gefächert. Ihre Wahlmöglichkeit ist allerdings dadurch eingeschränkt, dass auf Arbeitgeberseite eine Selektion des geeigneten Bewerbers stattfindet, die eine Passung von Person und Organisation bzw. Tätigkeit sicherstellen und letztlich zur Zufriedenheit aller Beteiligten beitragen soll. Insofern wird versucht, durch Personalauswahl sowohl der Wichtigkeit der Berufswahl gerecht zu werden, als auch den „Zufallsanteil“ an der beruflichen Entscheidung zu minimieren.

1.1 Historischer Abriss der Eignungsdiagnostik

Erste Ansätze des Selektionsgedankens bei Mitarbeitereinstellungen tauchten bereits vor 4000 Jahren in China auf (Wiggins, 1973, S.515). Rund 3000 Jahre später, während der Ming-Dynastie (1368-1644 v. Chr.), wurden bereits Vorformen von Assessment Center zur Führungskräfteauswahl durchgeführt (DuBois, 1970, S.3ff.). Das chinesische Auswahlverfahren wurde Anfang des 18. Jahrhunderts von britischen Missionaren und Diplomaten in Teilen übernommen und bildete später den Ausgangspunkt für die Auswahl britischer Staatsbeamter (Wiggins, 1973, S.516). Auch in der Bibel wird eine zweistufige Auswahlsituationen geschildert (Buch der Richter, 7, 1-8).

Während die geschilderten Beispiele auf keinerlei mess- und testtheoretischen Erkenntnissen fußten, wurden Ende des 19. Jahrhunderts erstmals systematischere Vorgehensweisen angewandt (Schuler, 2000, S.15). Zu jener Zeit beschäftigte sich der britische Forscher Sir Francis Galton mit dem Zusammenhang u.a. physischer Merkmale und der Erledigung bestimmter Aufgaben. Teilweise darauf aufbauend begann James McKeen Cattell etwas später (1890) in Amerika mit der Messung motorischer und sensorischer Fähigkeiten mithilfe von Testbatterien und der Untersuchung des menschlichen Denkvermögens. Zudem forschten Hermann Ebbinghaus, Alfred Binet und William Stern im Grundlagenbereich der Intelligenz. Basierend auf den Erkenntnissen und den entwickelten Methoden dieser For-

scher, zeigte Hugo Münsterberg (1912), dass zwischen im Labor erhobenen Daten und Leistungen am Arbeitsplatz Zusammenhänge bestehen. Er legte somit den Grundstein der heutigen Eignungsdiagnostik (DuBois, 1970).

Die wissenschaftlich-psychologisch fundierte und nach psychometrischen Prinzipien konstruierte Eignungsdiagnostik in Deutschland begann Anfang des 20. Jahrhunderts (Schuler, 2000, S.15). Sie wurde hauptsächlich von staatlichen Institutionen sowie großen Unternehmen (z.B. Siemens, AEG) angewandt und ging mit einer Vielzahl von Gründungen psychotechnischer Forschungsinstitute einher. Schwerpunkte lagen dabei auf Papier- und Bleistifttests sowie apparativen Verfahren und Arbeitsproben (Giese, 1924, S.3). Zu jener Zeit hatten bereits in den Vereinigten Staaten, hauptsächlich im militärischen Bereich, biographische Fragebögen, Persönlichkeits- und Intelligenztests ihre Anwendung gefunden (Barthel, 1988, S.26; Schuler, 2000, S.15). Ende der Zwanziger Jahre wurden in der Weimarer Republik bei der Offiziersauswahl multiple Auswahlverfahren, die als Vorläufer der heutigen Assessment Center zur Führungskräfteauswahl gesehen werden können und eine umfassende Persönlichkeitserfassung zum Ziel hatten, angewandt (Wiggins, 1973, S.516).

Basierend auf dem deutschen Auswahlverfahren wurde zur Zeit des Zweiten Weltkrieges die Auswahl britischer Offiziere neu entwickelt und später auch bei der Auswahl von britischen Staatsbeamten eingeführt. Der gestiegene Bedarf an Geheimdienstmitarbeitern nach Eintritt der Vereinigten Staaten in den Zweiten Weltkrieg führte wiederum jenseits des Atlantiks zu einer Übernahme des ehemals deutschen, von den Briten adaptierten Auswahlverfahrens. Hier erreichte das Testen einen ersten Höhepunkt, da zudem nahezu alle 14 Millionen amerikanischen Kriegsteilnehmer getestet wurden (Barthel, 1988, S. 28). Auch im industriellen Bereich wurden zunehmend Tests und systematische Verfahren zur Bewerberauswahl angewandt. Aus Amerika kam Mitte der Fünfziger die wahrscheinlich bahnbrechendste Studie in diesem Bereich, die *Management Process Study* der amerikanischen Telefongesellschaft AT&T (Bray, Campbell & Grant, 1974). Die dabei verwendeten Aufgaben und Tests finden größtenteils heute noch Anwendung. Von 1954 bis 1966 wurden darüber hinaus durch die *APA* Standards für die Entwicklung und Evaluation von Tests festgeschrieben, die in den Vereinigten Staaten ein gewisses Maß an Qualität sicherten. In Deutschland fokussierte man zu jener Zeit hingegen hauptsächlich auf die Anwendung explorativer Gesprächstechniken, ohne Gütekriterien zu berücksichtigen (Barthel, 1988, S.29; Hartmann, Haubl, Neuberger, Peltzer & Wakenhut, 1984, S. 79).

Aus methodischer Perspektive war damals die Entwicklung des Korrelationskoeffizienten nach Pearson (Hull, 1928, S. 9) ein entscheidender Schritt für die Beurteilung der

Brauchbarkeit von Auswahlverfahren. Sie bildete zudem die Grundlage für das Verfahren der Faktorenanalyse (Thurstone, 1931), das für die Erstellung zahlreicher Persönlichkeitstaxonomien (Herrmann, 1991, S.102ff.; Thurstone, 1935) ausschlaggebend war. Darüber hinaus erweiterten Cronbach und Gleser (1965) in ihren Arbeiten den bisherigen Validitätsfokus auf Aspekte der Relevanz diagnostischer Entscheidungen und der Folge von Fehlentscheidungen. Dies war Ausgangspunkt zahlreicher späterer Arbeiten zum Nutzen eignungsdiagnostischer Verfahren (insb. Schmidt, Hunter, McKenzie & Muldrow, 1979).

In den Sechzigern und Anfang der Siebziger herrschte international eine verstärkte gesellschaftskritisch motivierte oder fachwissenschaftlich begründete Ablehnung der Berufseignungsdiagnostik (Krise der Berufseignungsdiagnostik). Gründe hierfür lagen u.a. in der Diskussion um *Test-Fairness* (Kirkpatrick, Ewen, Barrett & Katzell, 1968) und der damit verbundenen Diskriminierung von Minderheiten (im Wesentlichen Farbige und Bürger lateinamerikanischer Abstammung in den Vereinigten Staaten). Zusätzlich entstand bei den Forschern zunehmend das Bewusstsein über die hohe Komplexität der Umwelt (Barthel, 1988, S.30), die das Denken in Systemen (unter Berücksichtigung von Fähigkeit, Motivation und Training) und die Abkehr von einfachen Auswahlmodellen verstärkte. In der Bundesrepublik dauerte die ethisch-politische Diskussion, begünstigt durch die Einführung von Hochschulzulassungstests, bis in die Achtziger Jahre hinein an (Brambring, 1983, S.425). Seitdem hat die Verwendung eignungsdiagnostischer Verfahren bei der Auswahl von Mitarbeitern in der Wirtschaft an Bedeutung gewonnen (Hossiep, 1995, S.1).

Mit Einzug der Informationstechnologie ist seit den Achtzigern auch die *computerunterstützte* Eignungsdiagnostik immer stärker eingesetzt worden. Computer werden dabei sowohl bei der Konstruktion, der Erfassung, der Auswertung als auch zur Entscheidungsfindung angewandt (Schuler, 2000, S.18). Zunehmend werden auch Psychologen an Personalentscheidungen in Unternehmen beteiligt (Schuler, Frier & Kaufmann, 1993). Ihre Rolle hat sich dabei vom Spezialisten hin zu breiter eingesetzten Mitarbeitern im Personalwesen entwickelt, die den Anforderungen der Verzahnung der Diagnostik mit dem Gesamtkonzept der Personalarbeit gerecht werden müssen (Wottawa, 1991, S.1ff.; de Wolff, 1989, S.81ff.).

Durch die von Schmidt und Hunter (Hunter & Schmidt, 1990; Schmidt, et al., 1993) entwickelten metaanalytischen Techniken zur *Validitätsgeneralisierung* wurden außerdem völlig neue Perspektiven eröffnet, die die weitgehende Situationsunabhängigkeit prognostischer, berufseignungsdiagnostischer Entscheidungen betrafen. Die psychologische Forschung ging zudem bis in die Achtziger Jahre von einem passiven Menschenbild aus und Berufserfolg wurde als Passungsproblem von Arbeitsplatz und Person aufgefasst. Seit den

Neunzigern wird Berufsentwicklung hingegen immer mehr als Ergebnis zielgerichteter Handlungen und der auf diese Ziele ausgerichteten Selbstregulation der Person verstanden. Dass heißt auch, dass der auszuwählende Mitarbeiter stärker in den Betrachtungsfokus gerückt ist.

Damit geht einher, dass sich Unternehmen von patriarchalischen, verantwortungszentrierten Strukturen hin zur Aufgaben- und Verantwortungsdelegation entwickelt haben (Mitarbeiter als Unternehmer im Unternehmen) (Stief, 2001, S.4). Die Gesellschaft und die Arbeitswelt befinden sich, nicht zuletzt beschleunigt durch die Globalisierung und der Entwicklung hin zu einer Dienstleistungsgesellschaft, seit einigen Jahren in einem tief greifenden Wandlungsprozess. Die Abkehr vom reinen Produktivitätsdenken vergangener Jahre hin zum Innovationsdenken und der Wissensgesellschaft (Fiedler-Winter, 2004, S. 32), bei dem zweifelsohne gerade europäische Unternehmen ihren Wettbewerbsvorteil sichern müssen, verdeutlicht die zunehmende Bedeutung des Humankapitals und damit seiner Auswahl (Etzel & Küppers, 2002, S.11).

1.2 Humankapital bei Unternehmensberatungen

Die geschilderten Entwicklungen und Anforderungen aus Kapitel 1.1 treffen auf die Branche der Unternehmensberater schon seit Jahren zu. Hohe Kundenorientierung, flache Hierarchien, Verantwortungsdelegation oder Projektarbeit sind nur einige der Rahmenbedingungen, die bei Beratungsunternehmen seit langem vorherrschen. Der *New-Economy-Boom* und das damit verbundene rasante Wachstum des Beratungsgeschäftes um das Jahr 2000 konnten allerdings die hohen Anforderungen an die Mitarbeiter von Beratungsunternehmen zeitweilig verdecken. Zur jener Zeit sind bei der Einstellung von Mitarbeitern, oftmals mangels geeigneter Bewerber, zur Sicherung des Wachstums Kompromisse eingegangen worden (Gross, 2003, S.42). Der starke Bedarf an Beratern, gepaart mit einem Mangel an qualifizierten Bewerbern („War for Talents“, Siemann, 2004, S.6), wurde oftmals über die „Stellschraube“ der Anforderungen geregelt und führte damit zu erheblichen Qualitätseinbußen beim Humankapital. Die starke Beratungsnachfrage auf Unternehmensseite ermöglichte dabei problemlos den Einsatz auch weniger qualifizierter Mitarbeiter.

Seit der Wende vom Verkäufer- zum Käufermarkt hat sich allerdings die Beziehung zwischen Klienten und Beratern vor dem Hintergrund starker Preiskämpfe und etlichen Skandalen stark gewandelt (Niewiem & Richter, 2004, S.9). Beratungskunden sind kritischer, restriktiver und verhandlungsstärker bei der Projektvergabe geworden, ihre Anforderungen an die Person des Beraters sind stark gestiegen (Lechner, Gesing, Kreutzer & Müller-

Stewens, 2005). Interne Projektorganisationen sowie Inhouse-Beratungen (Sandberg & Werr, 2003) haben den Beratungsunternehmen einen Teil ihrer Besonderheit und Einzigartigkeit genommen. Zudem ist der Wissensvorsprung bezüglich neuer Managementtechniken, Organisationsformen o.Ä., bei den Beratungen nicht mehr vorhanden. Dazu beigetragen hat ebenfalls der technologische Wandel. Im Zeitalter der Computertechnologie können Unternehmen zahlreiche Analysen und Modellrechnungen mit Leichtigkeit selbst durchführen. Via Internet und Wissensmanagement kommen sie an Informationen heran, die vor nicht allzu langer Zeit noch die Hilfe von Beratern erfordert hätten. Die hohe Geschwindigkeit von Wissenszuwachs und Innovation bringt zusätzlich eine kurze Lebenszeit einmal erworbener fachlicher Kompetenzen mit sich. Die Antizipation von Marktentwicklung und ständige Anpassung an Trends (Gross, 2003, S.42) verschärft darüber hinaus die Anforderungen an die Schlüsselqualifikationen eines Beraters (Eilles-Matthiessen, elHage, Janssen & Osterholz, 2002).

Unternehmensberatungen werden entscheidend von ihren Mitarbeitern geprägt. Neben dem fachlichen Know-how bestimmen die Innovationsfähigkeit, die Kreativität und nicht zuletzt die Sozialkompetenz der beschäftigten Berater langfristig den Erfolg. Sie sichern somit die Marktstellung des Unternehmens (Ashton, 1998; Hunt, 1995, S.248ff.). Fehleinstellungen können besonders in Unternehmensberatungen das Wohlbefinden und die Leistungsfähigkeit von ganzen Teams beeinflussen. Daneben erfordern Qualifikationsdefizite einen erhöhten Führungsaufwand. Hohe Fluktuationsraten vermindern zudem den Wissenstand des Beratungsunternehmens. Dabei sind Unternehmen nur begrenzt in der Lage, die Persönlichkeitsstrukturen und Grundeinstellungen ihrer Mitarbeiter zu verändern.

Vor diesem Hintergrund kommt einer professionellen Personalauswahl, die sowohl die fachlichen als auch persönlichen Komponenten der Beratertätigkeit treffend ermittelt, eine zentrale Bedeutung zu. Sie eröffnet die Chance, die Bewerber zu identifizieren, die den Anforderungen des Arbeitsplatzes am besten entsprechen, und wird so zur wichtigsten Einflussgröße auf die Mitarbeiterqualität im Unternehmen. Nötig hierfür ist eine moderne Personalarbeit, deren Diagnostik all diesen Anforderungen in der Auswahl, aber auch in der Entwicklung der Mitarbeiter gerecht wird.

1.3 Die ökonomische Perspektive der Personalauswahl

Bereits Münsterberg (1912, S.24) wies auf die ökonomische Bedeutung der Personalauswahl hin. Fehlentscheidungen bei der Personalauswahl von Beratern, die auf Grund des Aufgabenspektrums und der Qualifikation vergleichbar mit Führungskräften sind, führen zu

relativ gesehen erheblich höheren Kosten als bei der Einstellung von Mitarbeitern, die keine Führungsposition innehaben. Unternehmensberater haben neben dem Einfluss auf die wirtschaftliche Entwicklung des Beratungsunternehmens zudem einen Einfluss auf die wirtschaftliche Entwicklung des zu beratenden Unternehmens. Dies hat eine durchaus weiter gehende wirtschaftliche Tragweite als die klassische Führungskräfteauswahl (Hochholdinger, 2002, S.2ff.). Leistungsfähige Personalauswahlverfahren stellen somit einen monetär bezifferbaren Konkurrenzvorteil dar (Cook, 1998).

Obwohl also personalpolitische Entscheidungen ohne Zweifel von bedeutender ökonomischer Tragweite gekennzeichnet sind, nutzen dennoch viele Unternehmen nicht die aus wissenschaftlich-psychologischen Erkenntnissen entwickelten Verfahren. Dies ist insofern erstaunlich, da sie sich als besonders erfolgreich hinsichtlich Validitätsaspekten und somit letztlich auch der ökonomischen Vorteilhaftigkeit erwiesen haben (Cook, 1998, S.293ff.). Vielmehr ist das Verhalten zahlreicher Unternehmen nach wie vor von „Ignoranz und Hilflosigkeit“ (Etzel & Küppers, 2002, S. 12) gekennzeichnet. Augenscheinlich valide Methoden, die sich bereits seit Jahren als ungeeignet erwiesen haben, finden nach wie vor ihre Anwendung (Risch & Sommer, 1996, S.220). Neben der mangelnden Zusammenarbeit von Forschung und Praxis liegt dies sicher auch an den Verständnis- und Kommunikationsschwierigkeiten zwischen Psychologen und Betriebswirten sowie der geringen Kenntnis der Personalverantwortlichen über die Möglichkeiten der Nutzenberechnungen (Macan & Highhouse, 1994). Zu nennen ist in diesem Zusammenhang das sog. *Wissenschafts-Praxis-Dilemma*, das in einer mangelnden Zusammenarbeit der universitären Forschungseinrichtungen mit Unternehmen aus der Praxis begründet liegt. Ursache hierfür sind die Gegenpole der wissenschaftlichen Akribie und langwierigen Entwicklung von Verfahren auf der universitären Seite, sowie die Forderung nach praktikablen, oftmals den wissenschaftlichen Standards nicht genügenden Lösungen für die Praxis. Dies begünstigt ein Auseinanderdriften der Eignungsdiagnostik in Forschung und Praxis (Etzel & Küppers, 2002, S.14). Pragmatische Ansätze, die diese beiden Extreme verbinden und sowohl die Wissenschaft als auch die Praxis befruchten, sind hier wünschenswert. Die ökonomische Vorteilhaftigkeit valider Personalauslese sollte zudem Kern der Kommunikation der Psychologen mit dem Management innerhalb der Unternehmen sein. Die wissenschaftliche Sprache, bei den Psychologen gekennzeichnet durch zahlreiche statistisch-methodologische Begrifflichkeiten, muss hier der ökonomischen Ausdrucksweise weichen, denn, wie Cascio (1982, S.V) bereits treffend formulierte „...like it or not, the language of business is dollars, not correlation coefficients“. Dies kommt auch der in der betriebswirtschaftlichen Wissenschaft erkannten Notwendigkeit

der ökonomischen Messung immaterieller Werte, wie dem Humankapital, entgegen (Boudreau & Ramstad, 1999; Canibano, Garcia-Ayuso & Sanchez, 2000). Dem Psychologen wird es somit ermöglicht, seine Entscheidungen in ähnlicher Weise zu begründen, wie Betriebswirte oder Techniker es tun (Wottawa, 1991, S.4). Er kann so den komplexen Sachverhalt der Personalauswahl in einer Form ausdrücken, die im Controlling und bei der Unternehmensleitung gebräuchlich ist (Wickel-Kirsch, 2002, S.50ff.). Auch Beratungsunternehmen offerieren zunehmend Produkte, die den ökonomische Nutzen von Personalauswahlverfahren ermitteln (Fitz-Enz, 2000; Grossmann, 2000; Stamps, 2000).

Leider ist allerdings in Deutschland nach wie vor von einem Qualitätsproblem auszugehen. Zwar stellt die Entwicklung der DIN 33430 (Westhoff, et al., 2004) als weltweit erste Norm für Qualitätsstandards bei der berufsbezogenen Eignungsbeurteilung einen Schritt in die richtige Richtung dar. Die empirischen Bewährungskontrollen der meisten Auswahlverfahren sind bisher aber weitgehend ausgeblieben (Hossiep, 1995, S.1). Personaldiagnostische Verfahren sollten deshalb häufiger evaluiert werden (Hochholdinger, 2002, S.3). Entscheidend hierbei ist die Transparenz von Zusammenhängen der Ausgestaltung des Auswahlverfahrens mit dem späteren beruflichen Erfolg. Ebenso bedeutsam ist die Beurteilung des monetären Nutzens angewandter Instrumente. Dies sind Aspekte, die in den meisten Unternehmen auch auf internationaler Ebene bisher vernachlässigt wurden (Ryan, McFarland, Baron & Page, 1999). Die vorliegende Arbeit beschäftigt sich anhand des Beispiels einer Unternehmensberatung mit diesen Themen.

1.4 Zielsetzung, Inhalt und Struktur der vorliegenden Arbeit

In der vorliegenden Arbeit wird das Auswahlverfahren einer Unternehmensberatung evaluiert. Oberstes Ziel ist dabei die Reduktion personeller Fehlentscheidungen bei der Einstellung der Mitarbeiter und die Maximierung des Nutzens für das untersuchte Unternehmen. Das Vorgehen orientiert sich dabei an *Abbildung 1*. Im ersten Schritt werden die Prädiktoren und verwendeten Instrumente des vorliegenden Auswahlverfahrens analysiert. Techniken sind hier u.a. Reliabilitäts-, Beurteiler- sowie Profilanalysen mit dem Ziel, potenzielle Fehlerquellen auf der Prädiktorenseite zu reduzieren und Redundanzen aufzudecken. In einem zweiten Schritt steht die *tatsächliche* Validität des derzeit angewandten Auswahlverfahrens im Mittelpunkt. Hier werden Validitäten berechnet und über Korrelations-, Regressions- und Pfadanalysen Zusammenhänge der Prädiktoren mit den Kriterien (dem, was derzeit im Unternehmen als Beratungserfolg verstanden wird) aufgedeckt. Aus diesen auf der Messebene gewonnenen Ergebnissen werden erste praktische Empfehlungen zur Validitätserhöhung

abgeleitet. Darüber hinaus wird aus den Erkenntnissen der Wissenschaft untersucht, ob die im Unternehmen geltenden Kriterien das komplette Spektrum dessen abdecken, was man unter Beratungserfolg verstehen kann. Zudem wird analysiert, inwiefern sich die verwendeten Instrumente zur Messung des Beratungserfolges eignen. Die Evaluation des Personalauswahlverfahrens tangiert an dieser Stelle also auch Aspekte der Mitarbeiterbeurteilung (als Kriterium für erfolgreiche Beratung). Des Weiteren werden für die identifizierten Kriterien erfolgreicher Beratung Prädiktoren und zugehörige geeignete Messinstrumente aus dem derzeitigen Forschungsstand der Wissenschaft abgeleitet. Schließlich münden die Erkenntnisse in einer integrierten Darstellung eines möglichen neuen Gesamtkonzeptes und seiner ökonomischen Nutzenberechnung.

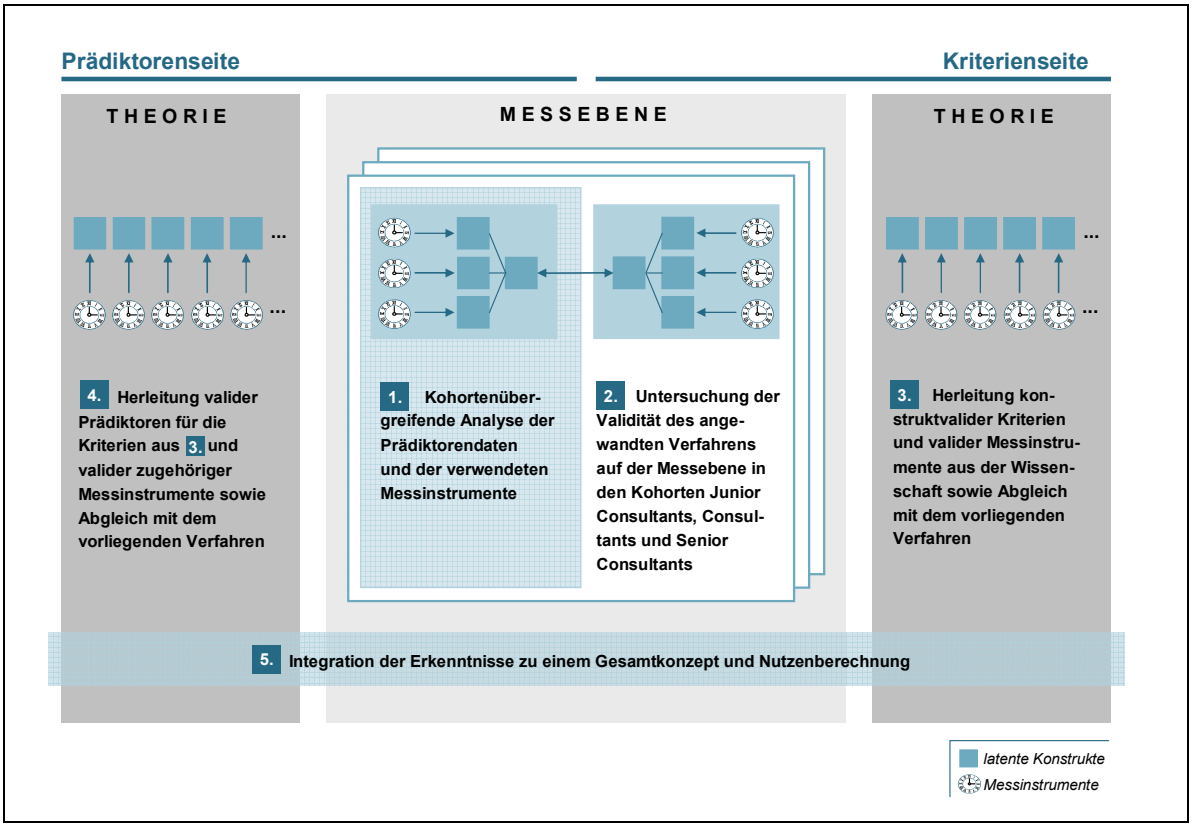

Abbildung 1: Gesamtvorgehen der Evaluation des Personalauswahlverfahrens

Für die Struktur der Arbeit, die in einen theoretischen und einen empirischen Teil gegliedert ist, hat das geschilderte Vorgehen folgende Auswirkungen: Zunächst werden in dem Kapitel 2.1 des theoretischen Teils Grundlagen zur Evaluation, zur Personalauswahl und zur Berufeignungsdiagnostik abgehandelt. Im Kapitel 2.2 wird anhand des eignungsdiagnostischen Prozesses auf zentrale Fragestellungen der Berufseignungsdiagnostik eingegangen. Das Kapitel 2.3 beinhaltet neben der Erörterung ausgewählter gängiger Nutzenmodelle die Implikationen für ein Nutzenmodell in der Unternehmensberatung. Kapitel 2.4 befasst sich

mit der Besonderheit des Berufserfolges in Unternehmensberatungen und erarbeitet Kriterien erfolgreicher Beratung. Im Kapitel 2.5 werden Prädiktoren des Berufserfolges beschrieben sowie ausgewählte diagnostische Verfahren erörtert. Zudem wird eine Übersicht der empirischen Befunde zu den Determinanten des Berufserfolges gegeben. Der dritte Teil der Arbeit beinhaltet die Empirie der vorliegenden Untersuchung. Zuerst werden die zentralen Fragestellungen erörtert und das Untersuchungsvorgehen skizziert. Anschließend werden die deskriptiven und inferenzstatistischen Ergebnisse dargestellt und interpretiert. Kapitel 3.4 setzt sich mit der Operationalisierung von Prädiktoren und Kriterien sowie den Folgen der diagnostischen Entscheidung auseinander. Im Anschluss wird der monetäre Nutzen des derzeit angewandten Auswahlverfahrens und möglicher Alternativen berechnet. Kapitel 3.6 erörtert darauf folgend die sich aus den Ergebnissen ergebenden Implikationen für die Praxis. Die Arbeit schließt im vierten Teil mit einer Zusammenfassung und einem Ausblick.

1.5 Einordnung der Arbeit in die Evaluationsforschung

Im Folgenden wird unter Evaluationsforschung im Sinne Suchmanns (1967) die explizite Verwendung wissenschaftlicher Forschungsmethoden und –techniken verstanden, die zur Durchführung einer Bewertung vorgenommen wird. Diese Auffassung, die die Möglichkeit des Beweises eines bestimmten Wertes einer Aktivität betont, wird von zahlreichen Evaluationsexperten geteilt (Bortz & Döring, 1995; Wittmann, 1985; Wottawa & Thierau, 2003). Die Einordnung der vorliegenden Arbeit in den Bereich der (auftraggebenden) Evaluationsforschung lässt sich durch folgende Aspekte rechtfertigen: Das Erkenntnisinteresse im vorliegenden Fall ist auf den Erfolg oder Misserfolg einer Maßnahme, nämlich des untersuchten Personalauswahlverfahrens begrenzt. Ziel ist es, diese Evaluationsfrage möglichst präzise zu beantworten. Die Generierung von für diese Frage nicht unmittelbar funktional wertvollem Hintergrundwissen ist hier von nachrangiger Bedeutung. Hierdurch grenzt sich auch Evaluationsforschung von angewandter oder Grundlagenforschung ab (Bortz & Döring, 1995, S.97). Darüber hinaus ist die Analyse vor dem Hintergrund der Weiterentwicklung der *technologischen Theorie* anzusiedeln. Sie liefert konkrete Ansatzpunkte, um wissenschaftliche Erkenntnisse in effiziente standardisierte Handlungsanleitungen umzusetzen. Die praktische Nutzbarmachung hat hier Priorität (Chen, 1990). Dieser Anwendungskontext soll nicht zuletzt auch dazu beitragen, das unter 1.1 geschilderte Kommunikationsproblem zwischen Psychologen und Betriebswirten zu neutralisieren und die Akzeptanz erarbeiteter Gestaltungsvorschläge zu erhöhen. Was man unter Evaluation, Personalauswahl und Berufeignungsdiagnostik genau versteht und welche Ausgestaltungsmöglichkeiten es gibt, soll nachfolgend im ersten Kapitel des theoretischen Teils näher betrachtet werden.

2 Theoretischer Teil

„Nichts ist geeigneter, uns den rechten Weg zu weisen,
als die Kenntnis der Vergangenheit.“
Polybios (200 v. Chr. - etwa 120 v. Chr.), griechischer Staatsmann und Feldherr

Ein zentraler Aspekt dieser Arbeit ist es, aus Vergangenheit gewonnene Erkenntnisse zu nutzen, um die zukünftige Ausgestaltung eines Auswahlverfahrens zu verbessern, also auf den „rechten Weg“ zu gelangen. Zunächst werden dafür in Kapitel 2.1 begriffliche und thematische Grundlagen abgehandelt, um zu einem eindeutigen Verständnis der Problemstellungen zu gelangen. Kapitel 2.2 zeigt anhand des eignungsdiagnostischen Prozesses die Abfolge von Schritten, die Auswahlverfahren vor ihrer Anwendung (also in der Vergangenheit) im Rahmen ihrer Entwicklung klassischerweise durchlaufen. Diese Abfolge dient als Grundlage zur Erklärung problematischer Aspekte, möglicher Fehler und zentraler Fragestellungen, die retrospektiv betrachtet für mögliche unerwünschte Ergebnisse verantwortlich sein können. Kapitel 2.3 beschäftigt sich mit dem „rechten Weg“, d.h. der Frage, wann man vom „rechten Weg“ sprechen kann, nämlich dann, wenn man seinen Nutzen maximiert. Inhalt sind deswegen verschiedene Nutzenmodelle und die Erarbeitung von Implikationen für ein Nutzenmodell bei Unternehmensberatungen. Das darauf folgende Kapitel 2.4 beschreibt im Anschluss die Besonderheiten der Erfolgsmessung bei Unternehmensberatungen und mündet in der Ableitung von Kriterien erfolgreicher Beratung. Im Rahmen des Kapitels 2.5 wird neben den Instrumenten und Prädiktoren des Berufserfolges auf empirische Befunde eingegangen, also die „Kenntnis der Vergangenheit“ erweitert, oder vielmehr Erkenntnisse aus vergangenen Untersuchungen bzw. Metaanalysen herangezogen, um den gegenwärtigen Zustand zu verbessern.

2.1 Grundlagen

In diesem Kapitel sollen die Begriffe *Evaluation*, *Berufseignungsdiagnostik* und *Intervention* vor dem Hintergrund der Personalauswahl näher betrachtet und Voraussetzungen der Eignungsdiagnostik zum besseren Verständnis der weiteren Arbeit abgehandelt werden.

2.1.1 Evaluation

In den unterschiedlichen wissenschaftlichen Disziplinen wird der Begriff der Evaluation vor vielfältigen Hintergründen verwendet, so dass eine einheitliche Definition bisher ausgeblieben ist (Wottawa & Thierau, 2003, S.13). Die unterschiedlichen Definitionsversuche unterscheiden sich hauptsächlich darin, wie weit oder eng der Evaluationsbegriff aufge-

fasst wird (siehe dazu auch Bortz & Döring, 1995, S.95ff.; Guba, 1969; Rossi & Freeman, 1985; Wentling & Lawson, 1975; Wittmann, 1985, S.17ff.; Wottawa, 1990). Im Folgenden orientiert sich die Arbeit an Suchmans (1967) Auffassung, die unter Evaluation den „Prozess der Beurteilung eines Produktes, Prozesses oder Programms, ohne expliziter Notwendigkeit der Verwendung systematischer Verfahren oder datengestützter Beweise“ versteht. Zusätzlich zu den von Suchmann (1967) definierten Begriffen der *Evaluationsforschung* (siehe 1.5.) und der *Evaluation* wurde von Abramson (Abramson, 1979) der Begriff der *Programmevaluation* eingeführt. Darunter versteht man eine Serie von Tätigkeiten, die rigoros wissenschaftliche Forschung in allen ethisch vertretbaren Situationen einsetzt (Wittmann, 1985, S. 21) und somit immer anwendungsbezogen ist. Programmevaluation ist somit weiter gefasst als der Begriff Evaluationsforschung; oftmals wird auch ihr interdisziplinärer Charakter betont (Sechrest & Figueredo, 1993).

Über die besprochene Frage hinaus, was Evaluation genau ist, lassen sich mehrere Paradigmen identifizieren und diese, je nach Zielsetzung und Bewertungskriterien, kontrovers diskutieren (Hochholdinger, 2002, S.11). Zum einen betrifft dies die Frage, ob *qualitative* oder *quantitative* Erhebungsverfahren eingesetzt werden sollten. Extrempositionen, die den Alleinvertretungsanspruch eines dieser beiden scheinbaren Gegenpole reklamieren, sind heutzutage allerdings eher selten. Vielmehr sind sogar Integrationsversuche auf methodischer und methodologischer Ebene unternommen worden (Bortz & Döring, 1995, S.281). Eine weitere, kritisch diskutierte Kontroverse betrifft die Wahl des Untersuchungsdesigns (quasiexperimentell vs. experimentell) (Cook & Campbell, 1976; Rossi & Freeman, 1985, S. 34). Diese beiden forschungsbetonten Positionen werden ergänzt durch Cronbachs (1982) idiosynkratische Ansicht, die die Einzigartigkeit jeder Untersuchung herausstellt und die Maximierung des Nutzens für die Entscheidungsträger unter Berücksichtigung gegebener Rahmenbedingungen (Ressourcen, Politik, Interessensvertreter, Sponsoren, usf.) fordert. Darüber hinaus gibt es noch Evaluatoren in der Praxis, die jegliche Forschungsansätze ablehnen, sich bei der Evaluation auf ihre Weisheit, ihre Intuition und ihren Erfahrungsschatz berufen und in jedem Fall die absolute Einzigartigkeit sehen (Rossi & Freeman, 1985, S.37). Das letzte Dilemma betrifft die Frage, ob Evaluationen inhaltsbasiert, also theoriegeleitet sein sollten oder ob die rein methodische Perspektive ausreichend ist (Weiss, 2001). Wortmann (1983, S.24) bezeichnet in diesem Zusammenhang Evaluation auch als eine multidisziplinäre, atheoretische Aktivität. Oft wird auch von „black box evaluation“ gesprochen (Chen, 1990, S.18).

In der Literatur gibt es zahlreiche Versuche, Evaluationsvorhaben wie das der vorliegenden Arbeit zu klassifizieren oder Typologien zu bilden, um eine Systematik in die Vielzahl der Erscheinungsformen zu bringen. Dabei werden, je nach Ansatz, verschiedene Komponenten unterschiedlich stark gewichtet (Wittmann, 1985, S. 42). Scriven (1991) unterscheidet zwischen formativer und summativer Evaluation. Stufflebeam (2000) entwickelte das CIPP-Modell. Hinter dem Akronym verbergen sich die Kürzel aufeinander folgender Phasen, denen sich unterschiedliche Evaluationstypen zuordnen lassen: Kontext-, Input-, Prozess- und Programmevaluation. Weitere Klassifikationsansätze gibt es von House (1980; siehe Wittmann, 1985, S. 43), Glass und Ellet (1980), Stufflebeam (2001) und Wottawa und Thierau[1] (2003)). Für die vorliegende Arbeit ist das *Modell der fünf Datenboxen* von Wittmann (1990) relevant.

Das Modell der vier Datenboxen (Wittmann, 1985, S. 90), das später um eine Datenbox erweitert, zu einem *Fünf-Datenboxen-Modell* wurde, ist im Rahmen der Entwicklung der *Reliabilitätstheorie* von Wittmann zur Systematisierung von Evaluationsstudien entworfen worden. Ausgangsbasis ist hierbei die zu berücksichtigende Information, die bei den Studien anfällt; das Modell eignet sich deshalb gut zur Darstellung methodischer Fragestellungen. *Abbildung 2* gibt die Modellstruktur wieder. Anlehnend an das Konzept der Cattellschen Datenwürfel mit den Dimensionen Person, Situationen und Variablen (Cattell, 1966) besteht das Modell aus einer Prädiktor- (PR-), einer experimentellen (ETR-), einer nichtexperimentellen (NTR-), einer Kriterien-(KR-) und einer Evaluationsbox (EVA).

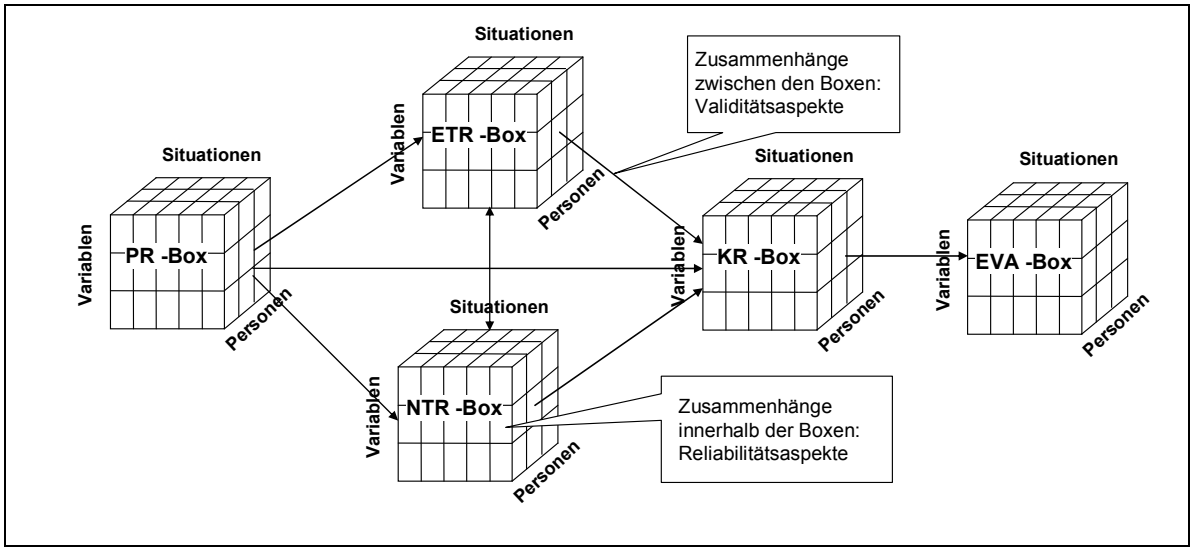

Abbildung 2: Modell der fünf Datenboxen (Wittmann, 1990)

[1] Obwohl Wottawa und Thierau die Möglichkeit eines idealtypischen Rasters zur Klassifikation von Evaluationsstudien verneinen, strukturieren sie unterschiedliche Evaluationsklassifizierungen anhand eines typischen Projektverlaufs

In der *PR-Box* sind alle Daten enthalten, die vor Untersuchungsbeginn vorhanden sind und als unabhängige Variablen zur Vorhersage der Kriterien herangezogen werden können. Die *KR-Box* enthält alle abhängigen Variablen, die zur Ermittlung der Wirksamkeit einer Intervention verwendet werden können. Die sich dazwischen befindende *ETR-Box* beinhaltet Informationen über Maßnahmen, die im Rahmen eines randomisierten Kontrollgruppenexperimentes durchgeführt wurden. In einem solchen Fall können die Effekte stärker auf Grund hoher *interner Validität* dem Treatment zugeordnet werden. Analog dazu befindet sich in der *NTR-Box* jene Treatment-Information, die nicht unter vollständig randomisierten Umständen stattgefunden und auf die der Evaluationsforscher keinen Einfluss hat. In diesen Fällen kann es zu einer Konfundierung der Effekte der *PR-* und *KR-Box* kommen (Wittmann, 1985, S.91). Die *EVA-Box* schließlich enthält Informationen zu der Forschungsfrage und den Evaluationszielen, die Grundlage der Untersuchung aller anderen Boxenbeziehungen sind. Das Konzept von Wittmann kann, wie unter 2.2.4 gezeigt wird, herangezogen werden, um verschiedene Varianzquellen beim Vergleich der Boxen zu ermitteln bzw. Varianzen in ihre Einzelbestandteile zu zerlegen. Eine sinnvolle Interpretation ist allerdings nur möglich bei Einhaltung der Brunswik'schen Symmetrieprinzipien (Brunswick, 1955).

Ein weiterer neuerer Systematisierungsversuch geht auf Chen (1990) zurück. Ausgehend von unterschiedlichen normativen und kausativen Theorien[2] (theoriegeleitete Evaluation) klassifiziert er Evaluationsstudien nach ihren Zielsetzungen. Daraus ergeben sich sechs verschiedene Evaluationstypen. Die *normative Treatmentevaluation* bewertet dabei, wie gut die Intervention umgesetzt wurde. Die *normative Evaluation der Implementierungsumgebung* betrachtet das Umfeld und die Rahmenbedingungen unter denen die Intervention stattfindet. Die *Outcome-* oder *Zielevaluation*, als letzte der normativen Evaluationsarten, beschäftigt sich schließlich mit den Kriterien einer Evaluation und ihrer Zielkonkordanz. Die *kausative Impact-Evaluation* hingegen untersucht interne Validitätsaspekte, d.h. inwiefern Veränderungen bei den Kriterien auch auf den Prädiktor zurückzuführen sind. Die *kausative Evaluation intervenierender Maßnahmen* geht der Frage nach, welche Prozesse hinter dem Zusammenhang von Prädiktor und Kriterium liegen. Die letzte Evaluationsart nach Chen ist die *kausative Generalisierungsevaluation*. Sie behandelt die externe Validität, also inwiefern gefundene Ergebnisse generalisierbar sind. Dabei legt sich Chen bei keiner Evaluationsart bezüglich der Wahl des Untersuchungsdesigns oder der Methoden fest. Vielmehr sollen diese fallspezifisch, in enger Abstimmung mit den Stakeholdern aus der Praxis, gewählt werden.

---

[2] Chens Theorieverständnis bezieht sich auf ein generelles substanzielles Wissen über Zusammenhänge und das Funktionieren gesellschaftlicher Vorgänge, wie im vorliegenden Fall beispielsweise das Zusammenspiel der agierenden Kräfte am Arbeitsmarkt

2.1.2 Die Evaluation der Personalauswahl als Intervention

Die Disziplin des Personalwesens beschäftigt sich mit Problemlösungen, die der Erfüllung der wirtschaftlichen Ziele der Organisation dienen (Kompa, 1989, S.2). Entscheidend ist hierbei die Frage der Gestaltung der Beziehung zwischen der Organisation und der Person. Pawlik (1976, S. 14 ff.) hat in seinem Konzept der Interventionsstrategien diese Gestaltungsansätze systematisiert (*Abbildung 3*). In Klammern sind jeweils praktische Anwendungsbeispiele genannt. Die Personalauswahl ist nach dieser Systematik also als eine Zuordnung von Personen zu vorgegebenen Arbeitsplätzen (Bedingungen) zu verstehen, oder genauer als der Entscheidungsprozess, an dessen Ende die Bestimmung derjenigen Kandidaten steht, die sich für eine bestimmte Position aus dem Kreise von Bewerbern als die Bestqualifizierten herausstellen (Wöhe, 1990, S.255). Personalauswahl als Intervention setzt damit an diagnostischen Feststellungen mit dem Ziel an, Veränderungen auf organisatorischer Ebene herbei zu führen (Amelang & Zielinski, 2002, S.433). Es ist allerdings auch möglich, umgekehrt vorzugehen, d.h. die Bedingungen auszuwählen, die sich für das vorhandene Personal als optimal erweisen.

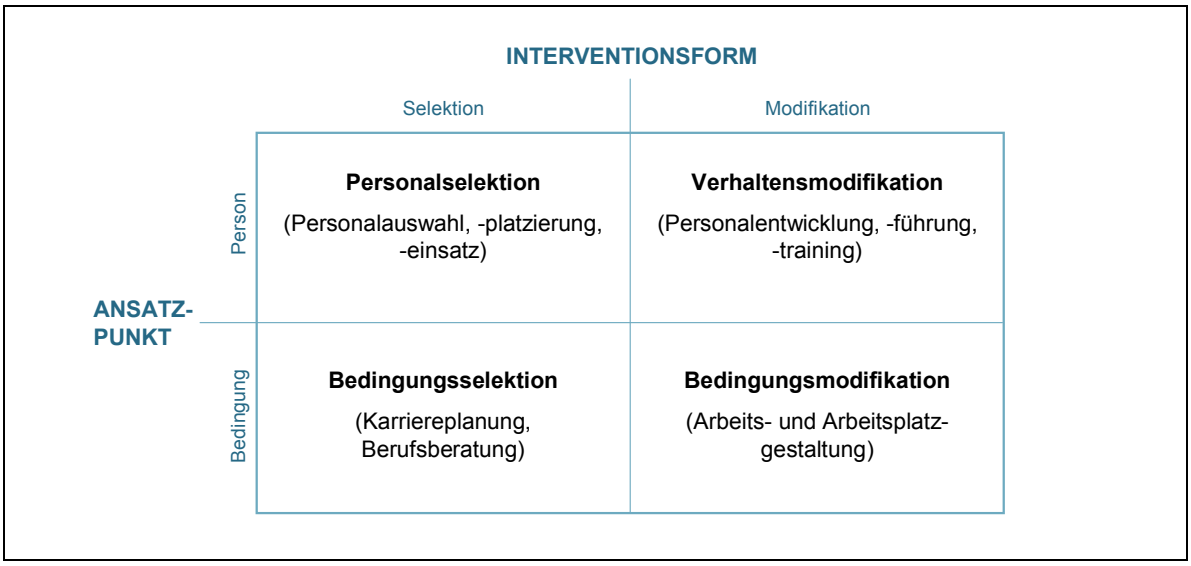

Abbildung 3: Schema der Interventionsstrategien (Pawlik, 1976, nach Krapp, 1986, S.573)

Ein klassisches Beispiel ist die Karriereplanung der Mitarbeiter, bei der je nach gegebenen Merkmalen Personen auf bestimmte Stellen (Bedingungen) befördert werden. Die beiden Selektionsstrategien gehen folglich von stabilen Persönlichkeits- und Situationsmerkmalen aus. Die zugrunde liegenden statischen Menschen- und Organisationsbilder sind dabei allerdings nicht ohne Kritik geblieben, da die Entwicklungsmöglichkeiten der einzelnen Person bei der sich stark verändernden Umwelt außer Acht gelassen werden. Zudem ist die Annahme der Konstanz organisatorischer Bedingungen vor diesem Hintergrund mehr als

fraglich (Boerlijst & Meijboom, 1989). Stattdessen gewinnt die Flexibilität von Mensch und Organisation immer mehr an Bedeutung. Diesem dynamischen Flexibilitätsgedanken tragen die Modifikationsstrategien Rechnung. Hier geht man von der Veränderbarkeit von Personen- und Umweltbedingungen aus. Beispiele hierfür sind auf der Personenseite Personaltrainings und auf der Organisationsseite die Arbeitsplatzgestaltung. In der Praxis tauchen die geschilderten Strategien nicht in Reinform, sondern als Mischstrategien auf (Kompa, 1989, S.5). Eine solche wäre die Auswahl von Mitarbeitern, die im Anschluss im Rahmen der Personalentwicklung betrieblich weitergebildet werden.

Die Intervention der Personalselektion, die den Schwerpunkt dieser Arbeit darstellt, lässt sich wiederum in unterschiedliche personelle Entscheidungsarten[3] einbetten. Dabei unterscheidet Wiggins (1973) zwei Dimensionen: die Möglichkeit der Ablehnung bzw. Zurückweisung der Teilnehmer sowie die Anzahl unterschiedlich zu besetzender Positionen. Daraus ergeben sich vier Arten von Personalentscheidungen (*Abbildung 4*).

		Zurückweisung der Bewerber	
		erlaubt	nicht erlaubt
Anzahl unterschiedlich zu besetzender Positionen	eine	**einfache Selektion**	**(Akzeptanz)**
	mehr als eine	**multiple Selektion**	**Klassifikation**

Abbildung 4: Arten der Personalentscheidungen (übersetzt aus Wiggins, 1973, S.231)

Personalauswahl kann also als psychologische Intervention angesehen werden. In der vorliegenden Arbeit wird dieses Interventionsverständnis allerdings weiter gefasst als bei Amelang und Zielinski (1997). Vielmehr wird der gesamte Personalauswahlprozess als Intervention verstanden. Die Eignungsdiagnostik ist dabei genauso ein Teilbereich wie verschiedene Handlungen, die als Eingreifen im System zu verstehen sind und Veränderungen hervorrufen. Diagnostik und Intervention werden demnach nicht als Abschnitte desselben psychologischen Prozesses im Sinne Fissenis (1990, S.5) verstanden, sondern Diagnostik als ein *Teilbereich* der Intervention, d.h. im vorliegenden Fall des Personalauswahlprozesses. Dieser Ansicht ist auch Pawlik (1976; S.18) bei der Beschreibung seiner Strategien: „Ände-

[3]Wiggins spricht in diesem Zusammenhang von „personnel decision problems“. Probleme sind hier eher als grundsätzliche Fragestellungen und weniger als Problematik zu verstehen

rungen in der Strategie angewandt-psychologischer Interventionen können Konsequenzen auch für die Entwicklung psychologischer Diagnostik haben, die *Teil dieser Interventionen* [Hervorhebung v. Verf.] sind“

Wie kann nun eine Evaluation der Personalauswahl aussehen? Die Taxonomie der Evaluationsarten von Chen (1990) hat sich bereits im Rahmen der Managementdiagnostik bei der Klassifikation von Studien bewährt (Hochholdinger, 2002). Nachfolgend sollen deshalb exemplarisch und anlehnend an diese Taxonomie evaluative Ansatzpunkte bei der Personalauswahl dargestellt werden. Dabei wird die Personalauswahl anhand eines exemplarischen Prozessmodells visualisiert.

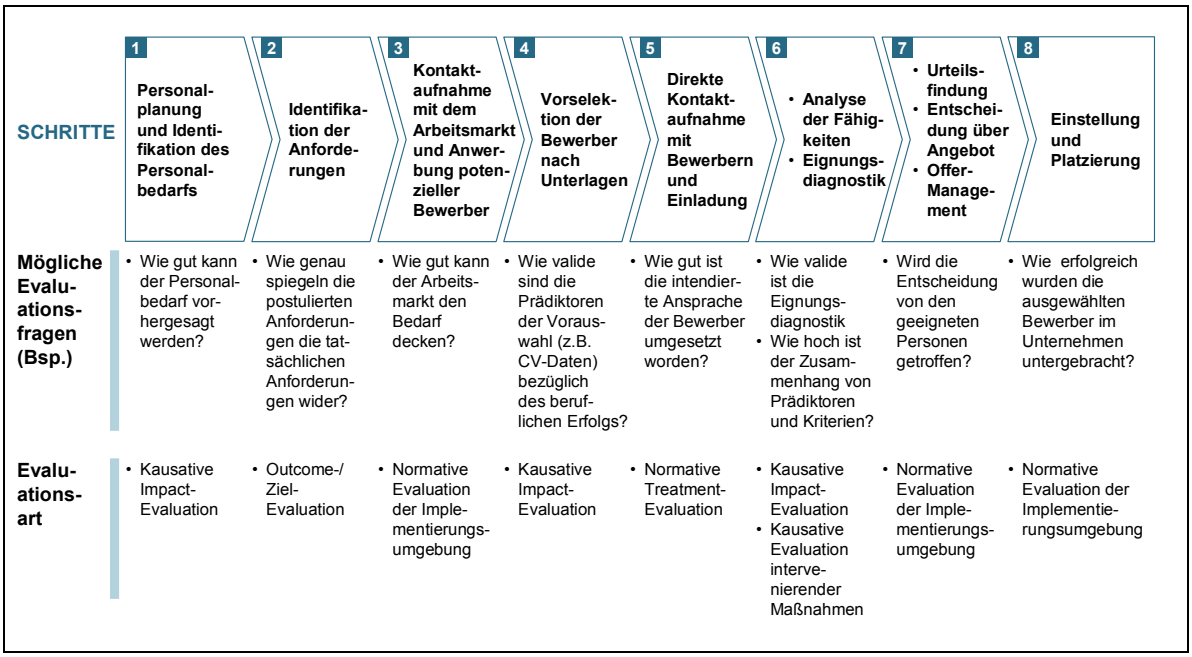

Abbildung 5: Beispiel für Evaluationsansätze bei der Personalauswahl (eigene Darstellung)

Abbildung 5 verdeutlicht, dass es eine Vielzahl von Evaluationsmöglichkeiten auf jeder Stufe des Personalauswahlprozesses gibt und dass Personalauswahl sehr weit gefasst werden kann, indem auch Aspekte der Personalbeschaffung sowie des Personalmarketings einbezogen werden. Dies trifft auch auf die Evaluation zu, die umfassender sein muss als die reine Fokussierung auf die Eignungsdiagnostik. Betrachtet man erneut *Abbildung 1*, so wird ebenso klar, dass die vorliegende Arbeit sich auf bestimmte Teilbereiche beschränkt. Diese betreffen im Wesentlichen die Schritte 2, 4, 6 und 7, die aus eignungsdiagnostischer Perspektive die entscheidenden Ansatzpunkte darstellen. Auf darüber hinaus gehende, für eine umfassende Bewertung des gesamten Verfahrens notwendige weitere Evaluationen wird aus Komplexitätsgründen verzichtet.

2.1.3 Psychologische Diagnostik und Berufseignungsdiagnostik

Der Begriff *Diagnostik* findet in vielen wissenschaftlichen Disziplinen Anwendung (z.B. in der Medizin oder im Technikbereich). In der Psychologie wird unter Diagnostik allgemein die Lehre von den Methoden und den Verfahren der sachgemäßen Durchführung einer Diagnose verstanden. Im Folgenden orientiert sich das Verständnis psychologischer Diagnostik an der Definition von Jäger und Petermann (1992):

> „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 gegebenen [sic] Daten zu einem Urteil (Diagnose, Prognose). Als Merkmalsträger gelten Einzelpersonen, Personengruppen, Institutionen, Situationen, Gegenstände, etc.." (Jäger & Petermann, 1992, S.11)

In der sehr detaillierten Definition von Jäger und Petermann kommt die Vielfalt der Objekte zum Ausdruck, die, je nach praktischer Zielsetzung Merkmalsträger psychologischer Diagnostik sein können, ohne diese zu begrenzen (z.B. Produkte, Organisationen, Personen, Personengruppen usf., siehe auch Petermann, 1992, S.493f.). Zur detaillierten Betrachtung der Rahmenbedingungen, Methoden, Anwendungen und Grundlagen der Diagnostik sei auf das diagnostische Dreieck von Hossiep und Wottawa (1993, S. 132) verwiesen.

Tabelle 1: Vier Alternativdimensionen psychodiagnostischer Zielsetzungen (Pawlik,1976)

Dimension	Diagnostische Zielsetzung
Statusdiagnostik vs. Prozessdiagnostik	Ist-Zustand (diagnostisch oder prognostisch) vs. Veränderungsmessung (diagnostisch oder prognostisch)
normorientierte Diagnostik vs. kriterienorientierte Diagnostik	Individuelle Unterschiede vs. Individuelle Position relativ zu einem Verhaltenskontinuum
Testen vs. Inventarisieren	Schließen von einer Verhaltensstichprobe auf einen Eigenschaftswert vs. Verhaltensrepertoire (Hierarchie)
Diagnostik als Messung vs. Diagnostik als Information für und über Behandlung	Schätzung eines Eigenschaftswertes vs. Entscheidungs- und Behandlungsoptimierung

Betrachtet man die Zielsetzungen der psychologischen Diagnostik, so lassen sich diese nach *Tabelle 1* in verschiedene Alternativdimensionen gruppieren (Pawlik, 1976, S.23). Bei der *Statusdiagnostik* liegt eine Stabilitätsannahme des zur erfassenden Merkmals zugrunde. Sie ist Grundlage der meisten Selektionsstrategien (Jäger, 1992, S.200; Pawlik, 1976,

S.24). Die *Prozessdiagnostik* ist in einer Vielzahl von Psychologiebereichen von großer Bedeutung, allerdings geht sie mit grundlegenden Problemen bei der Messung von Veränderungen einher (Jäger & Scheurer, 1992, S.202). Die *kriterienorientierte* Diagnostik zielt darauf ab, die einzelne Person unabhängig von Informationen über die Verteilung der Ausprägungsgrade der Charakteristika bei einer Menge von Personen (Schott, 1992, S. 227) zu beschreiben. Sie ist im Rahmen von Modifikationsstrategien deshalb häufig vorzufinden. Aus diagnostischer Sicht bilden Kriterium- und Normorientierung weniger Gegensätze. Vielmehr stellen sie sinnvolle Ergänzungen dar (Rettler, 1992, S. 225). So wird man beispielsweise bei der Personalauswahl nur Bewerber einladen, die mindestens zwei Auslandspraktika absolviert haben (*kriterienorientiert*) und von diesen jene auswählen, die im Interview zu den besten zehn Prozent gehören (*normorientiert*).

Bei der Diagnostik als *Testung* wird von einer Homogenitätsannahme ausgegangen, d.h. die Stichprobe spiegelt die Population der Verhaltensweisen in ausreichendem Maße *inhaltsvalide* wider (wie z.B. bei Simulationsspielen im Rahmen der Personalauswahl) (Scheurer, 1992, S.260ff.)[4]. Will man eine exakte Arbeitsanweisung formulieren, so handelt es sich bei der zugrunde liegenden Arbeitsanalyse um die Variante des *Inventarisierens* (Pawlik, 1976, S.32; Schmidtke, 1992, S.263ff.). Bei der Diagnostik als *Messung* sind in der Psychologie die zu messenden Merkmale häufig latent und nur indirekt messbar über manifeste Merkmalsindikatoren (Conrad, 1992, S.245). Diagnostik als *Information* für und über Behandlung ist hingegen weiter gefasst und konzentriert sich nicht alleine an Messgütekriterien. Stattdessen orientiert sie sich an der Angemessenheit der Entscheidungen, die auf Grundlage der Diagnose getroffen werden, sowie deren Effektivität (Pawlik, 1976, S. 34).

Die dargestellten Dimensionen spiegeln zwei grundsätzliche, verschiedene Modellvorstellungen der Diagnostik wider: *Eigenschafts*- und *Verhaltensdiagnostik* (Amelang & Zielinski, 1997, S.8 ff.). Die *Eigenschaftsdiagnostik* geht davon aus, dass sich das Erleben und Verhalten von Menschen in Form von Eigenschaften („traits“) beschreiben lässt. Es wird dabei ein Analogieschluss vom Verhalten auf die Eigenschaften gezogen, um wiederum eine Vorhersage für künftiges Verhalten vorzunehmen. Entscheidend dabei ist die Repräsentativität des Verhaltens für die Eigenschaft. Der Bezugsrahmen ist die *normorientierte* Messung. Demgegenüber geht der *verhaltenstheoretische* Ansatz davon aus, dass sich zukünftiges Verhalten durch vergangenes Verhalten am besten vorhersagen lässt. Dafür sind Stich-

[4] Das geschilderte Stichprobenproblem trifft je nach Fragestellung auch auf Situationen oder Personen zu (vgl. Scheurer, 1992, S.260ff).

proben des vorherzusagenden (Kriterium-)Verhaltens für einen Induktionsschluss auf allgemeines Verhalten notwendig (z.B. beim Rollenspiel im Rahmen der Personalauswahl).

Eignungsdiagnostik ist als jener Teilbereich psychologischer Diagnostik zu verstehen, der in der Entwicklung und Verwendung psychologischer Methoden zum Zwecke eignungsbezogener Prognosen und Entscheidungshilfen besteht (Schuler & Funke, 1991b, S.V). *Berufseignungsdiagnostik*, als Teilbereich davon, ist demnach:

> „... ein Oberbegriff für alle Arten der (gedanklichen) Zuordnung von Personen und beruflichen Situationen. Wissenschaftliche Basis der Berufseignungsdiagnostik sind vor allem Theorien der Anforderungen, Fähigkeiten und Leistungen sowie Methoden zu deren Messung und Modelle der Klassifikation. Praktische Einsatzbereiche sind alle Arten berufsbezogener Einschätzung, Beratung, Selektion und Zuordnung zum Zwecke der individuellen Berufs-Organisations- und Arbeitsplatzwahl, der institutionellen Berufsberatung sowie der Auswahl und Entwicklung von Mitarbeitern in Wirtschafts- und Verwaltungsorganisationen." (Schuler, 1988, S. 201)

Die von Schuler in seiner Definition angesprochenen praktischen Einsatzbereiche erinnern stark an die Interventionsstrategien nach Pawlik *(Abbildung 3)*, die Amelang und Zielinski (1997, S.8) auch zur Unterteilung der Arten psychologischer Diagnostik verwenden. Sie verdeutlichen das breite Spektrum der Einsatzmöglichkeiten der Berufseignungsdiagnostik. Sie hat dabei nicht alleine die Aufgabe, zur Erreichung organisatorischer Ziele beizutragen, sondern soll auch die Ziele des Bewerbers berücksichtigen. Maukisch (1978, S.107) spricht in diesem Zusammenhang auch von „Prognose der auf Optimierungskriterien bezogenen Interaktionseffekte". Kompa (1989, S. 35) sieht in ihr deshalb eine methodologische Grundlage der Personalauswahl, aus der Gestaltungshinweise für konkrete Auswahlverfahren abgeleitet werden können. Dabei versteht er unter Eignung den Grad der Entsprechung zwischen den Merkmalsträgern (Person, Organisation), der anhand von Optimierungskriterien beurteilt werden muss. Dies entspricht dem ursprünglich aus der Arbeits- und Organisationspsychologie stammenden *Person-Job-Fit-Ansatz* (Amelang & Zielinski, 1997, S.89).

In der Praxis wird allerdings die eignungsdiagnostische Problemstellung in erster Linie die Frage betreffen, inwiefern der Bewerber sich für die zu besetzende Stelle eignet. Von dem Unternehmen wird eine ausreichende Information über die zu besetzende Stelle und ihr Umfeld erwartet. Die Eigenverantwortlichkeit des Bewerbers vorausgesetzt, wird i. d. R. aber niemand vom Unternehmen eine Bedingungsselektion erwarten, die auf den Bewerber zutrifft. Dies entspricht auch der Annahme der Selbstregulation der Person bei der beruflichen Entwicklung aus den Neunziger Jahren (siehe 1.1).

2.1.4 Anforderungen an die Berufseignungsdiagnostik

Die Frage der Vorhersagbarkeit von Berufserfolg durch den Einsatz berufseignungsdiagnostischer Verfahren vor dem Hintergrund einer Selektionsstrategie setzt eine Reihe von Grundannahmen voraus. Diese betreffen die individuelle Streuung der Optimierungskriterien innerhalb der Gesamtpopulation pro Arbeitsplatz und Variation der Streuung der Rangordnungen von Arbeitsplatz zu Arbeitsplatz. Zudem sollte es eine regelhafte, gesetzmäßige Beziehung zwischen differenzialpsychologischen Merkmalen und den Optimierungskriterien geben (Kompa, 1989, S.36ff.). Berufseignungsdiagnostik ist, wie berichtet, der Teilbereich der Eignungsdiagnostik, der sich mit der gedanklichen Zuordnung von Personen zu Situationen und Situationen zu Personen beschäftigt. Dabei werden teilweise komplexe Messinstrumente eingesetzt, die wiederum Maßzahlen liefern, die mit dem Sachverhalt in Beziehung stehen. Messtheoretische Grundlage der meisten Operationalisierungen ist dabei die auf fünf Axiomen basierende klassische Testtheorie (KTT; Lord & Novick, 1968). Die KTT ermöglicht die Berechnung von Testgütekriterien, zu denen die Validität, Reliabilität und Objektivität als die wichtigsten gehören (siehe Amelang & Zielinski, 1997, 2002; Fisseni, 1990; Lord & Novick, 1968; Michel & Conrad, 1982. Aufbauend auf der Kritik an der KTT wurde als Weiterentwicklung die probabilistische Testtheorie (PTT) entwickelt, die auf Wahrscheinlichkeitszusammenhängen von manifesten, messbaren Variablen und latenten Konstrukten basiert. In der Berufseignungsdiagnostik spielt sie allerdings noch keine bedeutende Rolle.

Die Berufseignungsdiagnostik endet gängigerweise mit der Feststellung des Zustandes des Zielobjektes, also der Person, die die Grundlage für die Ableitung von Behandlungsmaßnahmen (Annahme oder Ablehnung) darstellt (Fisseni, 1990, S.4). Nach Barthel (1988, S.54ff.) gibt es dabei drei bedeutende Anforderungen, denen die Diagnostik genügen muss: die unterschiedlichen Interessen aller Betroffenen im Auswahlkontext, die rechtlichen Rahmenbedingungen sowie die genannten Gütekriterien aus der Testtheorie. Bezüglich Letzterer ist Wittmann (1987) der Ansicht, *Konstruktvalidität* sollte häufig besser als *Konstruktreliabilität* bezeichnet werden und nur dann der Validitätsbegriff Anwendung finden, wenn tatsächlich begrifflich getrennte Konstrukte korreliert werden. *Konstruktreliabilität* soll hingegen bei der Frage der Genauigkeit der Abbildung eines Konstruktes, also dem proportionalen Grad der Überlappung zwischen Indikatormenge und Konstrukt verwendet werden. Einen Ansatz, mit dessen Hilfe Konstruktreliabilität gemessen werden kann, stellt der *Multitrait-Multimethod-Ansatz* (MTMM) von Campbell und Fiske (1959) dar, nach dem jede Fähigkeit mit mehreren Methoden gemessen werden soll. Dies ermöglicht eine Separierung der Varianzanteile der Messwerte, die auf den Test, und solcher, die auf die Eigenschaft zu-

rückzuführen sind. Im Rahmen des *MTMM-Ansatzes* haben Campbell und Fiske zwei neue Validitätsarten eingeführt: *konvergente* und *diskriminante* Validität. Konvergente Validität meint dabei das Ausmaß, in dem theoretische Zusammenhänge von Konzepten auch in der Realität wieder zu finden sind. Diskriminante Validität ist hingegen das Ausmaß, in dem theoretisch nicht zu erwartende Zusammenhänge in der Realität tatsächlich auftauchen. Beide Validitäten kann man mit Hilfe der MTMM quantifizieren. *Konvergente Validität*, oder *Konstruktreliabilität* nach Wittmann, ist dann die Korrelation der Messwerte mehrerer Instrumente, die dieselbe Eigenschaft messen. Diskriminante Validität kennzeichnet demgegenüber die Korrelation zwischen gleichen Methoden bei Messung unterschiedlicher Eigenschaften. Für eine effektive MTMM-Validierung sollten sich die konvergenten Validitäten signifikant von Null unterscheiden und höher sein als die diskriminanten Validitäten bei gleicher und unterschiedlicher Methode. Darüber hinaus sollten die Muster diskriminanter Validitäten äquivalent sein (Fisseni, 1990, S.70).

2.1.5 Zusammenfassung

Die Darstellung der zahlreichen Begriffsdefinitionen der Evaluation führt zu der Erkenntnis, dass es kein allumfassendes von allen einheitlich akzeptiertes Konzept gibt, das allen Anforderungen gerecht wird. Vielmehr scheint es nahezu genauso viele Definitionen wie Evaluationswissenschaftler zu geben (Franklin & Trasher, 1976, S.20 zitiert nach Wottawa & Thierau, 2003, S.13). Begleitet wird die Definitionsvielfalt von Diskussionen um wissenschaftliche Paradigmen. Dazu gehören neben der Wahl des geeigneten Erhebungsverfahrens (qualitativ vs. quantitativ) die Gestaltung des Untersuchungsdesigns (quasiexperimentell vs. experimentell) und die Frage, ob Evaluation theoriegeleitet sein muss. Ähnlich stellt sich die Situation für die Klassifikationsansätze dar, die durch eine ebensogroße Vielfalt charakterisiert werden. Neben der Unterteilung von Scriven (1991) in formativ und summativ, gehören das CIPP-Modell von Stufflebeam (2001), das Fünf-Boxen-Modell von Wittmann (1990) und das neuere Modell von Chen zu den bedeutenden Klassifikationsansätzen. Während sich der Ansatz von Chen zur Einordnung allgemeiner Fragestellungen eignet, hat der Ansatz von Wittmann seine Stärken bei der Diskussion methodischer Fragen. Zur Einordnung der Personalauswahl vor dem Hintergrund der Personalarbeit lassen sich die zentralen Konzepte von Pawlik (1976) und Wiggins (1973) heranziehen. Es kann anhand eines typischen Verlaufes eines Personalauswahlverfahrens gezeigt werden, dass es eine Vielzahl an Fragestellungen und damit Evaluationstypen gibt, die an den einzelnen Schritten ansetzen,

so dass die reine Fokussierung der Evaluation auf die Eignungsdiagnostik zu kurz greifen würde.

Der Begriff der Diagnostik wird vor zahlreichen Kontexten verwendet und hat viele Anwendungsgebiete. Unter psychologischer Diagnostik versteht man dabei die Lehre von den Methoden und Verfahren der sachgemäßen Durchführung einer Diagnose. Dabei kann psychologische Diagnostik je nach Zielsetzung in unterschiedliche Alternativdimensionen klassifiziert werden (Pawlik, 1976). Diese sind wiederum Ausdruck zweier unterschiedlicher Modellvorstellungen der Diagnostik: der Eigenschafts- und der Verhaltensdiagnostik. Bei der Berufseignungsdiagnostik wird eine gedankliche Zuordnung von Personen zu Situationen und Situationen zu Personen vorgenommen. Die KTT ist dabei messtheoretische Grundlage der meisten und ermöglicht die Berechnung von Testgütekriterien, zu denen die Validität, Reliabilität und Objektivität als die wichtigsten gehören.

2.2 Der eignungsdiagnostische Prozess

In diesem Kapitel werden anhand eines exemplarischen eignungsdiagnostischen Prozesses die zentralen Fragen und Schritte der Berufseignungsdiagnostik im Rahmen der Personalauswahl abgehandelt (*Abbildung 6*).

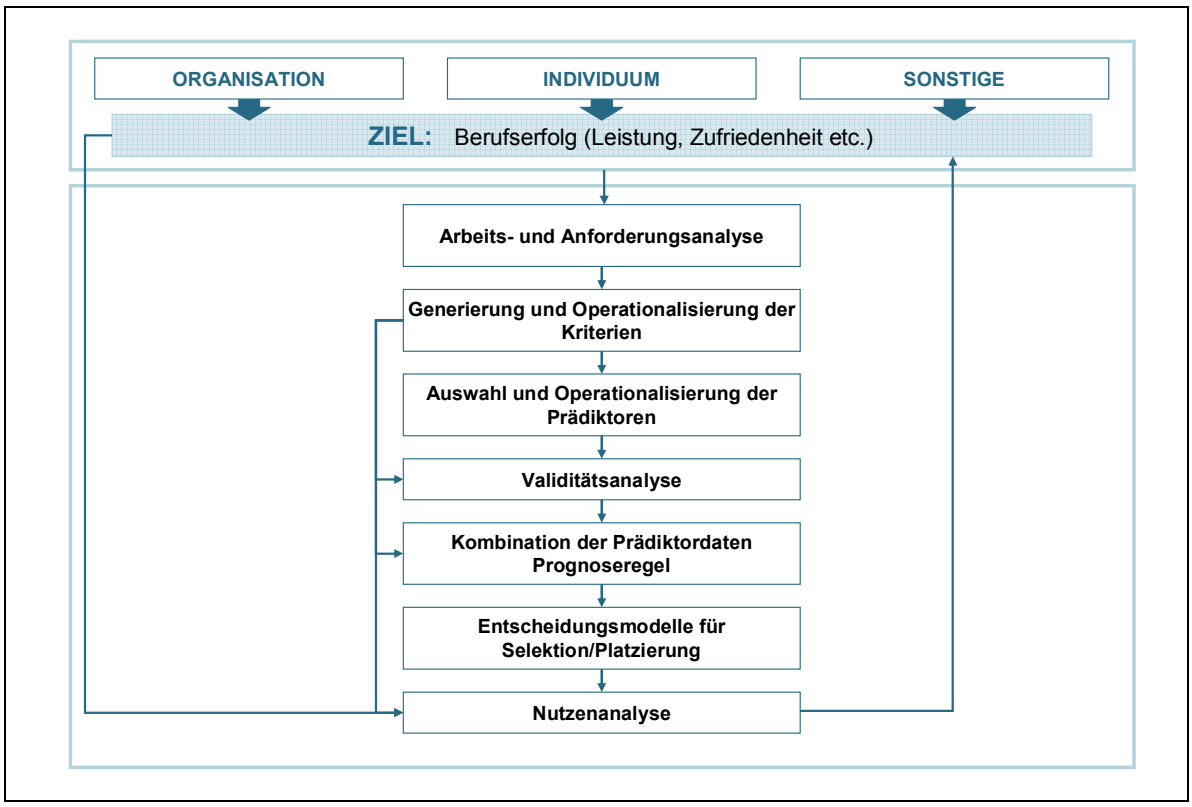

Abbildung 6: Eignungsdiagnostischer Prozess (in Anlehnung an Maukisch (1978, S.124f.))

Es handelt sich bei dem Phasenmodell um eine Variante eines analytischen Ansatzes nach Maukisch (1978, S. 124f.). Ausgangspunkt des Prozesses sind die unterschiedlichen Interessen aller Beteiligten, die dem Oberziel Berufserfolg (hier zu verstehen als „Passung“ von Organisation und Person) dienen sollen. Erster Schritt zur Zielerreichung ist dabei eine Arbeits- und Anforderungsanalyse (AA) der zu besetzenden Stelle. Aus ihr werden sowohl Kriterien gewonnen, die Berufserfolg bei der Arbeit in ausreichendem Maße widerspiegeln, als auch geeignete Personenmerkmale zur Vorhersage der festgelegten Kriterien identifiziert. Diese Kriterien und Prädiktoren müssen weiterhin operationalisiert, also messbar gemacht werden. Im Anschluss wird die Validität, d.h. der Zusammenhang von Prädiktor und Kriterium, überprüft und über die Festlegung einer Kombinationsvariante der Prädiktoren (Prognoseregel) maximiert. Daraus resultiert ein Entscheidungsmodell für die Selektion bzw. die Platzierung. Schließlich wird in einer Nutzenanalyse unter Abgleich mit den festgelegten Zielen der Gesamtnutzen des vorliegenden Modells untersucht.

2.2.1 Arbeits- und Anforderungsanalyse als Ausgangspunkt

Arbeits- und *Anforderungsanalysen* (AA), die von einer Vielzahl an Autoren als Synonyme oder Begriffspaar verwendet werden, sind nach Gatewood und Field (1990) der zielgerichtete systematische Prozess der Informationssammlung der wichtigen arbeitsrelevanten Aspekte einer Stelle (übersetzt aus Schuler, 2001a, S.47). Die gewonnenen Informationen betreffen generelle Daten zur Stelle, Daten zum Wesen der Organisation, zum Arbeitsinhalt, zum Arbeitsumfeld sowie zu den erfolgskritischen Personencharakteristika des Stelleninhabers (Algera & Greuter, 1998, S.144). Schuler (2001a) schreibt die zuletzt genannte Personenbeschreibung der Anforderungsanalyse zu, während er die reine Situationsbeschreibung zur Arbeitsanalyse zählt. Wichtige Quellen arbeitsanalytischer Informationen sind dabei Arbeitsplatzinhaber, Vorgesetzte und Arbeitsanalytiker (Schuler, 2001a, S.48). Die Anwendungsmöglichkeiten der AA gehen über die der Personalauswahl hinaus (Schuler, Moser, Funke, Donat & Barthelme, 1995, S.11). Bei der Personalauswahl liefert sie nicht nur die Grundlage zum Entwurf des Auswahlverfahrens, sondern generiert auch Informationen während des Auswahlprozesses. So kann sie beispielsweise bei der Anzeigengestaltung und der Information der Bewerber bezüglich der Arbeitsinhalte unterstützen oder als Anhaltspunkt im Rahmen der Vorstellungsgespräche dienen. Bei der Entwicklung des Auswahlverfahrens hilft sie bei der Generierung und Auswahl wesentlicher Kriterien des Erfolges (als spätere Basis für die Überprüfung der Zuordnungsentscheidung) und der Identifikation individueller Eignungsmerkmale erfolgreicher Aufgabenbewältigung der zu besetzenden Stelle (Kompa,

1989, S. 44). Diese Eignung betrifft nicht nur tätigkeitsspezifische (z.B. Fertigkeiten, Fähigkeiten), sondern auch tätigkeitsübergreifende Anforderungen (z.B. Entwicklungspotenzial) (Schuler, 2001a, S.45). Dabei muss die AA den Anforderungen der Vollständigkeit, Präzision und Bedeutsamkeit genügen (Kompa, 1989, S. 45).

Die genannten Anforderungen lassen sich nach Schuler (Schuler, 2000, S.59) in *Eigenschafts*-, *Verhaltens*-, *Qualifikations*- und *Ergebnisanforderungen* unterteilen. Zur ihrer Bestimmung stehen grundsätzlich drei Ansätze zur Verfügung (Eckardt & Schuler, 1992, S.536f.): die *erfahrungsgeleitet-intuitive* Methode basiert auf Expertenurteilen über die Anforderungen der Berufe und der korrespondierenden Personenmerkmale. Die Experten beschäftigen sich dabei detailliert mit den Eigentümlichkeiten der Berufe (Tätigkeiten, Material, Umweltbedingungen usf.). Dieses Vorgehen setzt allerdings große Erfahrung im arbeitsanalytischen Bereich voraus. Die *personenbezogene-empirische* Methode versucht hingegen, Merkmale von im Beruf tätigen Personen mit Kriterien wie Leistungsbeurteilungen, Mitarbeiterzufriedenheit oder Produktivität in Bezug zu setzen. Die statistischen Zusammenhänge sollen Rückschlüsse darüber ermöglichen, welche Merkmale erfolgreiche Personen in einem Beruf aufweisen. Die am häufigsten angewandte *arbeitsplatzanalytisch-empirische* Methode schließlich untersucht die beruflichen Tätigkeiten und Situationen mittels formalisierter Vorgehensweisen und aggregiert die gewonnenen Ergebnisse zu Aussagen. Dabei werden ebenfalls Schätzungen von Tätigkeitselementen auf Personenmerkmale vorgenommen, die aber auf Grund ihres stark formalisierten Charakters relativ robust gegen subjektive Einflüsse sind.

Zur Kategorisierung der Arbeitsverfahren können die Beschreibungsdimensionen eignungsdiagnostisch relevanter Instrumente nach *Abbildung* 7 herangezogen werden. Die Arbeitsanalyse auf der *Verhaltens*- und *Ergebnisebene* beinhaltet die objektive Erfassung der Tätigkeiten oder Tätigkeitselemente. Dies kann mit Hilfe von Aufgabeninventaren, deren Items eine Beschreibung der Tätigkeiten (z.B. über Häufigkeit, Schwierigkeit und Bedeutung) ermöglichen, erfolgen. Über den Detaillierungsgrad solcher Inventare herrschen kontroverse Meinungen (Schmidt, Hunter & Pearlman, 1981; Algera & Greuter, 1998, S.158). Die beliebtesten AA-Verfahren finden auf der *Verhaltensebene* statt, da sie den vielfältigen Einsatzmöglichkeiten der AA gerecht werden (Schuler, 2000, S.63). Zu den bekanntesten Verfahren gehören neben dem aus dem englischen *Position Analysis Questionnaire* (*PAQ*, McCormick & Jeanneret, 1988; Cook, 1993, S.36ff.) abgeleiteten *Fragebogen zur Arbeitsanalyse* (*FAA*, Frieling & Hoyos, 1978), die *Critical Incident Technique* (*CIT*, Flanagan, 1954).

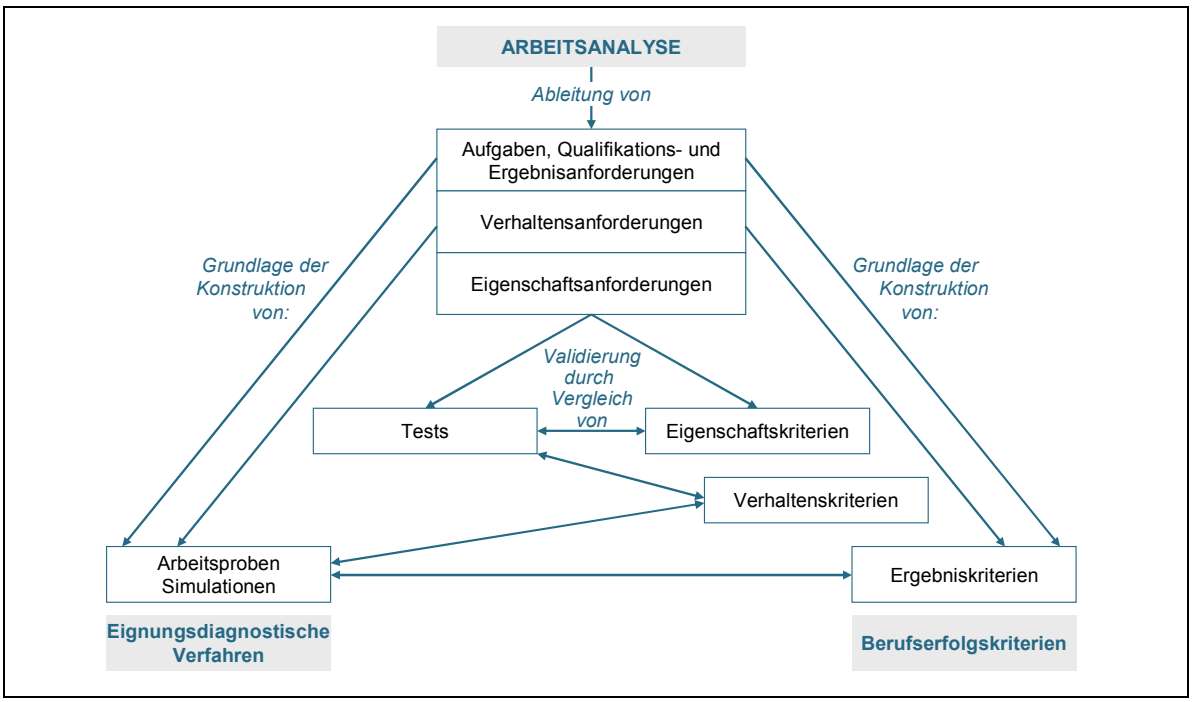

Abbildung 7: Beschreibungsebenen eignungsdiagnostisch relevanter Instrumente (aus Schuler, 2000, S.63)

Auf der Eigenschaftsebene geht es um Formulierung von Fähigkeiten und weiteren Eigenschaften, d.h. Prädiktoren, die für eine erfolgreiche Aufgabenerledigung zuträglich sind. Dabei unterscheiden Schuler und Funke (1993) zwischen *holistischem* und *dekomponierendem* Vorgehen. Holistisch meint die ganzheitliche Betrachtung des Arbeitsplatzes (z.B. *Ability Requirement Scales* (*ARS*) von Fleishman & Quaintance, 1984, *Personality-Related Position Requirements Form* (*PPRF*) von Raymark, Schmit und Guion, 1997), dekomponierend hingegen die separate Beurteilung der einzelnen Eigenschaften und anschließende Aggregation der Urteile (z.B. die Synthetisierung von Anforderungsprofilen auf Basis der Ergebnisse des PAQ bzw. FAA). Die Bestimmung von Anforderungen auf der Eigenschaftsebene ist nicht ohne Kritik geblieben. Dies betrifft hauptsächlich die in der Praxis am häufigsten angewandte, *erfahrungsgeleitet-intuitive* Methode. Die bedeutendsten Probleme sind das interpersonal unterschiedliche Verständnis der Eigenschaften, die Subjektivität von Fremd- und Selbsteinschätzung sowie mangelnde Kenntnis über die Zusammenhänge zwischen den Eigenschaftskonstrukten. Darüber hinaus gibt es vereinzelt die Ansicht, dass ein solches Vorgehen nicht mehr zur AA gezählt werden kann, da diese nur Beobachtbares beinhalten sollte (Harvey, 1991, siehe Schuler, 2001a, S.57). Neben den dargestellten eher allgemeinen AA-Verfahren gibt es auch in Deutschland neuere, spezifischere Verfahren, die eine differenziertere Betrachtung des Arbeitsplatzes vornehmen (z.B. *REBA 6.0,* Raymark., et al., 1997) oder RHIA/VERA-Produktion (Oesterreich, Leitner & Resch, 2000)).

Die AA bringt neben der durch die zahlreichen Instrumente angedeuteten Anwendungsvielfalt für die Personalarbeit im Allgemeinen und die Eignungsdiagnostik im Besonderen, eine große Anzahl von Vorteilen mit sich (Cook, 1993, S.34ff.). So kann die AA durch Kenntnis der relevanten Arbeitstätigkeiten zu einer hohen Inhaltsvalidität bei der Testkonstruktion und letztlich auch zu einer höheren Validität des Auswahlverfahrens beitragen (Tett, Jackson & Rothstein, 1991). Darüber hinaus können mit Hilfe der AA durch Faktorenanalyse unterschiedliche Tätigkeiten zu einer Hauptaufgabe zusammengefasst werden. Außerdem können über Cluster-Analysen Gruppen von unterschiedlichen Arbeitsstellen gebildet werden, die sich in ihren Inhalten stark ähneln (Krystofiak, Newman & Anderson, 1979). Dies verursacht eine Erleichterung bei der Auswahl und hohe Synergieeffekte bei dem Einsatz der Personalauswahlinstrumente und hilft bei der Identifikation von passenden Stellen bei Arbeitsplatzwechseln. Zudem unterstützt die AA bei der Formulierung präziser Arbeitsplatzbeschreibungen sowie der Entwicklung von Arbeitserfolgskriterien bei der Personalbeurteilung. Doch mit der AA gehen auch Probleme einher. So sind viele Arbeitsaufgaben schlecht definierbar, da sie einem ständigen Wandel unterliegen (Schuler & Funke, 1993). Diesem wird mit der Einschätzung von allgemeinen, generell als erfolgsrelevant bezeichneten Eigenschaften entgegengewirkt. Dies entspricht auch den oben erwähnten Ansichten von Schmidt, Hunter und Pearlman (1981). Daneben gibt es auch inhaltliche Defizite zahlreicher Verfahren, bei denen den erforderlichen kognitiven Fähigkeiten und dem Arbeitskontext wenig oder keine Beachtung geschenkt wird (Algera & Greuter, 1998, S. 143). Zudem sind viele Anforderungen (z.B. soziale Kompetenz, Charisma) schwierig zu beurteilen oder können durch Subjektivität in der Urteilsfindung fehlerbehaftet sein (Schmitt & Chan, 1998, S.57).

AA müssen genauso wie psychologische Tests den Gütekriterien der Objektivität, Reliabilität und Validität genügen, was allerdings mit einigen methodischen Problemen behaftet ist. Zu diesen zählt, dass die Reliabilität von AA durch Faktoren mitbestimmt wird, die mit der Einschätzung der Arbeitsplatzmerkmale korrelieren, z.B. mit der Arbeitszufriedenheit (Schuler, et al., 1995, S.18). Probleme kann es ebenso bei der Validitätsbestimmung geben. Die Kriteriumsvalidität einer AA beispielsweise, die durch Validierung an einer Expertenschätzung bestimmt werden kann, ist dem methodischen Problem der Varianzeinschränkung unterworfen, da bei dem Vergleich der Merkmalsdimensionen die Differenzen über die Validität entscheiden. Die nicht differenzierenden Dimensionen hingegen, auch oftmals von Interesse, sind nicht kriteriumsvalide abbildbar (Schuler, et al., 1995, S.18). AA können außerdem Diskriminierungstendenzen bei der Personalauswahl begünstigen. Dies

kann auf indirekte Weise über Ableitung falscher Anforderungen oder direkt über Ableitung richtiger Anforderungen, die eine bestimmte Personengruppe nicht besitzt (Algera & Greuter, 1998, S. 156f.) geschehen[5]. Zuletzt bleibt der immense Zeit- und Ressourcenaufwand bei der Durchführung von AA zu nennen, der dazu führt, dass diese aus Gründen der Ökonomie meistens nur bei der Neukonstruktion von eignungsdiagnostischen Verfahren angewandt werden.

Zusätzlich zu den angesprochenen deskriptiven Verfahren gibt es noch eine große Anzahl an teilweise auch länderspezifischen, enumerativen Berufs-Klassifikationssystemen, wie beispielsweise die *International Standard Classification of Occupation* (ISCO-88) oder das *Dictionary of Occupational Titles* (*DOT*). Eine gute Übersicht findet sich beim Commitee on Techniques for the Enhancement of Human Performance (CTEHP,1999, S.164ff.). Aus jüngerer Zeit sind Vorschläge gekommen, die AA in unterschiedliche Richtungen weiter zu entwickeln. Campion (1994) fordert beispielsweise, Arbeitsgestaltung und AA stärker aneinander zu koppeln und liefert Ansätze zur Berücksichtigung des immer wichtiger werdenden Aspektes der Teamfähigkeit in der Arbeit. Schmitt und Chan (1998) betonen die Notwendigkeit einer umfassenden Arbeitstheorie, die Theorien über das Zustandekommen von Leistungen mit Arbeitsstrukturtheorien verbindet. Eine solche Arbeitsstrukturtheorie ist das von dem *U.S. Employment Service* entwickelte und auf der *Functional Job Analysis* (FJA) (Fine & Wiley, 1971) basierende *Dictionary of Occupational Titles* (*DOT*). Ursprünglich als reines Klassifikationssystem gedacht, ist das *DOT* in der neuesten Fassung integrativer, da es ebenso teilweise Eigenschaftsanforderungen (z.B. Interesse, physische Voraussetzungen u.Ä.) enthält. Das DOT ist in der Vergangenheit allerdings besonders auf Grund der mangelnden Nähe zur reellen Arbeitswelt und den veralteten Inhalten kritisiert worden (Spenner, Otto & Call, 1980). Dies führte zu der Entwicklung eines Ersatzinstrumentes, dem *Occupational Information Network* (O*Net). Das *O*Net* wurde als Unterstützung für eine Vielzahl von Personen (Bewerber, Personalberater, Recruiter, usf.) für zahlreiche Anwendungen entwickelt (CTEH, 1999, S.169ff.). *Abbildung 8* zeigt die betrachteten Dimensionen. Dabei wird bei den einzelnen Inhalten der beschreibenden Dimensionen auf bekannte wissenschaftliche Modelle zurückgegriffen, d.h., es liegen größtenteils validierte Inhalte vor. So basiert ein Unterbereich von „Interessen“ auf Hollands Sechs-Faktoren-Typologie (Holland, 1985, siehe auch 2.5.1.3). Detaillierte Dimensionsbeschreibungen fin-

[5] Aus Sicht des Autors bedeutet dies allerdings in letzter Konsequenz, dass jegliche Form der Personalauswahl, die nicht dem Zufallsprinzip folgt, diskriminative Aspekte enthält.

den sich bei CTEHP (1999), Peterson et. al. (2001) und O*Net-Center (2005). Für die Arbeitsstelle *Management Consultant* findet sich diese in *Anhang I.*

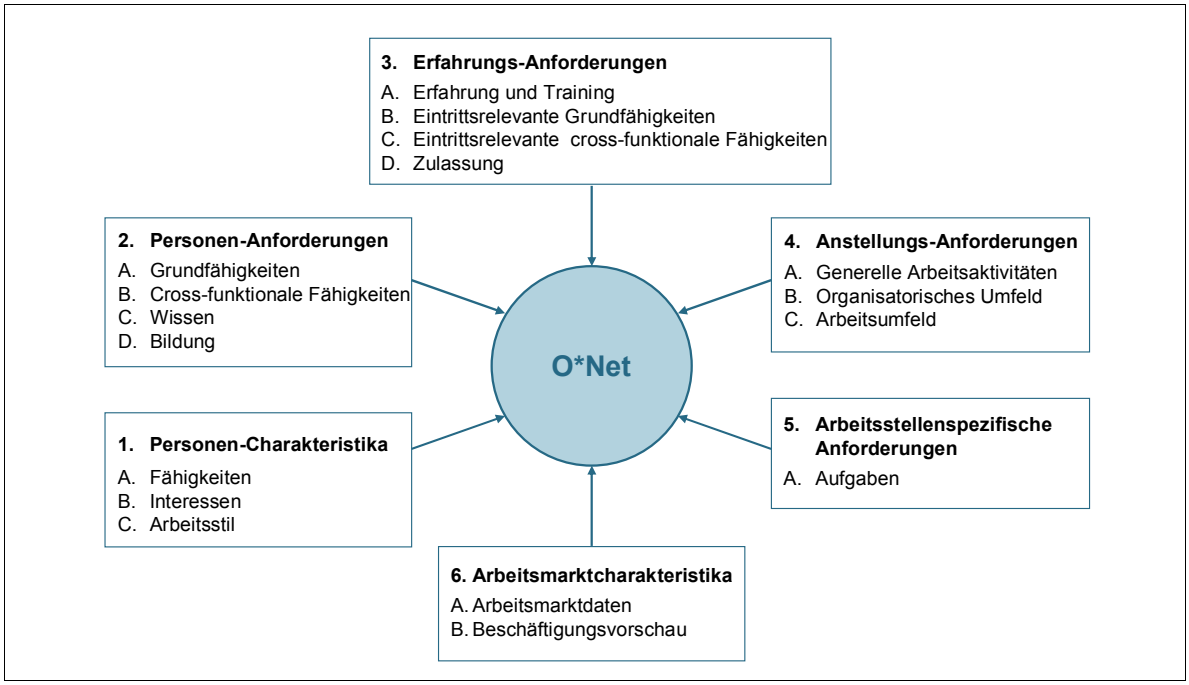

Abbildung 8: Beschreibungsdimensionen des O*Net (übersetzt O*Net-Center (2005))

Neben der hohen Anwendungsvielfalt und der umfassenden Beschreibung liegen die Vorteile des O*Net in der Integration enumerativer, deskriptiver Komponenten und Eigenschaftsanforderungsableitungen, der vollelektronischen Ausgestaltung und der damit verbundenen Auswertungsmöglichkeit aus zahlreichen Perspektiven. Zudem ist es kompatibel mit zahlreichen anderen Systemen und über das Internet online verfügbar (http://online.onetcenter.org/). Neuere Entwicklungen wie O*Net sind also wesentlich besser in der Lage, bei der Beschreibung der Anforderungen, die eine Arbeitstelle mit sich bringt und Voraussetzung für eine erfolgreiche Bewältigung der Arbeit sind, zu unterstützen, als vergleichbare ältere Instrumente. Doch welches sind die Kriterien, an denen eine erfolgreiche Arbeit gemessen werden kann? Wie kann man diese gewinnen und auf geeignete Weise operationalisieren? Mit diesen Fragen beschäftigt sich das nachfolgende Kapitel.

2.2.2 Generierung und Operationalisierung der Kriterien

Sinn und Zweck einer Personalauswahl ist die Gewinnung von Mitarbeitern, die für die zu besetzende Stelle die am besten Geeigneten sind, oder anders ausgedrückt, die sich im Beruf bewähren. In diesem Fall handelt es sich aber bei dem Begriff Bewährung nicht um eine festgelegte Größe, sondern um ein hypothetisches Konstrukt, das einer Operationalisie-

rung bedarf (Kompa, 1989, S.48). Im Folgenden soll vor dem Hintergrund der Vielzahl an Möglichkeiten, wie man Berufsbewährung bzw. Berufserfolg definieren kann (Schuler & Funke, 1993, S.235; Schuler, 2000, S.154), auf den Leistungsaspekt, dass heißt den Beitrag des Einzelnen zu den Zielen der Organisation als Berufsbewährung fokussiert werden.

2.2.2.1 Taxonomie der Leistungskriterien

Es gibt zunächst unterschiedliche Möglichkeiten, Leistungskriterien einzuteilen. Thorndike (1949) unterscheidet zwischen *Actual*- und *Ultimate*-Kriterien. Das *Ultimate*- oder auch ultimative Kriterium ist das tatsächliche Leistungskonstrukt, das man messen will, nämlich die endgültige Berufsbewährung einer Person. Dafür müsste man aber bis zum Ende des Berufslebens der Person warten, um ein ultimatives Urteil unter Berücksichtigung aller relevanten Aspekte abzugeben (Guion, 1961, zit. nach Cook, 1993). Da dies nur schwer realisierbar ist, bedient man sich einer konkret messbaren Ausprägung, dem *Actual*- oder aktuellen Kriterium. Cook (1993, S.211) unterscheidet darüber hinaus *objektive* (z.B. Verkaufszahlen) und *subjektive* Kriterien, wobei letztere eine intuitive Einschätzung durch einen Beurteiler voraussetzen. Des Weiteren schlägt Guion (1965) vor, zwischen Verhaltensweisen als primärem Kriterium und Verhaltensergebnissen als sekundärem Kriterium zu trennen. Schuler, et. al. (1995) unterscheiden hingegen analog *Abbildung 7* zwischen *Eigenschafts*-, *Verhaltens*- und *Ergebniskriterien*.

Ergebniskriterien werden bevorzugt dort Anwendung finden, wo gute Ergebnisse durch eine Vielzahl von Verhaltensweisen erzielt werden können, wo Verhaltensmessungen nicht möglich sind oder wo verhaltensbezogene Interventionen unangemessen erscheinen. Verhaltenskriterien werden direkt aus Tätigkeitsbeschreibungen abgeleitet und sind dann sinnvoll, wenn Wissen über den Zusammenhang von Verhaltensweisen und Berufserfolg vorliegt. Eigenschaftskriterien wird man schließlich bei Unmöglichkeit der Messung der beiden Ersteren oder bei Potenzialschätzungen einsetzen. Entscheidend für die Kriterienentwicklung ist dabei die strenge Orientierung der Leistungsbeurteilung an den in der AA sich als wesentlich herausgestellten Aufgaben und Anforderungen. *Tabelle 2* gibt eine Übersicht dieser Kriterien, einige Beispiele und ihre wichtigsten Vor- und Nachteile. In der Regel wird man aber Berufserfolg oder -bewährung nicht mit einem einzigen Kriterium messen, sondern, je nach Position, mehrere Kriterien anwenden. So postulierte beispielsweise Campbell (1994) auf oberster Ebene acht Faktoren, aus denen Leistung besteht. McCormick und Ilgen (zit. nach Kompa, 1989, S. 50) unterscheiden mehrdimensionale Kriterien in Abhängigkeit ihrer Validierungsmöglichkeit. Findet eine separate Verwendung jedes Kriteriums statt, so sprechen sie von *multiple criteria*. Werden Gewichtungsschemata angewandt, um die Subkrite-

rien zu einem Oberkriterium zusammenzufassen, so ist dieses ein *composite criterion*. Wird jedoch ein Gesamtkriterium etwa durch Beurteilerschätzung gewonnen, erhält man ein *overall criterion.* Eine Verwendung der letzteren beiden Kriterien verwehrt einem jedoch die Möglichkeit, detailliert einzelne Verhaltensfacetten der Person zu betrachten. Die Bestimmung der Gewichte unterliegt zudem einer gewissen Willkür. Schmidt, Hunter und Ones (1992, S.656) halten die kontroverse Diskussion um die geeignete Zusammensetzung der Kriterienwahl allerdings für obsolet, da sich kognitive Tests sowohl für *Overall*-Kriterien als auch für *Composite*-Kriterien als valide erwiesen haben.

Tabelle 2: Leistungskriterien und ihre Vor- und Nachteile (in Anlehnung an Schuler, 1993 und Schuler, et al., 1995)

Kriterien	Beispiele	Vor- (+) und Nachteile (-)
Ergebniskriterien	Verkaufs- und Umsatzzahlen, Produktionseinheiten, Fluktuationsdaten und Fehlzeiten, Ausschusszahlen, Reklamationen, Marktanteile	+ Deutliche Repräsentation der eigentlichen Leistungsziele + Verfahrensspielraum + Anwendbarkeit auf individuelle, Gruppen- und Unternehmensziele - Schwierige individuelle Zurechnung der Ergebnisse - Überbewertung des Zähl- und Messbaren - Unzulängliche Steuerungsmöglichkeiten bei negativen Zielabweichungen
Verhaltenskriterien	Bringt gute Vorschläge, findet sich schnell in neuen Fachgebieten zurecht, bleibt auch in Stresssituationen motiviert, arbeitet zuverlässig	+ Hoher Informationsgehalt für den Beurteilten + Anhaltspunkte für Steuerung und Verbesserung der Leistung - Beobachtungsschwierigkeit des tatsächlichen Verhaltens - Zielbezug oft unklar - Einengung des Verhaltensspielraums - Beurteiler neigen zu globalen Eigenschaftseinschätzungen
Eigenschaftskriterien	Zuverlässigkeit, Problemlösungsfähigkeit, soziale Kompetenz, Selbstvertrauen, Ausdrucksfähigkeit, Sprachliche Fähigkeit, Konzentration, Leistungsbereitschaft	+ Hoher Allgemeinheitsgrad und daher Vergleichbarkeit + Gute individuelle Zuordnung - Hohe Urteilssubjektivität - Exzessivität (Urteile sind nicht nur auf leistungsrelevante Aspekte beschränkt) - Geringe Tauglichkeit für Feedback und Verhaltenssteuerung

Ein neuerer Ansatz zur Struktur beruflicher Leistung auf Verhaltensebene beruht auf Campbell et. al. (1993) und Campbell (1994) und sieht die Leistung als Interaktion der Determinanten *deklaratives Wissen*, *prozedurales Wissen* und *Fertigkeiten* sowie *Motivation*, die multiplikativ verknüpft sind. Borman & Motowidlo (1993) wiederum erweitern die reine Aufgabensicht um den Begriff der „umfeldbezogenen" Leistung, die Leistungsaspekte enthält, die über die formalen Arbeitsaspekte hinausgehen, wie z.B. „Unterstützung ungeübter Kollegen".

2.2.2.2 Zur Problematik der Erfassung und Bewertung der Kriterien

Betrachtet man die Messgüte der Kriterien, so ist zunächst die Konstruktvalidität von größter Relevanz. Leistungskriterien sollten die jeweils bedeutsamen Aufgaben- und damit Leistungsbereiche widerspiegeln und auf den in der AA ermittelten Aufgaben und Anforderungen inhaltlich basieren (Schuler, et al., 1995, S.55). Hohe Konstruktvalidität bedeutet also eine hohe Überlappung zwischen dem *actual*- und dem *ultimate*-Kriterium im Sinne Thorndikes. Diese Validität kann durch den Aspekt der *Kontamination* (Erfassung irrelevanter Aspekte, (Cook , 1995)) oder *Defizienz* (Fehlen relevanter Aspekte) verringert werden. Schuler (2004b) weist beispielsweise darauf hin, dass auch oftmals nicht betrachtete qualitative Aspekte, wie das Agieren in einem Netzwerk sozialer Beziehungen, als Leistungskriterium bedeutsam sind. Eine Vernachlässigung dieser könnte also zu einer Defizienz führen.

Problematisch im Zusammenhang mit der Validität der Kriterien ist, dass diese prinzipiell nicht geprüft werden kann (Hossiep, 1995, S.50), da sie die Wirklichkeit selbst darstellen, die als Prüfkriterium herangezogen wird. Das Heranziehen eines anderen Kriteriums führt wiederum zu einem Zirkularitätsproblem der Validierung (Welches ist nun das „wahre“ Leistungskriterium?) (Marcus & Schuler, 2001, S.417). Bezüglich des Zusammenhangs zwischen objektiven und subjektiven Kriterien liegen unterschiedliche metaanalytische Ergebnisse vor. Heneman (1986) zeigte, dass Vorgesetztenurteile mit objektiven Leistungskriterien bestenfalls zu r=.27 korrelieren. Viswesvaran (1993, zit. nach Schuler, 2004a) kam in seiner Metaanalyse auf eine höhere Konkordanz von r=.57 mit Kollegenurteilen. Conway, Lombardo und Sanders (2001) im gleichen Zusammenhang nur auf Werte zwischen r=.25 und r=.29. Cook (1993, S.214) kommt zu dem Schluss, dass gerade bei Unternehmen, die keine objektiven Leistungskriterien anwenden, die subjektiven Kriterien die geringste Validität haben.

Neben der Validität ist die Reliabilität entscheidend. Das gemessene, also aktuelle Kriterium, muss hinreichend zuverlässig sein (Auf mögliche Verletzungen der Reliabilität und ihre Auswirkungen auf die Validität wird unter 2.2.4 detaillierter eingegangen). Zudem setzt die Kriterienauswahl Praktikabilität voraus, d.h. der Aufwand der Erfassung sollte sich in vertretbaren Grenzen halten (Cook, 1995). Kompa (1989) fordert als praxisbezogene Anforderungen an das Kriterium zusätzlich die Messbarkeit, die Vorhersagbarkeit, die Äquivalenz (d.h. das Vorhandensein äquivalenter Kriterien), die Akzeptanz und die Transparenz. Außerdem sollte das Kriterium Differenzierungsvermögen besitzen, d.h. in ausreichendem Maße Unterscheidungen zwischen Bewerbern mit guter und schlechter Leistung treffen.

In einem weiteren Schritt gilt es nun, die unterschiedlichen möglichen Leistungskriterien in ihrer Ausprägung zu bewerten. *Abbildung 9* verdeutlicht anhand eines Kriterienwürfels, wie diese ausgestaltet werden können. Unterschieden werden dabei die Dimensionen Kriterientyp (s.o.), Urteilsquelle und Skalierungs- (oder Bewertungs-)Verfahren. Die Wahl des Skalierungsverfahrens sollte auf den Anwendungszweck abgestimmt sein (Schuler, et al., 1995) und einer Reihe von Anforderungen genügen (z.B. Ökonomie, Transparenz, Eindeutigkeit.; siehe Schuler, 2004b, S.20).

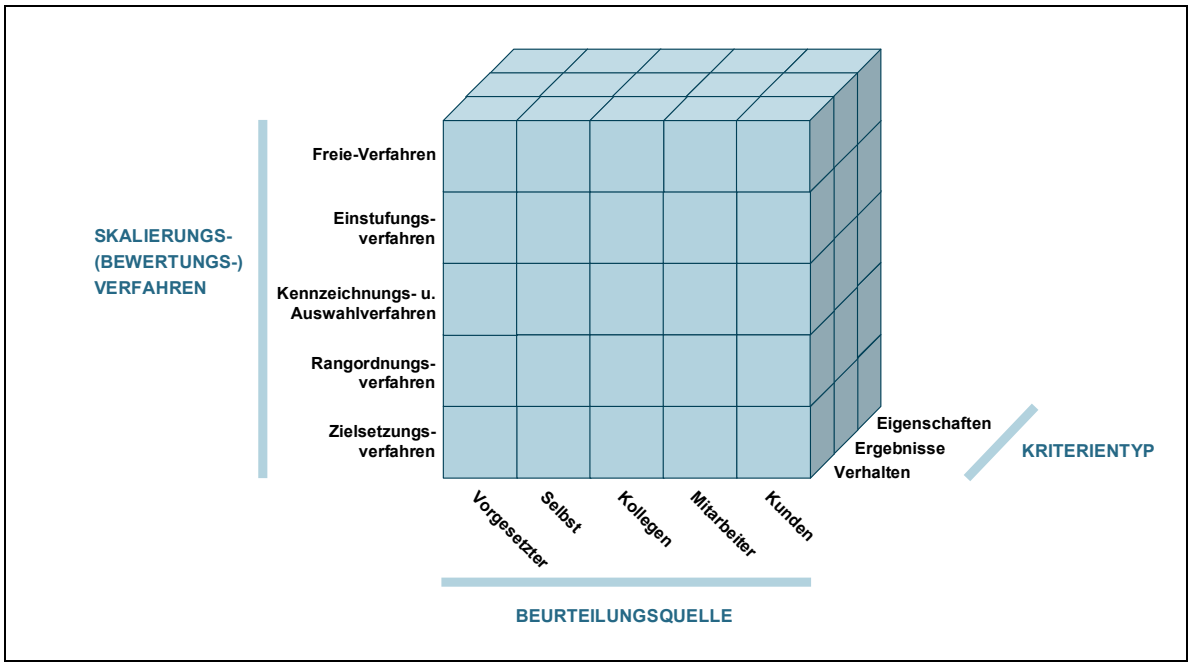

Abbildung 9: Kriterienwürfel (eigene Darstellung)

Eine gute Übersicht der Skalierungsverfahren und ihrer Vor- und Nachteile findet sich bei Schettgen (1996). Dabei kann die Bewertung selber bereits einen leistungsfördernden Effekt verursachen (Guzzo, Jette & Katzell, 1985). Sie geht aber dennoch mit einer Vielzahl von zu bewältigenden Problemen einher. So haben beispielsweise subjektive Kriterien, wie die am häufigsten verwendeten Vorgesetztenbeurteilungen (siehe 2.2.2.3), geringe Reliabilitäten (Cook, 1993, S.212). Reliabilitätsprobleme sind aber auch bei objektiven Kriterien vorzufinden (Hunter, Schmidt & Judiesch, 1990), besonders wenn es sich um indirekte Leistungsdaten (z.B. Gehalt) handelt (Hossiep, 1995, S.52). Außerdem können diese irreführend sein, wenn das scheinbar objektive Kriterium, wie z.B. Verkaufszahlen bei einem Außendienstmitarbeiter, auch von weiteren Faktoren, wie z.B. Markt- oder Produkteigenschaften, abhängt. Hossiep (1995, S.55) weist darauf hin, dass Ergebniskriterien immer aus einem Zusammenspiel von Person und Situation entstehen und somit nicht nur der Person zuzu-

schreiben sind. Zudem können sie aus zahlreichen unterschiedlichen Verhaltensweisen entstehen. Schuler (2004b) stellt dazu fest, dass der Teamaspekt, d.h. die Leistung als Ergebnis einer Teamarbeit, wie sie in zahlreichen Berufen der Fall ist, selten Berücksichtigung findet. Weitere Schwierigkeiten verursachen dynamische, also sich über die Zeit verändernde Kriterien, die erstmals durch Ghiselli und Haire (1960) thematisiert wurden. Dynamik hat Auswirkungen auf die Validität zwischen Prädiktoren und Kriterium und stellt die Anforderung an ein Auswahlinstrument, Leistungen zu jedem Zeitpunkt gut vorherzusagen. Barrett, Coldwell und Alexander (1985) stellten erstmals die Existenz dynamischer Kriterien in Frage und wiesen auf die Möglichkeit hin, dass es sich in den Fällen um unreliable Kriterien handeln könnte. Deadrick und Madigan (1990 zit. nach Schmidt, et al., 1992, S.659) legten in einer Längsschnittstudie dar, dass die Leistungsstabilität mit zunehmendem Messabstand abnimmt, während die Validität kognitiver Fähigkeiten als Prädiktor zunahm und die der psychomotorischen Fähigkeiten konstant blieb. Damit zeigten sie, dass dynamische Kriterien nicht zwangsläufig mit einem Validitätsverlust einhergehen. Insgesamt dürften sie zudem eine Ausnahmeerscheinung bleiben (Greuter & Algera, 1989, S.151).

Zusammenfassend lässt sich feststellen, dass die Wahl und Operationalisierung der Leistungskriterien von einer Vielzahl von Faktoren abhängen und mit zahlreichen potenziellen Fehlerquellen einhergehen, weshalb sie sich letztlich, wie in Metaanalysen bestätigt, auch entscheidend auf die Konstruktvalidität und somit auf die prognostische Validität des Auswahlverfahren auswirken (siehe Schmitt, Gooding, Noe & Kirsch, 1984). Im Anschluss soll das im empirischen Teil dieser Arbeit verwendete Instrument der *Vorgesetztenbeurteilung* als Kriteriumsoperationalisierung kritisch durchleuchtet werden.

2.2.2.3 Das Vorgesetztenurteil im Kreuzfeuer der Kritik

Bei komplexen Leistungen wie die der Unternehmensberatung stehen keine einfachen Maßzahlen, die die Leistung adäquat widerspiegeln, zur Verfügung (siehe dazu 2.4.2). Vielmehr wird hier, wie in der allgemeinen Praxis bei anderen Unternehmen auch, auf das Vorgesetztenurteil als subjektives Kriterium zurückgegriffen, obwohl vielfach auf die eingeschränkte Sichtweise bei der Beurteilung leistungsrelevanten Verhaltens hingewiesen wurde (Schuler, 2004b). Vorgesetztenbeurteilungen beruhen in einem hohen Maß auf subjektiven Eindrücken. Wie solche Eindrücke über andere Personen zustande kommen, ist Gegenstand der Forschung zur *sozialen Urteilsbildung*, in deren Kontext zahlreiche theoretische Ansätze hervorgebracht wurden (z.B. Borman, 1978; Feldmann, 1981). In Anlehnung an Schuler (2004a) sollen nachfolgend anhand des Verhaltens-Eindrucks-

Aussagemodells (VEA-Modell; Brandstätter, 1983) wesentliche Aspekte dargestellt werden. Das VEA-Modell postuliert dabei drei Ebenen, auf denen unterschiedliche Quellen des Einflusses auf das Urteil wirken können und somit für Differenzen im Urteil zweier Personen über eine Dritte verantwortlich sein können. *Abbildung 10* zeigt die Ebenen und wichtigsten Quellen. Zusätzlich unterliegen subjektive Kriterien zahlreichen weiteren Urteilsfehlern bei dem Informationserwerb, der -verarbeitung, der -ausgaben und der -rückkoppelung, auf die im Rahmen der klinischen Urteilsbildung (siehe dazu 2.2.5.1) eingegangen wird.

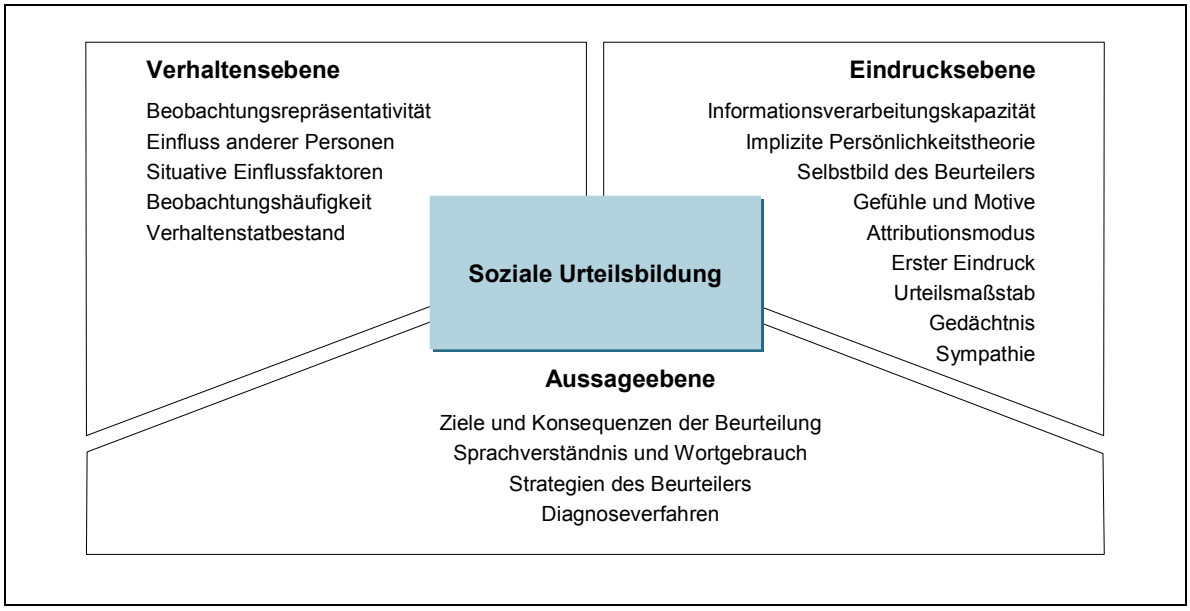

Abbildung 10: Drei-Ebenen-Modell der sozialen Urteilsbildung (eigene Darstellung in Anlehnung an Schuler, 2004a, S.36)

Die Vielzahl der Verzerrungsquellen erklärt, warum es zu erheblichen Urteilsabweichungen zwischen Beurteilern kommen kann (zu einer detaillierten Diskussion der Reliabilitätsproblematik siehe Murphy & De Shon, 2000a; Murphy & De Shon, 2000b und Schmidt, Viswesvaran & Ones, 2000). Zur Quantifizierung dieser Abweichungen wird im Rahmen der Leistungsbeurteilung als *Objektivitätsmaß* die *Interrater-Reliabilität* (z.B. das Korrelationsmaß zwischen Beurteilern) herangezogen (Schuler, 2004a). Hossiep (1995, S.52) berichtet von einer Interrater-Reliabilitätshöhe von Vorgesetztenbeurteilungen von maximal r=.50, was ihre Verwendung als Berufsbewährungskriterium für die Evaluation von Auswahlverfahren kritisch erscheinen lässt. Auch Viswesvaran, Ones und Schmidt (1996) zeigten in einer Metaanalyse, dass die Reliabilitäten zwischen Beurteilern durchweg geringer sind als die innerhalb eines Beurteilers, was in der Vergangenheit wohl zu erheblichen Unterschätzungen der Validitäten geführt hat. Die interindividuelle Urteilskonkordanz für das Vorgesetztenurteil wurde mit r=.52 angegeben. Vorgesetztenurteile zeigten sich allerdings

durchgehend reliabler als solche von Personen gleicher Stufe (siehe auch Conway & Huffcut, 1997). In einer weiteren Metaanalyse wiesen sie nach, dass die mangelnde Übereinstimmung der Urteile dieser beiden Gruppen nur zu einem geringen Teil auf ein uneinheitliches Konstruktverständnis und hauptsächlich auf Beurteilungsschwierigkeiten (Messfehler) zurückzuführen waren (Viswesvaran, Schmidt & Ones, 2002). Mount, Judge, Scullen, Sytsma und Herzlett (1998) kamen sogar zu noch geringeren Reliabilitätswerten. Ähnlich geringe Werte der Urteilskonkordanz finden sich bei dem Vergleich der Urteile über verschiedene Quellen. Allgemein zwischen Selbst- und Fremdurteil liegen diese bei Conway und Huffcut (1997) zwischen r=.26 und r=.31 und beim Vergleich von Vorgesetzten- und Selbsturteil nach Schuler et. al. (2004) zwischen r=.19 und r=.34.

Der Großteil der Vorgesetztenurteile wird anhand von Einstufungsskalen vorgenommen. Mit diesem Vorgehen gehen systematische Urteilsverzerrungen einher, die sich in drei Urteilstendenzen niederschlagen: Mittelwert-, Streuungs- und Korrelationstendenzen (Schuler, 2004a). *Tabelle 3* erklärt diese.

Tabelle 3: Urteilstendenzen (in Anlehnung an Schuler, 2004a, S.37ff.)

Tendenz	**Wesen**	**Detaillierung**
Mittelwerts-tendenz	Urteilerspezifische Lokalisierung von Werten. Wichtige Einflussgrößen: Adaptionsniveau, Skalenverankerung, Anforderungsnorm, Position des Beurteilers auf Merkmalskontinuum und situationsspezifische Konventionen über den Wortgebrauch	*allgemein*: Milde, Strenge *differentiell*: Wertschätzung oder Ablehnung einzelner Mitarbeiter
Streuungs-tendenz	Einschätzungen verschiedener Beurteiler streuen unterschiedlich durch unterschiedliche Einschätzung der Streuung der Norm	Häufung der Urteile am Mittelwert
Korrelations-tendenzen	Beurteilerbedingte Korrelation unterschiedlicher, auch unabhängiger Beurteilungsdimensionen (Überstrahlungseffekt), bedingt durch Neigung zu einem individuellen Maß an Unterscheidung zwischen Merkmalen	Halo-Effekt

Neben der angesprochenen Objektivität der Urteile, die sich über die Interrater-Reliabilitäten berechnet, ist die Reliabilität, also Zuverlässigkeit der Vorgesetztenurteile relevant. Viswesvaran, et al. (1996) errechneten metaanalytisch eine interne Konsistenz (Cronbachs Alpha) bei Gesamtbeurteilungen der Mitarbeiter durch Vorgesetzte von r=.86 und eine intraindividuelle Urteilskonkordanz (bei Messwiederholung) von r=.81. Die Reliabilität der Vorgesetztenurteile ist dabei bei Verhaltens- und Ergebnisdimensionen leicht höher, als bei Eigenschafts- und Fähigkeitsdimensionen (Schuler, et al., 1995).

Die besprochenen Urteilsfehler haben ebenfalls Auswirkungen auf die Konstruktvalidität des Vorgesetztenurteils. Faktoranalytisch zeigt sich diese in der Extraktion eines *Generalfaktors*, der einen Großteil der Varianz aufklärt (Viswesvaran, Schmidt & Ones, 2005). Er wird neben den klassischen Urteilsfehlern (insb. *Halo-Effekt*) darauf zurückgeführt, dass sehr generell angesiedelte Faktoren (wie *allgemeine Intelligenz* oder generelle Persönlichkeitseigenschaften) den Berufserfolg auf nahezu allen Dimensionen beeinflussen. Dies führt zu hohen Interkorrelationen (Feldmann, 1981; Motowidlo, Borman & Schmit, 1997). Hulin (1982) weist zudem darauf hin, dass die meisten Fähigkeiten positiv zusammenhängen und somit Personen, die Stärken in einer Fähigkeit haben, automatisch auch in anderen Bereichen stark sind.

Die Ausführungen verdeutlichen, dass die Verwendung von Vorgesetztenurteilen als Validierungskriterium bei der Evaluation von Personalauswahlinstrumenten auf Grund ihrer, durch Urteilsfehler bedingt eingeschränkten Objektivität (Interrater-Reliabilität) und Reliabilität, mit Vorsicht vorzunehmen ist. Dennoch erfreut sie sich großer Beliebtheit, nicht zuletzt bedingt durch die ökonomische Erhebung, die scheinbare Vergleichbarkeit von verschiedenen beruflichen Tätigkeiten und der Möglichkeit, umfassende Bewährungseinschätzungen vorzunehmen (Hossiep, 1995, S.51f.). Zudem sind sie meistens in den Unternehmen sowieso vorhanden, da sie als personalpolitisches Instrument in regelmäßigen Abständen durchgeführt werden.

2.2.3 Auswahl und Operationalisierung der Prädiktoren

Nach dem Feststehen der Kriterien des Berufserfolges besteht der nächste Schritt darin, geeignete Vorhersagevariablen (Prädiktoren) für die Kriterien zu identifizieren. Grundsätzlich kommen dafür alle Informationen über die Person in Frage, die zur Reduzierung der Unsicherheit bezüglich der zu treffenden Entscheidung beitragen können (Hossiep, 1995, S.29). *Abbildung 11* gibt eine Übersicht möglicher Determinanten des Berufserfolges. Die Wahl der geeigneten Prädiktoren(konstrukte) erfordert aber neben dem aus der AA gewonnenen spezifischen Wissen über die Arbeitsstelle ebenso wissenschaftliches Know-how (Guion, 1998, S. 155). Dies betrifft vor allem das Wissen um Erkenntnisse über das Wirkungsgefüge von Prädiktor- und Kriteriumskonstrukten aus Studien und Metaanalysen (siehe 2.5). Die Auswahl der Prädiktoren sollte sich aber nicht alleine an einer übertrieben praxisfernen Wissenschaftsperspektive, sondern letztlich an einem „gesunden“ Mix aus wissenschaftlichen Erkenntnissen und dem konkreten Erfahrungswissen in der Organisation orientieren. Kompa (1989) zeigt zudem, dass hohe Vorhersagevalidität alleine nicht die An-

gemessenheit eines Prädiktors für die Auswahlentscheidung bestimmen kann. So gibt es in vielen Fällen Personeneigenschaften, die Grundvoraussetzung guter Leistungserbringung sind (z.B. Sehfähigkeit), allerdings nicht zu Erklärung von Leistungsunterschieden beitragen. Zudem ist die subjektive Bedeutsamkeit eines Prädiktors, also die nomologische Einordnung in das kognitive System des Entscheidungsträgers, für seine Verwendung bedeutsam. Problematisch ist außerdem, dass viele Prädiktoren Maximalleistungen erfassen (z.B. Leistungstests), während Kriterien sich an typischen Leistungskriterien, also der Normalleistung orientieren. Letztere hängt aber in nicht unerheblichen Maße auch von motivationalen Faktoren und interindividuellen Unterschieden ab, so dass es zu einer mangelnden Kongruenz zwischen Prädiktor und Kriterium kommt (Hossiep, 1995, S.81). Kontroversen gibt es zusätzlich darüber, ob durch zunehmende Komplexität und Schwierigkeit der Tätigkeit, die durch psychologische Verfahren aufklärbaren Varianzanteile am Kriterium sinken (Hunter & Hunter, 1984; Jäger, 1986).

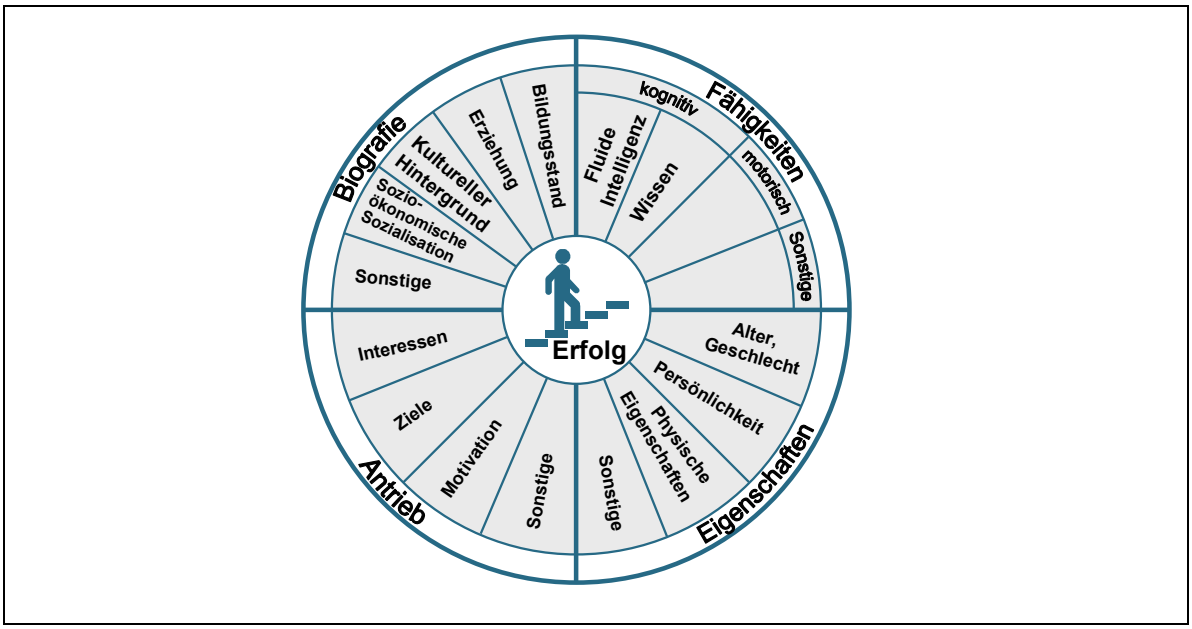

Abbildung 11: Mögliche personelle Determinanten des beruflichen Erfolges (eigene Darstellung in Anlehnung an Hossiep, 1995, S.29)

Nach der Festlegung der relevanten Prädiktoren besteht die nächste Hürde in der geeigneten Operationalisierung. Es gibt dabei nicht *das* eine geeignete Instrument. Instrumente zur Operationalisierung haben vielmehr anwendungsspezifische Vor- und Nachteile (Jones, 1989, S.275). Die Bandbreite, mit der Messinstrumente angewandt werden, ist erstmals von Cronbach (1949) unter dem Begriff *Bandbreiten-Fidelitäts-Dilemma* angesprochen worden. Gemeint ist damit, dass zunehmende Anwendungsflexibilität bei Instrumenten auf Kosten der Fidelität (Zuverlässigkeit, Präzision der inhaltlichen Abgrenzung) geht. Für die Entwick-

lung geeigneter eignungsdiagnostischer Verfahren stehen unterschiedliche Konstruktionsprinzipien zur Verfügung, die sich damit beschäftigen, welche Items zu einem Test zusammengefasst werden sollen. Nach der *rationalen* oder auch *deduktiven* Methode ist das Vorliegen einer Theorie die Ausgangsbasis für die Konstruktion oder die Auswahl eines Instrumentariums. Das theoretische Postulat sollte allerdings zur Sicherung des Erfolges empirisch validiert werden (Amelang & Zielinski, 1997, S. 98ff.; Wiggins, 1973). Die *externale* oder *kriteriumsbezogene* Methode setzt als Basis der Konstruktion das Vorliegen verschiedener Gruppen voraus. Die Auswahl der einzusetzenden Instrumente verläuft nach rein statistisch-empirischen Befunden. Die Nachteile dieses Vorgehens liegen darin, dass Differenzierungen von Einzelpersonen nur auf der Gruppenebene möglich sind. Das heißt, dass zwar Wahrscheinlichkeiten bezüglich der Zugehörigkeit zu einer Gruppe, nicht jedoch eine dimensionale Interpretation beim Vorliegen distinkter Klassen ohne kontinuierliche Übergänge angegeben werden können. Zudem ist eine inhaltliche Interpretation der eingesetzten Verfahren ohne theoretisch-rationale Überlegungen unzulässig und somit eine die Validität steigernde Weiterentwicklung schwer (Amelang & Zielinski, 1997, S.103; Wiggins, 1973). Bei der *induktiven oder auch internalen* Methode werden Items, basierend auf ihren Korrelationen und faktoranalytischen Prinzipien, zu Skalen zusammengefasst (Fisseni, 1990, S. 30). Der *Prototypenansatz* schließlich geht davon aus, dass kognitive Kategorien als Einteilungsraster dienen können, mit denen Zuordnungen von Items auf bestimmte Skalen überprüft werden können. Zunächst galt dies für Objekte, dann für Menschentypen, später sogar für Verhaltensweisen (Amelang & Zielinski, 1997, S. 109).

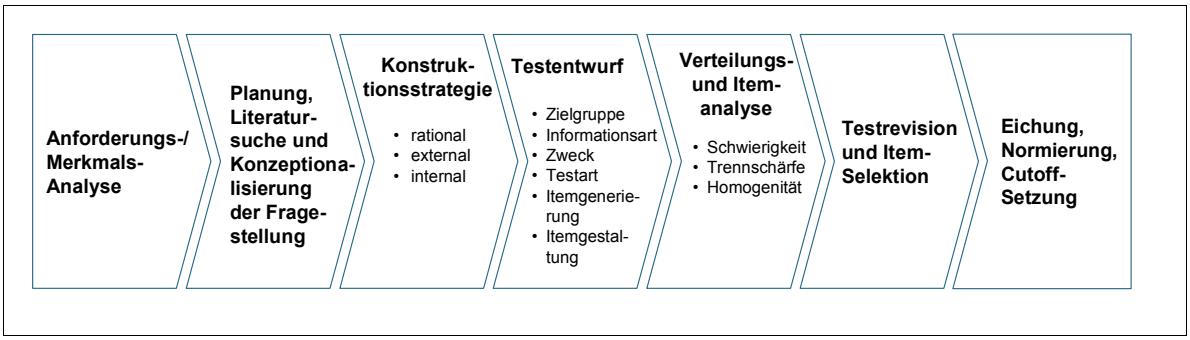

Abbildung 12: Ablauf einer Testkonstruktion nach der KTT (in Anlehnung an Bühner (2004) und Fisseni (1990))

In vielen diagnostischen Untersuchungen werden Tests verwandt, d.h. Verfahren, bei denen in standardisierten Situationen Verhaltensmerkmale von Personen erfasst werden, die als Indikatoren für bestimmte Personeneigenschaften dienen sollen. Dabei kann ein Test als jegliche Form der Messung von Daten, also auch beispielsweise als ein Beurteilerrating auf

bestimmten Eigenschaften verstanden werden (Moser & Schuler, 1989). *Abbildung 12* gibt abschließend eine Übersicht der Schrittabfolge bei der Testkonstruktion am Beispiel der KTT. Mit dem zentralen Gütekriterium für die Beurteilung von Tests, insbesondere bei der Stufe der Testrevision, der *Validität*, beschäftigt sich das nachfolgende Kapitel.

2.2.4 Validitätsanalyse

Die *Kriteriumsvalidität* ist, vor dem Hintergrund der Personalauswahl, die wichtigste Validitätsart und soll in diesem Kapitel näher betrachtet werden. Doch zunächst soll für die dabei auftretenden Probleme, Fragestellungen und Zusammenhänge mit der Reliabilität, ein analytisches Rahmenkonzept, das sich an die KTT und die Reliabilitätstheorie von Wittmann (1988) anlehnt, dargestellt werden. Betrachtet man die Datenboxen aus *Abbildung 2*, so betreffen Fragen der Validität die Beziehungen der Boxen untereinander und die der Reliabilität solche innerhalb einer Box. Die Varianz einer Messung (also einer Box) kann nach Wittmann dann in folgende Einzelkomponenten zerlegt werden:

$$(1) \qquad \sigma^2_{tot} = \sigma^2_{ZwP} + \sigma^2_{ZwS} + \sigma^2_{ZwPxS} + \sigma^2_{Fehler}$$

Demnach setzt sich die Gesamtvarianz σ^2_{tot} einer Messung (einer Variablen) aus der Varianz zwischen den Personen σ^2_{ZwP}, der Varianz zwischen den Situation σ^2_{ZwS}, der Interaktionsvarianz σ^2_{ZwPxS} dieser beiden und der Fehlervarianz σ^2_{Fehler} zusammen. Folglich konstituieren die ersten drei Komponenten, ohne den Fehlerterm, die *wahre* Varianz. Die Reliabilität lässt sich dann analog zu den Axiomen der KTT darstellen durch:

$$(2) \qquad r_{tt} = \frac{\sigma^2_{ZwP} + \sigma^2_{ZwS} + \sigma^2_{ZwPxS}}{\sigma^2_{ZwP} + \sigma^2_{ZwS} + \sigma^2_{ZwPxS} + \sigma^2_{Fehler}}$$

Überträgt man nun das Vorhersageproblem der Personalauswahl auf das Konzept der fünf Datenboxen aus *Abbildung 2*, so zeigt sich, dass es sich um eine Boxenkombination der PR-, NTR- und der KR-Box handelt (auf die Evaluationsbox wird unter 2.3 separat eingegangen). Die NTR-Box spiegelt dabei all die „ungewollten" Treatments, denen Personen im Lauf der Zeit, die von der Einstellung bis zum Zeitpunkt der Beurteilung (als KR-Box) vergangen ist, ausgesetzt waren. Diese können beispielsweise je Person unterschiedliche Aufgaben sein, die zu unterschiedlich starken Lerneffekten geführt haben und somit die Leistung auf dem Kriterium beeinflussen. Da bei der *PR-NTR-KR*-Boxenkombination auf Grund des nichtexperimentellen Charakters keine Randomisierung vorliegt, kommt es zu einer Konfundierung der Effekte von NTR- und PR-Box auf dem Kriterium. Dass bedeutet, dass die aufgeklärten Varianzanteile von Kriterium und Prädiktor nicht als einfache Summe darstellbar

sind. Entsprechend *Gleichungen (1)* und *(2)* lassen sich für jede einzelne der drei Boxen, die Varianzkomponenten und ihre Reliabilitäten separat darstellen:

$$(3a) \quad \sigma^2_{tot}{}^{PR} = \sigma^2_{ZwP}{}^{PR} + \sigma^2_{ZwS}{}^{PR} + \sigma^2_{ZwPxS}{}^{PR} + \sigma^2_{Fehler}{}^{PR}$$

$$(3b) \quad r_{tt}{}^{PR} = \frac{\sigma^2_{ZwP}{}^{PR} + \sigma^2_{ZwS}{}^{PR} + \sigma^2_{ZwPxS}{}^{PR}}{\sigma^2_{ZwP}{}^{PR} + \sigma^2_{ZwS}{}^{PR} + \sigma^2_{ZwPxS}{}^{PR} + \sigma^2_{Fehler}{}^{PR}}$$

$$(4a) \quad \sigma^2_{tot}{}^{KR} = \sigma^2_{ZwP}{}^{KR} + \sigma^2_{ZwS}{}^{KR} + \sigma^2_{ZwPxS}{}^{KR} + \sigma^2_{Fehler}{}^{KR}$$

$$(4b) \quad r_{tt}{}^{KR} = \frac{\sigma^2_{ZwP}{}^{KR} + \sigma^2_{ZwS}{}^{KR} + \sigma^2_{ZwPxS}{}^{KR}}{\sigma^2_{ZwP}{}^{KR} + \sigma^2_{ZwS}{}^{KR} + \sigma^2_{ZwPxS}{}^{KR} + \sigma^2_{Fehler}{}^{KR}}$$

$$(5a) \quad \sigma^2_{tot}{}^{NTR} = \sigma^2_{ZwP}{}^{NTR} + \sigma^2_{ZwS}{}^{NTR} + \sigma^2_{ZwPxS}{}^{NTR} + \sigma^2_{Fehler}{}^{NTR}$$

$$(5b) \quad r_{tt}{}^{NTR} = \frac{\sigma^2_{ZwP}{}^{NTR} + \sigma^2_{ZwS}{}^{NTR} + \sigma^2_{ZwPxS}{}^{NTR}}{\sigma^2_{ZwP}{}^{NTR} + \sigma^2_{ZwS}{}^{NTR} + \sigma^2_{ZwPxS}{}^{NTR} + \sigma^2_{Fehler}{}^{NTR}}$$

Die, die Auseinandersetzungen um prädiktive und konkurrente Validität flankierende Diskussion über die Validitätsstabilität lässt sich anhand dieser Darstellung gut erklären. Erfahrungsgemäß sinkt die Prognosefähigkeit mit zunehmendem zeitlichem Abstand zwischen Prädiktor und Kriterium. Gründe dafür können, wie in 2.2.2 bereits angesprochen, instabile, also unreliable Kriterien und, wie Tack (1980) postuliert, ebenso unreliable Prädiktoren sein. Reliabilitätseinbußen auf Kriterienseite können dabei durch unvorhersehbare Interventionen (NTR-Box, $\sigma^2_{ZwP}{}^{NTR}, \sigma^2_{ZwS}{}^{NTR}, \sigma^2_{ZwPxS}{}^{NTR}$) oder beispielsweise sich verändernde Aufgaben und Umweltveränderungen ($\sigma^2_{ZwS}{}^{KR}, \sigma^2_{ZwPxS}{}^{KR}$) bedingt sein. Auf der Prädiktorenseite wiederum können sich Personenvariablen beispielsweise durch situative Bedingungen ($\sigma^2_{ZwS}{}^{PR}$) bei mehrmaliger Testvorgabe oder bei zeitlich nachgelagerter Prädiktorenerfassung, durch berufliche Erfahrung (NTR-Box, $\sigma^2_{ZwP}{}^{NTR}, \sigma^2_{ZwS}{}^{NTR}, \sigma^2_{ZwPxS}{}^{NTR}$) verändern.

In der Regel wird man nicht in der Lage sein, alle Varianzkomponenten aus *Gleichung (3)* bis *(5)* voneinander zu trennen. Stattdessen werden immer bestimmte Bestandteile in ihrer Höhe verborgen bleiben bzw. nur bestimmte Varianzkomponenten dem Interesse des Evaluationsforschers dienen. Dies bedeutet, dass einzelne Komponenten der *wahren* Varianz, sofern sie nicht in ihrer Höhe identifizierbar sind, zu *unerwünschter* Varianz werden und somit die unsystematische Fehlervarianz erhöhen. Damit verringert sich aber auch die Reliabilität der Beobachtungen *erwünschter* Varianz. Dies bedeutet, dass je nach Fragestellung unterschiedliche Varianzquellen zur Fehler- bzw. zur wahren Varianz beitragen. Man nehme als Beispiel den hypothetischen Fall an, dass mit Hilfe der Intelligenz von Bewerbern der Berufserfolg vorhergesagt werden soll. Nun werden alle Bewerber im Rahmen des Auswahlverfahrens an unterschiedlichen Terminen einem Intelligenztest unterzogen, um die

„wahren“ Unterschiede in dem Intelligenzniveau festzustellen ($\sigma^2{}_{ZwP}{}^{PR}$=.4). Allerdings gibt es ungewollte Variationen bei der Durchführung des Tests durch eine unterschiedlich starke Lärmbelästigung (*unerwünschte* Varianz $\sigma^2{}_{ZwS}{}^{PR}$=.2). Nicht alle Bewerber reagieren gleich empfindlich auf diese Belästigung, manche von ihnen sind bei gleichem Lärmpegel in ihrer Leistungsfähigkeit stärker beeinträchtigt (*unerwünschte* Varianz $\sigma^2{}_{ZwPxS}{}^{PR}$=.2). Da Werte über die unterschiedlichen Lärmpegel und die Sensitivität der Bewerber nicht vorliegen, lassen sich diese Varianzanteile nicht bestimmen. Sie sind Teil der unsystematischen Fehlervarianz, die den ursprünglichen Messfehler ($\sigma^2{}_{Fehler}{}^{PR}$=.2) erhöht. Die Reliabilität der Messung der „*wahren“*, gewünschten Werte der Intelligenz, die als Prädiktor für die Berufserfolg verwendet werden sollen (Fall (6b)), halbiert sich gegenüber der Reliabilität der „wahren“, gewünschten und ungewünschten Werte (Fall (6a)):

$$(6a) \qquad r_{tt}{}^{PRa} = \frac{\sigma^2{}_{ZwP}{}^{PR} + \sigma^2{}_{ZwS}{}^{PR} + \sigma^2{}_{ZwPxS}{}^{PR}}{\sigma^2{}_{ZwP}{}^{PR} + \sigma^2{}_{ZwS}{}^{PR} + \sigma^2{}_{ZwPxS}{}^{PR} + \sigma^2{}_{Fehler}{}^{PR}} = \frac{.4+.2+.2}{.4+.2+.2+.2} = .8$$

$$(6b) \qquad r_{tt}{}^{PRb} = \frac{\sigma^2{}_{ZwP}{}^{PR}}{\sigma^2{}_{ZwP}{}^{PR} + \sigma^2{}_{ZwS}{}^{PR} + \sigma^2{}_{ZwPxS}{}^{PR} + \sigma^2{}_{Fehler}{}^{PR}} = \frac{.4}{.4+.2+.2+.2} = .4$$

Die KTT sieht für den Fall der Validitätseinbußen durch mangelnde Reliabilität die Attenuations- oder Minderungskorrektur vor. Dabei wird die gemessene Korrelation zwischen Kriterium und Prädiktor an der Wurzel der Reliabilität der beobachteten Kriterienwerte (einfache Minderungskorrektur) oder zusätzlich auch noch an der Reliabilität der beobachteten Prädiktorwerte (doppelte Minderungskorrektur) relativiert bzw. korrigiert, um die Korrelation der „wahren“ Werte zu erhalten. Damit ergibt sich:

$$(7) \qquad r_{PR,KR}{}^{wahr} = \frac{r_{PR,KR}{}^{beobachtet}}{\sqrt{r_{tt}{}^{PR}\, r_{tt}{}^{KR}}}$$

Gleichung (7) gibt also den Gesamtzusammenhang der Validität zwischen Prädiktor und Kriterium und der Reliabilität ihrer beobachteten Werte wieder. Sie zeigt das bekannte *Verdünnungsparadoxon*, nach dem bei Konstanthaltung der Validität und Erhöhung der Reliabilität die „wahre“ hochgerechnete Validität sinkt (Fisseni, 1990, S.73). Für den Kontext der Evaluation von Personalauswahlverfahren sollte in die Korrektur lediglich die Kriterienreliabilität einfließen, da nur sie die tatsächlichen Ergebnisse beeinflusst, während die Prädiktorunreliabilität ein Faktum darstellt (Guion, 1998, S.313). Gleichungen (3) bis (5) verdeutlichen zudem, dass diese wiederum von einer Vielzahl potenzieller Varianzquellen abhängen. Wittmann (1985) zeigte, dass durch systematische, theoriegeleitete Aggregation über die verschiedenen Varianzquellen, unsystematische (unerwünschte) Varianz eliminiert

werden kann, der Anteil systematischer (gewünschter) Varianz an der Gesamtvarianz größer wird und somit die Reliabilität steigt. Die Auswirkungen von Aggregation auf die Reliabilität eines Tests findet sich ebenso in der im Rahmen der KTT bekannten *Spearman-Brown-Formel* wieder:

$$r_{tt_k} = \frac{k r_{tt}}{1 + (k-1) r_{tt}} \tag{8}$$

Gleichung (8) errechnet den Reliabilitätsgewinn eines Tests, bei Verlängerung um den Faktor *k*. Wie man erkennen kann, wird der Test bei Zunahme der Anzahl an Testitems reliabler. Setzt man nun *Gleichung (2)* für r_{tt} ein, so erhält man nach Umformung:

$$r_{tt_k} = \frac{\sigma^2_{ZwP} + \sigma^2_{ZwS} + \sigma^2_{ZwPxS}}{\sigma^2_{ZwP} + \sigma^2_{ZwS} + \sigma^2_{ZwPxS} + \frac{1}{k}\sigma^2_{Fehler}} \tag{9}$$

Gleichung (9) verdeutlicht, dass die Fehlervarianz um den Faktor *1/k* bei Verlängerung um *k* verringert wird. Das heißt, dass mit zunehmender Anzahl zu aggregierender Items sich der Messfehler bei der Korrelation der Items (unter der Voraussetzung $r_{tt} \neq 0$) verringert (Wittmann & Matt, 1986, S. 316f.).

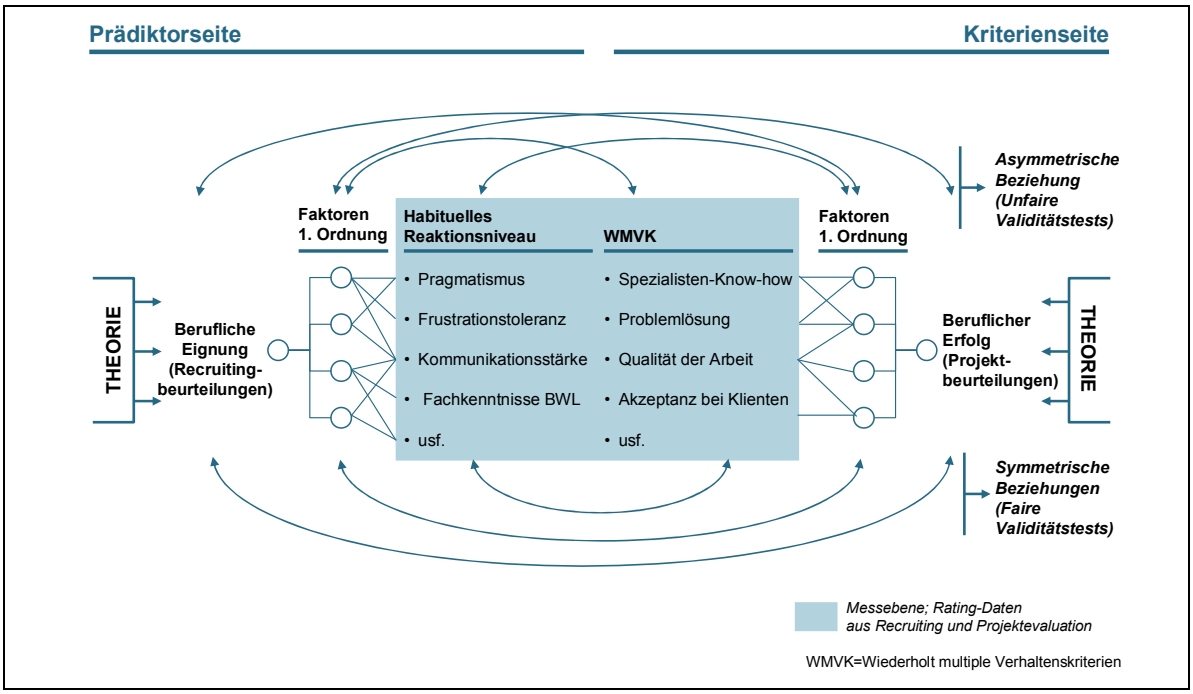

Abbildung 13: Beispiel für ein hierarchisches Brunswik'sches Linsenmodell

Wittmann (1985; 1990) zeigt außerdem, dass Kriterium und Prädiktor hinsichtlich ihrer Breite und inhaltlichem Bezug übereinstimmen sollten, da es ansonsten ebenfalls zu er-

heblichen Validitätseinbußen kommen kann. Anhand des von Brunswik (1955) im Kontext der Perzeption entwickelten hierarchischen Linsenmodells, weist er darauf hin, „dass ein rein empirischer Ansatz, der nur die Betrachtung des Korrelationskoeffizienten im Auge hat, nicht die alleinige Lösung sein kann, da er nichts über die Generalitätsebene aussagt“. Das bedeutet auf die oben dargestellten Varianzkomponenten bezogen, dass sie einander entsprechen (symmetrisch sein) müssen. *Abbildung 13* zeigt diese Symmetrieprinzipien am Beispiel des Auswahlverfahrens der vorliegenden Evaluationsstudie. Geht man nun von der Annahme aus, die dargestellte Struktur gäbe die tatsächlichen faktoranalytisch bestätigten Konstrukte wieder, so würden Vergleiche unterschiedlicher Generalitätsebenen (Asymmetrie) zu *unfairen* Validitätstests führen. *Faire* Validitätstest wären hingegen Vergleiche auf gleicher Generalitätsebene (Symmetrie). Asymmetrien, also mangelnde Konstruktvalidität, können selbst bei völlig reliablen Instrumenten wie Unreliabilität wirken und die Validität verringern oder gar zu einer Validität von Null führen (Wittmann, 1985, S.66).

Zudem können bei Asymmetrie Prädiktoren mit Kriterien trotz hoher Konstruktvalidität nur sehr gering korrelieren, auch wenn die „wahre“ Korrelation der Konstrukte hoch ist. Basierend auf der Linsengleichung von Tucker (1964) hat Wittmann die Symmetrieaspekte in das bisher Erläuterte integriert und, von der Notation hier leicht verändert, wie folgt dargestellt:

$$(10) \qquad r_{PR,KR}^{\ beobachtet} = S\sqrt{r_{tt}^{\ PR} r_{tt}^{\ KR}}\ r_{PR,KR}^{\ wahr} R_{PR} R_{KR} + C\sqrt{1-R^2{}_{PR}}\sqrt{1-R^2{}_{KR}} + e$$

Betrachtet man den Gleichungsteil links vom Summenzeichen, so ist die Multiplikation des wahren Zusammenhangs $r_{PR,KR}^{\ wahr}$ mit $\sqrt{r_{tt}^{\ PR} r_{tt}^{\ PR}}$ die Auflösung der Minderungskorrekturformel (7) nach $r_{PR,KR}^{\ beobachtet}$. Sie berücksichtigt also die Unreliabilität von den Prädiktor- und Kriteriumsoperationalisierungen analog zu den Gleichungen (3b) und (4b). Hinzukommen noch die Größen R_{PR} und R_{KR}, die die Konstruktvalidität von Prädiktor und Kriterium unter intendierten symmetrischen Bedingungen und somit auch die Minderung der Validität durch Asymmetrie ausdrücken. Sie sind als multiple Korrelationskoeffizienten dargestellt, die den linearen Anteil des Zusammenhangs der Operationalisierungen von Prädiktor und Kriterium mit den entsprechenden Konstrukten widerspiegeln. Die Größe S ist Ausdruck des potenziellen Selektionseffektes. Sie kann Werte größer und kleiner Eins annehmen und ist immer dann ungleich Eins, wenn die Varianz der Stichprobe ungleich der Varianz der Population, auf die generalisiert werden soll, ist. Eine Überschätzung (S >1) kann beim Vorhandensein von Extremgruppen vorliegen, eine Unterschätzung (S <1), bei der Personalauswahl der gängigere Fall, verdeutlicht das Phänomen der *restriction of range* (Wittmann,

1990, S.248). Betrachtet man nun den Gleichungsteil rechts vom Summenzeichen ($C\sqrt{1-R^2_{PR}}\sqrt{1-R^2_{KR}}$), so beinhaltet dieser die nichtlinearen Komponenten zwischen Kriterium und Prädiktor. Aus Vereinfachungsgründen und auf Grund der generellen Stärke allgemeiner linearer Modelle, soll aber die Attenuation durch nicht lineare Varianzanteile im Folgenden vernachlässigt und die nichtlinearen Varianzanteile wie Fehlervarianz *e* betrachtet werden (siehe auch Wittmann, 1987, S.217).

$$r_{PR,KR}^{\ beobachtet} = S\sqrt{r_{tt}^{\ PR} r_{tt}^{\ KR}}\ r_{PR,KR}^{\ wahr} R_{PR} R_{KR} + e \qquad (11)$$

Gleichung (11) zeigt, dass dann die Korrelation von Prädiktor und Kriterium durch insgesamt sechs Faktoren bestimmt wird. Neben der „wahren" linearen Korrelation von Prädiktor- und Kriteriumskonstrukt bei perfekter Symmetrie ($r_{PR,KR}^{\ wahr}$), ist dies die Konstruktvalidität oder -reliabilität von Prädiktor und Kriterium bei intendierter Symmetrie ($R_{PR}R_{KR}$). Darüber hinaus beeinflussen die Reliabilitäten der jeweiligen Messinstrumente von Prädiktor ($r_{tt}^{\ PR}$) und Kriterium ($r_{tt}^{\ KR}$) sowie des Selektionsfehlers *S*, der als einzige Position die Korrelation erhöhen kann, die beobachtete Korrelation. Die dargestellten Zusammenhänge sind auch auf andere korrelative Zusammenhänge (z.B. zwischen NTR und KR) übertragbar.

Abbildung 14 fasst die geschilderten Aspekte zusammen und zeigt links die Kriterien- und rechts die Prädiktorenseite. Die Varianzen der einzelnen Variablen sind durch Kreise dargestellt. Auf der unteren Ebene der Population ist die tatsächliche insgesamt vorhandene Verteilung bzw. Varianz der Gesamtpopulation, d.h. im vorliegenden Beispiel die Gesamtzahl der Bewerber, abgebildet. Durch den Selektionseffekt (symbolisiert durch *S*+*e*) kann diese, wie oben erwähnt, die Varianz der gemessenen Variablen auf der manifesten Ebene beeinflussen. Die Schnittmenge zwischen Kriterium und Prädiktor auf der manifesten Ebene stellt die tatsächlich gemessene Validität dar, die auch durch die NTR-Box (s.o.) beeinflusst wird. Die Schnittmengen von Kriterium und Prädiktor zur latenten Ebene visualisieren die Konstruktvalidität. Ihre Schnittmenge untereinander auf der latenten Ebene ergibt die tatsächliche „wahre" Validität. Das Schaubild zeigt, dass es erhebliche Differenzen zwischen der gemessenen und tatsächlichen prädiktiven Validität geben kann und wo die potenziellen Varianzquellen, die diese Unterschiede verursachen, vorzufinden sind.

Ein Beispiel soll die geschilderten Zusammenhänge verdeutlichen: Im Rahmen der Validierung eines Auswahlverfahrens wird Berufserfolg durch das aktuelle Kriterium der Geschwindigkeit, mit der die Personen ihnen übertragene Aufgaben erledigen, im Rahmen einer Vorgesetztenbeurteilung bestimmt (Leistungskriterium). Als Prädiktor soll Intelligenz

dienen, die im Rahmen eines Tests, der ausschließlich auf verbale Fähigkeiten fokussiert, gemessen wird. Die erhaltenen Daten auf Prädiktoren- und Kriterienseite sind durch die Datenboxen visualisiert, ihre Varianz als Kreise. Als Stichprobe dienen jene Personen, die mindestens drei Jahre im Unternehmen arbeiten und von denen Intelligenzwerte des Testes aus dem Auswahlverfahren vorliegen. Da nicht die Gesamtpopulation aller Personen, die das Verfahren durchlaufen haben, herangezogen wird, entsteht ein Selektionseffekt (*S*), der eine varianzverändernde Auswirkung haben kann. Der Test verbaler Fähigkeiten erfasst aber auch Aspekte, die nichts mit Intelligenz gemein haben, wie z.B. Englischkenntnisse. Dies bedeutet, dass Teile der Varianz zwischen den Personen auf der Prädiktorenseite nicht mit dem Konstrukt überlappen (teilweise symbolisiert durch die nicht gefärbten Anteil von $\sigma^2{}_{ZwP}{}^{PR}$). Außerdem mussten die Teilnehmer während des Test die o.g. Lärmbelästigungen ertragen ($\sigma^2{}_{ZwS}{}^{PR} + \sigma^2{}_{ZwPxS}{}^{PR}$). Zudem gibt es Fehler bei der Übertragung der Ergebnisse in einen Rechner ($\sigma^2{}^{PR}{}_{Fehler}$). Die dadurch beeinträchtigte Konstruktreliabilität (R_{PR}) zeigt die gefärbte Fläche zwischen manifester und latenter Ebene auf der Prädiktorenseite.

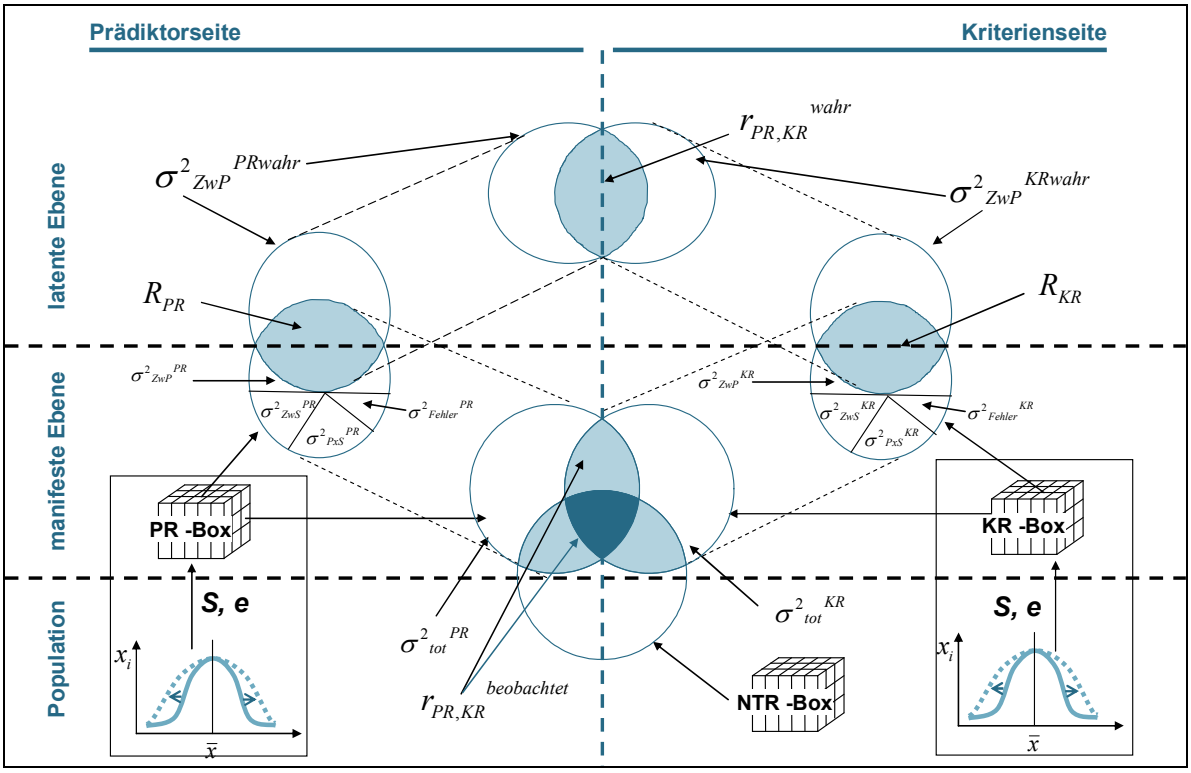

Abbildung 14: Darstellung des Gesamtzusammenhangs der Varianzen[6]

[6] Prämisse der Darstellung ist, dass die interessierenden Varianzkomponenten auf beiden Seiten die interindividuellen Differenzen in der Intelligenz und im Berufserfolg sind

Auf der Kriterienseite wiederum gibt es einige Personen, die vom Vorgesetzten lediglich bei unwichtigen Aufgaben beobachtet wurden (z.B. Kaffeekochen). Diese schnell zu erledigen, hat mit dem tatsächlichen Berufserfolg nichts zu tun. In die Messung des Kriteriums fließen aber auch diese Tätigkeiten ein (ein Teil davon ist wiederum der nicht gefärbte Teil von $\sigma^2{}_{ZwP}{}^{KR}$). Darüber hinaus ist der Vorgesetzte eine sehr launische Person, die sich bei der Urteilsbildung stark von ihrem augenblicklichen Stimmungszustand beeinflussen lässt und zahlreichen anderen Urteilsfehlern unterliegt ($\sigma^2{}_{ZwS}{}^{KR}, \sigma^2{}_{PxS}{}^{KR}, \sigma^2{}_{Fehler}{}^{KR}$). Auch hier zeigt der gefärbte Teil von $\sigma^2{}_{ZwP}{}^{KR}$ die Konstruktreliabilität R_{KR}. Außerdem hatten die Mitarbeiter während ihrer Berufstätigkeit unterschiedlich gute Lerngelegenheiten (NTR-Box), die die berufliche Leistung beeinflussen und nicht auf den Prädiktor zurückzuführen sind.

Für die Validierung gibt es noch ein zusätzliches Problem: Das Leistungskriterium (Geschwindigkeit) kann genauso wenig in vollem Umfang Berufserfolg (als eigentliches Kriterium) widerspiegeln wie verbale Fähigkeiten die Intelligenz. Damit liegt eine eingeschränkte Konstruktvalidität auf Kriterium- und Prädiktorseite vor. Dies bedeutet, dass die Symmetrieprinzipien einen korrelativen Vergleich auf Grund von Asymmetrie verbieten. Für eine faire Validierung wäre es empfehlenswert, zunächst eine Symmetrieebene zwischen Prädiktor und Kriterium herzustellen, beispielsweise, indem man ein anderes Kriterium, wie *Kommunikation am Arbeitsplatz,* heranzieht. Außerdem könnte man die Beobachtungsanzahl beim Kriterium pro Person erhöhen und die Werte aggregieren. Damit würde ein Teil der unerwünschten Varianz (Kontamination des Kriteriums) reduziert. Zuletzt könnte man auch durch die Beobachtung einer Person durch mehrere Vorgesetzte einen Teil des Messfehlers herausmitteln. Auf Prädiktorenseite könnte man die Teilnehmer in schalldichten Räumlichkeiten den Test direkt am Rechner durchführen lassen, um unerwünschte Varianzquellen auszuschalten. Die Darstellung verdeutlicht, dass der Verweis auf die Reliabilität von Messinstrumenten alleine wenig aussagefähig ist. Perfekt reliable Auswahlverfahren können bei mangelhafter Konstruktvalidität prognostisch wertlos sein. Die gleiche Auffassung vertreten Michel und Conrad (1982, S.53f.), wenn sie meinen, dass die Bedeutung der Reliabilität insbesondere hinsichtlich ihres Beitrages zur Validität betrachtet werden sollte.

Die in den vorangehenden Abschnitten dargestellte Vielfalt varianzbeeinflussender Aspekte könnte zu der Annahme verleiten, dass Validitätsvergleiche nur zwischen Verfahren sinnvoll sind, die auf gleiche Weise an der gleichen Zielgruppe mit dem gleichen Kriterium durchgeführt wurden. Diesen Glaube an differenzielle Validität und Moderatoreffekte teilten und teilen immer noch zahlreiche Forscher (Schuler, 2000, S. 55). Allerdings konnten Schmidt und Hunter (Hunter & Schmidt, 1990; Schmidt, et al., 1985) in ihrem Konzept der

Validitätsgeneralisierung und der Metaanalysen zeigen, dass die Situationsspezifität größtenteils auf methodische Artefakte (Stichprobenfehler) zurückzuführen ist und Validitätskoeffizienten in hohem Maße, unter der Voraussetzung hinreichend großer Stichproben, generalisierbar sind. Anders ausgedrückt bedeutet dies, dass die tatsächlichen, auf den „wahren" Varianzen beruhenden „wahren" Validitäten generell und nicht differenziell sind. Hunter und Schmidt (1990) entwickelten Methoden, mit denen zahlreiche Studien in ihren Ergebnissen zusammengefasst bzw. generalisiert und dadurch Fehlerquellen eliminiert werden können. Auf einige mit Hilfe ihrer metaanalytischen Techniken gewonnenen Erkenntnisse wird unter 2.5 eingegangen.

2.2.5 Kombination der Prädiktordaten

Die im vorherigen Abschnitt dargestellte *Abbildung 14* geht aus Vereinfachungsgründen vom Vorliegen eines einzigen Prädiktors aus. In der berufseignungsdiagnostischen Wirklichkeit wird man aber auf mehrere Prädiktoren zurückgreifen, so dass sich die Frage stellt, wie man diese kombinieren muss, um eine maximal gültige Vorhersage des Kriteriums zu bekommen. Grundsätzlich muss man dabei zwei Ansatzpunkte unterscheiden, die in den bahnbrechenden Arbeiten von Meehl (1954), Sawyer (1966) und Wiggins (1973) ausführlich thematisiert wurden: der intuitive, im Kopf des Beurteilers stattfindende Ansatz und das statistische, auf mathematischen Modellen basierende Vorgehen. Beide sollen nachfolgend näher betrachtet werden.

2.2.5.1 Klinische vs. statistische Urteilsforschung

Wiggins (1973, S. 122ff.) unterteilt die Phasen der Kriteriumsprognose in die Datenerhebung und die -kombination, aus denen die Vorhersage des Kriteriums entsteht. Bei dem Schritt der Datenerhebung (der Prädiktorendaten) unterscheidet er zwischen *beurteilender Messung* (*judgemental measurement*) und *mechanischer Messung* (*mechanical measurement*). Basis der *beurteilenden Messung*, die durchaus auch numerische Größen generieren kann, sind subjektive Urteilsprozesse im Kopf des Diagnostikers. Er übernimmt dabei also die Funktion des Messgerätes. Die *mechanische Messung* verläuft hingegen standardisiert und objektiv. Es finden keine subjektiven Urteilsprozesse statt. Unabhängig davon, ob die Daten nun mechanisch oder klinisch erhoben wurden, setzt die Datenkombination das Vorhandensein von quantifizierten Daten voraus. Hier kann man ebenso eine klinische sowie eine statistische (oder mechanische) Variante unterscheiden (Kompa, 1989, S. 77). Der erste Fall, von Schuler (2000) auch *Expertenmodell* genannt, basiert wiederum auf einer menschlichen Entscheidung, die über eine optimale Datenkombination entscheidet. Beim zweiten

Fall, dem *Präzisionsmodell*, werden zur Datenkombination statistische Modelle angewandt, die auf Basis der in der Empirie ermittelten Beziehungen zwischen Prädiktor und Kriterium zu einer optimalen Kombinationsregel gelangen. Inwiefern die statistische der klinischen Datenkombination über- oder unterlegen ist, wurde kontrovers diskutiert (Wiggins, 1973). Kritiker der klinischen Urteilsbildung sehen den Kliniker bestenfalls als „second-rate IBM machine“, da die mathematische Modellierung der Vorhersage des Kriteriums optimale Lösungen liefert, denen sich der Kliniker nur annähern kann (Sarbin, 1944). Demgegenüber argumentieren Verfechter der klinischen Variante, dass der Kliniker es mit individuellen Fällen zu tun hat und nicht Entscheidungen über Gruppen fällt (nomothetische vs. idiographische Perspektive). Zudem betonen sie die Fähigkeit des Klinikers, einzigartige Muster in der Kombination der Prädiktorendaten zu erkennen und diese in das Urteil mit einfließen zu lassen (Kompa, 1989).

Tabelle 4: Ergebnisse der Sawyer'schen Reanalyse und Beispiele (Wiggins, 1973, S. 197**)**

Art der Datenerhebung	**Klinische Datenkombination**	**Mechanische Datenkombination**
Klinisch	Rein klinisch (20%) (Einstellungsentscheidung basierend auf Vorstellungsgespräch und Bewerbungsunterlagen)	Eigenschaftsbeurteilung (43%) (Kombination von aus Interviewratings gewonnenen Daten nach festgelegten mathematischen Entscheidungsregeln)
Mechanisch	Profilinterpretation (38%) (Persönlichkeitsprofile aus Persönlichkeitsfragebögen werden vom Diagnostiker interpretiert)	Rein mechanisch (63%) (Werte aus Intelligenztests werden nach empirisch ermittelten Regeln kombiniert)
Beide	Klinisch zusammengesetzt (26%) (Eindrücke aus einem Interview sowie Werte aus einem Intelligenztest fließen in das Urteil durch den Diagnostiker ein)	Mechanisch zusammengesetzt (75%) (Werte aus einem Interviewrating und einem Intelligenztest werden mit Hilfe mathematischer Modelle kombiniert)
Eine oder beide	Klinische Synthese (50%) (Werte aus einem Interviewerrating und einem Intelligenztest werden mit Hilfe mathematischer Modelle kombiniert und inkl. Prognose dem Diagnostiker vorgelegt, der die Einstellungsentscheidung trifft)	Mechanische Synthese (75%) (Werte aus einem Interviewerrating und einem Intelligenztest werden dem Diagnostiker vorgelegt. Seine Prognose und die Ausgangsdaten werden mit Hilfe mathematischer Modelle kombiniert)

( ) Überlegenheit in % gegenüber den anderen Methoden

Die Diskussionen um die Güte statistischer (mechanischer) vs. klinischer Datenkombination finden ihren Ursprung in Untersuchungen von Meehl (1954). Er postulierte die Überlegenheit des Klinikers im Entdeckungszusammenhang, da dieser neben der Erfahrung

auch in der Lage sei, Persönlichkeitstheorien in seine Vorhersage einzubeziehen. Wiggins (1973, S.156ff.) zeigte anhand vorliegender Studien basierend auf dem Brunswik'schen Linsenmodell, dass die beste Modellierung des klinischen Urteils durch multiple Regressionsmodelle möglich ist (paramorphe Abbildung des klinischen Urteils). Nonlineare Komponenten beim klinischen Urteil seien zudem eher unwahrscheinlich. Meehl (1954,1965) selbst stellte darüber hinaus in seinen Box-Score-Studien, in denen er Ergebnisse aus zahlreichen Studien zusammenfasste, die Überlegenheit statistischer Vorhersagen fest. Diese Erkenntnis fand durch Grove und Meehl (1996) auch in einer neueren Metaanalyse mit 136 Studien Bestätigung.

Sawyer (1966) reanalysierte 45 Untersuchungen, die größtenteils von Meehl stammten und unterschied, wie oben erwähnt, zwischen Messung und Datenkombination, so dass acht Klassen unterschiedlicher Prognosemethoden entstanden. Die Ergebnisse des Vergleichs der einzelnen Prognosemethoden sind in *Tabelle 4* dargestellt. Der Prozentsatz bei den einzelnen Methoden gibt an, in wie viel Prozent der Fälle diese den anderen Methoden, die in den Vergleich eingegangen sind, überlegen sind. In Klammern ist für jede Methode ein Beispiel aus der Berufseignungsdiagnostik genannt.

Obgleich die Prozentzahlen auf Grund unterschiedlicher Anzahl von Vergleichen und der Tatsache, dass jeweils unterschiedliche Methoden verglichen wurden, mit Vorsicht zu interpretieren sind, kommt Wiggins (1973, S.197f.) zu Schlüssen, die grundsätzlich die Überlegenheit statistischer Vorgehensweisen postulieren. Allerdings unterstreicht er auch die Bedeutung des Klinikers als Datensammler. Unter dem Begriff „Automated clinical prediction“ behandelt Wiggins (1973, S. 199ff.) zudem anhand einiger Studien verschiedene Möglichkeiten, klinische Urteile nachzumodellieren. Dabei werden klinische Interpretationen von Daten standardisiert und dienen Computerprogrammen als Basis für nachfolgende Urteilsbildung. Die Daten können nach klinischem Wissen oder nach rein empirisch abgeleiteten Regeln kombiniert werden. Kleinmutz (1963) zeigte, dass die Simulation der Urteile eines Klinikers, eine höhere Vorhersagegenauigkeit als der Kliniker selbst erreichen. Goldberg (1970) kam bei der Modellierung durch multiple Regression zu vergleichbaren Schlüssen. Neuere Ansätze zur Replikation subjektiver Theorien und der Einschätzung der Bewährung verschiedener Maßnahmen sind die Struktur-Lege-Technik nach Groeben und Scheele, HYPAG und TYPAG (Hollmann, 1991) sowie das Goal-Attainment-Scaling (GAS) (Wittmann, 1985; Wottawa & Thierau, 2003).

Die zahlreichen Untersuchungen brachten Wiggins dazu, Empfehlungen zur Entscheidungsfindung beim Vorhandensein oder Fehlen von Kriterieninformation zu geben, die

sich im Wesentlichen an der Überlegenheit statistischer Vorgehensweisen orientieren (siehe dazu Wiggins, 1973 und Wittmann, 1985, S.292). Doch warum ist das klinische dem statistischen Urteil so unterlegen? Wittmann (1985, S.300ff.) listet tabellarisch mögliche Fehler bei der intuitiven Urteilsbildung auf, die ursächlich für die von Wiggins festgestellte Unterlegenheit sein können. *Abbildung 15* nennt diese. Für eine ausführliche Erklärung mit Beispielen sei auf Wittmann (1985, S.300ff.) verwiesen.

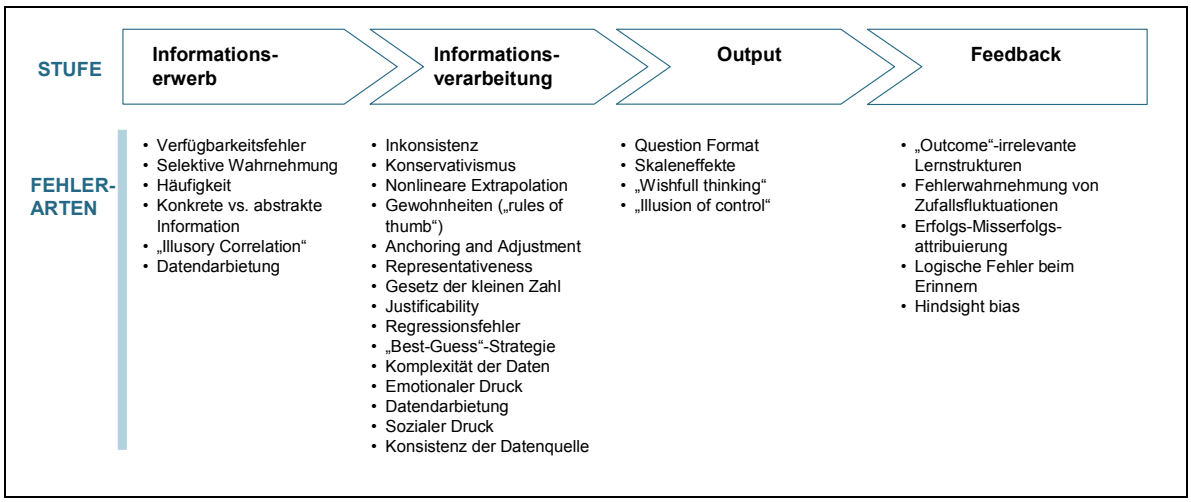

Abbildung 15: Urteilsfehler (in Anlehnung an Wittmann, 1985, S.300ff.)

Für die eignungsdiagnostische Praxis der Personalauswahl bleibt zu vermerken, dass die statistischen Methoden sich zwar als die Überlegenen gezeigt haben, diese aber wiederum das Vorhandensein großer Stichproben erfordern (Amelang & Zielinski, 2002, S.419). Dieser Hintergrund dürfte aber gerade bei kleineren Unternehmen oder Unternehmen mit geringem Personalbedarf nicht gegeben sein, so dass hier die Verwendung des klinischen Vorgehens mangels Alternativen vertretbar ist. Allerdings haben, wie oben erwähnt, zahlreiche metaanalytische Studien (z.B. Schmidt & Hunter, 1998) Hinweise auf die Generalisierbarkeit prognostischer Validitäten von vielen Prädiktoren geliefert. Hier könnten Ansatzpunkte für die Verwendung der statistischen Methode auch bei kleineren Stichproben vorzufinden sein. Doch selbst bei großen Unternehmen mit ausreichend großen Stichproben wird nach wie vor im Rechtfertigungskontext intuitive Information verarbeitet, statt auf statistische Prognosen zu vertrauen (Risch & Sommer, 1996, S.220).

Die Wahl des statistischen Prognoseverfahrens hängt nicht zuletzt, wie unter 2.2.6 zu sehen sein wird, auch mit der zugrunde liegenden Entscheidungsstrategie zusammen. Oftmals wird dabei in der Literatur zwischen dem Prognoseverfahren und der Strategie nicht unterschieden (z.B. Kompa, 1989; Guion, 1998). Nachfolgend sollen zunächst die statistischen Prognoseverfahren methodisch genauer betrachtet werden, bevor unter 2.2.6 auf die

Entscheidungsstrategien und das damit zusammenhängende Entscheidungsverhalten eingegangen wird.

2.2.5.2 Statistische Prognoseverfahren und methodische Perspektiven

Betrachtet man zunächst den bivariaten Fall, so zeigt *Abbildung 16* beispielhaft mögliche Zusammenhänge zwischen Prädiktor und Kriterium.

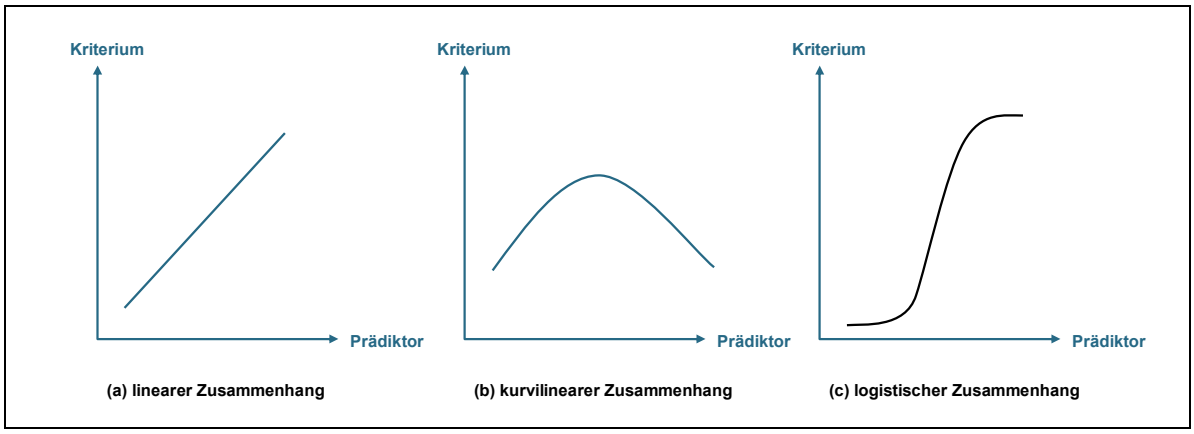

Abbildung 16: Beispiele bivariater Zusammenhänge von Prädiktor und Kriterium

In der personaldiagnostischen Praxis haben sich, wie auch in vielen anderen Bereichen, die linearen Modelle wie Fall (a) durchgesetzt. Neben der Robustheit linearer Verfahren gegenüber Verletzungen ihrer Prämissen und dem relativ seltenen Vorkommen nichtlinearer Zusammenhänge, gibt es noch weitere, eher technische Gründe für die Beliebtheit linearer Verfahren (Guion, 1998, S.300). Dennoch sollte man nicht blind auf Linearität vertrauen, sondern auch die Möglichkeit nichtlinearer Zusammenhänge bei den Analysen in Betracht ziehen, da diese durchaus vorkommen können (Ghiselli, 1964). Fall (b) kann beispielsweise dann auftreten, wenn Personen mit sehr hohen Prädiktorwerten aus Missgunst schlechtere Beurteilungen von ihrem Vorgesetzten bekommen, da sie eventuell besser oder intelligenter sind als er selbst. Fall (c), anlehnend an die PTT als *job characteristic curve* benennbar, wurde im Rahmen einer Untersuchung zum Berufserfolg von Raju, Steinhaus, Edwards und DeLessio (1991) erfolgreich angewandt.

Die einfachsten bivariaten linearen Zusammenhänge werden mit dem häufig verwendeten Produkt-Moment-Korrelationskoeffizienten von Pearson sowie der einfachen linearen Regression dargestellt. Der Korrelationskoeffizient ist dabei das Zusammenhangsmaß, das die Nähe der Kriteriumswerte an der einfachen linearen Regressionsgeraden quantifiziert. Eine Verletzung der Linearitäts-Prämisse führt zu einer Unterschätzung des Zusammenhangs

der Variablen (Guion, 1998, S.311). Das gleiche gilt für die Verletzung der Homoskedastizität der Residuen, d.h. der Gleichheit der Varianzen der Regressionsresiduen bei unterschiedlichen Abschnitten der Prädiktorenwerte (Cohen & Cohen, 1983, S.126ff.). Eine weitere Prämisse des Korrelations- und Regressionsansatzes ist die Unkorreliertheit der Messfehler (keine Autokorrelation), d.h., dass die Regressionsresiduen in der Grundgesamtheit unkorreliert sind. Autokorrelation tritt verstärkt bei Zeitreihenanalysen auf und führt zu Verzerrungen bei der Ermittlung des Standardschätzfehlers der Regressionskoeffizienten, der die Streuung der Kriterienwerte um die Regressionsgerade wiedergibt. Er verzerrt somit auch die Ermittlung von Konfidenzintervallen für die Regressionskoeffizienten (Backhaus, Erichson, Plinke & Weiber, 1996, S.39). Zusätzlich gehen die linearen Modelle von einer Normalverteilung der Störgrößen aus.

Weitere problematische Aspekte im Rahmen der Anwendung statistischer Verfahren bei der Personalauswahl gehen mit dem Signifikanztest einher. Die Nullhypothese geht dabei von der Annahme aus, dass es keinen Zusammenhang in der Population gibt, aus der die Stichprobe gezogen wurde, d.h., die Korrelation exakt Null ist. Guion (1998, S. 321) weist allerdings darauf hin, dass das Verwerfen der Nullhypothese nicht mit der Annahme verbunden sein kann, dass die Alternativhypothese mit der Höhe des errechneten Korrelationskoeffizienten für die Population gilt. Vielmehr wird dabei getestet, dass unter der Annahme nicht vorhandener Korrelation in der Population, eine Korrelation, die nicht Null ist, in der Stichprobe mit einer bestimmten Wahrscheinlichkeit zufallsbedingt auftritt (siehe dazu auch Cohen, 1990, S. 1308). Der in 2.2.4 beschriebene Selektionseffekt schränkt diese Interpretierbarkeit zusätzlich ein. Darüber hinaus kann das Verwerfen einer Annahme nicht als Negierung eines Zusammenhangs interpretiert werden, zumal das Fehlen eines Zusammenhangs sehr schwer zu beweisen ist (Guion, 1998, S. 322).

Auch die Setzung des Signifikanzniveaus für die Annahme oder Verwerfung der Nullhypothese ist zwar allgemein anerkannt bei 5% (signifikant), respektive 1% (sehr signifikant), dennoch in seiner Höhe willkürlich. Darüber hinaus beeinflusst die Setzung des Signifikanzniveaus nicht nur den Fehler 1. Art (α-Fehler, Verwerfung der Nullhypothese, obwohl diese richtig ist), sondern auch den gegenläufig abhängigen Fehler 2. Art (β-Fehler, Beibehaltung der Nullhypothese bei Richtigkeit der Alternativhypothese). Die Wahrscheinlichkeit, tatsächlich vorhandene Effekte aufdecken zu können (Teststärke $1-\beta$, siehe Cohen & Cohen, 1983), hängt zudem von der Stichprobengröße, der Effektstärke in der Population sowie dem α-Fehler-Niveau ab. Cohen (1988) ermöglicht mit seinen *power tables* die Festlegung von Stichprobengrößen in Abhängigkeit der genannten Größen, allerdings vor dem

Hintergrund perfekter Kriteriumsreliabilität und fehlender Varianzeinschränkung. Schmidt, Hunter und Urry (1976) liefern eine erweiterte Version unter Berücksichtigung dieser beiden Aspekte. Die geschilderten Aspekte beziehen sich auf den auf kontinuierlichen, intervallskalierten Variablen basierenden Fall. Bei fehlender Notwendigkeit differenzierter Vorhersagen auf einem Kontinuum, also dem Vorliegen kategorialer Variablen, können die biseriale Korrelation oder der tetrachorische ϕ-Koeffizient Anwendung finden. Einsatzmöglichkeiten solcher Vorgehensweisen im Auswahlkontext zeigt Hollmann (1991).

Zur Erklärung methodischer Prämissen wurde der Einfachheit halber zunächst auf den bivariaten Fall eingegangen. Eine Kombination von Prädiktordaten führt aber logischerweise zu einer multivariaten Betrachtung. Dabei kann man *kompensatorische* und *nicht-kompensatorische* Modelle unterscheiden. Erstere gehen wiederum von der Linearität aus und ermöglichen den Ausgleich (Kompensation) von niedrigen Werten auf einem Prädiktor durch bessere Werte auf einem anderen Prädiktor. Dies ist auch gleichzeitig ihr größter Kritikpunkt im Kontext der Personalauswahl, da bei Stellenbesetzungen oftmals Mindestausprägungen auf unterschiedlichen Prädiktoren Voraussetzung sind. So kann beispielsweise ein Unternehmensberater mangelnde Rechenkenntnisse nicht durch exzellente verbale Fähigkeiten ausgleichen, wenn er in einem Projekt einen Geschäftsplan rechnen soll. *Tabelle 5* gibt eine kurze Übersicht ausgewählter kompensatorischer Prognoseverfahren. Für eine detaillierte Beschreibung sei auf die einschlägige Literatur (Cohen & Cohen, 1983; Bortz, 1993; Backhaus, et al., 1996) verwiesen.

Tabelle 5: Ausgewählte kompensatorische Prognoseverfahren

Prognoseverfahren	**Inhalt**
Einfaches additives Modell	Addition oder Durchschnittsbildung der einzelnen Ausprägungen auf den Prädiktoren. Ausgangspunkt ist die Gleichgewichtung aller Prädiktoren.
Multiple Regression	Addition gewichteter Prädiktorenwerte als Schätzung der Kriteriumsausprägung. Gewichtungskoeffizienten werden mathematische nach dem Kriterium der kleinsten Quadrate ermittelt, d.h. nach der Minimierung der quadrierten Abweichung der Prognose vom tatsächlichen Kriteriumswert.
Profilmethode	Vergleich zweier Profile (Bewerberprofil mit Idealprofil) hinsichtlich ihrer Ähnlichkeit. Angewandte Methoden sind neben dem aufwendigen Centour-Verfahren Distanzmaße oder Verlaufsähnlichkeit.

Die multiple Regression ist das gängigste angewandte Verfahren, für das die im vorherigen Abschnitt besprochenen methodischen Prämissen der einfachen Regression ebenso gelten. Multiple Regressionen können als simultane, schrittweise oder hierarchische Variante durchgeführt werden. Während die simultane Variante alle signifikanten Prädiktoren gleich-

zeitig in die Regressionsgleichung aufnimmt, werden bei der schrittweisen Regression Prädiktoren sequentiell nach ihrer maximalen inkrementellen Varianzaufklärung am Kriterium aufgenommen[7] (Bortz, 1993, S.427). Im Kontext des Entdeckungszusammenhangs ist dieses Vorgehen sehr beliebt, obwohl die Interpretation der Ergebnisse hinsichtlich möglicher kausaler Zusammenhänge sehr schwer ist. Zudem ist bei einer großen Anzahl an Prädiktorvariablen, auf Grund der separaten Testung jedes einzelnen Prädiktors, eine Kapitalisierung des Zufalls, also eine α-Fehler-Inflation, nicht auszuschließen (Cohen & Cohen, 1983, S.124). Die hierarchische Regressionsvariante geht von einer vorhandenen Forschungslogik bei der Auswahl der Prädiktoren aus und gibt somit die Reihenfolge der Prädiktoren kausalitäts- und theoriegeleitet vor. Nur sie ermöglicht eine Zuordnung aufklärender Varianzanteile zu den einzelnen Prädiktoren (*variance partioning*), da die Position eines Prädiktors in der Sequenz die Höhe seines inkrementellen Varianzanteils erheblich beeinflusst. Allerdings liegt die Schwierigkeit hier gerade darin, geeignete theoretische Gerüste zu finden, die eine alleingültige, kausale Sequenz ermöglichen (Guion, 1998, S.343).

Eine wichtige Voraussetzung für die Signifikanzprüfung multipler Regressionen ist die multivariate Normalverteilungsannahme für alle beteiligten kontinuierlich verteilten Variablen, wenn der Stichprobenumfang *N* nicht wesentlich größer als die Variablenanzahl *k* ist ($N>40$ bei $k<10$, siehe Bortz, 1993, S.417). Wichtig ist zudem die Vollständigkeit des Modells (Backhaus, et al., 1996, S.37). Die Vernachlässigung relevanter Variablen *(underfitting)* führt zu nicht zufälligen Schwankungen in den Störgrößen. Ein Einbeziehen zu vieler Variablen (*overfitting*) führt wiederum zu Zufallskorrelationen und ineffizienten Schätzungen. Darüber hinaus sollte keine *Multikolinearität* vorliegen, d.h. eine hohe wechselseitige Abhängigkeit der Prädiktoren, die sich in Form hoher Interkorrelation zeigt. Diese führt bei Nichtbeachtung zu einer Missinterpretation der Regressionsgewichte im simultanen Regressionsfall. *Multikolinearität* gefährdet die rechnerische Genauigkeit der Regressionsgewichte (Cohen & Cohen, 1983, S.116), erschwert ihre Interpretation und kann zu Verzerrungen in den Teststatistiken führen. Ein Ausweg bietet hierbei die Verwendung von Strukturkoeffizienten, als Division der Einzelkorrelation der Variablen mit der multiplen Korrelation (Bortz, 1993, S.420). Liegt die Vermutung eines die hohe Korrelation bedingenden, dahinter liegenden Konstruktes nahe, so kann man sich auch durch faktoranalytische Ansätze behelfen und Variablen zu einem Faktor zusammenfassen. Auch die Anwendung einer hierarchischen Regression kann, bei Angemessenheit der hierarchischen Sequenz, helfen (Cohen &

[7] Dies gilt für die Vorwärtsvariante, bei der Rückwärtsvariante werden zunächst alle Prädiktoren in die Regressionsgleichung aufgenommen und Variablen nach geringster Varianzaufklärung sukzessive herausgenommen.

Cohen, 1983, S.115). Einen Ansatz, um die Stabilität von Regressionsgewichten zu überprüfen, ist die *Kreuzvalidierung*, bei der die an einer Stichprobe gewonnene Regressionsgleichung an einer anderen Stichprobe zur Vorhersage verwendet wird (Bortz, 1993, S.420). Multiple Regressionen sollten immer kreuzvalidiert werden, bevor auf ihrer Grundlage Entscheidungen getroffen werden (Guion, 1998, S.353).

Tabelle 6: Ausgewählte nichtkompensatorische Prognoseverfahren

Prognoseverfahren	**Inhalt**
Multiple Schwellenwerte	Für jeden Prädiktor werden Schwellenwerte festgelegt, die definieren, ab welcher Prädiktorausprägung ein erfolgreiches Kriteriumsverhalten zu erwarten ist. Der Bewerber wird nur dann akzeptiert, wenn er auf oder über allen Schwellenwerten liegt.
Sequentielle Hürden	Bewerber müssen bestimmte Hürden überspringen, um in die endgültige Auswahlsituation zu gelangen. Die Schwellenwerte je Hürde können durch einen festen Ausprägungsgrad oder durch einen Anteil an Kandidaten, die die nächste Stufe erreichen sollen bestimmt sein.
Konjunktive und disjunktive Verfahren	Konjunktive Verfahren orientieren sich an der schlechtesten Prädiktorausprägung aus allen Prädiktorenausprägungen eines Bewerbers und basieren darauf die Entscheidung. Disjunktive Verfahren orientieren sich demgegenüber an dem besten Prädiktorwert.

Ein weiteres, nicht selten auftretendes Phänomen im Kontext multipler Regression stellen *Suppressoreffekte* dar. Sie tauchen häufig bei aggregierten Daten auf. Suppressoren im Rahmen der Personalauswahl sind Variablen, die nicht mit dem Kriterium, wohl aber mit der Prädiktormessung korrelieren und somit irrelevante Varianz dem Prädiktor hinzufügen. Dies verringert die Korrelation des Prädiktors mit dem Kriterium, d.h. der eigentliche Zusammenhang wird verdeckt (zu den unterschiedlichen Suppressorkonstellationen siehe Cohen & Cohen, 1983, S.95f). Die Aufnahme des Suppressors in die Regressionsgleichung unterdrückt die irrelevanten Varianzanteile des Prädiktors und erhöht somit seine Varianzaufklärung am Kriterium. Ähnliches gilt für die bereits angesprochenen Moderatoreffekte. Moderatoren sind Variablen, die mit der Korrelation zweier Variablen korrelieren, d.h. ihren Zusammenhang beeinflussen (Guion, 1998, S.345). Zum Nachweis des Moderatoreffektes kann man im Rahmen sog. *moderierter multipler Regression* neben dem Prädiktor und dem Moderator eine weitere, aus dem Produkt dieser beiden gebildete Variable in die Regression einbeziehen. Die Größe und das Vorzeichen dieser neuen Prädiktoren geben dann Auskunft über die moderierende Wirkung (Bortz, 1993, S.428).

Kompensatorische Modelle wie die multiple Regression gehen also von der Möglichkeit aus, Schwächen in einem Bereich durch Stärken in einem anderen Bereich auszugleichen. Nichtkompensatorische Modelle lassen eine solche Möglichkeit nicht zu. *Tabelle 6*

gibt eine Übersicht. Die genannten nichtkompensatorischen Prognoseverfahren kann man auch als Entscheidungsstrategien auffassen und werden deshalb im nächsten Kapitel genauer betrachtet.

Ein letztes zu erwähnendes Verfahren, mit dessen Hilfe Zusammenhänge zwischen Prädiktoren und Kriterien hinsichtlich Kausalität analysiert werden kann, stellt der SEM (**S**tructural **E**quation **M**odelling)-Ansatz dar. Er ermöglicht die Analyse der Beziehungen von latenten, d.h. nicht beobachteten Variablen oder auch hypothetischen Konstrukten. Dabei werden zunächst in einem Strukturmodell auf Grund hypothetischer Überlegungen aufgestellte Beziehungen zwischen diesen Konstrukten abgebildet. Im Anschluss wird ein Messmodell für die endogenen (zu erklärenden) und exogenen (als gegebene Größe akzeptierte) Variablen, auch Indikatorvariablen genannt, spezifiziert. Dieses enthält empirische Indikatoren für die latenten Größen (Backhaus, et al., 1996, S.394). Der SEM-Ansatz führt nun eine Kausalanalyse auf der Ebene von aggregierten Daten (Kovarianz- und Korrelationsdaten zwischen den Indikatorvariablen) durch und überprüft mit Hilfe des Instrumentariums der Regressionsanalyse, ein gegebenes Hypothesensystem in seiner Gesamtheit. Das Vorliegen genau spezifizierter, theoretisch hergeleiteter Hypothesen ist eine wichtige Voraussetzung für die Kausalitätsprüfung. Fehlen diese, so handelt es sich um einen explorativen Ansatz. Gewöhnlicherweise wird der Gesamtzusammenhang (die Struktur) durch ein Pfaddiagramm dargestellt.

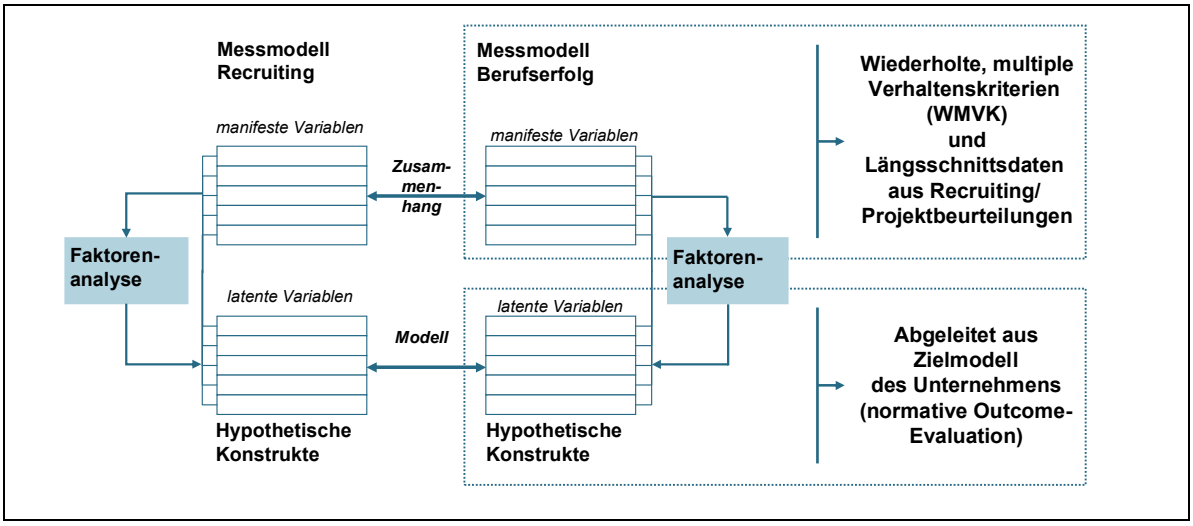

Abbildung 17: Beispiel für ein lineares Strukturgleichungsmodell

Die dabei formulierten Gleichungen bezeichnet man deshalb als Strukturgleichungen. Bei dem regressionsanalytischen Sonderfall der Pfadanalyse werden standardisierte und unstandardisierte Pfadkoeffizienten ermittelt. Mit Hilfe des sog. *Fundamentaltheorems der*

Pfadanalyse lässt sich dann die Korrelation zweier Variablen in eine kausale und eine nichtkausale Komponente zerlegen. Die kausalen Effekte lassen sich wiederum als direkte oder indirekte (d.h. über eine andere Variable gehende) Effekte darstellen. Pfadanalysen werden bei Personalauswahlverfahren üblicherweise dann angewandt, wenn man herausfinden will, wie einzelne Bestandteile oder Elemente des Auswahlverfahrens mit der späteren Leistung zusammenhängen (Guion, 1998, S.359). Die regressionsanalytischen Ansätze werden gleichermaßen für die latenten (nicht beobachtbaren) Variablen appliziert. Voraussetzung hierfür ist allerdings, dass die latenten Variablen aus den Indikatorvariablen erschlossen werden. Dies wird im LISREL-Ansatz mit Hilfe der Faktorenanalyse erreicht. Dabei stellen die Faktorladungen die standardisierten Regressionskoeffizienten der Indikatoren auf die latenten Variablen dar. *Abbildung 17* zeigt abschließend das Vorgehen linearer Strukturgleichungsmodelle am Beispiel eines Personalauswahlfahrens.

2.2.6 Entscheidungsmodelle und Ergebnisse der Auswahlentscheidung

In dem vorherigen Abschnitt wurde auf statistisch-methodische Besonderheiten ausgewählter prognostischer Verfahren im Kontext der Personalauswahl eingegangen. Diese prognostischen Verfahren liefern Informationen, auf deren Basis Entscheidungen über die Zuordnung der Bewerber zu einer Intervention (z.B. Annahme oder Ablehnung) getroffen werden. Cronbach und Gleser (1965, S.16ff.) unterscheiden dabei Dimensionen, nach denen Entscheidungsarten klassifiziert werden können (*Tabelle 7*).

Tabelle 7: Entscheidungsarten nach Cronbach & Gleser (1965, S.16ff.)

Dimension	**Entscheidungsarten**
Nutzen der Entscheidung geht zugunsten von	Institution vs. Individuum
Annahme	festgelegt vs. variabel
Behandlung	singulär vs. multipel
Möglichkeiten von Ablehnung	ja vs. nein
Informationsdimension	univariat vs. multivariat
Entscheidung	terminal vs. investigatorisch

Im Falle der Personalauswahl, wie sie in der vorliegenden Arbeit untersucht wird, handelt es sich um eine Entscheidungsart der Institution mit einer festgelegten Annahmequote und einer multiplen Behandlung (im Sinne einer mehrstufigen Testung). Dabei werden terminale Entscheidungen mit Möglichkeiten der Ablehnung auf Basis multivariater Infor-

mationen getroffen. Es gibt dabei verschiedene formelle Regeln oder auch Entscheidungsstrategien, mit denen man zu einer Entscheidung gelangen kann (Wiggins, 1973, S.226). Grundsätzlich kann man dabei zwei, im vorherigen Kapitel bereits angesprochene Dimensionen unterscheiden: kompensatorische vs. konjunktive, sowie ein- vs. mehrstufige Strategien. Kompensatorische Strategien, wie sie bei linearen Regressionen Ausgestaltung finden, werden in der Praxis sehr häufig angewandt (Amelang & Zielinski, 2002, S.437). In ihrer kombinatorischen Variante werden die Summen aus Teilkompetenzen gebildet. Beim „Oder-Konzept" genügt ein entsprechend hoher Wert auf einem Prädiktor. Kompensatorische Strategien sind immer dann dysfunktional, wenn, wie oben erwähnt, Mindestausprägungen auf unterschiedlichen Prädiktoren erforderlich sind. Liegt so ein Fall vor, sind so genannte „Und-Strategien" oder auch konjunktive Entscheidungsstrategien, wie das Prognoseverfahren der multiplen Schwellenwerte, notwendig.

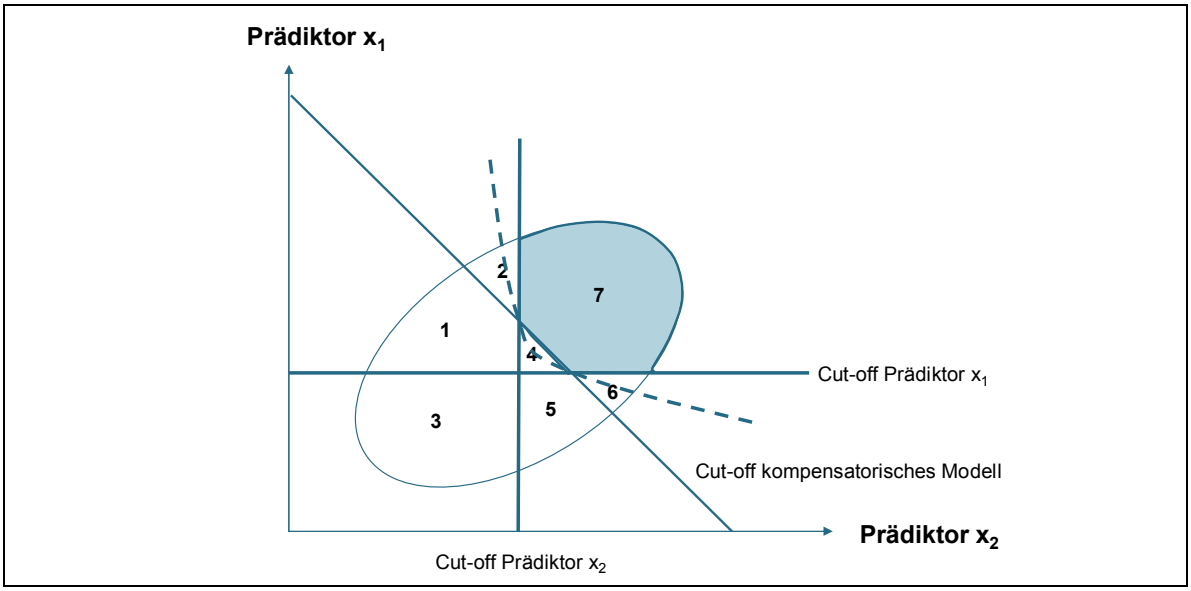

Abbildung 18: Konjunktive und kombinatorisch-kompensatorische Strategie (in Anlehnung an Kompa, 1989, S. 88)

Vergleicht man die konjunktive mit der kombinatorisch-kompensatorischen Strategie am Beispiel zweier Prädiktoren (*Abbildung 18*), so zeigen sich deutliche Unterschiede im Entscheidungsverhalten. Bei beiden Strategien werden Bewerber aus den Feldern 1, 3 und 5 abgelehnt und aus dem Feld 7 akzeptiert. Während man aber bei der konjunktiven Strategie die Bewerber aus den Feldern 2 und 6 ebenfalls ablehnt und jene aus dem Feld 4 akzeptiert , ist dies bei der kombinatorisch-kompensatorischen Strategie genau umgekehrt. Dies verdeutlicht, dass Erstere mittelmäßig und Letztere einseitig befähigte Mitarbeiter in Kauf nimmt (Kompa, 1989, S.89). Die gestrichelte Linie stellt einen Kompromiss zwischen beiden Vor-

gehensweisen dar (partiell kompensatorische Strategie). Das Hauptproblem der konjunktiven Strategie liegt in der Festlegung der Höhe der Trennwerte, da es hier keine eindeutigen Verfahren gibt. Zudem ist dieser Ansatz nur vor dem Hintergrund perfekt reliabler Prädiktoren vertretbar (Guion, 1998, S.349).

Einstufige Entscheidungsvorgehen finden in der Praxis der Personalauswahl nicht selten statt. Bei anspruchsvollen Tätigkeiten wird man aber i.d.R. mindestens zwei Stufen haben: die Vorauswahl anhand von Bewerbungsunterlagen und das Auswahlverfahren selber, welche beide natürlich auch wiederum in unterschiedliche Sequenzen unterteilt sein können. Ein einstufiges Verfahren ist die nichtsequentielle Batterie, bei dem die gesamte Batterie (z.B. ein Intelligenztest) an alle Bewerber vorgegeben wird und jene ausgewählt werden, die in dem optimal gewichteten Summenwert, die höchste Ausprägung haben. Beim „single-screen"-Vorgehen fußt die Entscheidung auf einem einzigen Test (Amelang & Zielinski, 2002, S. 240).

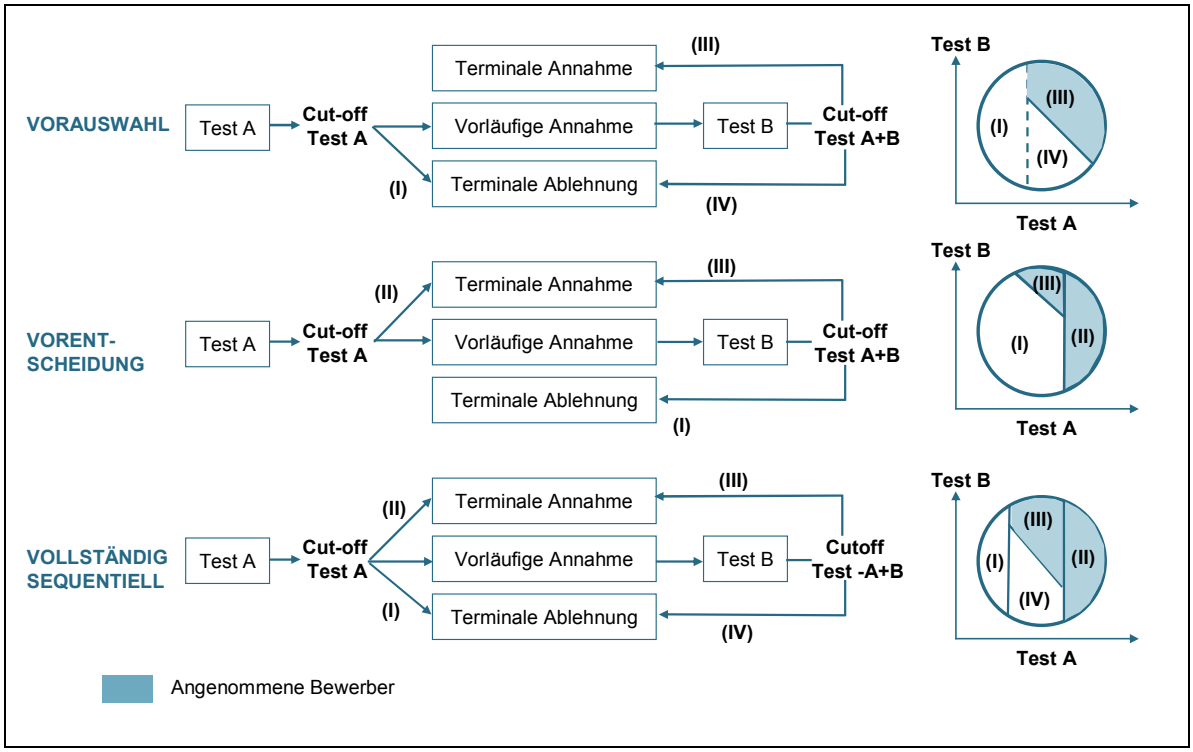

Abbildung 19: Sequentielle Strategien (in Anlehnung an Cronbach & Gleser, 1965)

Sequentielle Hürden werden oftmals eingesetzt, wenn man die Anzahl der Bewerber reduzieren will, z.B. bei kostspieligen Verfahren. Eine klassische sequentielle Hürde ist die Vorselektion nach Bewerbungsunterlagen. Zwar ist die Genauigkeit von sequentiellen Prognoseverfahren höher, die Prädiktorenvarianz verringert sich allerdings mit zunehmender Stufe, so dass eine Korrekturformel angewandt werden muss, wenn die Validität der zeitlich

später zutreffenden Prädiktoren abgeschätzt werden soll. Zudem sind sie aufwendiger und teurer als nichtsequentielle Verfahren (Kompa, 1989, S. 93). Sequentielle Verfahren unterscheidet man in drei Grundmuster (Wiggins, 1973, S. 262ff.). *Abbildung 19* visualisiert die drei Vorgehensweisen und die Auswirkung der Entscheidung anhand zweier Prädiktoren.

Wiggins (1973) vertritt die Ansicht, vollständig sequentielle Strategien seien den anderen Strategien meistens überlegen, vor allem, wenn man die Validität und die Kosten des Verfahrens betrachte. So müssen bei sequentiellen Verfahren nicht alle Bewerber sämtliche Tests durchlaufen und die Validität des ersten Tests wird vollkommen ausgeschöpft. Neben der Validität spielt aber auch die Anzahl ausgewählter Bewerber eine Rolle. Cronbach und Gleser (1965, S.77ff.) weisen darauf hin, dass bei extremen Selektionsraten die Überlegenheit sequentieller Strategien abnimmt. Das Ergebnis eines solchen „Strategielaufes“ ist eine Zuordnung der Personen zu einer Merkmalsklasse (Annahme vs. Ablehnung) (Fisseni, 1990, S.243). Auch sequentielle Entscheidungsstrategien lassen sich auf eine Folge einstufiger Klassenzuordnungen reduzieren (Amelang & Zielinski, 2002), weshalb auch im Folgenden die einstufige Betrachtungsweise zu Erklärungszwecken herangezogen wird. *Abbildung 20* zeigt beispielhaft die Klassenzuordnung von drei Auswahlverfahren.

Obwohl in Rahmen der Personalauswahl bisher immer von der *Selektion* die Rede war, so handelt es sich im vorliegenden Fall um eine *Klassifikation*, da die Bewerber als geeignet oder nicht geeignet eingestuft bzw. klassifiziert werden. Im Rahmen der Beurteilung von Entscheidungsergebnissen ist die Betrachtung aller Klassifikationsfelder von entscheidender Bedeutung. Von *Selektion* spricht man hingegen dann, wenn es gilt, für eine dieser Merkmalsklassen den geeigneten Bewerber zu finden. Insofern sind *Selektion* und *Klassifikation* nicht disjunkt trennbar (Fisseni, 1990, S.247). Cronbach und Gleser (1965) sehen in der Selektion auch einen Sonderfall der Klassifikation, bei der Ablehnung als Möglichkeit der Entscheidung zugelassen wird. Die dargestellte Klassifikation basiert auf Cronbach und Glesers (1965) im Rahmen der Entscheidungstheorie entwickelter Vierfelder-Matrix. Die innerhalb der Bewerberausprägungen dargestellten Klassen stellen die möglichen Fehler bei der Klassenzuordnung dar. Ein Fehler der ersten Art sind die falsch Positiven (FP), d.h. Bewerber, die zu Unrecht eingestellt wurden. Der Fehler 2. Art sind die falsch Negativen (FN), d.h. Bewerber, die zu Unrecht abgelehnt wurden. Bei den verbleibenden zwei Klassen, den richtig Positiven (RP) und richtige Negativen (RN), hat keine Falschzuordnung stattgefunden. Die Flächen dieser beiden Klassen gilt es nun zu maximieren, die der anderen beiden zu minimieren. Im linken Beispiel haben wir ein Auswahlverfahren, das gar keine Validität besitzt. Dementsprechend ist die Bewerberverteilung kreisförmig. Die Auswahl entspricht

daher dem Zufallsprinzip. In der mittleren Abbildung wird ein Verfahren mit einer Validität größer als Null angewandt. Dementsprechend sind die Flächen der falsch zugeordneten Bewerber kleiner. Die rechte Abbildung zeigt gegenüber der mittleren Abbildung wiederum eine Validitätsverbesserung.

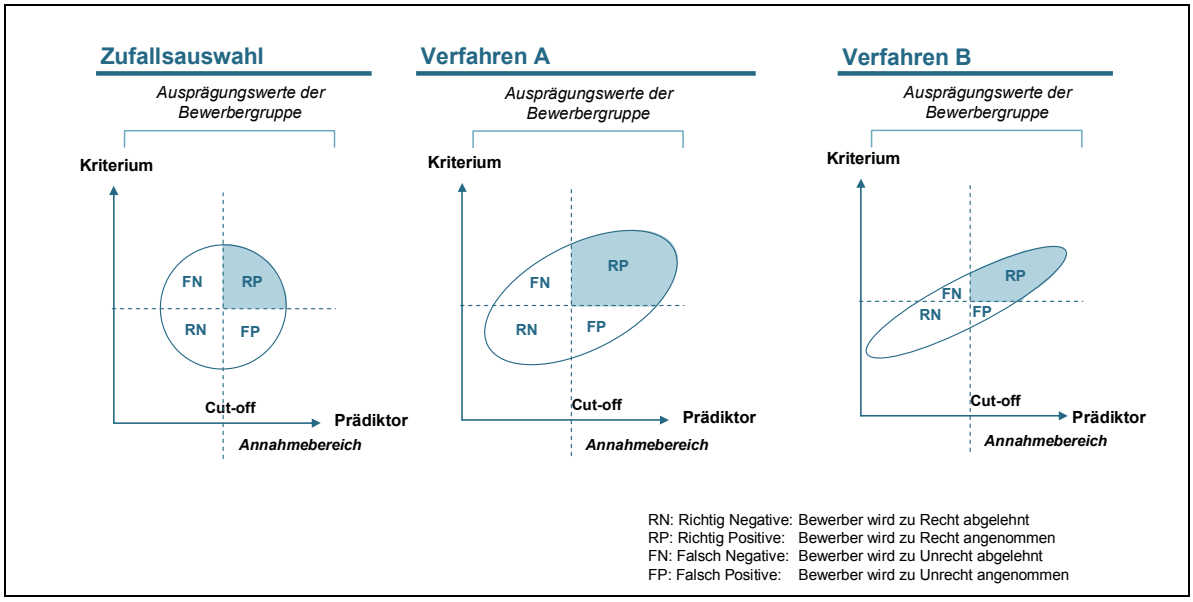

Abbildung 20: Klassenzuordnungen bei der Personalauswahl (eigene Darstellung)

Um unterschiedliche Verfahren und Untersuchungsergebnisse miteinander vergleichen zu können, berechnet man gängigerweise Wahrscheinlichkeiten für die einzelnen Felder. *Abbildung 21* zeigt die Berechnung dieser Wahrscheinlichkeiten. Der Begriff der *Basisrate (BR)* meint dabei, mit welcher Wahrscheinlichkeit ein Bewerber aus der Gesamtpopulation der Bewerber geeignet ist, und wird auch *natürlicher Eignungsquotient* genannt. Die *Selektionsrate (SR)* ist die Wahrscheinlichkeit, mit der eine Person ausgewählt wurde. Die dargestellte Klassifikation beruht auf einer an Trennwerten von Prädiktor und Kriterium vorgenommenen Dichotomisierung. Gängigerweise kann man den Zusammenhang zwischen dichotomen Prädiktor und Kriterium anhand des Gütemaßes des Korrelationskoeffizienten ϕ (s .o.) berechnen. Diese Perspektive aus der klassischen Testtheorie, die auf Messung und Vorhersage fokussiert, greift aber im Rahmen der Entscheidungstheorie zu kurz, da sie keine Information über die Wahrscheinlichkeiten in den Feldern beinhaltet. Noack und Petermann (1992, S. 299) nennen einige spezifischere Gütekriterien, die dieser Forderung entgegenkommen und in *Tabelle 8* dargestellt sind.

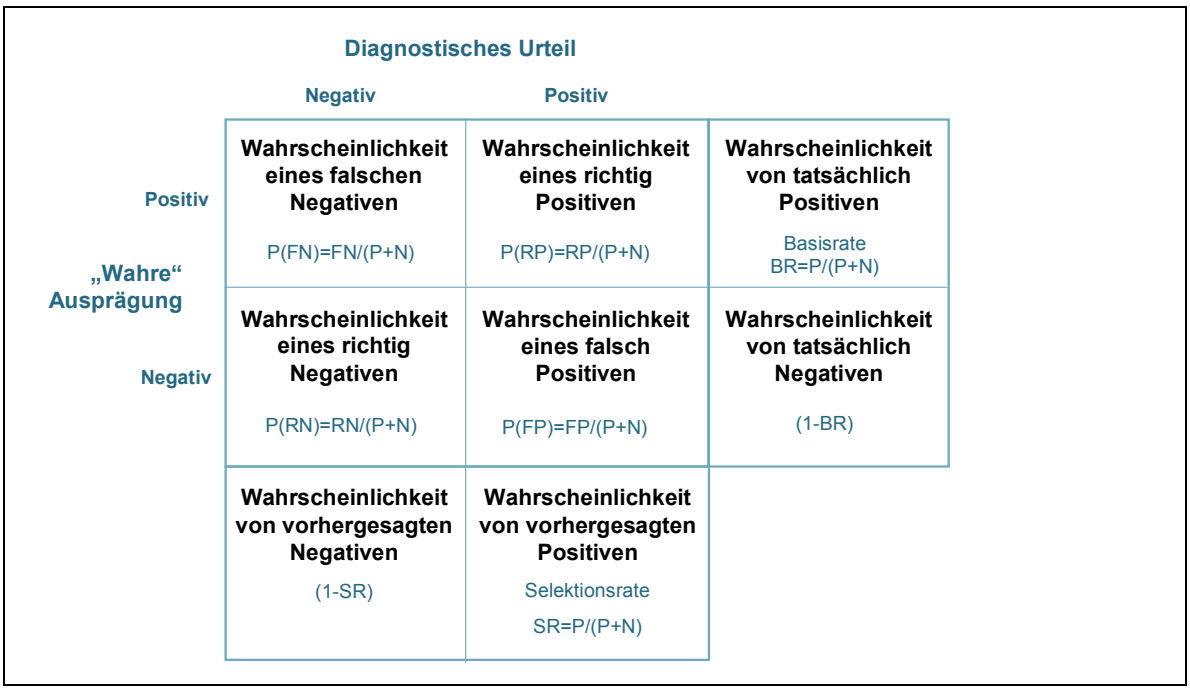

Abbildung 21: Erweiterte Vierfeldertafel (in Anlehnung an Wittmann, 1985, S.391)

Tabelle 8: Gütekriterien diagnostischer Entscheidungsstrategien (in Anlehnung an Noack & Petermann, 1992, S. 299)

Gütekriterium	**Inhalt**
Sensitivität	Wahrscheinlichkeit, mit der ein vorliegender positiver Zustand als solcher erkannt wird (RP/(RP+FN))
Spezifität	Wahrscheinlichkeit, mit der ein vorliegender negativer Zustand als solcher erkannt wird (RN/(RN+FP))
Positiver Vorhersagewert	Wahrscheinlichkeit, mit der eine positive Diagnose zutreffend ist (RP/(RP+FP))
Negativer Vorhersagewert	Wahrscheinlichkeit, mit der eine negative Diagnose zutreffend ist (RN/(RN+FN))

Die Basisrate und die Selektionsrate können die Bedeutung der Validität für die Sicherheit, mit der Entscheidungen getroffen werden, beeinflussen, denn das Treffen richtiger Entscheidungen hängt nicht alleine von der Validität des Instrumentes ab. *Abbildung 22* verdeutlicht den Einfluss der Selektionsraten und der Basisrate ausgehend von einer gegebenen Validität. Die Abbildung zeigt, dass es bei niedriger Selektionsrate auch interessant sein kann, weniger valide Auswahlinstrumente einzusetzen, da diese trotzdem zu effizienten Entscheidungen führen können. Effizient meint in diesem Zusammenhang, dass Bewerber ausgewählt werden, die später auch erfolgreich sind (Im Rahmen der Nutzenberechnung wird allerdings erläutert, wie auch die FN dem Unternehmen Probleme bereiten können). Die Basisrate in der Bewerberpopulation hingegen sollte aus Sicht des diagnostischen Instrumenta-

riums durch geeignete Maßnahmen (z.B. Personalmarketing) so beeinflusst werden, dass sie eine mittlere Größe besitzt, da unter diesen Umständen das diagnostische Verfahren den größten Nutzen bringt[8] (Kompa, 1989, S.102f.).

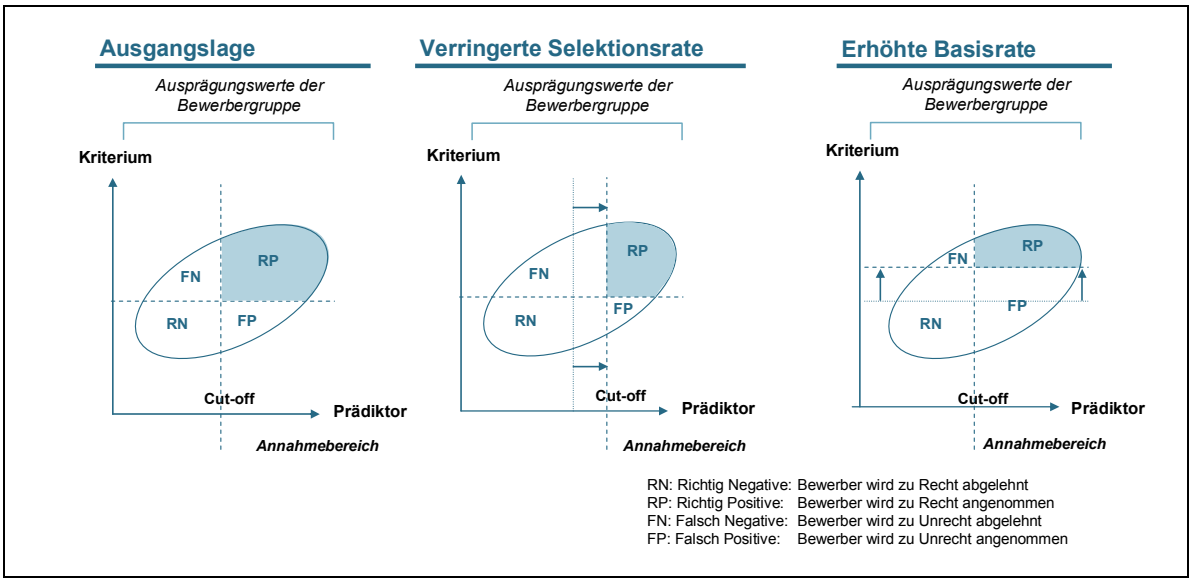

Abbildung 22: Einfluss von Basisrate und Selektionsrate bei gegebener Validität (in Anlehnung an Kompa, 1989)

Mathematisch lässt sich der Zusammenhang zwischen dem tetrachorischen ϕ-Koeffizienten, SR und BR wie folgt darstellen:

$$\phi = \frac{P(RP) - BR\,SR}{\sqrt{BR\,(1-BR)\,SR\,(1-SR)}} \tag{12}$$

Die P-Werte geben die Wahrscheinlichkeiten in den unterschiedlichen Feldern an. Durch Auflösung von (12) lassen sich, bei Kenntnis von SR und BR, die Wahrscheinlichkeit P(RP) berechnen und die fehlenden drei Zellen additiv ableiten (Wiggins, 1973, S.247f.). Taylor und Russel (1939) haben einen Ansatz entwickelt, mit dem man das Zusammenwirken von Basisrate, Selektionsrate und Validität von Prädiktoren beschreiben kann. Ausgangspunkt ihres Konzeptes ist der Zuwachs an korrekten Ergebnissen bei Einführung eines Verfahrens im Vergleich zur Zufallsauswahl. Korrekte Ergebnisse werden als Bewerber, die nach der Einstellung erfolgreich sind, verstanden. Die Basisrate ist dabei der Anteil der erfolgreichen unter den ausgewählten Mitarbeitern, d.h. der positive Vorhersagewert oder auch Erfolgsquote. Taylor und Russel haben das Ausmaß dieser Erfolgsquote eines neuen Verfah-

[8] Dieser Ansicht von Kompa muss insofern widersprochen werden, als dass das Hauptanliegen des Diagnostikers nicht in einem optimalen Einsatz eines Verfahrens liegt, sondern in einer optimalen Auswahl, so dass möglichst hohe Basisraten wünschenswert sind.

rens in Abhängigkeit seiner Selektions-, Basisrate und Validität in ausführlichen Tabellenwerken dargestellt und somit eine Beurteilung der Effektivität einer Entscheidungsstrategie ermöglicht. *Abbildung 23* verdeutlicht die Zusammenhänge. Dabei zeigt sich, dass man bei niedriger Selektionsrate, auch mit Verfahren mäßiger Validität eine inkrementelle Prognosegüte erreichen kann. Hohe Selektionsquoten hingegen machen auch hoch valide Testverfahren unwirksam. Hier spiegelt sich der Effekt der Qualitätsverluste beim Humanpotenzial, der zu Zeiten der „War for Talents" zu bemängeln war, wider. Zudem ist die Zunahme der Erfolgsquote mit abnehmender Selektionsquote bei Validitäten unter r=.70 positiv beschleunigt. Außerdem führt eine Steigerung der Validität um einen bestimmten Betrag zu einer unterschiedlichen Änderung der Erfolgsquotensteigerung in Abhängigkeit der Selektionsrate. Allerdings beschränken die Annahme der Linearität von Prädiktor und Kriterium sowie die Unabhängigkeit des neuen Prädiktors von den bestehenden Prädiktoren, die Interpretierbarkeit der Ergebnisse (Kompa, 1989, S. 104).

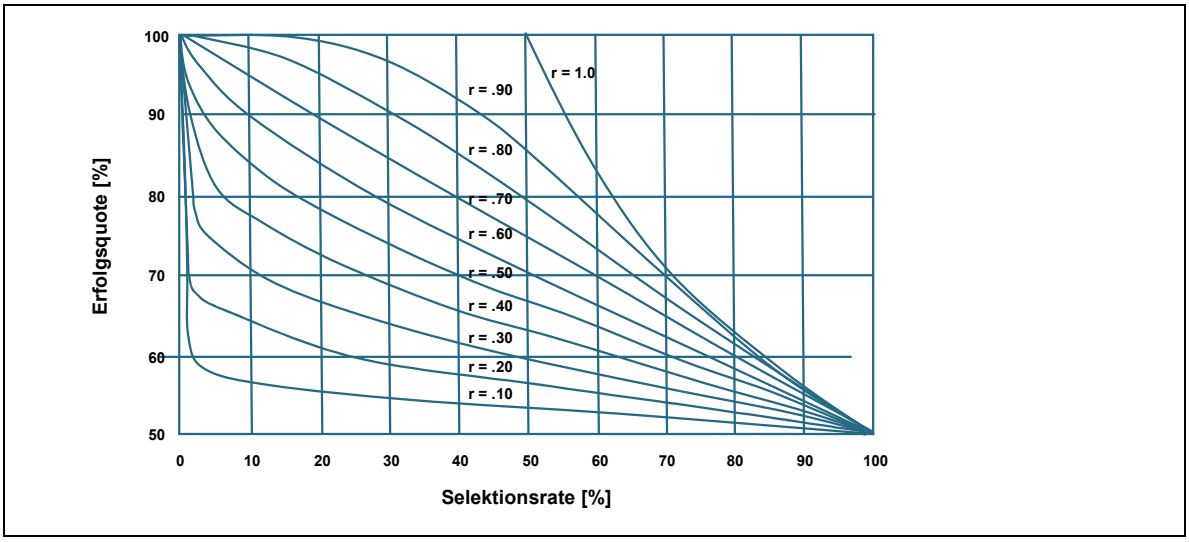

Abbildung 23: Wert eines Testverfahrens in Abhängigkeit situativer Parameter (Taylor & Russel, 1939, S.569)

Wie gezeigt betrachten die besprochenen Gütekriterien aus *Tabelle 6* sowie der Validitätskoeffizient ϕ oder der Produkt-Moment-Korrelationskoeffizient *r*, nur Teilaspekte der Vierfeldermatrix aus *Abbildung 20.* Zu dieser Erkenntnis kommt auch Hollmann (1991, S.61):

> „Zusammenfassend lässt sich nach diesen Darstellungen festhalten, dass die in einem Korrelationskoeffizienten ausgedrückte Kriteriumsvalidität als allgemeine Eigenschaft eines Tests allenfalls eine mittlere Größe bei verschiedenen Cut-off-Werten bzw. Selektionsquoten darstellen kann, die selbst bei korrekter Berechnung durch die biseriale Rangkorrelation und nicht durch die Produkt-Moment oder punkt-biseriale Korrelation, für den Anwender von äußerst begrenzter Relevanz ist."

Auch die Sicht von Taylor und Russel beschränkt sich auf eine Optimierung des positiven Prädiktionswertes, da sie davon ausgeht, dass ein hoher Anteil erfolgreicher unter den ausgewählten Mitarbeitern oberste Priorität besitzt. Somit werden lediglich die Ergebnisse der Annahme von Bewerbern berücksichtigt und die der Ablehnung ignoriert, da sie keinen unmittelbar spürbaren Einfluss auf die Produktivität besitzen. Bezieht man allerdings die Bewerbersicht in die Betrachtung mit ein, so kann ein hoher positiver Prädiktionswert dysfunktional sein, wenn er auf Kosten des negativen Prädiktionswertes entsteht. In einem solchen Fall würde die Anzahl an FN ebenfalls zunehmen. Dies bedeutet, dass ein „ganzheitlicher" Ansatz den gesamten Zuordnungsfehler in allen Feldern möglichst gering halten muss. Zu diesem Zweck werden die Randwahrscheinlichkeiten aus *Abbildung 21* herangezogen, die die absoluten Wahrscheinlichkeiten in dem Vierfelderschema an den jeweiligen Grundraten relativieren. Für die Minimierung des Zuordnungsfehlers bieten sich mehrere Entscheidungsregeln an, die in *Tabelle 9* beschrieben sind.

Tabelle 9: Entscheidungsregeln zur Minimierung des Zuordnungsfehlers (in Anlehnung an Amelang & Zielinski, 2002, S.446f.)

Gütekriterium	**Inhalt**
Neyman-Pearson-Kriterium	Anwendung des Modells der statistischen Hypothesenprüfung auf die Klassenzuordnungsentscheidung. Verschiebung des Entscheidungskriteriums nach Maßgabe eines festgelegten α-Niveaus. Anwendbar bei geringer Relevanz des β-Fehlers.
Minimax-Prinzip	Minimierung des maximalen Zuordnungsfehlers über alle Klassen hinweg
Minimun-Loss	Minimierung des Zuordnungsfehlers über alle Klassen. Dies kann zu relativ hohen Fehlern in einer Klasse führen

Die bisherigen Ausführungen gingen von einer bestehenden Dichotomisierung auf Prädiktor- und Kriteriumsseite aus. In der Praxis wird man aber häufig kontinuierliche Variablen vorfinden, bei denen erst Trennwerte zu setzen sind. Dies ist ein entscheidender Aspekt, der die Vorhersagewerte beeinflusst.

Die *ROC-Kurve* (*Receiver-Operating-Characteristic*) ermöglicht eine simultane, von der Basisrate unabhängige Bestimmung von Spezifität und Sensitivität für verschiedene Testtrennwerte. Voraussetzung ist die empirische Ermittlung von Verteilungskennwerten der unterschiedlichen Gruppen (Amelang & Zielinski, 2002, S.448). *Abbildung 24* zeigt die Verteilung von zwei Gruppen unterschiedlicher Eignung und die Implikationen für die unterschiedlichen Testtrennwerte. Im Gegensatz zu dem Ansatz von Taylor und Russel, berücksichtigt die ROC-Kurve neben den FP auch FN. Je nach Art der diagnostischen Fragestellung wird der Diagnostiker eine der dargestellten Vorgehensweisen zur Minimierung des

Zuordnungsfehlers wählen oder sich an den Kennzahlen aus *Tabelle 8*, die keine ganzheitliche Fehlerbetrachtung vornehmen, orientieren (Hollmann, 1991, S.29). Dies hängt von der Anwendungssituation, den Grundraten und den Bewertungen für die vier Möglichkeiten (RP, FP, RN, FN) ab. Liegen diese vor, so kann man den situationsspezifisch besten Trennwert wählen. Doch die Güte einer diagnostischen Entscheidung kann nicht alleine an der Gesamtstatistik aus dem Vierfelderschema erschlossen werden, da auch die Forderung nach einer möglichst fehlerfreien Entscheidungsstrategie nicht ausreicht (Noack & Petermann, 1992, S.301). Vielmehr müssen erweiterte Überlegungen zu dem Nutzen der einzelnen Entscheidungsergebnisse angestellt werden, auf die in Kapitel 2.3 genauer eingegangen wird.

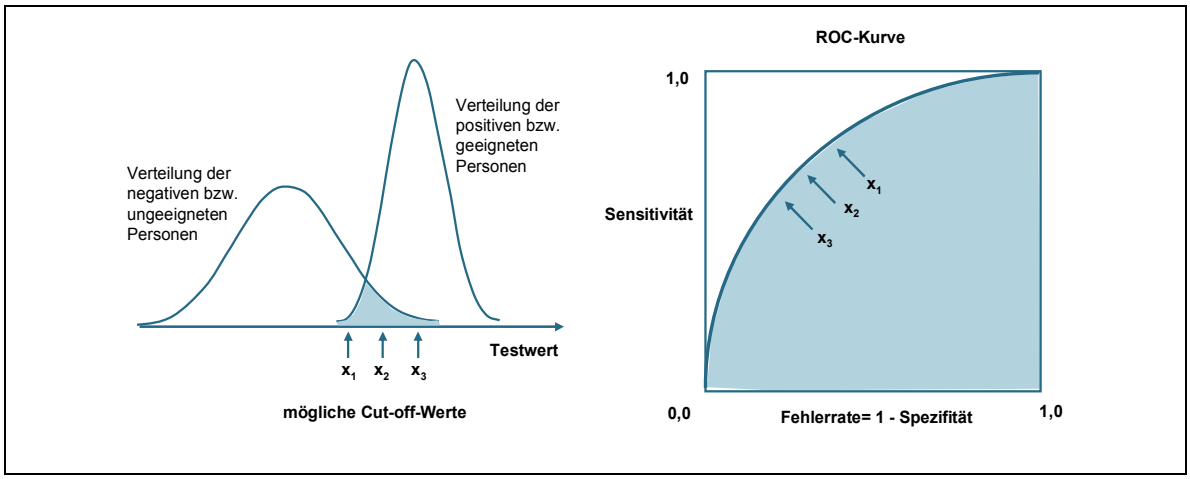

Abbildung 24: Verteilung von 2 Gruppen unterschiedlicher Eignung und ROC-Kurve (Noack & Petermann, 1992, S.300)

2.2.7 Zusammenfassung

Die Schritte, die bei der Berufseignungsdiagnostik im Rahmen der Personalauswahl vorgenommen werden, lassen sich anhand des *eignungsdiagnostischen Prozesses* nach Maukisch (1978) erklären. Ausgangspunkt des eignungsdiagnostischen Prozesses ist die *Arbeits-* und *Anforderungsanalyse* (AA), die die zielgerichtete Informationssammlung der arbeitsrelevanten Aspekte einer Stelle zum Inhalt hat. Dabei können *Eigenschafts-*, *Verhaltens-*, *Qualifikations-* und *Ergebnisanforderungen* ermittelt werden. Grundsätzlich stehen dazu *erfahrungsgeleitet-intuitive*, *personenbezogen-empirische* und *arbeitsplatzanalytisch-empirische* Methoden zur Verfügung. Das neuere, aus den USA stammende Klassifikationsverfahren *O*Net* ermöglicht die Beschreibung einer Arbeitsstelle anhand einer Vielzahl von Dimensionen.

Der nachfolgende Schritt besteht in der *Generierung* und *Operationalisierung* der Kriterien. Schuler und Funke (1993) unterscheiden dabei *Ergebnis-*, *Verhaltens-* und *Eigenschaftskriterien*. Dabei gibt es sowohl für objektive als auch für subjektive Kriterien zahlrei-

che potenzielle Fehlerquellen, die die Gütemaßstäbe beeinträchtigen können. Für die festgelegten Kriterien gilt es im Anschluss geeignete Prädiktoren zu finden. Man unterscheidet dabei eine große Anzahl möglicher Determinanten des Berufserfolges. Ihre Auswahl hängt von der Kenntnis über ihr Wirkungsgefüge mit dem Kriterium zusammen. Für die *Operationalisierung der Prädiktoren* stehen die Prinzipien der *rationalen* (deduktiven), der *externalen* (kriteriumsbezogenen), der *induktiven* (internalen) *Methode* und der *Prototypenansatz* zur Verfügung. Im Rahmen der darauf folgenden *Validitätsanalyse* geht es hauptsächlich darum, die Kriteriumsvalidität der Prädiktoren zu ermitteln, genauer die Vorhersagevalidität. Ihre Höhe kann aber durch mangelnde Reliabilität und mangelnde Konstruktvalidität erheblich beeinträchtigt werden. Das *Konzept der fünf Datenboxen*, die *Reliabilitätstheorie* sowie das *Brunswik'sche Linsenmodell* (Wittmann, 1987; 1990) können als Rahmenkonzepte zur systematischen Analyse dieser Beeinträchtigung herangezogen werden.

Sind nun hinreichend valide Prädiktoren vorhanden, stellt sich die Frage, wie diese kombiniert werden müssen, um eine optimale Vorhersage zu erzielen. Generell unterscheidet man dabei, bei der Datenerhebung und der Datenkombination, *klinische* und *statistische* Varianten. Zahlreiche Untersuchungen von Wiggins (1973), Sawyer (1966) und Meehl (1954) deuten dabei auf eine Überlegenheit statistischer Verfahren hin. Statistische Verfahren können in *kompensatorische* und *nichtkompensatorische* Verfahren unterteilt werden. Obwohl sie überlegen sind, so sind sie in ihrer Anwendung auf Grund ihrer Prämissen nicht unproblematisch. Kausalanalytische Betrachtungen lassen sich mit Hilfe des auf regressions- und faktoranalytischen Prinzipien basierenden *LISREL-Ansatzes* und der *Pfadanalyse* durchführen. Diese prognostischen Verfahren liefern Information, auf deren Basis Entscheidungen über die Zuordnung der Bewerber zu einer Intervention (z.B. Annahme oder Ablehnung) getroffen werden. Cronbach und Gleser (1965) unterscheiden dabei vier Dimensionen, nach denen Entscheidungsarten klassifiziert werden können. Unterschieden werden zudem *kompensatorische* und *konjunktive* sowie *ein-* und *mehrstufige Strategien*. Bei mehrstufigen Strategien unterscheidet man *Vorauswahl-*, *Vorentscheidungs-* und *vollständig sequentielle Strategien.*

Mithilfe von Cronbach und Glesers *Vierfelderschema* (1965) kann gezeigt werden, dass durch einen Validitätskoeffizienten eines Prädiktors in einem nur sehr geringen Umfang die Güte einer Entscheidung beurteilt werden kann. Noack und Petermann (1992) berichten spezifischere Gütekriterien, die die Wahrscheinlichkeiten richtiger Entscheidung felderspezifisch ermitteln. Taylor und Russel (1939) kombinieren in ihrem Ansatz Validität, BR und SR und ermöglichen die Beurteilung der Effektivität einer Entscheidungsstrategie. Sie fokussie-

ren jedoch dabei lediglich auf die Ergebnisse der Annahme der Bewerber. Nach dem *Neyman-Pearson-Kriterium*, dem *Minimax*- und dem *Minimum-Loss-Prinzip* können allerdings die Zuordnungsfehler über alle Felder verringert werden. Je nach diagnostischer Fragestellung wird der Diagnostiker ein anderes Gütekonzept wählen. Die reine Sicht auf die Vierfelderstatistik betrachtet jedoch nicht die Folgen einer Entscheidung, sondern lediglich ihre Eintrittswahrscheinlichkeit. Dies erfordert die Anwendung weiter gehender Nutzenmodelle.

2.3 Nutzenmodelle bei der Personalauswahl

Im Abschnitt 2.2.6 wurde gezeigt, dass der Validitätskoeffizient nicht ausreicht, um eine Entscheidung über ein zu verwendendes Auswahlverfahren zu treffen. Vielmehr sind die Ergebnisse der aus dem Auswahlverfahren resultierenden Entscheidungen sowie deren Bewertung hinsichtlich des Nutzens für die Institution (oder der Person) heranzuziehen. Im Folgenden soll auf die Perspektive der Institution, also die des Anwenders, fokussiert werden. Nutzenmodelle haben dabei neben einer Evaluationsfunktion, wie im Falle der vorliegenden Arbeit, eine Legitimations- und Kommunikationsfunktion. Letztere beide dienen als Rechfertigung für den Einsatz eines Verfahrens und zum Verständnisgewinn bei den Entscheidungsträgern durch Verwendung einer betriebswirtschaftlichen Sprache (Funke & Barthel, 1990). Die direkte Erfassung eines solchen Nutzens bleibt den Anwendern allerdings oftmals verwehrt, so dass die Nutzenfestlegung immer auch einer gewissen Willkür unterliegt. Problematisch sind dabei hauptsächlich die Quantifizierung und die Messung bestimmter Ergebnisse (Bsp.: Zunahme an Kreativität) sowie die Transformation mehrerer Ergebnisse in eine einheitliche Maßeinheit, wie z.B. das „dollar-criterion“ (Brogden & Taylor, 1950). Dennoch ist eine Nutzenbestimmung der Folgen personeller Entscheidungen durch die Institutionen unumgänglich, wenn eine rationale Entscheidung über den Einsatz eines Auswahlverfahrens getroffen werden soll. Im folgenden Kapitel sollen zunächst die gängigen, teilweise schon angesprochenen Ansätze dargestellt werden.

2.3.1 Übersicht historischer Ansätze

Unter 2.2.6 wurden bereits erste Ansätze zur Nutzenermittlung eignungsdiagnostischer Verfahren angesprochen. Zur Kalkulation des Nutzens wurden in der Organisationspsychologie zahlreiche Modelle entwickelt, die sich in chronologischer Reihenfolge als jeweilige Weiterentwicklung des Vorgängermodells abgelöst haben (Schuler, et al., 1995, S.177ff.). Neben dem Validitätskoeffizienten sind bereits in frühen Monographien der *Determinationskoeffizient*, der *Alineationskoeffizient* und der *Index der Vorhersageeffektivität* als einfache Maßzahlen des Nutzens vorgeschlagen worden (Hull, 1928; Kelley, 1923). All

diese Maße orientieren sich aber lediglich an der Enge des Zusammenhanges zwischen Prädiktor und Kriterium. Ihre Aussagefähigkeit ist deshalb, wie oben berichtet, stark eingeschränkt. Zudem herrscht Unsicherheit, welche der Maßzahlen zur Beurteilung der Kriterienvorhersagekraft herangezogen werden soll, da diese die Nützlichkeit, unter sonst gleichen Bedingungen, unterschiedlich einschätzen (Brogden, 1946). Sie stellen letztlich eher Präzisionsindices als Nutzenindices dar. Brogden (1946) konnte jedoch zeigen, dass der Validitätskoeffizient den Anteil der Einsparung wiedergibt, die auf Grundlage eines Tests an der maximal möglichen Einsparung erzielt werden kann. Die maximale Einsparung ist bei einer Selektion, die auf dem Kriterium selbst basiert, erreichbar. Die Interpretation im Sinne Brogdens ist Grundlage nahezu aller modernen Nutzenmodelle und somit Ausgangspunkt der nachfolgend beschriebenen Ansätze.

Der im vorherigen Kapitel dargestellte Ansatz von Taylor und Russel (1939) nimmt zwar eine situationsspezifische Betrachtung vor, indem er Selektions- und Basisrate berücksichtigt. Auch dieses Modell gerät aber schnell an seine Grenzen. Diese betreffen seine Normalverteilungs- und Linearitätsannahme, die reine Anteilsbetrachtung der Klassen (und fehlende Wertzuweisung) (Cascio, 1982) sowie die problematische Anwendung bei subjektiven Beurteilereinschätzungen der Eignung (Hunter & Schmidt, 1982). Darüber hinaus ist nicht geklärt, ob zur Verbesserung des Auswahlverfahrens die Validität erhöht oder die Selektionsquote verringert werden sollte. Hier fehlt die dafür notwendige Kosteninformation. Zudem geht, durch die grobe Dichotomisierung des Leistungskriteriums, bei der Erfolgsquote Information verloren, was Naylor und Shine (1965) zu einer Weiterentwicklung veranlasste. Sie orientieren sich in ihrem Modell, das auf dem Ansatz von Kelley aus dem Jahr 1923 beruht (Barthel, 1988, S.91), an der durchschnittlichen Leistung der mit unterschiedlich validen Methoden ausgewählten Mitarbeiter unter Verwendung standardisierter Kriterienwerte. In seiner Anwendung, etwa der Berechnung von Leistungszuwächsen oder der Festlegung von Auswahlquoten zur Erreichung einer bestimmten Leistung, ist das Modell breiter. Allerdings gelten auch hier die gleichen Prämissen und Einschränkungen wie bei dem Modell von Taylor und Russel (Barthel, 1988, S.92).

Sands (1973) erweitert in seinem CAPER-Modell (**C**ost of **A**ttaining **Pe**rsonnel **R**equirements) die Sicht, indem er die zu minimierenden Kosten der Personalauswahl und -einarbeitung in die Nutzenschätzung mit einbezieht. Er ergänzt dabei das Taylor und Russel-Modell um weitere Parameter und zeigt damit die Grenzen der Nutzensteigerung bei Verkleinerung der Auswahlquote (Barthel, 1988). Die Stärken des CAPER-Modells liegen in seiner leichten Kommunizierbarkeit und in der Vermeidung einer Über- bzw. Unteroptimie-

rung von Einzelbereichen. Zudem kann es Zusammenhänge von Personalentscheidungen und verursachenden Kosten aufdecken (Cascio, 1982). Die Beschränkungen des Ansatzes von Taylor und Russel bleiben aber auch hier bestehen.

Betrachtet man den rein dichotomen Fall, so kann mit der allgemeinen Formel von Wiggins (1973, S.258) die Vergleichbarkeit alternativer Entscheidungsstrategien auf Basis der jeweiligen Gesamtnutzenwerte hergestellt werden. Liegen allerdings kontinuierliche Variablen vor, so müssen diese künstlich dichotomisiert werden, was ebenso mit einem Informationsverlust einhergeht. Ausgehend von der Vierfeldermatrix aus *Abbildung 21* ergibt sich:

$$(13) \quad EU = U_1 P(RP) + U_2 P(FP) + U_3 P(RN) + U_4 P(FN) - U_t$$

EU	= Erwarteter Gesamtnutzen der Auswahlstrategie
U_i	= Nutzen der Ergebnisse der einzelnen Felder (VP, VN, FP, FN)
$P( )$	= Wahrscheinlichkeit des Eintretens des in der Klammer stehenden Ergebnisses
U_t	= Kosten für das Auswahlverfahren

Die Orientierung an der Vierfelder-Tafel ist als Grundmodell zu verstehen. Es sind aber auch mehr Abstufungen denkbar. Wiggins Modell sieht keine implizite Berücksichtigung von Nutzen im Sinne von Geldäquivalenten vor, obgleich der Nutzen auch monetär, etwa durch Schätzungen, bestimmt werden kann. Wittmann (1985, S.361) zeigt dies exemplarisch mit fiktiven Werten für eine Evaluationsstudie im klinischen Bereich. Eben gerade im klinischen Bereich kommt den fälschlicherweise abgelehnten Patienten (FN) eine besondere Bedeutung zu. Ein Vorenthalten notwendiger Therapien kann sogar lebensbedrohlich sein. Für den Fall der Personalauswahl wird allerdings eher die institutionelle Perspektive ausschlaggebend sein, die eine Betrachtung der abgelehnten Bewerber nicht vorsieht. Entsprechend ist diese Perspektive in den nachfolgenden Modellen, die den Versuch unternehmen, den Nutzen direkt in Geldeinheiten auszudrücken, nicht enthalten.

Cronbach und Gleser (1965) haben diese ökonomische Perspektive in ihrem entscheidungstheoretischen Modell formalisiert. Sie schlagen vor, das Credo zahlreicher Diagnostiker aufzugeben, „wahre“ Werte zur Vorhersagezwecken zu schätzen. Stattdessen sehen sie den Zweck von Testverfahren in der Generierung richtiger Entscheidungen und zwar in Relation zu möglichen Alternativentscheidungen (Cronbach & Gleser, 1965, S.1ff.) . Dies dürfte allerdings bis heute eher Wunschdenken geblieben sein, da nach wie vor ein Großteil der Psychologen von einer den physikalischen Messinstrumenten vergleichbaren Güte psychologischer Messinstrumente träumt oder in seinem Verhalten von der Tauglichkeit der Verfahren für die Vorhersage auch bei individuellen Entscheidungen ausgeht. Wie unter 2.2.6 dargestellt, berücksichtigen die Autoren neben dem Validitätskoeffizienten und den situati-

ven Parametern (wie Taylor und Russel) auch die Art der Entscheidung. Das heißt, dass das Wertsystem der Beurteilung in die Nutzenbetrachtung integriert wird. Zudem fließen die Handlungsergebnisse und Kosten des Verfahrens in die Betrachtung mit ein. Das Modell von Cronbach und Gleser ermöglicht die Berechung des Gesamtnutzens einer Entscheidungsstrategie. Das bedeutet, dass für eine gegebene Behandlung, bei Vorliegen kontinuierlicher Testwerte, eine Funktion gebildet werden kann, die die erwarteten, bewerteten Ergebnisse in Abhängigkeit der Testwerte darstellt (Cronbach & Gleser, 1965, S.25). Gibt es nun mehrere Behandlungsmöglichkeiten, so können auf Basis der univariaten Information Entscheidungen optimiert werden. In der Praxis jedoch hat das komplexe Modell von Cronbach und Gleser bisher keine Relevanz erlangt (Schuler, et al., 1995, S.179).

Werden für die Informations- und Kriteriumskategorien Kontinuitätsannahmen gemacht, konstante Kosten für alle Probanden (Bewerber) unterstellt und eine lineare Beziehung zwischen Testwert und Nutzen postuliert, so geht dass Modell in das von Brogden (1949) über (Schmidt, et al., 1979; Cascio, 1980; siehe Amelang & Zielinski, 2002). Brogdens Modell berücksichtigt den Auswahlkontext und ermittelt den Nutzen einer Entscheidungsstrategie für den einfachen Selektionsfall über die Abhängigkeit von Auswahlquote, Validität eines Tests und den Wert der Leistung der ausgewählten Mitarbeiter. Dabei setzt es voraus, dass Prädiktor und Leistungskriterium in gleicher Weise verteilt sind und das die Regression der Leistung auf den Prädiktor linear ist (Barthel, 1988, S.109). Das bedeutet, dass die Neigung der Regressionsgeraden von der Korrelation zwischen Prädiktor und Kriterium, also von der Validität des Tests bestimmt wird *(Gleichung (14))*.

(14) $$\overline{Z}_y = r_{xy}\,\overline{Z}_x$$

Gleichung (14) verdeutlicht, dass der durchschnittliche standardisierte Kriteriumswert $\overline{Z}_y$ der selektierten Gruppe sich aus dem durchschnittlichen standardisierten Prädiktorwert $\overline{Z}_x$ multipliziert mit der Validität des Tests r_{xy}, ergibt. Bei fehlender Kenntnis der Prädiktorwerte abgelehnter Bewerber kann man dabei, unter der Annahme der Normalverteilung, der Auswahl der Bewerber anhand des Prädiktorwertes von „oben nach unten“ und der Kenntnis der Selektionsrate Q $\overline{Z}_x$ schätzen *(Gleichung 15)*.

(15) $$\overline{Z}_x = \lambda(Q)/Q$$

λ *(Q)* ist dabei die Ordinatenhöhe an der Stelle des Cutoffs, die bei der Standardnormalverteilung der Selektionsquote entspricht. De Corte (2000a;2000b) weist darauf hin, dass

die Schätzung der standardisierten Kriterienwerte von der Annahme unerschöpflicher Bewerberzahlen ausgeht. Er entwickelte Korrekturformeln für die Berechnung eines Schätzwertes mit Konfidenzintervallen für den Fall endlicher Bewerberzahlen. Will man nun den Gesamtnutzen in Geldeinheiten ausdrücken, so muss man zusätzlich die Standardabweichung der Produktivität der Mitarbeiter in Geldeinheiten (SD_y) und die Kosten des Auswahlverfahrens berücksichtigen. Der Nutzen pro ausgewähltem Individuum ergibt sich dann aus:

(16) $$U = r_{xy}\, SD_y\, \lambda(Q) / Q - C_A / Q$$

U:	Nutzen eines Verfahrens pro ausgewählten Bewerber
r_{xy}:	Prognostische Validität des Prädiktors
SD_y:	Standardabweichung der Produktivität der Mitarbeiter
$\lambda(Q)$:	Ordinatenhöhe an der Stelle des Cutoffs bei der Standardnormalverteilung
Q:	Selektionsrate
C_A:	Testkosten pro ausgewählten Bewerber

Brogden kommt zunächst zu den gleichen Schlussfolgerungen wie Taylor und Russel. Dies beinhaltet, dass geringe Auswahlquoten einen gleichen Effekt wie hohe Validitäten bezüglich Einsparungen haben und auch Tests mit geringen Validitäten nennenswerte Nutzenzuwächse erzielen können. Allerdings weist er auch darauf hin, dass Testkosten diese Einsparungen wieder zunichte machen können und es sogar zu einem negativen Nutzen kommen kann. Dies führt zu einer Relativierung der Bedeutung der Auswahlquote, die immer im Zusammenhang mit den Testkosten gesehen werden muss (Brogden, 1949, S.174; Cronbach & Gleser, 1965, S.38ff.). Die Ermittlung eines „Break-Even“ des Auswahlverfahrens erfordert dabei eine Quantifizierung der Mehrleistung in der gleichen Einheit wie die Kosten, also in Geldeinheiten. Cronbach und Gleser orientieren sich für den Spezialfall des einstufigen Testens mit fester Behandlung an Brogdens Modell und berechnen den Nutzen pro getesteten Bewerber. Berücksichtigt man die durchschnittliche Betriebszugehörigkeit im Modell, so ergibt sich für den Betrag, um den der Nutzen bei Anwendung eines Verfahrens im Vergleich zu einer Zufallsauswahl steigt (vgl. Hunter & Schmidt, 1982):

(17) $$\Delta U = N_A\, T\, r_{xy}\, \bar{Z}_x\, SD_y - CN_B$$

ΔU:	Nutzenzuwachs eines Verfahrens bei einjähriger Anwendung
N_A:	Zahl angenommener Bewerber
N_B:	Zahl der Bewerber
T:	Durchschnittliche Betriebszugehörigkeit in Jahren
r_{xy}:	Prognostische Validität des Prädiktors
SD_y:	Standardabweichung der Produktivität der Mitarbeiter
C:	Kosten des Auswahlverfahrens pro Mitarbeiter
$\bar{Z}_x$:	Durchschnittlich standardisierter Prädiktorwert von RP und FP

Gleichung (17) ist das Grundmodell moderner Nutzenmodelle und wird in Anlehnung an Boudreau (1993) nach den Autoren Brogden, Cronbach und Gleser *BCG-Modell* genannt. Das *BCG-Modell* beruht auf einer Vielzahl von teilweise bereits angesprochenen Annahmen (Cascio, 1980; Boudreau, 1991). Diese betreffen u.a. die Standardisierung der Testwerte, die Rangreihenbildung bei der Behandlungsentscheidung, die innerhalb des Auswahlverfahrens dichotom ist, die lineare Regression der Handlungsergebnisse auf den Prädiktorwert und die Annahme des fehlenden Zusammenhangs zwischen abgelehnten Bewerbern und Nutzen. All diese Prämissen sowie das Fehlen weiterer Parameter schränken die Anwendung eines solchen Verfahrens erheblich ein. Die komplexe Realität wird durch das *BCG-Modell* also in nur unzureichendem Maße abgebildet. Darauf basierend sind modernere Nutzenmodelle entwickelt worden, die unter näher 2.3.3 betrachtet werden. Doch zunächst soll nachfolgend auf das Problem der Operationalisierung der Modellparameter, die mit erheblichen Schwierigkeiten behaftet ist, eingegangen werden.

2.3.2 Zur Problematik der Parameteroperationalisierung

Im vorherigen Abschnitt wurde die Notwendigkeit der Operationalisierung zahlreicher Parameter im *BCG-Modell* angesprochen, die nachfolgend exemplarisch für die weiteren Modelle dargestellt werden soll. Betrachtet man *Gleichung (16)*, so betrifft dies zunächst die *Selektionsquote Q*, die leicht über das Verhältnis N_A/ N_B ermittelt werden kann. Auf Basis der *Gleichung (15)* kann, mit Hilfe der von Naylor und Shine (1965) veröffentlichten Tabellen, der Wert der Ordinatenhöhe abgelesen werden. Allerdings setzt dies die Normalverteilung der Prädikorwerte und die Selektion der Bewerber von „oben nach unten“ voraus. Gerade letztere Regel dürfte durch die nicht immer strikte Anwendung in der Praxis (Härtefälle, Bewerber mit guten Beziehungen, Ablehnung von Stellen durch Bewerber) selten vorzufinden sein. Dies führt zu einer Verminderung von $\bar{z}_x$. Empfehlenswert ist deshalb die empirische Ermittlung der Ordinatenhöhe.

Die Probleme bei der Ermittlung des *Validitätskoeffizienten* r_{xy} sind unter 2.2.4 ausführlich betrachtet worden. Folgendes soll dazu noch angemerkt werden: Oftmals wird zwischen der empirisch ermittelten und der tatsächlichen Validität unterschieden. Dabei kann die empirische Validität, wie gezeigt, systematisch von der tatsächlichen „wahren“ Validität abweichen. Gründe hierfür können sowohl Varianzeinschränkungen auf Prädiktor- und Kriterienseite als auch eine mangelnde Reliabilität des Kriteriums sein. Eine Korrektur des Validitätskoeffizienten hängt davon ab, ob das Risiko einer irrtümlichen Überschätzung der Validität der möglichen Unterschätzung des Nutzens vorgezogen wird. Eine Überschätzung

ist etwa bei niedriger Reliabilität des Prädiktors oder des Kriteriums und Ausreißerwerten in der Stichprobe denkbar. Eine einzelfallbezogene Entscheidung über Korrekturen ist hier sicherlich empfehlenswert.

Eine weitere Schwierigkeit besteht in der Operationalisierung der Kosten des Auswahlverfahrens. Eine Unterteilung nach dem betriebswirtschaftlichen Prinzip in variable und fixe Kosten kann hier hilfreich sein. Fixe Kosten eines Auswahlverfahrens sind die anteiligen Gemeinkosten der Personalabteilung (Sachkosten, Personalkosten) und die direkt dem Personalauswahlverfahren zuordenbaren Kosten (Personal-, Entwicklungs- und Evaluationskosten usf.). Variable Kosten sind solche, die mit der Häufigkeit der Durchführung oder der Anzahl der Bewerber in Verbindung stehen (z.B. Kosten der Durchführung, Reise- und Unterkunftskosten der Bewerber, Lizenzgebühren für Tests usf.). Boudreau (1983a; 1983b) und Cronshaw und Alexander (1985) berücksichtigen diese Überlegungen in ihren Modellen (siehe 2.3.3).

Der wohl kritischste Parameter, den es zu schätzen gilt, ist die *Standardabweichung der Produktivität* SD_y. In ihrem Zusammenhang wird deshalb auch von der „Achillesferse“ des entscheidungstheoretischen Ansatzes gesprochen (Cronbach & Gleser, 1965, S.121). Man benötigt sie, um die durch ein besseres Auswahlverfahren verursachten Produktivitätsgewinne monetär zu quantifizieren. Die dabei auftretenden Probleme betreffen sowohl die Definition des bewerteten Arbeitsergebnisses y als auch die Schätzung seiner Höhe und seiner Standardabweichung SD_y. Boudreau (1991) unterscheidet dabei drei Ansätze zur Ermittlung des „pay-off“: Frühere Studien fokussierten dabei auf die *Kostenreduktion* als Nutzenergebnis (Lee & Booth, 1974; Naerssen, 1965; Eaton, Wing & Mitchell, 1985; Arnold, Rauschenberger, Soubel & Guion, 1982). Doch die Anwendungsschwächen dieser Ansätze führten zu einer erweiterten Betrachtung (Boudreau, 1991, S.241). Die Gruppe der Autoren um Schmidt und Hunter (Schmidt, et al., 1979) definiert y als den *Verkaufswert einer Leistung*, die durch einen Mitarbeiter erbracht wird. Als Vergleichsgrößen dienen hier der Preis, der für den externen Zukauf (Outsourcing) dieser Leistung bezahlt werden müsste, bzw. das Gehalt des Mitarbeiters (Bobko, Karren & Parkington, 1983; Burke & Frederick, 1984; DeSimone, Alexander & Cronshaw, 1986; Cascio, 1982; Cascio & Ramos, 1986).

Der *Verkaufswertansatz* ist insbesondere wegen der mangelhaften Wiedergabe des institutionellen Nutzens heftig kritisiert worden (Boudreau, 1983a; 1983b), gerade bei hohen variablen Kosten. Dies kann zu einem tatsächlich negativen Nutzeneffekt bei einem positiven rechnerischen Ergebnis führen, wenn die (variablen) Grenzkosten, zuzüglich der Interventionskosten, größer sind als der Nutzenzuwachs. Boudreau (1983a) schlägt deshalb die

Berücksichtigung der variablen Kosten bei der Berechnung der SD_y-Größe bei Verwendung des Verkaufswertes vor, indem er sie mit dem Faktor (1-Anteil variable Kosten am Verkaufswert) multipliziert. Bobko, et al. (1983) verglichen in ihrer Studie zur Schätzung von SD_y bei Versicherungsvertretern, den *Verkaufswert*- mit dem *Gewinnbeitragsansatz* und kamen zu erheblichen Unterschieden in den Schätzwerten, was die Problematik der *y*-Schätzung untermauert. Boudreaus Ansatz kann als Deckungsbeitragsbetrachtung verstanden werden und orientiert sich an dem *Gewinnbeitrag* des Mitarbeiters am Unternehmen (Cronbach & Gleser, 1965; Roche, 1965; Hunter, Schmidt & Coggin, 1988). Er berücksichtigt also alle vorher genannten Ansätze und kann deshalb als das in den meisten Fällen geeignete Vorgehen betrachtet werden.

Es gibt derzeit keine einfache Methode zur Schätzung der Standardabweichung SD_y. Schuler, et. al. (1995) teilen die vorhandenen Methoden in vier Gruppen ein (*Tabelle 10*). Das in der Praxis am häufigsten angewandte Verfahren ist das *globale Schätzverfahren* (Schuler & Funke, 1993). Arvey und Murphy (1998) und Boudreau (1991) sehen die Gefahr, dass auf Basis solcher Schätzungen die Nutzenanalyse niemals eine hohe Präzision erreichen wird. Boudreau und Ramstad (2002) zeigen zudem, dass SD_y unternehmensübergreifend kaum vergleichbar ist, da bestimmte Tätigkeiten in einem Unternehmen sehr erfolgskritisch sind, während in einem anderen Unternehmen die Variabilität in den gleichen Tätigkeiten kaum Einfluss auf das Unternehmensergebnis hat. Dies erfordert eine Identifikation der relevanten unternehmensspezifischen Schlüsselpositionen. Bei Verfahren, die direkt SD_y-Ausprägungen schätzen, indem sie sich an beobachtbarer Leistung orientieren und zugrunde liegende Annahmen berücksichtigen (z.B. Raju, Burke & Normand, 1990), bleibt die Wahl der geeigneten Zusammenhangsfunktion zwischen der Leistung und ihrem Wert derzeit immer eine Herausforderung (Boudreau & Ramstad, 2002). Problematisch sind in diesem Zusammenhang auch die zahlreichen Faktoren, die die SD_y-Schätzung beeinflussen können. Cesare, Blankenship und Gianetto (1994) und Judiesch, Schmidt und Mount (1992) zeigten, dass bis zu 50% der Variabilität von SD_P (als prozentuale Standardabweichung, die noch nicht in Geldeinheiten transformiert wurde) auf nicht-„output"-bezogene Faktoren zurückzuführen ist. Diese können neben Urteilsfehlern, der Kontext der Beurteilung (Bobko, Shetzer & Russel, 1991), Gehälter, Beziehungsgeflechte oder auch demographische Eigenschaften sein (Roth, Pritchard, Stout & Brown, 1994).

Tabelle 10: Methoden zur Schätzung von SDy (Schuler, et al., 1995; Boudreau, 1991)

Verfahren	Vorgehen	Kritik
Kostenrechnungsmethode	Verwendung von betriebswirtschaftlichen Kennzahlen, wie z.B. Deckungsbeiträgen von Außendienstmitarbeitern und Berechnung der Standardabweichung (Brogden & Taylor, 1950; Roche, 1965).	Die kausale Rückführung der Unterschiede in der Produktivität auf die Mitarbeiter ist oftmals fraglich. Das Vorgehen ist sehr aufwendig und teilweise willkürlich (Roche, 1965; Greer & Cascio, 1987).
Globales Schätzverfahren	Expertenbefragung bezüglich des geldwerten Vorteils von Mitarbeitern, die bestimmte Leistungen erbringen (Orientierung der Schätzungen des Verkaufswertes an der Normalverteilung) (Schmidt, et al., 1979; Eaton, et al., 1985).	Vorgehen ist einfach handhabbar, geht aber mit einer starken Subjektivität der Schätzungen und oftmals große Streuung in den Urteilen einher (Bobko, et al., 1983). Zudem ist von schwachen Reliabilitäten der Urteile auszugehen (DeSimone, et al., 1986).
Proportionale Regeln	Pauschal 40-70% des Bruttogehaltes (Hunter & Schmidt, 1982).	Das Vorgehen führt häufig zu einer Unterschätzung der SD_Y (Boudreau, 1993). Zudem verstößt es gegen die Unabhängigkeit der Varianz vom Mittelwert (Bobko, et al., 1983, S.175) und ist empirisch widerlegt worden (Boudreau, 1991).
Mehrstufiges dekomponierendes Verfahren (CREPID)	Ausgehend von einer nach Wichtigkeit gewichteten AA werden Mitarbeiter von ihren Vorgesetzten auf diesen Dimensionen bewertet und ein gewichteter Gesamtwert gebildet. Multiplikation des Gesamtwertes mit durchschnittlichem Bruttojahresgehalt und Bestimmung der SD_Y des Geldwertes (Cascio, 1982; Cascio & Ramos, 1986).	Sehr aufwendiges Vorgehen, das ebenso einer Subjektivität der Schätzungen unterliegt.

Die Kritikpunkte der geschilderten Verfahren aus *Tabelle 10* verdeutlichen, dass die Schätzung von SD_y mit erheblichen Unsicherheitsaspekten einhergeht. Boudreau (1993; Rich & Boudreau, 1987) nennt vier Vorgehensweisen, mit deren Hilfe sich das Ausmaß dieser Unsicherheit abschätzen lässt. Bei der *Sensitivitätsanalyse* werden bei den Berechnungen die Variablen von ihren höchsten bis zu ihrem niedrigsten Wert variiert, um abzuschätzen, welche Variablen den größten Einfluss auf den Nutzen haben. Bei der *Break-Even-Analyse* wird der niedrigste Wert einer Variablen ermittelt, der noch einen positiven Gesamtnutzen generiert. Die *algebraische Berechnung* der Varianz des Nutzens basiert auf Berechnungsformeln zur Varianz der Produkte randomisierter Variablen. Über verschiedene, sehr komplexe Formeln und der Annahmen der Normalverteilung des Nutzens ist eine Berechnung der Varianz des Nutzens möglich. Bei der *Varianzberechnung des Nutzens mit Monte-Carlo-Analysen* werden zufällige Berechnungen mit unterschiedlichen Kombinationen der Parameter durchgeführt, so dass eine Verteilung der Nutzenwerte mit entsprechenden Parametern (Mittelwerte, Streuung) entsteht. Voraussetzung hierfür ist das Vorhandensein dieser Streuungen und Mittelwerte der Variablen.

2.3.3 Neuere Entwicklungen bei Nutzenmodellen

Die bisher dargestellten Modelle zeigen besonders in der Praxis ihre Schwächen, da es mit ihnen nicht möglich ist, Wirtschaftsorganisationen von der Notwendigkeit der Investition in Auswahlverfahren zu überzeugen. Sie tragen somit nicht ausreichend zur Überbrückung der unter 1.3 geschilderten Kluft zwischen Psychologen und Management bei (Barthel, 1988, S.149). Hauptgründe liegen in einer ungenauen Operationalisierung verwendeter Parameter und einem Mangel an Berücksichtigung weiterer relevanter Größen, wie der Diskontierung zukünftig erwarteter Einnahmen. Das Modell von Boudreau (1983a; 1983b) greift diese Kritikpunkte auf und kombiniert die Nutzenberechnung mit dem Kapitalwertmodell der dynamischen Investitionsrechnung. Damit entspricht er teilweise einer Forderung von Cronshaw und Alexander (1985), die die Ausgaben der Personalabteilung, wie beispielsweise für ein Auswahlverfahren, als Investition betrachten. Dies bedeutet, dass ökonomische Konzepte der Investitionsrechnung, wie Amortisationsdauer, Barwert sowie Verzinsung oder Steuern, in die Nutzenberechnung einfließen sollten. Es ermöglicht eine Vergleichbarkeit der Investitionen im Personalbereich mit den Investitionen in anderen Unternehmensbereichen (Cascio & Morris, 1990; Cronshaw & Alexander, 1985). Dies wird somit der Legitimations- und Kommunikationsfunktion der Nutzenanalyse gerecht. Boudreaus ausgebaute Variante des Modells, in dem auch die Anwendung von Verfahren über mehrere Perioden sowie variable und fixe Testkosten berücksichtigt werden, zeigt *Gleichung (18)*. Boudreau spaltet dabei das Konzept der Behandlungswirkungsdauer T (siehe *Gleichung (17)*), wie es von Schmidt et. al. (1979; Hunter & Schmidt, 1982) eingeführt wurde, in einzelne Wirkperioden *k* auf. Die Gleichung ermöglicht dabei für verschiedene Wirkperioden die Verwendung unterschiedlicher Parameter. Zusätzlich berücksichtigt er den Mitarbeiterfluss über die Anzahl akzeptierter (N_{at}) und die Anzahl ausgeschiedener Mitarbeiter (N_{st}) je Wirkperiode. Problematisch ist allerdings, dass in dem Modell, anders als von Cronshaw und Alexander (1985) gefordert, keine Fixkosten berücksichtigt werden. Die Annahme gleichbleibender Fixkosten bei allen Verfahren mag dieses Vorgehen rechtfertigen. Allerdings können im Rahmen der Entwicklung und Implementierung der Auswahlverfahren durchaus Kosten in beträchtlicher Höhe entstehen, die sich verfahrensspezifisch u. U. in ihrer Höhe unterscheiden. Die berechnete Nutzendifferenz sollte deshalb am Ende diesen Einmalkosten gegenübergestellt werden. Ein Ansatz, der die Unterteilung in variable und fixe Kosten berücksichtigt, erweitert die *Gleichung (18)* zu *Gleichung (19)*.

$$(18)\qquad \Delta U = \sum_{k=1}^{F}\left[\sum_{t=1}^{k}\left(N_{at} - N_{st}\right)\right]\left\{\left[1/(1+i)^{k}\right] r_{xy}\ \bar{z}_x\ SD_Y\ (1+v)(1-s)\right\} - \sum_{k=1}^{F}\left\{C\,(1-s)\left[1/(1+i)^{k-1}\right]\right\}$$

ΔU:	Nutzendifferenz (Kapitalwert) im Vergleich zu einem alternativen Verfahren
F:	Wirkungsdauer des Verfahrens (Anwendungsperioden + ϕ-Betriebszugehörigkeit)
k:	Zeitabschnitt der Wirkungsdauer (gängigerweise 1 Jahr)
N_{at}:	Anzahl akzeptierter Bewerber in der Wirkperiode t
N_{st}:	Anzahl ausgeschiedener Mitarbeiter in der Wirkperiode t
$1/(1+i)^k$:	Diskontierungsfaktor mit Kalkulationszinsfuss i auf den Kapitalwert
r_{xy}:	Validitätsdifferenz des neuen Prädiktors gegenüber dem alten Prädiktor
$\bar{z}_x$:	Durchschnittlich standardisierter Prädiktorwert von RP und FP
SD_y:	Standardabweichung des Leistungskriteriums
(1+v):	Einfluss variabler Kosten der Mehrleistung auf den Gewinn (v ist negativ bei positivem Zusammenhang von Mehrleistung und variablen Kosten)
(1-s):	Einfluss der Steuern s auf den Gewinn
C:	Kosten des Auswahlverfahrens pro Mitarbeiter
$1/(1+i)^{k-1}$:	Diskontierungsfaktor mit Kalkulationszinsfuss i auf die anfallenden Kosten

$$(19)\qquad \Delta U = \sum_{k=1}^{F}\left[\sum_{t=1}^{k}\left(N_{at} - N_{st}\right)\right]\left\{\left[1/(1+i)^{k}\right] r_{xy}\ \bar{z}_x\ SD_Y\ (1+v)(1-s)\right\}$$
$$- \sum_{k=1}^{F}\left\{C_{vt}\ (N_{at}/Q_t)\,(1-s)\left[1/(1+i)^{k-1}\right]\right\}$$
$$- \sum_{k=1}^{F}\left\{C_{ft}\,(1-s)\left[1/(1+i)^{k-1}\right]\right\}$$

C_{ft}:	Fixe Kosten des Auswahlverfahrens in Periode t
C_{vt}:	Variable Kosten in der Periode t

Im Übrigen gelten die Bezeichnungen von *Gleichung (18)*

Boudreaus Modell überzeugt durch seinen logischen Aufbau und seine klare Definition. Die Brauchbarkeit des Ansatzes hängt allerdings von der Realitätsnähe seiner Annahmen und der Konstanz der Bedingungen (z.B. der Kosten) ab. Zudem gilt, wie bei allen anderen Modellen auch, dass die Operationalisierung der Annahmen hinreichend objektiv sein sollte (siehe 2.3.2). Außerdem fokussiert es genau wie das BCG-Modell lediglich auf die Nutzenmaximierung der Institution. Das heißt, dass die Folgen der Fehlklassifikation bei den abgelehnten Bewerbern nicht, wie in dem Nutzenansatz von Wiggins, berücksichtigt werden. Bemühungen, diese Aspekte ebenso monetär zu quantifizieren, sind, wie bereits angesprochen, bisher nahezu ausgeblieben (Schuler, 2000, S.179). Auf deren Bedeutung bei der Auswahl von Unternehmensberatern wird unter 2.3.5 separat eingegangen.

Das dargestellte Modell von Boudreau quantifiziert, wie die anderen ökonomischen Nutzenmodelle auch, den ökonomischen Nutzen für die Institution, indem es den Produktivitätszuwachs durch die neu ausgewählten Mitarbeiter ermittelt. Es sind aber darüber hinaus noch weitere ökonomische Vorteile für die Institution durch eine verbesserte Auswahl denk-

bar, die in diesen Überlegungen nicht enthalten sind, sicherlich nicht zuletzt, weil sie schwer zu erfassen sind. *Abbildung 25* gibt eine beispielhafte Übersicht.

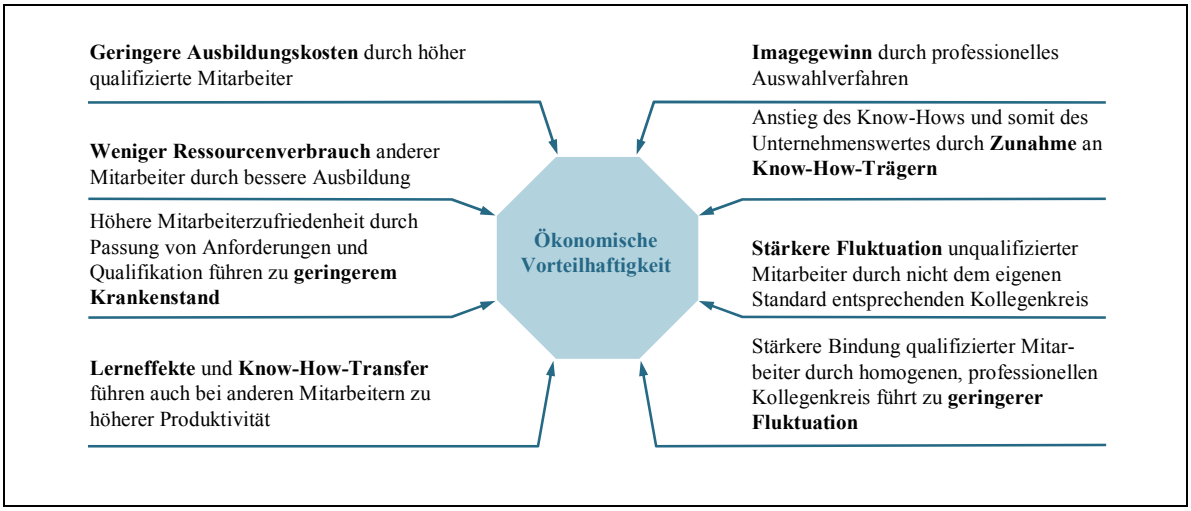

Abbildung 25: Mögliche weitere ökonomische Nutzenaspekte bei Auswahlverfahren

Aufbauend auf den Erkenntnissen von Boudreau sind einige Modellerweiterungen im Personalauswahlkontext vorgenommen worden, die weitere Parameter berücksichtigen. Murphy (1986) erweitert Boudreaus Modell um den Effekt abgelehnter Offerten, indem er Formeln zur Berechnung des standardisierten Prädiktorwertes ausgewählter Bewerber darstellt. Tziner, Meir, Dahan und Birati (1994) fügen einen Inflationsfaktor hinzu, der aber auch direkt im Zinssatz Berücksichtigung finden kann (Holling, 1998). Mabon (1996; 1998) wendete das Modell für die Berechnung des Wertes von Personalabbau an. Barrick und Alexander (1991) berechneten die Effekte beim Arbeitsplatzwechsel innerhalb von Unternehmen. De Corte (1994; 1997; 1998a; 1998b) erweiterte die Perspektive um den Aspekt der Probezeit und der damit verbundenen Möglichkeiten kurzfristiger Personalfreisetzung. Darüber hinaus entwickelte er Methoden zur Nutzenberechnung bei Klassifikationsentscheidungen bei mehreren unterschiedlichen zu besetzenden Stellen (De Corte, 1998a, 1998b, 2000a).

Die durch all diese Modelle zunehmende Komplexität der Berechnungen führte zu immer neuen Implikationen bezüglich Datensammlung und -analyse und letztlich dazu, dass all diese Entwicklungen sich bisher in der Praxis nicht durchgesetzt haben (Boudreau & Ramstad, 2002, S.10; siehe auch 2.3.4). Neben den dargestellten Nutzenmodellen, die einen mathematischen Charakter aufweisen, gibt es noch eine Reihe weiterer alternativer Vorgehensweisen, die in *Tabelle 11* dargestellt sind. Darüber hinaus nennen Boudreau und Ramstad (2002) noch weitere Ansätze, wie die Integration von Personalinformation und

Humankapitalquantifizierung ins Finanzberichtswesen, die Quantifizierung von Forschungs- und Wissensaspekten oder den Einsatz von Pfadanalysemodellen, um den kausalen Zusammenhang von Personalmaßnahmen und ökonomischen Auswirkungen zu untersuchen. Zusätzlich gibt es noch eine Reihe weiterer Modelle, mit deren Hilfe Expertenwissen zu einer Entscheidungsfindung strukturiert werden kann. Dazu gehören die *Gruppendiskussion* zwischen Experten (Wottawa & Hossiep, 1987, S.43), die *Delphi-Methode* (Wittmann, 1985, S.317ff.) oder die *Szenario-Technik* (Wottawa & Hossiep, 1987, S.44). Zur Feststellung des ökonomischen Nutzens eines Auswahlverfahrens scheinen diese aber eher ungeeignet, da sie von stark subjektiven Einflüssen abhängig sind. Vielmehr ist ihr Einsatz aber zur Beurteilung und Entscheidungsfindung der Eignung und Akzeptanz einzelner Bewerber im Rahmen des Auswahlverfahrens denkbar.

Tabelle 11: Mögliche Alternativen zur Nutzenanalyse (in Anlehnung an Boudreau & Ramstad, 2002**)**

Verfahren	**Beispiele**	**Vor- (+) und Nachteile (-)**
Effizienzmessung im Personalbereich (Cascio (2000); Fitz-Enz und Davison (2002))	Berechnung von Einstellungs- oder Trainingskosten pro Mitarbeiter	+ Gute Quantifizierbarkeit und Vergleichbarkeit + Direkte Anbindung an die unternehmerische Kostenrechnung und Berichtssysteme möglich + Ansatzpunkt für Kostenreduktion und -kontrolle + Faktenbasiert ohne Schätzungen - Verfahren sind in der Regel zu kostenlastig - Keine geeignete Quantifizierung des Mitarbeiterwertes
Personalbereichsaktivitäten- und „Best Practice"-Indices (Becker und Huselid (1998); Delery und Doty (1996); Huselid, Jackson und Schuler (1997); Ichniowski, Shaw und Prennushi (1997))	Messung des Zusammenhangs von Personaltrainings und Unternehmensprofit, Benchmarks zum Humankapital	+ Fokussierung auf Aktivitäten sehr operational getrieben + Ökonomischer Effekt wird in direkt anerkannten Größen gemessen (z.B. Unternehmensprofit) - Kausalitätsrichtung der Ergebnisse oftmals unklar
Multiattributive Nutzenschätzung (MAUT) (Roth (1994); Roth und Bobko (1997); Wittmann (1985))	Organisationsspezifisch entwickelte MAUT-Modelle zur Entscheidungsfindung	+ Integration weiterer nicht monetärer, nicht quantifizierbarer oder nicht-linearer Aspekte in die Nutzenbetrachtung + Hohe Akzeptanz durch hohe Beteiligung der Entscheidungsträger an der Nutzenbeurteilung - Hohe Komplexität und hoher Aufwand - Großes Wissen über Ursache-Wirkungszusammenhänge bei Anwendern Voraussetzung
Balanced Scorecard (Becker, Huselid und Ulrich (2001); Kaplan und Norton (1992))	Inwieweit werden durch Personalbereichsaktivitäten Kundenziele erreicht?	+ Bekanntes Instrument bei Entscheidungsträgern + Vielzahl an Maßnahmen abbildbar + Software ermöglicht Teilanalysen für verschiedene Fragestellungen - Bei mangelnder Kenntnis Gefahr der Fehlinterpretation - Instrument wurde nicht für den Personalbereich entwickelt - Oftmals Anwendungsfehler im Personalbereich

2.3.4 Empirische Arbeiten zu Nutzenberechnungen im Auswahlkontext

Im Kontext der Personalauswahl sind bisher wenige Untersuchungen zur ökonomischen Vorteilhaftigkeit von Auswahlverfahren durchgeführt worden. Diese unterscheiden sich zudem stark hinsichtlich ihrer Inhalte und der verwendeten Parameter. Bis Mitte der Achtziger Jahre basierten die Untersuchungen zum Nutzen von Auswahlverfahren größtenteils auf dem Nutzenmodell von Brogden (1949). Später wurden einige Untersuchungen mit dem erweiterten Modell von Boudreau durchgeführt. Die dabei untersuchten Tätigkeiten nahmen in ihrer Komplexität zu (vom Arbeiter zum Manager). In den Neunzigern lag der Fokus der Veröffentlichungen eher auf statistischen und methodischen Problemen. Seitdem gibt es, trotz der steigenden Bedeutung der Wertermittlung von Humankapital, einen stetigen Rückgang an Publikationen zum Nutzenthema (Boudreau & Ramstad, 2002, S. 5). Ein möglicher Grund ist die immer noch fehlende Akzeptanz dieser Modelle auf betriebswirtschaftlicher Seite, die der Durchführung solcher Studien nicht zuträglich ist. Diese mag teilweise ihre Ursache in der zu hohen Komplexität und Realitätsferne der Modelle haben (Rauschenberger & Schmidt, 1987). Zudem ist der Mangel an klaren Paradigmen bei den Entscheidungsmodellen zum Humankapital, die eine Vielzahl von Zugängen zu der Thematik ermöglichen, problematisch (Boudreau & Ramstad, 2002). Die Anwender in der Praxis sind dadurch in der Wahl des geeigneten Ansatzes, die teilweise auch durch modische Erscheinungen geprägt ist, überfordert. Zusätzlich ist sie stark durch Erfahrungen und Erfolge, die der Anwender bereits im Vorhinein mit diesen Ansätzen gewonnen hat, beeinflusst (Macan & Highhouse, 1994; Roth, Segers & Wright, 2000). Ein entscheidender Erfolgsfaktor für die Akzeptanz von Nutzenanalysen beim Management ist dabei die Kommunikation (Cascio, 1996). Die Untersuchungen von Latham und Whyte (1994; Whyte & Latham, 1997) und Cronshaw (1997) deuten auf einen häufigen Mangel bezüglich der sachgemäßen Verwendung von Nutzenanalysen im geeigneten Kontext hin. Carson, Becker und Henderson (1998) sehen hingegen die leichte Verständlichkeit von Kommunikationsinhalten als entscheidenden Akzeptanzaspekt an. Hazer und Highhouse (1997) zeigten auch die Bedeutung des „framing effect“ bei der Beurteilung der Nutzenanalyse[9].

Anhang II gibt eine Übersicht der bekanntesten ökonomischen Nutzenstudien. Sie beruhen teilweise auf fiktiven Daten und rechnen innerhalb eine Studie mit variierenden Parametern. Die Übersicht zeigt, dass fast alle Untersuchungen auf Schätzungen der SD_y-Größe mithilfe unterschiedlichster Vorgehensweisen basieren. Das dabei am häufigsten eingesetzte

[9] Der „framing effect“ ist ein Teil von individuellen Entscheidungsheuristiken, die eine kognitiv vereinfachende Wirkung auf die Urteilsfindung besitzen und Such- sowie Abwägungsprozesse umfassen. Sie werden häufig dann angewendet, wenn der Entscheider nur unzureichende Information besitzt, und sind gewissermaßen eine Art „Faustregel“ .

Verfahren ist die globale Schätzung. Darüber hinaus werden die Kosten sowie die Selektionsraten der vergleichenden Alternativen konstant gehalten, das Boudreau-Modell also nicht in seinen Möglichkeiten vollkommen ausgeschöpft. Die berichteten Nutzengewinne sind in ihrer Höhe außerdem höchst unterschiedlich, was die Validität der Untersuchungen in Frage stellt (Holling, 1998, S.15f.). Zudem fallen einige der Studien durch ihre vorsichtige, oft nicht nachvollziehbare Schätzung von SD_Y sowie einer wenig konservativen Schätzung der Parameter (Validität, Kosten, Auswahlquoten) auf, weshalb sie nicht ohne Kritik geblieben sind (siehe Barthel, 1988, S.114ff.).

2.3.5 Implikationen für die Nutzenberechnungen bei Unternehmensberatungen

Das vorherige Kapitel hat eine Übersicht der im Rahmen der Personalauswahl durchgeführten empirischen Nutzenstudien gegeben. Da bisher nach Kenntnis des Autors noch keine derartige Untersuchung in einer Unternehmensberatung durchgeführt wurde, soll in diesem Kapitel auf die Besonderheiten dieses Kontextes eingegangen werden. Dabei soll der Schwerpunkt auf zwei Aspekten liegen: den Konsequenzen personeller Fehlentscheidungen und der Ermittlung von $SD_{y.}$

Unternehmensberatungen sind Dienstleistungsunternehmen, deren Unternehmenserfolg und Unternehmenswert in einem sehr hohen Maße von der Qualität der Mitarbeiter (als Wissensträger und Netzwerker) abhängig sind. Insofern kommt dem Humankapital eine wesentlich größere Bedeutung zu, als beispielsweise bei einem Produktionsunternehmen, dessen Wert in nicht unerheblichem Ausmaß auch von materiellen Vermögensgegenständen bestimmt wird. Dies bedeutet, dass der Personalauswahl eine Schlüsselrolle zum Erfolg des Beratungsunternehmens zukommt (siehe 1.2). Doch warum haben Fehlentscheidungen bei der Auswahl von Beratern gravierende Folgen? Die Vierfeldertafel aus *Abbildung 21* soll helfen, die einzelnen Aspekte zu erörtern.

Bewerber, die zu Recht angenommen wurden (RP), haben logischerweise eine positive Wirkung auf das Beratungsunternehmen. Sie besitzen oder entwickeln wertvolles Know-how und notwendige Netzwerke für Projektakquisitionen und erledigen die ihnen im Rahmen von Projekten übertragenen Aufgaben zur Zufriedenheit aller Beteiligten (siehe auch 2.4). Anders natürlich bei den zu unrecht angenommenen Mitarbeitern (FP). Im günstigsten Fall wird man diese relativ schnell erkennen und innerhalb der Probezeit freisetzen, so dass mit Ausnahme der damit verbundenen Kosten (Gehalt für die Zeit im Angestelltenverhältnis, Einarbeitungskosten) keine weiteren Schäden für das Beratungsunternehmen zu beklagen sind. Tritt dieser Fall nicht ein, so können die Konsequenzen allerdings gravierend sein.

Aus interner Perspektive kann eine Fehleinstellung eine erhebliche Mehrarbeit für die Beraterkollegen bedeuten. Insbesondere bei der Projektarbeit in einem Team kann dies schnell kritische Ausmaße annehmen. Gewöhnlich besitzen im Beratungsgeschäft Teams aus Kostengründen eine Mannstärke, die wenig zeitlichen Spielraum für zusätzliche Tätigkeiten zulässt. Somit ist jedes Teammitglied bis an seine Grenzen ausgelastet. Wird von einer Person nun die geforderte Leistung nicht erbracht, muss diese folglich von den restlichen Teammitgliedern geleistet werden. Dies kann zu Missstimmungen, Unruhe und Unzufriedenheit im Team führen. Das gilt natürlich gleichermaßen für projektfremde Tätigkeiten, wie Akquisitionsvorbereitungen oder Produktentwicklungen. Zudem können hier auch fehlerhafte Produkte und Akquisitionen entstehen. Aus Opportunitätssicht muss auch der entgangene interne Nutzen, der durch jemand erbracht worden wäre, der den Anforderungen entspricht, aber nicht eingestellt wurde, da die Stelle bereits besetzt ist, berücksichtigt werden.

Aus externer Perspektive zeigen sich weitere potenzielle Risiken. Der zu Unrecht eingestellte, unqualifizierte Berater kann in der Interaktion mit dem Kunden zunächst einen schlechten Eindruck hinterlassen. Dieser kann den Kunden das für diesen Berater veranschlagte Honorar in Frage stellen lassen. Im Rahmen seiner Arbeit für den Kunden können ihm auch gravierende Fehler unterlaufen, die das gesamte Projekt in Gefahr bringen und dem Kunden finanziellen Schaden zufügen. Im schlimmsten Fall führt dies dazu, dass auf Grund der unterlaufenen Fehler von diesem und anderen potenziellen Kunden weitere Projekte abgelehnt werden oder das Beratungsunternehmen für den entstandenen Schaden haftbar gemacht wird. In einem solchen Fall kann dieser schnell hohe Millionensummen betragen. Zudem sind Berater auch Repräsentanten des Beratungsunternehmens. Ein unqualifizierter Repräsentant kann zu einem erheblichen Imageschaden im Außenverhältnis (bei Kunden, potenziellen Kunden, Bewerbern oder Medien) beitragen. Eine Trennung von dem entsprechenden Mitarbeiter nach Ablauf der Probezeit führt letztlich zu zusätzlichen Kosten (z.B. Abfindung).

Die Ausführungen zeigen, dass die Beratungstätigkeit eine äußerst heikle Angelegenheit ist. Berater arbeiten oftmals an vertraulichen, sehr sensiblen und hoch erfolgskritischen Themen im Managementumfeld des Klienten für eine relativ hohe Vergütung. Dieser Kontext lässt kaum Spielraum für mögliche Fehler. Es herrscht quasi eine „Null-Toleranz“. Doch nicht nur die akzeptierten, sondern auch die abgelehnten Bewerber sind für die Unternehmensberatungen erfolgskritisch. Die dargestellten Nutzenmodelle, die in den bisherigen Untersuchungen (2.3.4) verwendet wurden, betrachten die abgelehnten Bewerber allerdings nicht. Sie gehen von folgender Annahme von Cronbach und Gleser (1965) aus:

> „Die Entscheidung, einen Bewerber zurückzuweisen, bedeutet meist, dass er keinen weiteren Kontakt mit der Institution hat. Wir können daher das Ergebnis einer solchen Entscheidung so betrachten, als habe es einen Wert von Null." (Cronbach und Gleser (1965, S.36ff.) übersetzt von Amelang und Zielinski (2002, S.453)

Diese Ansicht kann für die Personalauswahl bei Unternehmensberatungen nur bedingt gelten. Ein umfassender Ansatz müsste im Sinne der Opportunitätskosten auch bei den zu Unrecht abgelehnten Bewerbern den entgangenen Gewinn geltend machen, den diese Bewerber für das eigene Unternehmen erbracht hätten. Zudem sollte auch der mögliche Verlust betrachtet werden, den die Bewerber verursachen, wenn sie zu einer Konkurrenzberatung wechseln und für diese Projektakquisitionen, in Konkurrenz zu der Beratung, die sie vorher abgelehnt hat, gewinnen. Dieses Phänomen dürfte bei der Personalauswahl für Vertriebsstellen anderer Branchen gleichermaßen zutreffen. Neu ist hingegen folgende Überlegung: Ein Getränkehersteller lehnt einen Bewerber zu Unrecht ab. Der Bewerber, von seinen Qualitäten sehr überzeugt, ist verärgert über das „schlechte" Auswahlverfahren und das unprofessionelle Verhalten des Herstellers. Er beschließt deshalb, in Zukunft auf die Produkte des Herstellers zu verzichten. Der Hersteller, der jedes Jahr Getränke zu Millionen verkauft, wird eine solche Entscheidung verkraften können, sie vielmehr gar nicht merken. Insofern trifft die Aussage von Cronbach und Gleser zu und der Fehlentscheidung kann diesbezüglich ein Wert von Null zugewiesen werden. Passiert nun das gleiche bei einem Beratungsunternehmen, so können solche Entscheidungen dramatischere Folgen haben. Bewerber, die zu einem Auswahlverfahren von Beratungen eingeladen werden, haben in der Regel hervorragende Lebensläufe. Die Wahrscheinlichkeit, dass solche Personen in entsprechende Entscheidungspositionen kommen und somit zu potenziellen Klienten werden, ist nicht bekannt. Sie ist aber nach Meinung des Autors durchaus höher, als dass der im Beispiel des Getränkeherstellers abgelehnte Kandidat eines Tages eine Führungsposition bei einem großen Abnehmer der Getränke innehat. Selbst wenn der abgelehnte Bewerber später bei einem anderen Arbeitgeber nicht die Entscheidung über eine Projektvergabe trifft, so kann er doch organisationsintern für eine negative Einstellung gegenüber dem ehemaligen „Wunscharbeitgeber" sorgen. Zudem kann er als Multiplikator seine negativen Vorstellungen über das Beratungsunternehmen verbreiten. Dies alles kann sowohl Akquisitionsbemühungen, als auch operative Projektarbeiten behindern und somit zu ökonomischen Nachteilen führen. Ähnliche Überlegungen haben in der Praxis dazu geführt, dass zahlreiche Unternehmensberatungen Alumni-Netzwerke für ehemalige Mitarbeiter betreiben, mit deren Hilfe eine nicht unbeträchtliche Anzahl an Projektaufträgen gewonnen werden.

Die zuletzt geschilderte Problematik dürfte gleichermaßen auf die zu Recht abgelehnten Bewerber zutreffen. Die mangelnde Eignung einer Person für die Tätigkeit eines Beraters

bedeutet nicht zwangsläufig, dass diese Person nicht zu einem Entscheidungsträger bei einem potenziellen Kunden werden kann. Hier kommt einer geeigneten Kommunikation, die im Anschluss an die Entscheidung aufklärend und deeskalierend auf den abgelehnten Bewerber wirkt, eine wichtige Rolle zu. Sie kann dazu führen, dass der Bewerber trotz Ablehnung positive Assoziationen mit dem Unternehmen verbindet.

Die bisher angesprochenen Aspekte zeigen, dass die reine Betrachtung der akzeptierten Bewerber, wie sie von den oben besprochenen Nutzenmodellen vorgenommen wird, bei Beratungsunternehmen zu kurz greift. Zwar könnte man mit Hilfe der unter 2.3.3 geschilderten erweiterten Ansätze, wie z.B. MAUT, zu einer monetären Quantifizierung dieser Aspekte gelangen. Diese wären aber wiederum auf Basis von unpräzisen Expertenurteilen entstanden und somit eine weitere „Achillesferse“ in den Berechnungsansätzen zum Nutzen. Für eine erweiterte Betrachtung fehlen also derzeit geeignete Verfahren, die zu einer präzisen Ermittlung der entstehenden Kosten und Erlöse führen. Vor diesem Hintergrund ist eine Verwendung der bisherigen Ansätze auch bei Beratungsunternehmen sinnvoll, allerdings immer nur dann, wenn der Anwender die geschilderten Aspekte im Hinterkopf behält und bei der Ergebnisdarstellung auf die Prämissen hinweist.

Die Verwendung des klassischen Vorgehens bei der Nutzenberechnung von Auswahlverfahren für Unternehmensberatungen führt zwangsläufig zu dem umstrittenen Thema der SD_y-Schätzung. Wie unter 2.3.2 gesehen, gibt es zahlreiche Schätzmöglichkeiten von SD_y, die allerdings nicht immer die notwendige Präzision mitbringen, um eine breite Akzeptanz zu erreichen. Problematisch ist vor allen Dingen die Ermittlung der Beziehung von erbrachter Leistung und der monetären Bewertung, die über Expertenschätzungen vorgenommen wird. Holling und Reiners (1994) zeigten anhand von Verkaufszahlen bei Außendienstmitarbeitern, wie eine empirische Ermittlung von SD_y vorgenommen werden kann. Ein ähnlicher betriebswirtschaftlicher Ansatz kann auch, wie nachfolgend gezeigt, bei Unternehmensberatungen angewandt werden.

In Unternehmensberatungen wird die Leistungserbringung der Berater direkt in Rechnung gestellt, da diese auf Projekten zu einem bestimmten Tagessatz verkauft wird. Dies ermöglicht eine in allen Beratungen übliche, direkte Ermittlung von Produktivitätskennzahlen für jeden Mitarbeiter, indem man die durch ihn gearbeiteten und gebuchten Tage auf einem Projekt ins Verhältnis zur insgesamt zur Verfügung stehenden Arbeitszeit setzt. Diese direkte Ermittlung von y und der über alle Mitarbeiter hochgerechneten SD_y stellt eine elegante Lösung dar, die eine hohe Akzeptanz beim Management innerhalb der Unternehmen genießen dürfte, da sie u.a. auch als Bemessungsgrundlage für variable Gehaltsbestandteile

herangezogen wird. Der Nutzenzuwachs durch Produktivitätsgewinne ist dabei nahezu voll ergebniswirksam, da die variablen, durch den Produktivitätsgewinn entstehenden Kosten durch Nebenkostenbudgets vom Kunden zusätzlich übernommen werden. Eine Ausnahme bildet dabei der variable Gehaltsbestandteil der Mitarbeiter, der zwar teilweise abhängig von der Produktivität, aber in seiner absoluten Höhe im Vergleich zu den Gesamterträgen eher vernachlässigbar ist. Die Nebenkostenbudgets dienen darüber hinaus zur Deckung von internen Leistungen der Serviceabteilungen (z.B. Research, Grafik), so dass durch sie sogar ein Teil der Fixkosten reduziert wird. Dies bedeutet, dass das Heranziehen der Tagessätze ohne Nebenkostenbudgets als Ergebnisverbesserung (vor Steuern) des Produktivitätszuwachses, eine konservative Schätzung darstellt. Zusätzlich zu den spezifischen Prämissen des angewandten Nutzenmodells müssen allerdings noch folgende zwei Annahmen gelten:

1. *Die Produktivität des Beraters wird ausschließlich durch seine Fähigkeiten bedingt*.

Diese Annahme lässt sich sicherlich in der Praxis nicht verifizieren. Die eigene Produktivität ist i.d.R. nur bedingt durch den Berater beeinflussbar und durch zahlreiche andere Faktoren bestimmt. Dazu gehören z.B. indirekt die Marktnachfrage nach den Leistungen des Fachbereiches, in dem der Berater tätig ist und die Akquisitionsleistung des Vorgesetzten (Mentors). Zu den direkten Faktoren zählen die interne Arbeitsverteilung, die Verfügbarkeit, vorherige Projekterfahrungen und der Ruf und Bekanntheitsgrad des Beraters im eigenen Unternehmen. Demgegenüber verkaufen sich erfolgreiche Berater auf den Projekten über Verlängerung des Auftrages selbst weiter und viele der oben genannten Aspekte sind auch von Beraterfähigkeiten abhängig. Es ist deshalb von einem Zusammenhang zwischen Beraterfähigkeiten (wenn auch nicht ausschließlich) und Produktivität auszugehen.

2. *Der ökonomischen Erfolgsbeitrag des Beraters ist ausschließlich durch seine externe fakturierte Leistungserbringung auf den Projekten bestimmt.*

Auch diese Aussage ist nur bedingt korrekt. Berater arbeiten oft an Produktentwicklungen und Akquisitionen, die essentiell für den späteren Verkauf von Projekten und damit für den ökonomischen Erfolg vom Unternehmen sind. Da diese Tätigkeit allerdings i.d.R. alle Berater gleichermaßen betrifft und nur einen geringen Teil der Arbeit ausmacht, ist eine Vernachlässigung vertretbar.

Alles in allem ermöglichen die dargestellten Überlegungen, unter Berücksichtigung der genannten Prämissen, eine Berechnung des Nutzens, die eine höhere Akzeptanz im Managementbereich erreichen dürfte als die Berechnungen in den bisherigen Studien. Weniger bedeutsam sind dabei allerdings die exakten absoluten Beträge des Nutzens, die durch die

dargestellten Restriktionen eher einen fiktiven Charakter aufweisen dürften. Als Orientierungsgröße möglicher Effekte und als Vergleichsmaßstab dürften sie hingegen, nach Meinung des Autors höchst brauchbar sein.

2.3.6 Zusammenfassung

Die Darstellungen des laufenden Kapitels haben gezeigt, dass der Wert eines Auswahlverfahrens von zahlreichen situativen Aspekten und den damit zusammenhängenden Parametern (Selektionsrate, Basisrate, SD_y usf.) abhängt und nicht nur das alleinige Validitätskriterium bei der Auswahl des Verfahrens gelten kann. Der Einsatz von weniger validen Verfahren kann, bei einer breiteren Einsatzmöglichkeit (Breitbandverfahren) u. U. sogar zu einem höheren Gesamtnutzen für die Organisation führen als die nur eng einsetzbaren hoch validen Schmalbandverfahren. Darüber hinaus dürfen Auswahlverfahren wie gezeigt nicht anhand ihres absoluten Nutzen, sondern immer relativ zu bereits vorhandenen Informationen und der besten a-priori-Strategie bewertet werden (Kompa, 1989). Dies bedeutet, dass der inkrementelle Nutzen der Verfahren für ihren Einsatz entscheidend ist.

Die über einfache Kennzahlen hinausgehenden und aufeinander aufbauenden, historischen Ansätze der Nutzenanalyse von Taylor und Russel (1939), Sands (1973), Cronbach und Gleser (1965) und Brogden (1949) bilden die Realität in nur unzureichendem Maße ab. Sie basieren auf zahlreichen, oftmals nicht überprüfbaren Annahmen und enthalten wesentliche Parameter nicht. Zudem ist die Operationalisierung der Parameter, insbesondere der Standardabweichung der Produktivität SD_y, der Achillesferse der Entscheidungstheorie, problematisch. Dabei werden nach Boudreau (1991) drei Ansätze zur Ermittlung des „payoff“ durch Produktivitätsgewinn unterschieden: die *Kostenreduktion*, der *Verkaufswert*- und der in den meisten Fällen adäquate *Gewinnbeitragsansatz*. SD_y kann dabei in den häufigsten Fällen nicht direkt gemessen, sondern muss anhand von Schätzverfahren bestimmt werden. Schuler et. al. (1995) gliedern diese in die *Kostenrechnungsmethode*, *proportionale Regeln*, dem *CREPID-Verfahren* und den am häufigsten angewandten *globalen Schätzverfahren*. Das neuere Modell von Boudreau (1983b) integriert in die Nutzenbetrachtung das Kapitalwertmodell der dynamischen Investitionsrechnung. Obgleich das Modell durch seine innere Logik und klare Definition überzeugt, hängt seine Brauchbarkeit von der Realitätsnähe der Annahmen und der Konstanz der Bedingungen ab. Das Boudreau-Modell quantifiziert zudem, wie alle anderen Modelle auch, den ökonomischen Nutzen lediglich anhand von Produktivitätsgewinnen. Es sind jedoch auch zahlreiche weitere ökonomische Vorteile durch eine bessere Personalauswahl denkbar (z.B. geringere Trainingskosten). Neben den statistischen Modellen gibt es noch eine Reihe weiterer Alternativen, die zur Entscheidungsfindung

herangezogen werden können. Dazu gehören z.B. *Benchmarking*, *MAUT* und *Balanced Scorecard*.

Im Kontext der Personalauswahl sind bisher eher wenig empirische Arbeiten veröffentlicht worden, hauptsächlich bedingt durch die hohe Komplexität und die mangelnde Akzeptanz der Nutzenmodelle. Keine der Arbeiten beschäftigt sich dabei mit Nutzenberechnung bei Unternehmensberatungen, die besondere Implikationen mit sich bringt. Diese betreffen hauptsächlich die Konsequenzen personeller Fehlentscheidungen, die wesentlich gravierender ausfallen können als bei anderen Unternehmen, und die vereinfachte Ermittlung der SD_y, die direkt gemessen werden kann. Allerdings müssen die Ergebnisse immer vor dem Hintergrund einiger beratungsspezifischer Prämissen beurteilt werden.

2.4 Berufserfolg in Unternehmensberatungen

Im vorherigen Kapitel wurde bereits angedeutet, dass die Tätigkeit der Unternehmensberatung eine Vielzahl an Facetten beinhaltet, an denen Erfolg gemessen werden kann. Die klassischerweise herangezogenen Kriterien der Beförderung oder des Gehaltes geben dabei nur bedingt Auskunft, wie erfolgreich ein Mitarbeiter ist. Die strenge Beförderungspolitik im Sinne eines "up or out", wie sie in vielen Beratungen betrieben wird, lässt alle verbliebenen Mitarbeiter als erfolgreich erscheinen, da keine Varianz mehr vorliegt. Dieses Phänomen wird durch eine relativ kurze durchschnittliche Verweildauer der Mitarbeiter zusätzlich verstärkt. Zudem sind innerhalb der Hierarchiestufen die Gehälter weitgehend homogen, so dass auch hier keine notwendige Differenzierung vorhanden ist. Eine Möglichkeit, den Erfolg zu quantifizieren, stellt immer dann, wenn die genannten Prämissen vorliegen, die Produktivität des einzelnen Beraters dar. Da die Einhaltung der Prämissen aber eher unwahrscheinlich ist, ist sie nach Meinung des Autors zwar ein guter Anhaltspunkt zur Nutzenberechnung (siehe 2.3), weil sie grundsätzlich die Leistung von Beratern quantifizierbar macht. Für die Beurteilung des Erfolges von Mitarbeitern allerdings ist es sinnvoller, „eine Ebene tiefer“ zu gehen und die Frage zu stellen, wann jemand eine erfolgreiche Beratungsleistung erbringt (zunächst unabhängig davon, ob er dabei im Sinne des Unternehmens produktiv ist oder nicht). Die dabei eingenommene Perspektive orientiert sich an der externen, durch den Berater erbrachten Leistung auf Projekten beim Kunden, da diese den Hauptteil seiner Tätigkeit ausmachen dürfte. Aus diesem Grund sollen im Folgenden auf das Wesen der Unternehmensberatung näher eingegangen und mögliche Modelle zur und Probleme bei der Bewertung von Beratungsleistungen skizziert werden. Das Kapitel 2.4 mündet in einer Ableitung von Kernelementen erfolgreicher Beratung. Diese Kriterien sollen im späteren empirischen Teil als Vergleichsbasis für das verwendete Kriterium der untersuchten Unter-

nehmensberatung dienen. Somit können Anhaltspunkte darüber gewonnen werden, ob inhaltlich alle relevanten Aspekte des Beratungserfolges abgedeckt sind und ob sich die zur Messung verwendeten Instrumente (in dieser Studie das Vorgesetztenurteil) dafür eignen.

2.4.1 Unternehmensberatung – Begriff, Wesen und Funktionen

Eine einheitlich akzeptierte Definition des Begriffes *Unternehmensberatung* ist bisher in der Literatur ausgeblieben. Vielmehr wird immer wieder auf die Vielzahl der Begriffssynonyme und Definitionen hingewiesen (Elfgen & Klaile, 1987; Ernst, 2002; Kröber, 1991; Schrädler, 1996; Steyrer, 1991). Die Gründe hierfür liegen in der großen Heterogenität der Ausprägungen der Unternehmensberatungen (Strasser, 1993) hinsichtlich ihrer *Tätigkeiten* (IT-, Marketing-, Strategie-, Restrukturierungsberatung usf.), ihrer *Größe,* der wahrgenommenen *Beraterrolle* (Gutachterberatung, systemische Beratung, Organisationsentwicklung und Expertenberatung, Walger, Achatzi & Benkert, 1999) sowie ihrer unscharfen, wenig abgegrenzten zersplitterten *Anbieterstruktur*. Zudem verhindert die uneindeutige disziplinäre Zuordnung der Thematik (BWL, Psychologie, Soziologie, Wirtschaftsgeschichte usw.) die Integration der Facetten in eine einheitliche „Theorie der Beratung“ (Ernst, 2002, S.10). Eine gute Übersicht möglicher Definitionen findet sich in Sangüesa (2003, S. 15ff.). Die vorliegende Arbeit orientiert sich für den Begriff der (Unternehmens-)Beratung (als Handlung) an folgender Definition:

> „Unter Unternehmensberatung verstehen wir eine von einem unabhängigen, eigenverantwortlichen, professionellen Berater individuell für die Klientenorganisation marktmäßig erbrachte Dienstleistung, welche darauf ausgerichtet ist, in einem interaktiven Prozess gemeinsam mit dem Klienten ein Lösungskonzept für eine komplexe betriebswirtschaftliche Problemstellung zu erarbeiten und auf Wunsch auch dessen Implementierung zu unterstützen“ (Hoffmann, 1991, S. 40)

Die Zuordnung zu den Dienstleistungen ermöglicht eine weitergehende Beschreibung der im Vergleich zu anderen Dienstleistungen komplexen Unternehmensberatung, die auch Konsequenzen für die Evaluation nach sich ziehen (siehe Ernst, 2002, S.10ff. und Sangüesa, 2003, S.39ff.). *Tabelle 12* gibt eine Übersicht der relevanten Charakteristika für die Unternehmensberatung.

Tabelle 12: Charakteristika der Dienstleistung Unternehmensberatung (in Anlehnung an Ernst (2003))

Charakteristik	Inhalt
Qualifikations-differenz	Berater verfügt über ein von Kundenseite her defizitäres, abstraktes Wissen, das bei der Generierung von Handlungsempfehlungen zur Problemlösung hilfreich ist (Selchert, 1997, S.30f.).
Interaktion	Probleme werden gemeinsam von Beratern und Klienten definiert und gelöst (Elfgen & Klaile, 1987, S.29f.). Die problemzentrierte Beratungssicht wird somit um eine soziale Komponente als Fundament der Glaubwürdigkeit der Qualifikationsdifferenz erweitert.
Immaterialität	Die vom Berater erbrachte Dienstleistung verfügt über keine physische Präsenz (Selchert, 1997, S.37f.).
Singularität	Der wandelnde Kontext der erbrachten Leistung macht sie einzigartig (Elfgen & Klaile, 1987, S.31; Selchert, 1997, S.36).
Indeterminier-barkeit	Beratungsleistungen sind in ihrer Wirkung von den Beteiligten nicht vollständig vorher-sehbar (McKinley & Scherer, 2000) und die beeinflussenden Faktoren größtenteils nicht kontrollierbar.

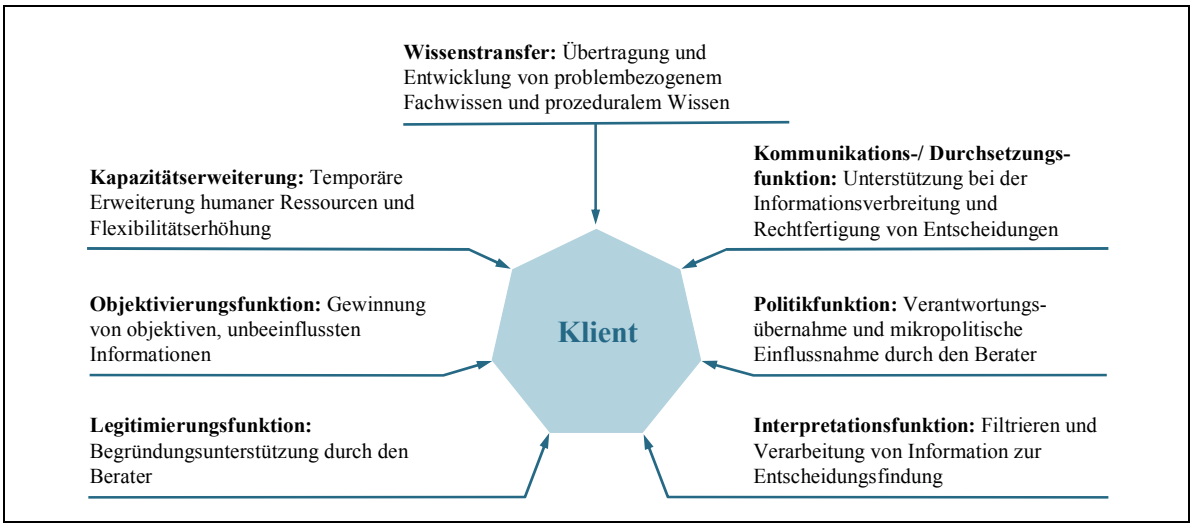

Abbildung 26: Funktionen von Beratungen (in Anlehnung an Ernst, 2002)

Betrachtet man nun die Wirkung von Unternehmensberatungsaktivitäten, so lassen sich diese nach Kieser (1998) in *offizielle* und *latente* Beratungsfunktionen unterscheiden. *Latent* sind diese dann, wenn sie in der offiziellen Begründung des Einsatzes der Berater nicht vorkommen, obwohl sie dennoch Bestandteil der Bewertung der Beratungsleistung durch die Beteiligten sind. *Abbildung 26* zeigt eine Übersicht der wichtigsten Beratungsfunktionen. Die dargestellten Charakteristika und Funktionen können als Anhaltspunkte für eine Bewertung der Beratungsleistung herangezogen werden. Welche grundsätzlichen Probleme dabei auftreten, soll nachfolgend dargestellt werden.

2.4.2 Zur Problematik der Bewertung von Beratungsleistungen

Die im Vergleich zu anderen Dienstleistungen höhere Komplexität der Beratung führt zu Besonderheiten, die es erschweren, diese Leistung angemessen zu evaluieren (Ernst, 2002, S.51ff.). Betrachtet man die im vorherigen Kapitel angesprochene *Qualifikationsdifferenz,* so steht der Klient als beurteilende Instanz bei mangelndem eigenem Know-how (das evtl. sogar dazu geführt hat, dass er einen Berater engagiert) zunächst vor der Schwierigkeit, die Güte eines Ratschlages des Beraters zu beurteilen. Auch ein Abwarten bis zur Umsetzung der Entscheidung kann, bei mangelnder Kenntnis der Folgen alternativer Entscheidungen, zu keinem objektiven Urteil führen (Starbuck, 1992). Ernst (2002, S.52) weist auf Parallelen dieser Problematik zur Informationsasymmetrie bei der Agenturtheorie hin (Eisenhardt, 1989), bei der der Prinzipal den Agenten mit einer Tätigkeit beauftragt, die dieser besser ausführen kann, als er selbst. Bei Informationsasymmetrie befindet sich der Prinzipal in einer Umwelt, über die er nicht vollständig informiert ist. Insbesondere die Eigenschaften, Handlungen, Informationen und bzw. oder die Absichten des Akteurs sind ihm nicht bekannt (z.B. durch „impression management“). Dies kann zu einem opportunistischen Verhalten des Agenten führen. Im Beratungsfall kann es sogar darin münden, dass der Berater, auf Grund der Informationsasymmetrie seine Aufgaben selbst definiert und der Beratungsbedarf durch den Prinzipal nicht abschätzbar ist (Ernst, 2002, S.52).

Die *Interaktionssituation* zwischen Berater und Klienten führt dazu, dass die Erfolgs- und Misserfolgsattribution nicht eindeutig vorgenommen werden kann. Der Anteil des Beraters am gemeinsamen Erfolg ist damit nicht isolierbar (Ernst, 2002, S.53). Zudem ist die Beurteilung des Beratungserfolges durch den Klienten maßgeblich durch die Qualität der Beziehung (wie etwa Sympathiewerte) zu dem Berater gefärbt (Clark, 1995). Ähnlich problematisch ist die *Immaterialität* der Leistung, die ein Aufstellen klarer und von allen Beteiligten geteilter Kriterien der Qualitätsbeurteilung erschwert. Zusätzlich betreffen durch die zeitliche und persönliche *Indeterminierbarkeit* der Leistung die Folgen einer Beratung eine derartige Vielzahl von Variablen, dass eine vollständige Identifizierbarkeit und Erfassung dieser nicht praktikabel ist. Zudem werden die Ergebnisse durch makroökonomische Faktoren überlagert. Die *Singularität* schließlich erschwert das Heranziehen von Referenzprojekten als Beurteilungsmaßstab, da ihre Einzigartigkeit Beurteilungsspielräume zulässt (Ernst, 2002, S.54).

Die geschilderten Aspekte sind sicherlich mit ein Grund dafür, dass systematische Beurteilungen (im Sinne einer Zielvereinbarung und späteren Überprüfung) der Beraterleistungen auf Kundenseite eher selten durchgeführt werden (Strasser, 1993) und Beratungser-

folge besonders auf Kundenseite als schwer messbar gelten (Kaas & Schade, 1995). Kailer und Scheff (2000) untersuchen die Häufigkeit verwendeter Evaluierungsmethoden auf Kundenseite bei kleineren und mittleren Unternehmen (KMU) (*Tabelle 13*).

Tabelle 13: Evaluierung des Beratungserfolges (Kailer und Scheff (2000), S.61) (N=372)

Methoden	**Häufigkeit [%]**
Abschlussbesprechung	98.1
Erarbeitung/Präsentation eines Beratungsberichtes	71.2
Nachfrage im Unternehmen einige Zeit nach dem Projekt	67.3
Kennzifferververgleiche	26.9
Interviews/Fragebogenerhebung	25.0
Interviews/Fragebogenerhebung einige Zeit nach dem Projekt	15.4
Erhebung durch Dritte	13.5
Andere Formen	11.5

Die in *Tabelle 13* genannten Instrumente und Vorgehensweise dürften, auch wenn sie sicherlich nicht systematisch allumfassend sind, bei ihrem Einsatz wenige Probleme verursachen. Problematischer hingegen ist, aus den oben genannten Gründen, die Bestimmung der Inhalte, also der Kriterien, an denen Projektfortschritte und Beratungserfolge gemessen werden sollen. Hier sind häufig Abstimmungsdefizite und Unstimmigkeiten zwischen Beratern und Klienten feststellbar (Kailer & Falter, 2000, S.144f.). Welche Kriterien überhaupt zur Evaluation von Beratungsleistungen herangezogen werden können, soll im nächsten Abschnitt näher betrachtet werden.

2.4.3 Ableitung von Kriterien erfolgreicher Beratung

Insgesamt ist der derzeitige Forschungsstand zur Evaluation von Beratungsleistungen als lückenhaft zu bezeichnen, stellenweise ist sogar von „Forschungswildwuchs“ die Rede (Steyrer, 1991, S.21). Die dabei abgehandelten Themen betreffen hauptsächlich die Zufriedenheit mit der Beratungsleistung, Auswahlkriterien und Charakteristika guter Berater sowie die Evaluierung der Beratung. Eine ausführliche Übersicht empirischer Studien, ihrer Inhalte und ihrer Kritik findet sich in Sangüesa (2003, S.71ff.). Ein Großteil der Untersuchungen steht in der Tradition funktionalistischer Ansätze, die die Perspektive der Feststellung von kausalen Zusammenhängen zwischen Beratungsleistung und Unternehmenseffektivität einnehmen (z.B. Fritz & Effenberger, 1998; Gable, 1996; Hoffmann, 1991; Klein, 1978). Diese betrachten nicht selten nur Teilaspekte und sind zudem häufig mit methodischen Fehlern

behaftet. Außerdem bleiben sie oftmals rein deskriptiv oder genügen nicht wissenschaftlichen Ansprüchen. Aus diesem Grund sind sie für die Übernahme von Kriterien nur bedingt tauglich. Im Gegensatz zur funktionalistischen Sicht orientiert sich z.B. die Arbeit von Ernst (2002) am Konstruktivismus und betrachtet die Evaluation aus sozial-kognitiver Beurteilungsperspektive des Klienten. Für die Ableitung von Kriterien erfolgreicher Beratung ist sie allerdings auch wenig geeignet, da sie, neben dem eingeschränkten Klientenfokus, auf das Zustandekommen von Bewertungen, also die Urteilsbildung fokussiert.

Ein umfassender Ansatz, der sowohl eine Prozess- als auch eine Ergebnisbetrachtung vornimmt, stammt von Sangüesa (2003). Mithilfe der Charakteristika einer Dienstleistung (*Tabelle 12*), der Qualitätsdefinition aus der ISO 9000 (DIN EN ISO 9000:2000) und verschiedener Modelle zur Dienstleistungsqualität (Donabedian, 1980; Grönross, 1982; Meyer & Mattmüller, 1987; Zeithaml, 1981), leitet sie Anforderungen für die Entwicklung eines Evaluierungsmodells für Beratungsleistungen, das *BPEK-Modell*, ab. Das Akronym steht dabei für die einzelnen zu betrachtenden Perspektiven bei der Evaluation (**B**erater, **P**rozess, **E**rgebnis, **K**unde (Klient)). Aus den oben genannten Studien extrahiert Sangüesa (2003, S.123ff.) in einem weiteren Schritt die verwendeten Kriterien, beurteilt ihre Brauchbarkeit und ordnet sie, wenn möglich, den vier Modellperspektiven zu. Zusätzlich gruppiert sie die Kriterien innerhalb der Perspektiven zu Oberbegriffen. Bei einer Phasenbetrachtung eines Problemlöseprozessmodells als postuliertem Beratungsprozess bedient sie sich der Technik des *Blueprinting* und *sequentiellen Ereignismethode* (SEM), um die gruppierten Kriterien zuzuordnen.

Sangüesas Kriterienzuordnung ist sehr subjektiv und die Oberbegriffsbildung verwirrend, da sie sich gegenseitig bedingen und teilweise überlappen. Außerdem bezeichnet sie zum Teil Hypothesen als Kriterien (Sangüesa, 2003, S.123ff.), was nicht im Einklang mit der wissenschaftlichen Kriteriendefinition steht. Hinzukommt eine, aus Sicht des Autors unglückliche Vermengung von *Eigenschafts*-, *Verhaltens*- und *Ergebniskriterien* innerhalb der einzelnen Perspektiven, die die Nachvollziehbarkeit des Vorgehens zusätzlich erschwert. Zudem führt die Anwendung von *Blueprinting* und der *SEM* zu einer sehr hohen Komplexität, die nicht an wissenschaftlichem Gehalt gewinnt. Innerhalb der Anwendung dieser Techniken verläuft die Zuordnung von Kriterien zu einzelnen Phasen des Beratungsprozesses offensichtlich nicht nach festgelegten Regeln, sondern willkürlich subjektiv. Der Ansatz, eine Kriterienbetrachtung aus der *BPEK*-Perspektive vorzunehmen, ist allerdings aus Sicht des Autors durchaus sinnvoll. Deshalb wird er auch im Folgenden vorgenommen und ebenso

auf dem vom American Supplier Institute (ASI) entwickelten *Problemlöseprozess*[10] (Sangüesa 2003, S.35) zurückgegriffen. *Abbildung 27* zeigt diesen im Zusammenhang mit den anderen vier Perspektiven und der vom Autor zusätzlich eingeführten Perspektive des *Kontextes*, da dieser eine moderierende Wirkung auf die anderen Dimensionen haben kann.

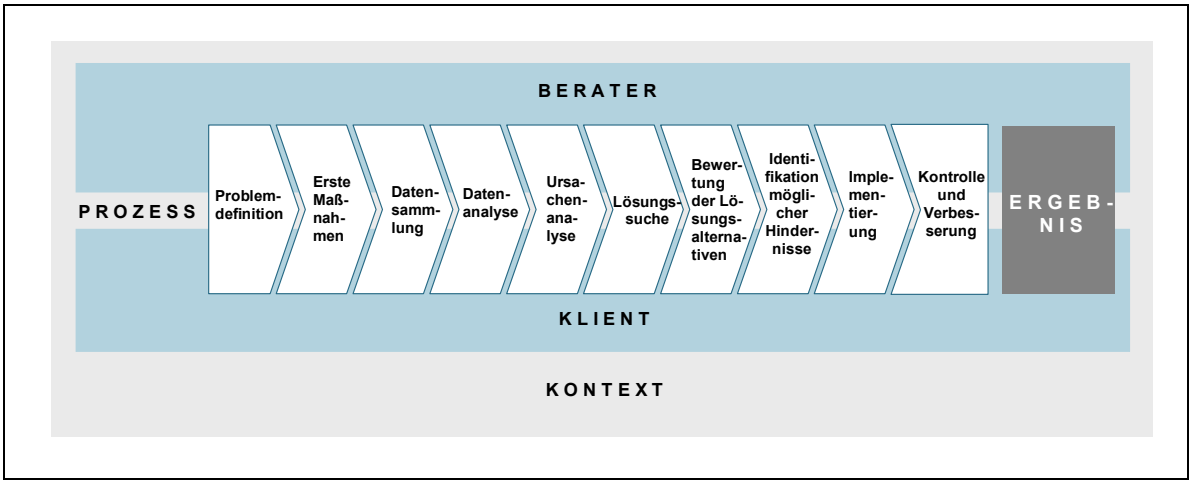

Abbildung 27: Modell zur Herleitung der Kriterien erfolgreicher Beratung

Die Darstellung verdeutlicht, dass der Prozess, an dessen Ende das Ergebnis der Beratung bzw. des Projektes steht, in der Interaktion zwischen Berater und Klient abgearbeitet wird. Dabei spiegeln nach Meinung des Autors diese Dimensionen exakt die Trilogie der Kriterien (*Eigenschafts*-, *Verhaltens*- und *Ergebniskriterien*) aus *Abbildung 7* wider. Dem Berater und dem Klienten können dabei *Eigenschaftskriterien* zugeordnet werden, dem Prozess die *Verhaltenskriterien* und dem Ergebnis konsequenterweise die *Ergebniskriterien*. Darüber hinaus spielt der projektspezifische Kontext eine moderierende Wirkung. So können beispielsweise gleiche Ergebnisse (wie z.B. eine Ergebnisverbesserung beim Klienten von 10 Mio. EUR) vor unterschiedlichen Hintergründen (z.B. geringe vs. angemessene Ressourcen auf Klientenseite) verschiedene Interpretationen zulassen. Der Autor hat die in der Arbeit von Sangüesa durchgeführte Zuordnung der Kriterien der untersuchten und zusätzlicher Studien nach dieser Systematik wiederholt vorgenommen. Darüber hinaus hat er weitere, auch aus dem O*Net abgeleitete Kriterien hinzugezogen. Gängigerweise sollten die Kriterien im Rahmen einer AA erstellt werden. Da dies in der vorliegenden Arbeit nicht möglich war, ist auf die geschilderte Alternative zurückgegriffen worden. Der Rückgriff auf die Literatur sollte dabei helfen, keine wesentlichen Aspekte bei der Genese der Kriterien nach dem

[10] In der von der ASI entwickelten Darstellung handelt es sich um einen Stern, weshalb der Ausdruck *Problemlösestern* verwendet wird. In der vorliegenden Arbeit wird aus Visualisierungsgründen die Prozessdarstellung gewählt.

Modell aus *Abbildung 27* zu vernachlässigen. Der Rückgriff auf das O*Net bei der Generierung von Kriterien stellt eine akzeptierte Vorgehensweise dar (Peterson, et al., 2001). Eine Übersicht der aus der Literatur übernommenen Kriterien findet sich in *Anhang III*, die Gruppierung, Erweiterung und Zuordnung der einzelnen Kriterienarten in *Anhang IV*. Die durch den Autor postulierten Oberdimensionen wurden mit dem Zweck entwickelt, einen umfassenden Blick der notwendigen Betrachtungswinkel bei der Evaluation von Beratungsleistungen zu ermöglichen und die Kriterienvielfalt sinnvoll zu gruppieren. Die Kategorisierung verlief dabei nach subjektiven, dem Autor sinnvoll erscheinenden dargestellten fünf Dimensionen (siehe *Abbildung 28*) unter Berücksichtigung relevanter Literatur (siehe *Anhang III*). Diese sollten in einem weiteren Schritt für die Eigenschaftskriterien faktoranalytisch verifiziert werden, um von tatsächlichen Leistungskonstrukten ausgehen zu können. Als Beurteilungsbasis, ob alle relevanten Aspekte bei einer vorliegenden Leistungsevaluation berücksichtigt wurden, wie im empirischen Teil vorgenommen, scheinen sie dem Autor jedoch höchst brauchbar. Die kann dabei helfen die Konstruktvalidität des Kriteriums (R_{KR} aus *Abbildung 18*) zu maximieren. *Abbildung 28* zeigt abschließend die so aus der Analyse entstandenen Kriterien. Bei der Kategorie *Prozess* sind lediglich die Oberbegriffe dargestellt. Für eine Kriteriendetaillierung sei erneut auf *Anhang IV* verwiesen.

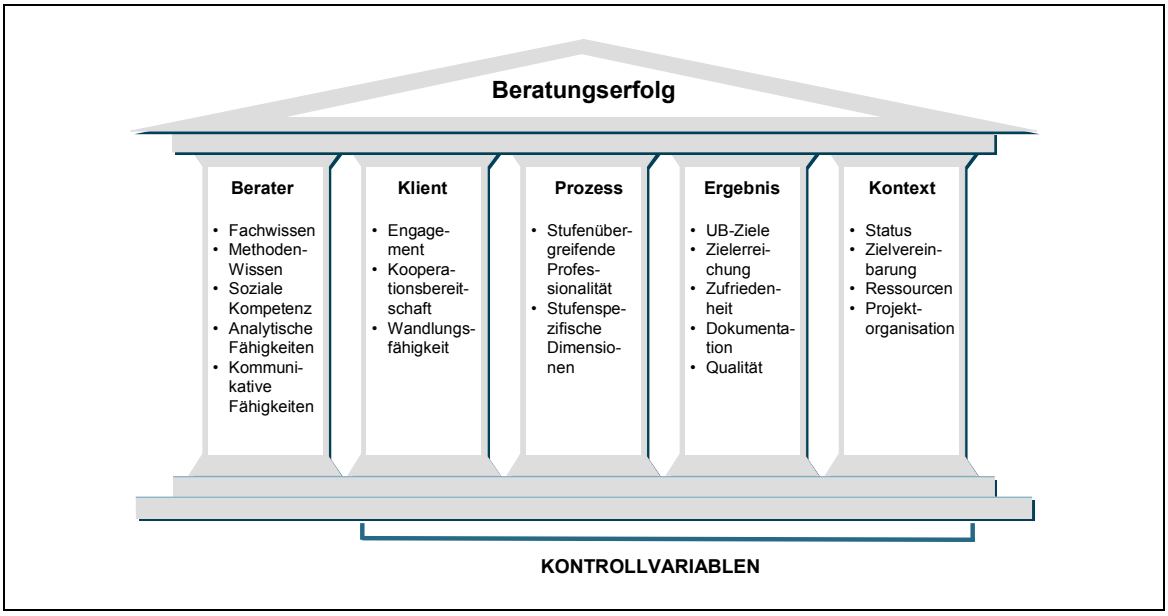

Abbildung 28: Kriterienmodell der fünf Säulen erfolgreicher Beratung

Als Surrogat für die beschriebenen, sehr facettenreichen Kriterien ist zu Evaluationszwecken auch das Kriterium des forcierten Ausscheidens möglich (*Forced Exit*). Im Gegensatz zu anderen Branchen wird in Unternehmensberatungen üblicherweise das strenge „up-

or-out“-Prinzip angewendet (s.o.). Dies bedeutet, dass Mitarbeiter dann das Unternehmen verlassen müssen, wenn sie nach einer bestimmten Zeit die nächste Beförderungshürde nicht überwinden. Üblicherweise ist das der Fall, wenn die Leistung der Stufe, die der Berater hat, nicht entspricht, oder dem Berater eine gute Leistung auf der nächsten Stufe nicht zugetraut wird. Folglich ist der Verbleib auf einem Hierarchielevel auf Dauer nicht möglich. Dies führt dazu, dass forciert ausgeschiedene Mitarbeiter erheblich häufiger bei untersuchten Unternehmensberatungen vorzufinden sind als bei anderen Unternehmen, so dass die Variable ausreichend Varianz in den Untersuchungsstichproben aufweist. Damit eignet sich diese Variable als dichotomes Kriterium für Beratungserfolg.

2.4.4 Implikationen für die vorliegende Studie

Betrachtet man erneut die Kriterienseite von *Abbildung 14*, lassen sich aus den gewonnenen Erkenntnissen der letzten Abschnitte Implikationen für Validitätsstudien bei der Personalauswahl von Unternehmensberatungen ableiten. Zur Erreichung einer hohen Konstruktvalidität des Kriteriums R_{KR}, ist die Berücksichtigung aller genannten Inhalte aus dem *Fünf-Säulen-Modell* empfehlenswert. Nur so kann sichergestellt werden, dass auch tatsächlich das erfasst wird, was unter Beratungserfolg verstanden werden kann. Umfasst das verwendete Kriterium nur Teilbereiche, so besteht das Risiko, dass es zu einer asymmetrischen Beziehung zwischen den untersuchten Prädiktoren und Kriterien kommt. Damit besteht die Gefahr der Verzerrung der empirisch ermittelten Validitäten. Zudem ist das Wirksamwerden zusätzlicher Moderatoreffekte nicht auszuschließen. Gleiche Leistungen von zwei Beratern können, bei sehr unterschiedlichen Klienten oder stark differierenden Kontexten, zu völlig unterschiedlichen Kriterienausprägungen führen. Eine Vernachlässigung solcher Gegebenheiten führt zu einem verdeckten Wirksamwerden dieser Effekte (*NTR-Box*) und somit zu Validitätsberechnungen, die nicht der Güte des angewandten Auswahlverfahrens entsprechen. Folglich sollte überprüft werden, inwieweit das verwendete Kriterium die fünf Säulen des Beratungserfolges angemessen widerspiegelt, bzw. Die in *Abbildung 28* genannten Kontrollvariablen berücksichtigt werden. Für den empirischen Teil dieser Arbeit bedeutet dies, dass das Kriterium *Vorgesetztenurteil* hinsichtlich seiner inhaltlichen Dimensionen mit den fünf Säulen abgeglichen werden muss. *Abbildung 29* visualisiert abschließend dieses Vorgehen .

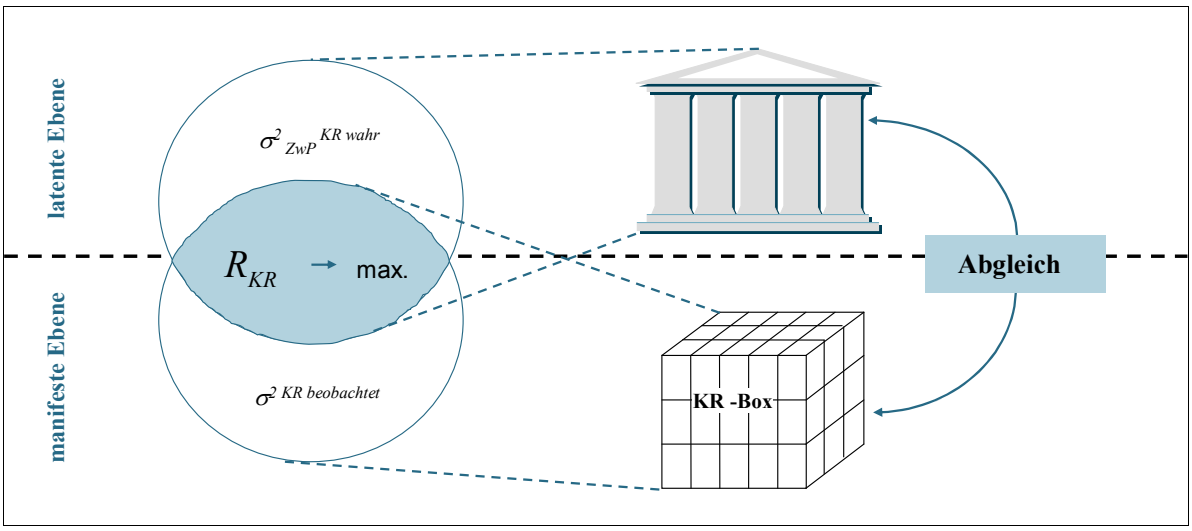

Abbildung 29: Implikationen für die Evaluationsstudie

2.4.5 Zusammenfassung

Die klassischen Kriterien des Berufserfolges wie Gehalt oder Beförderungslevel sind bei Unternehmensberatungen auf Grund ihrer „up-or-out"-Beförderungspolitik und der Homogenität der Gehälter innerhalb einer Stufe nur begrenzt einsetzbar. Eine Alternative stellt die Betrachtung projektbezogener Beratungstätigkeit zur Ableitung von Kriterien, die diese als erfolgreich einstufen lassen, dar. Beratungstätigkeiten sind Dienstleistungen und als solche sind sie durch die Charakteristika der *Qualifikationsdifferenz*, *Interaktion*, *Immaterialität*, *Singularität* und *Indeterminierbarkeit* beschreibbar. All diese Eigenschaften führen dazu, dass der Erfolg dieser Dienstleistung nicht exakt messbar oder zuordenbar ist und deshalb, wie zahlreiche Studien belegt haben, auch in der Praxis kaum berücksichtigt wird. Das Heranziehen des O*Net und die Analyse von zahlreichen Studien zur Evaluation von Beratungsleistungen helfen bei einer Ordnung der Kriterienvielfalt. Die verwendeten *Verhaltenskriterien* können dabei dem *Beratungsprozess*, die *Eigenschaftskriterien* dem *Berater* oder dem *Klienten* und die *Ergebniskriterien* dem *Ergebnis* zugeordnet werden. Zudem ist eine erfolgreiche Beratung auch vom *Kontext* abhängig, der als fünfte Kriteriumsperspektive das *Fünf-Säulen-Modell* erfolgreicher Beratung komplettiert. Innerhalb dieser fünf Perspektiven können nun, wie vom Autor gezeigt, die Kriterien zu Oberdimensionen zusammengefasst werden. Alternativ kann auch forciertes Ausscheiden als dichotomes Surrogat für Beratungserfolg herangezogen werden. Das Betrachten der Oberdimensionen innerhalb der Perspektiven soll bei der Beurteilung helfen, ob ein verwendetes Leistungsbeurteilungssystem alle relevanten Aspekte beinhaltet und auf geeignete Weise operationalisiert (hohe Konstruktvalidität). Im empirischen Teil der Arbeit wird genau dies für die vorliegende Studie vorgenommen.

2.5 Theoretische Perspektiven und empirische Befunde

Im letzten großen Abschnitt des theoretischen Teils soll auf die für die Evaluation des vorliegenden Auswahlverfahrens wesentlichen empirischen Befunde eingegangen werden. Dabei werden zunächst unter 2.5.1 auf die gängigen Erkenntnisse und Theorien zu den Merkmalen einiger, in *Abbildung 11* bereits genannter Prädiktoren, erläutert. Im Anschluss werden unter 2.5.2 ausgewählte empirische Befunde zu den Prädiktoren dargestellt, bevor unter 2.5.3 auf die Validität und die Eigenschaften ausgewählter Instrumente zur Operationalisierung der Prädiktoren eingegangen wird.

2.5.1 Theoretische Grundlagen zu ausgewählten Prädiktoren des Berufserfolges

Es gibt eine große Zahl möglicher Prädiktoren, die zur Vorhersage von Berufserfolg oder zur Hypothesenbildung über Zusammenhänge mit Berufserfolg herangezogen werden können. Aus der Erfahrung heraus lassen sich sicherlich mit Unterstützung einer AA, verschiedene relevante Prädiktoren ermitteln. Diese sind aber oftmals von hohem Allgemeinheitsgrad, nicht überschneidungsfrei und bedürfen einer genaueren Erklärung (Guion, 1998, S.122f.). Der Rückgriff auf theoretisch fundierte Konstrukte bietet hier einen besseren Ausgangspunkt und soll im Folgenden näher betrachtet werden. Am Ende jedes Abschnittes wird kurz erläutert, inwieweit diese Konstrukte auch im untersuchten Auswahlverfahren erfasst werden. Die erläuterten Konstrukte liefern außerdem den theoretischen Unterbau für die Interpretationen der Ergebnisse der Analysen dieser Arbeit.

2.5.1.1 Kognitive Fähigkeiten

Mit dem Begriff der *Intelligenz* als nicht direkt beobachtbarem Konstrukt kognitiver Fähigkeiten ist in den letzten 100 Jahren Forschung versucht worden, auf Verhalten von Personen in Leistungssituationen wie der Arbeits- und Berufswelt Rückschlüsse zu ziehen. Basis der dabei verwendeten Diagnostik (siehe 2.5.3) sind jeweils unterschiedliche Intelligenzmodelle. Die dabei verwendeten Konstrukte spiegeln bei empirischer Bewährung den aktuellen Forschungsstand wider und nehmen in ihrer Leistungsfähigkeit zu. Ausgangspunkt ist Spearmans (1904) Generalfaktorentheorie, in der er einen kognitiven Leistungen zugrunde liegenden, allgemeinen Intelligenzfaktor (*g*-Faktor) postulierte, der heute weitgehend anerkannt ist (Guion, 1998; Holling, Preckel & Vock, 2004; Lubinski, 2004). Hinzukommt eine spezifische, für die jeweilige Aufgabe erforderliche Fähigkeit (*s-Faktor*). Im Gegensatz zu Spearman ging Thurstone (1935) von insgesamt sieben, nicht völlig voneinander unabhängigen, aber empirisch wiederholt bestätigten Primärfaktoren der Intelligenz aus. Die Un-

terschiede zu Spearmans Faktoren lassen sich durch methodische Besonderheiten erklären (Holling, et al., 2004, S.20). Modernere hierarchische Ansätze integrieren beide Modelle.

Die *hierarchische* Anordnung verschiedener Faktoren ist als empirisches System entstanden (Guion, 1998, S.125). So berichtet Anastasi (1988) von weit über 100 kognitiven Faktoren. Carroll (1993) weist darauf hin, dass die Interpretation jeglicher Nuance als eigener Faktor zu einer unendlichen Faktorenzahl führen könnte. Die Faktorenvielfalt findet sich ebenfalls bereits in Guilfords dreidimensionaler Taxonomie bei seinem *Structure of Intellect-Model* (Guilford, 1956) mit 120 Faktoren wieder. Das Modell ist allerdings sowohl konzeptionell als auch methodisch stark kritisiert worden (Brody & Brody, 1976; Carroll, 1968; Horn & Knapp, 1974; Undheim & Horn, 1977). Cattells hierarchische Modellvorstellung geht von der Annahme aus, dass sich der *Generalfaktor g* in eine *fluide* und eine *kristalline* Fähigkeitskomponente aufteilen lässt. Die *fluide Intelligenz* g_f ist dabei die Fähigkeit, sich neuen Situationen anzupassen und neuartige Probleme zu lösen, ohne dass gelerntes Wissen dabei von Bedeutung wäre. Es ist also die Fähigkeit zum schlussfolgernden Denken und Problemlösen. Die *kristalline Intelligenz* g_c hingegen beinhaltet kognitive Fähigkeiten, die sich durch Kumulation von Lernerfahrung (Akkulturation) seit der Geburt entwickelt haben. Sie bezieht sich also auf die Verarbeitung vertrauter Information und die Anwendung von Wissen. Seine empirisch nicht bestätigte *Investmenttheorie* geht von der Annahme aus, dass *fluide Intelligenz* die Voraussetzung für *kristalline Intelligenz* darstellt und eine Person über die ganze Lebenspanne *fluide Intelligenz* in Lernen investiert. Neben diesen beiden Generalfaktoren, die Cattell (1971) auf Grund nicht vorhandener Unabhängigkeit um einen weiteren übergeordneten Faktor erweiterte, postulierte er auf erster Ebene, ähnlich wie Thurstone, verschiedene Primärfaktoren. Er legte somit den Grundstein moderner Strukturmodelle der Intelligenz.

Ein integratives Modell, das auf einem sehr umfassenden Inventar von Intelligenz- und Kreativitätsaufgaben basiert, ist das *Berliner Intelligenzstrukturmodell* (*BIS*) von Jäger (1967; 1982). Es geht davon aus, dass allgemeine Intelligenz allen intelligenten Leistungen zugrunde liegt und sich aus sieben Fähigkeiten zusammensetzt. Diese lassen sich in zwei Modalitäten (operativ und inhaltlich), unter denen kognitive Leistungen klassifiziert werden können, einteilen. Dabei wird jede intelligente Leistung durch alle diese Fähigkeiten zu unterschiedlichen Anteilen bestimmt. Das *BIS* enthält dabei nicht alle vorstellbaren unterscheidbaren Intelligenzkomponenten und ist als Modell zu verstehen, das für zukünftige Erweiterungen offen ist (Holling, et al., 2004, S.26). Die Validität des Modells wurde vor unterschiedlichen Kontexten nachgewiesen (Jäger, et al., 2004 (in Druck); Süß, Oberauer,

Wittmann, Wilhelm & Schulze, 2002). Auf einem ebenso umfassenden Datenmaterial basiert das sehr elaborierte *Three-Stratum-Modell* von Carroll (1993), das faktoranalytisch entwickelt wurde, drei Hierarchieebenen (Stratum I bis III) aufweist und bestehende Strukturmodelle erweitert. Sternbergs *Triarchische Theorie der Intelligenz* wiederum versteht Intelligenz als dynamisches Konzept und berücksichtigt Umgebungsbedingungen (Sternberg, 1997b). Sie geht damit über das klassischen Intelligenzkonzept, das mit Tests erfasst wird, hinaus und ist als Ergänzung zu den Strukturmodellen zu verstehen (Holling, et al., 2004, S.31).

Eine weitere integrierte Darstellung der Erwachsenenintelligenz wurde von Ackerman (1996) im Rahmen seiner *PPIK-Theorie* entwickelt. In ihr vereinigt er eine Vielzahl von Theorien über den Zusammenhang von Persönlichkeit, Interesse und intellektueller Entwicklung. Er unterscheidet dabei vier Komponenten, die einen Beitrag zur Entwicklung der Erwachsenenintelligenz leisten: Intelligenz als *Prozess*, als *Wissen*, als *Interesse* und als *Persönlichkeit*. Erstere beide ergeben, ähnlich wie bei Cattell, die allgemeine Intelligenz. Intelligenz als *Prozess* beinhaltet dabei Komponenten der Informationsverarbeitung (Verarbeitungskapazität, Wahrnehmungsgeschwindigkeit, Gedächtnisspanne und räumliche Rotation). Wissen versteht sich im Sinne der kristallinen Intelligenz von Cattell. Es lässt sich in spezifisches Berufswissen, dass sich bis zum mittleren Lebensalter ausbildet und unspezifisches nebensächliches Wissen (Allgemeinwissen, Hobby), das bis ins hohe Erwachsenenalter zunehmen kann, unterteilen. Die Entwicklung der Intelligenz als *Wissen* orientiert sich an der Investmenttheorie von Cattell (1971). Fähigkeiten und Wissen entwickeln sich dabei gemeinsam. Die anderen zwei Komponenten, *Interesse* und *Persönlichkeit*, sind dabei unabhängig von der Intelligenz. Ihre Interaktion leistet allerdings einen Beitrag zur intellektuellen Entwicklung im Lebensverlauf. Nach Ackerman hängt das Wissen eines Erwachsenen im frühen Alter von der Informationsverarbeitungsfähigkeit ab. Später allerdings gewinnt zunehmend der individuelle Aufwand, der, in Abhängigkeit der Persönlichkeit und der Interessen, zur Wissensaneignung betrieben wird, an Bedeutung. Individuelle Fähigkeiten haben dabei Einfluss auf den Erfolg bei der Erledigung bestimmter Aufgaben. Die Interessen und die Persönlichkeit hingegen beeinflussen die Motivation, mit der die Aufgabe in Angriff genommen wird.

In der vorliegenden Untersuchung wird nicht auf eine explizite Erfassung kognitiver Fähigkeiten, etwa durch Tests, zurückgegriffen, auch wenn dies wünschenswert wäre (siehe 2.5.2.1). Es gibt folglich auch kein hierarchisches oder sonstiges Intelligenzmodell, das dem Verfahren zugrunde liegt. Vielmehr werden einzelne, eher spezifische Komponenten durch

Interviewer und Beobachter mit Hilfe einer Skala eingeschätzt. Dazu gehören z.B. *logisch-analytisches Denken* oder *Kreativität*. Es ist deshalb anzunehmen, dass zumindest Teilbereiche des Intelligenz-Konstruktes erfasst werden (siehe auch 3.3).

2.5.1.2 Persönlichkeitskonstrukte

Persönlichkeit ist u.a. eine Mischung aus Werten, Temperament, Charakter und Motivation. Diese sind zwar als Anforderungen an Arbeitsplätze relativ leicht zu konzeptualisieren, aber umso schwerer zu operationalisieren (Guion, 1998). Als ein, in zahlreichen Studien auch kulturübergreifend faktoranalytisch immer wieder bestätigtes Modell, hat sich das *Fünf-Faktoren-Modell der Persönlichkeit* (*Big 5*) herauskristallisiert (Digman, 1990; Goldberg, 1993; McCrae & John, 1992). Obwohl ein großer Konsens über die Anzahl der Dimensionen, nämlich fünf, besteht, ist ihre Bezeichnung in der Literatur uneinheitlich. *Abbildung 30* zeigt einen Definitionsversuch für den deutschen Sprachgebrauch und die dazugehörigen charakteristischen Verhaltensmerkmale.

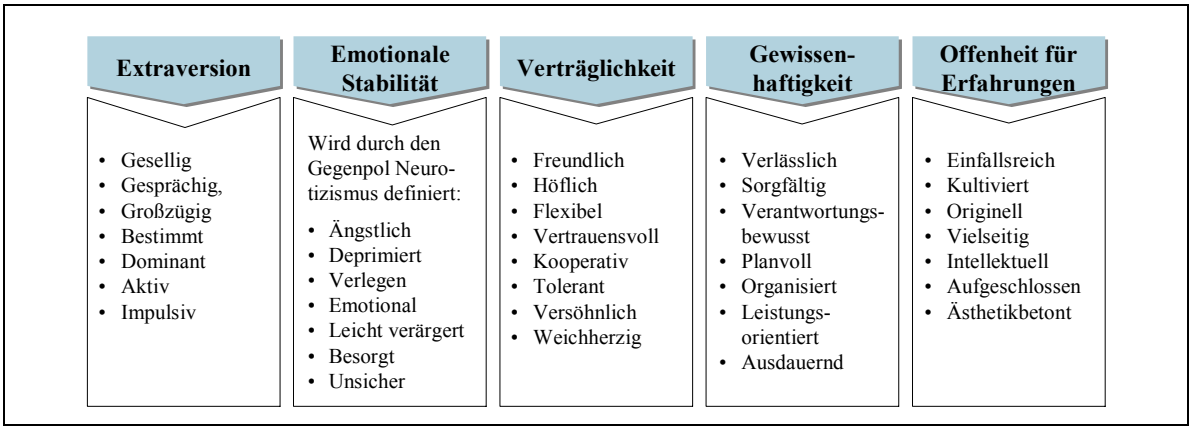

Abbildung 30: Das Fünf-Faktoren-Modell der Persönlichkeit (aus Schuler, 2000, S.29)

Neben den weitgehend anerkannten *Big 5* gibt es noch weitere Taxonomien. So nennt Eysenck (1991) lediglich drei Persönlichkeitsfaktoren: *Psychotizismus*, *Extraversion* und *Neurotizismus*. Cattell (1972) hingegen unterscheidet 16 Faktoren, die er auf psycholexikalischen Studien und faktoranalytischen Prinzipien gestützt postulierte, unter ihnen *Dominanz*, *Emotionale Stabilität* und *Sozialkompetenz*. Hough (1992) wiederum bietet eine Variante mit neun Persönlichkeitsfaktoren an. Sie unterteilt in ihrem Modell die Dimensionen *Extraversion* und *Gewissenhaftigkeit* in weitere Eigenschaften und ergänzt diese um zwei weitere Dimensionen. Neben den dargestellten Taxonomien gibt es noch eine große Menge an weiteren, im Rahmen von Tests empirisch gefundenen Konstrukten.

Modelle zur Erklärung der Zusammenhänge mit Berufserfolg sind bei den Persönlichkeitsmerkmalen bisher weitgehend ausgeblieben. Schuler (1991b) postuliert einen Zusammenhang zwischen Persönlichkeit und Führungserfolg anhand der zwei Variablen *Leistungsmotivation* und *Fähigkeiten.* Barrick, Mount und Strauss (1993) betonen in ihrem Berufserfolgsmodell den Einfluss von *Gewissenhaftigkeit* auf das Setzen von Zielen und der Verpflichtung gegenüber diesen Zielen bei Verkäufern. Schmidt und Hunter (1992) nennen neben einem direkten Einfluss der *Gewissenhaftigkeit* auf Vorgesetztenurteile auch indirekte Wege über ein sich stärker entwickelndes Arbeitswissen und die Anzahl disziplinarischer Verwarnungen. Goff und Ackermann (1992) schlagen das Konstrukt *Typisches Intellektuelles Engagement (TIE)* als Attraktion, die intellektuell anspruchsvolle Aufgaben auf Personen ausüben, vor. Sie unterscheiden zudem *Intelligenz als maximales* und *Intelligenz als typisches Verhalten* und postulieren Assoziationen von *TIE* mit Letzterem. Demnach wirkt sich *TIE* auf den Erwerb von Wissen aus (Rolfhus & Ackerman, 1999), das für bestimmte kognitive Aufgaben leistungsbestimmend ist. Goff und Ackermann wiesen empirisch nach, dass Personen mit hohen *TIE*-Werten schwer zu erreichende Ziele präferieren. Sie sind zudem offen und neugierig sowie intellektuell interessiert. Außerdem werden sie teilweise von Aufgaben so gefangen genommen, dass die Wahrnehmung von umgebender Zeit und Situation eingeschränkt ist (Wilhelm, Schulze, Schmiedek & Süß, 2003).

In der vorliegenden Untersuchung werden Persönlichkeitseigenschaften ebenfalls nur durch Beobachter und Interviewer mit Hilfe einer Skala eingeschätzt. Die eingeschätzten Dimensionen wurden nicht theoretisch abgeleitet. Sie beruhen folglich auf keinem der geschilderten Persönlichkeitsmodelle. Es gibt jedoch semantische Ähnlichkeiten zu einzelnen Konstrukten. So gleicht beispielsweise die Dimension *Intellektuelle Neugier* dem Konstrukt *TIE*. Es ist deshalb auch hier von einer Überlappung mit einigen, empirisch bestätigten Konstrukten auszugehen. Eine detailliertere Betrachtung findet sich unter 3.3.

2.5.1.3 Interessen

Obwohl mit den Persönlichkeitskonstrukten verwandt, sind *Interessenvariablen* nicht generelle Verhaltensdispositionen, sondern beziehen sich auf Präferenzen bei der Arbeit und dem Arbeitsumfeld (Schmitt & Chan, 1998, S. 119). Interessen werden dabei, bei Kongruenz zum Arbeitsplatzinhalt und Umfeld, im Zusammenhang mit *Arbeitszufriedenheit*, die wiederum zu einer höheren Produktivität führen soll, gesehen. Ihr Einsatz konzentrierte sich daher in der Vergangenheit hauptsächlich auf den Bereich der Berufsberatung. Die wichtigste und am weitesten verbreitete Interessentaxonomie in der Personalpsychologie geht auf

Holland (1959; 1973; 1985; 1996) zurück. Seine *RIASEC-Taxonomie* ist in *Abbildung 31* dargestellt und kann neben Personen, auch auf Organisationen angewandt werden (Schmitt & Chan, 1998). Hollands Hexagonalmodell geht davon aus, dass Personen bestimmte Muster auf den sechs Dimensionen haben. Dabei haben solche mit hohen Werten auf einer Dimension, auf der gegenüberliegenden Dimension eher niedrige Ausprägungen.

Nach Holland (1997) ist der Erfolg von dem Passungsprozess zwischen Person (Persönlichkeit und Berufsidentität) und Arbeitsumgebung abhängig und führt zu höherer beruflicher Stabilität und Zufriedenheit. Ein neuerer Ansatz von Kristof (1996) betrachtet ebenfalls Übereinstimmung von Organisation und Person als geeigneten Prädiktor für Berufserfolg (*Person-Organization-Fit*). Dabei versteht sie unter *Supplementary Fit* eine Übereinstimmung von grundlegenden Werten, Zielen, Kultur und Normen des Unternehmens mit den Zielen, Werten, der Persönlichkeit und den Verhaltensweisen der Person. Darüber hinaus nennt sie zusätzlich *Complementary Fit* als das Vorhandensein sich gegenseitig ergänzender Aspekte, die den Bedürfnissen des jeweils anderen entsprechen.

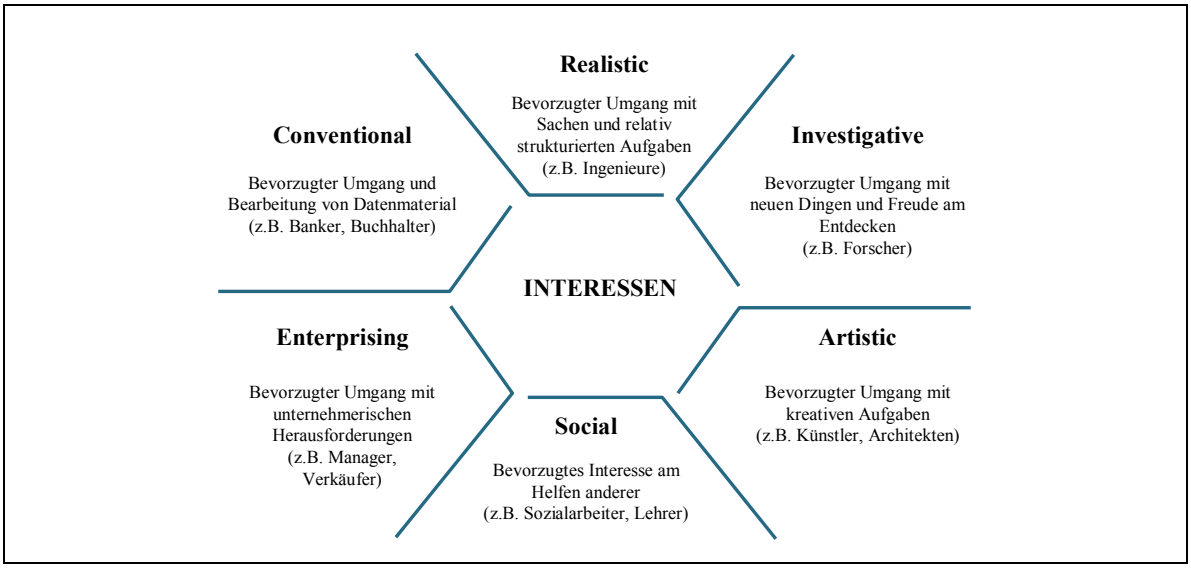

Abbildung 31: RIASEC-Taxonomie von Holland (adaptiert aus Schmitt & Chan, 1998)

Auch die Interessen des Bewerbers werden in dem untersuchten Auswahlverfahren durch nur wenige Variablen und nicht umfassend abgedeckt. Es gilt deshalb die gleiche mangelnde theoretische Fundierung, die bei den Intelligenz- oder den Persönlichkeitskonstrukten angesprochen wurde. Die Beobachter und Interviewer stufen die Bewerber z.B. auf der Dimension *Fit-ISB-Team* ein, die im Wesentlichen den *Person-Organization-Fit* im Sinne Kristofs widerspiegeln soll. Darüber hinaus haben einige der gemessenen Variablen semantische Ähnlichkeit mit einzelnen *RIASEC*-Dimensionen. So ist beispielsweise denkbar, dass

die Einschätzungen auf der Dimension *unternehmerisches Denken* Überlappungen mit der Dimension *Enterprising* besitzen. Außerdem können auch biographische Informationen über den Bewerber Hinweise auf bestimmte Präferenzen oder Interessen geben.

2.5.1.4 Motivation

Ein weiteres, augenscheinlich valides Konstrukt, das aber im Auswahlkontext eher selten Anwendung findet, ist die *Motivation*. Eine systematische valide Ermittlung der Ausprägungen bleibt aber bei Auswahlverfahren, insbesondere durch die Tendenz der sozialen Erwünschtheit, meistens verwehrt (Sieglen, 1998). Motive sind keine Eigenschaftskonstrukte mit Alleinerklärungsanspruch. Sie treten, bevor sie verhaltenswirksam werden, mit Umweltbedingungen in Wechselwirkung (interaktionistische Sicht, siehe Heckhausen, 1989). Die zwei, im Zusammenhang mit Berufserfolg bedeutendsten Motive sind das *Machtmotiv* und das *Leistungsmotiv*. Unter *Machtmotiv* versteht man allgemein die „Disposition zur absichtlichen Beeinflussung des Verhaltens, Denkens und Fühlens anderer Personen“ mit dem Ziel, eigene Ziele oder Entscheidungen durchzusetzen (zu einzelnen Definitionsbeispielen siehe Heckhausen, 1989, S.368ff.). Unterschieden werden dabei die Motivtendenzen *Hoffnung auf Macht* und *Furcht vor Schwäche (*Heckhausen, 1989). McClelland (1985; McClelland & Boyatzis, 1982) postuliert mindestens mittelstarke Ausprägungen des Machtmotivs bei effektiven Führungskräften, gepaart mit einer geringen Ausprägung des Anschlussmotivs und einer hohen Selbstkontrolle. Im Gegensatz zum *Machtmotiv* zielt das *Leistungsmotiv* auf Handlungsergebnisse ab, die bei Anlegung von Qualitäts- oder Schwierigkeitsmaßstäben positiv bewertet werden und zwar im Sinne positiver Selbstbewertungen. *Leistungsmotivation* kann einfach als „Auseinandersetzung mit dem Tüchtigkeitsmaßstab“ verstanden werden (McClelland, Atkinson, Clark & Lowell, 1953). Analog zum Machtmotiv unterscheidet man auch hier *Hoffnung auf Erfolg* und *Furcht vor Misserfolg* als Motivtendenzen (Atkinson, 1978). Forschungsleitendes Modell der Leistungsmotivationsforschung ist das *Risikowahl-Modell* von Atkinson (1957), nach dem anstehende Aufgabenstellungen die Bildung von Erfolgs- und Valenzerwartungen auslösen, die wiederum, je nach Motivtendenz, die Verhaltensrichtung erklären. Schuler und Prochaska (2000) weisen jedoch darauf hin, dass das Konstrukt *Leistungsmotivation* in seinem Homogenitätsgrad, seiner Stabilität, Universalität und Generalisierbarkeit nicht bestätigt wurde und auch die Frage der Konstruktvalidität wenig geklärt ist. Im Gegensatz zur klassischen Auffassung sehen sie Leistungsmotivation als globale Verhaltensorientierung, an der vielfältige Aspekte der Persönlichkeit beteiligt sind (*Zwiebelmodell der Leistungsmotivation*).

Motivation wird in der vorliegenden Studie nicht erfasst. Es ist allerdings nicht unwahrscheinlich, dass bestimmte motivationale Dispositionen im Rahmen der Beobachtung und der Interviews zum Vorschein kommen. Somit würden sie Eingang in die Einstufungen der Beurteiler und Interviewer finden. Eine Untersuchung dieses Konstruktes bleibt allerdings auf Grund des Datensatzes verwehrt.

2.5.1.5 Wissen

Eine gängige Unterscheidung von Wissen ist die in *deklaratives* („*knowing what*“) und *prozedurales* Wissen („*knowing how*“) (Anderson, 1982, Fortmüller, 1997). So entwickelten beispielsweise Weitz, Sujan und Sujan (1986) und Szymanski (1988) Rahmenmodelle zum Zusammenhang dieser Wissensarten mit Verkaufserfolg bei Verkaufsangestellten. Demnach sind drei Faktoren für die Verkaufsleistung bedeutsam: *Anpassungsfähigkeit*, *verkaufsrelevantes* Wissen und der Verkaufsprozess mit den Informationskategorien *deklaratives* Wissen (attributive Information über den Kunden) und prozedurales Wissen (Reihenfolge der Geschehnisse und Beeinflussungstechniken). McCloy, Campbell und Cudeck (1994) sehen in diesen beiden Wissenskomponenten, zusammen mit *Leistungsmotivation*, die direkten Determinanten des Berufserfolges. Diese werden wiederum durch indirekte Determinanten (wie allgemeine Intelligenz) beeinflusst. Hunter und Hunter (1984) nennen *Berufswissen* (*job knowledge*) als weitere Wissensart und meinen damit die Kenntnis der mit einer beruflichen Tätigkeit verbundenen Sachverhalte, Verantwortungen und Prozesse. Eine weitere Form des Wissens, die im Auswahlkontext oft genannt wird, ist *tacit knowledge* als das wichtigste Konstrukt, das *praktische Intelligenz* erfassen soll (Sternberg, et al., 2000). *Praktische Intelligenz* ist die Fähigkeit, alltägliche Probleme der realen Welt zu lösen bzw. sich an eine sich verändernde Umwelt anzupassen, diese zu verändern oder auszuwählen. Sternberg (1997a) sieht sie als unabhängig von dem allgemeinen Intelligenzfaktor *g* und in ihr einen besseren Prädiktor für den Erfolg von Führungskräften. Erfolgsrelevant sind demnach ebenso praktische und kreative Fähigkeiten, da praktische Probleme auch deren Erkennung und Formulierung erfordern. Problemlöseprozesse gehen dabei mit extensiven Informationssucheprozessen einher, bei denen frühere alltägliche Erfahrungen eingebunden werden. Sternberg et. al. (2000) unterscheiden dabei *akademisches* und *praktisches* (prozedurales) Wissen und kritisieren die Fokussierung der meisten Studien auf die erste Wissensart, da im alltäglichen Leben wie im Beruf dem praktischen Wissen eine größere Bedeutung zukommt. Sternberg und Hedlund (2002) weisen darauf hin, dass *g* nicht als alleiniges Konstrukt zu Vorhersage von Berufserfolg ausreicht, da es u.a. zu wenig Varianz aufklärt, subgruppenspezifische Unterschiede verursacht und testtechnisch auf realitätsfernen Problemen

beruht. Praktisches Wissen (*tacit knowledge)* hingegen ist das erfahrungsabhängige, an einen bestimmten Kontext gebundene prozedurale Wissen, das oftmals nicht verbalisiert und ohne Unterstützung des sozialen Umfelds angeeignet wird. Das *Tacit-Knowledge* Konzept ist allerdings wegen seines Konstruktverständnisses, seiner Operationalisierung und seiner unzureichenden faktoranalytischen Bestätigung stark kritisiert worden (Gottfredson, 2003a, 2003b). Sogar seine Existenz als eigenständiges Konstrukt ist fraglich (Schmidt & Hunter, 1993).

Bei dem untersuchten Auswahlverfahren hat die Erfassung von Wissen eine zentrale Bedeutung. So werden im Rahmen von Interviews sowohl *deklaratives Wissen* (z.B. *BWL-* oder *VWL-Kenntnisse*) als auch *prozedurales Wissen* (z.B. *Methodenkenntnisse*) erfasst. Teilweise beruht die Einstufung auf diesen und anderen Dimensionen, auf der Leistung bei der Bearbeitung gestellter Fallstudien. Es kann deshalb davon ausgegangen werden, dass auch *Berufswissen* in die Beurteilungen des Recruitings einfließt.

2.5.2 Ausgewählte empirische Befunde zu Prädiktoren des Berufserfolges

Nachfolgend soll eine Übersicht der wesentlichen Erkenntnisse zu den beschriebenen Prädiktoren gegeben werden. Zu einer detaillierten Beschreibung der geschilderten Befunde und Perspektiven sei auf die genannte Literatur verwiesen. Im Empirieteil der Arbeit wird auf einzelne Befunde zurückgegriffen. Es soll dabei geprüft werden, ob sich diese bei der vorliegenden Untersuchung replizieren lassen.

2.5.2.1 Intelligenz

Die Veröffentlichungen der letzten 15 Jahre wurden hauptsächlich durch das Thema der allgemeinen Intelligenz (*g*-Faktor, kurz *g*) und ihrer Bedeutung geprägt. Einigkeit herrscht mittlerweile darüber, dass Intelligenz hierarchisch aufgebaut ist (Lubinski, 2004; S.98), mit einem Allgemeinfaktor *g* auf oberster Ebene (Schmidt, 2002), der sich am kürzesten mit *Fähigkeit zum Lernen* beschreiben lässt (Hunter, 1986; Hunter & Schmidt, 1996). Zahlreiche Studien haben zudem gezeigt, dass *g*, sowohl für die Vorhersage von Ausbildungs- als auch Berufserfolg, einen entscheidenden Beitrag leistet (z.B. Ghiselli, 1966; Hunter, 1986; Ones, Viswesvaran & Dilchert, 2004; Ree & Earles, 1992; Ree, Earles & Teachout, 1994; Salas & Canon-Bowers, 2001; Salgado, Anderson, Moscoso, Bertua, de Fruyt, et al., 2003; Schmidt & Hunter, 1998; Schmidt-Atzert, Deter & Jaeckel, 2004; Vinchur, Schippmann, Switzer & Roth, 1998). *g* kann somit als wichtigster Prädiktor angesehen werden (Holling, et al., 2004). Durchschnittlich erreicht die prädiktive Validität für das Kriterium

Vorgesetztenbeurteilung eine Höhe von r=.50 (Gottfredson, 2002, S.44). *g* ist somit valider als nichtkognitive Prädiktoren. Objektive Leistungsmaße sind dabei ebenso besser vorhersagbar als subjektive Leistungsmaße wie harte, technische Fakten im Vergleich zu „weichen" Kriterien (z.B. Sozialverhalten gegenüber Kollegen). Dabei ist die prädiktive Validität von *g* unabhängig vom betrachteten Beruf, sofern dieser Informationsverarbeitungsprozesse beinhaltet (Campbell, 1990; Schmidt & Hunter, 1998; Schuler, 2000). Sie nimmt aber mit zunehmender Komplexität des ausgeübten Berufes (z.B. von r=.27 bis r=.53 bei Hunter & Hunter, 1984) zu. Komplexität kann dabei als Höhe des Informationsverarbeitungsbedarfes verstanden werden (Gottfredson, 2002).

Verschiedentlich wird auch auf einen Zusammenhang von *g* mit dem Prädiktor *Berufswissen* (siehe 2.5.1.5) hingewiesen. Intelligenz beeinflusst dabei in Verbindung mit gemachten Erfahrungen vermutlich Lernprozesse, die zu einer höheren beruflichen Kompetenz führen (Schmidt, Hunter, Outerbridge & Trattner, 1986; Schmidt & Hunter, 1993). Schmidt, Hunter, Outerbridge und Goff (1988) zeigen, dass bedingt durch Lernerfahrung, intellektuell hoch veranlagte Personen einer zeitlich zunehmenden Differenzierung der beruflichen Leistungsfähigkeit unterliegen (Divergenzhypothese). Bei intellektuell schwach Veranlagten ist hingegen genau das Gegenteil der Fall (Konvergenzhypothese). Deadrick und Madigan (1990) fanden Unterschiede in der prädiktiven Validitätsstabilität kognitiver Fähigkeitstests (steigend), Erfahrung (fallend) und Fertigkeitentests (gleich bleibend). Direkte Intelligenzeffekte auf die Berufsleistung sind demnach eher bei weniger routinierten Tätigkeiten zu erwarten (Gottfredson, 2002). Die frühere Annahme hingegen, der Zusammenhang zwischen allgemeiner Intelligenz und Berufserfolg sei artifizieller Natur und durch den soziökonomischen Status bedingt (McClelland, 1973), gilt heute weitgehend als überholt (Barret & Depinet, 1991; Murray, 1998; Sieglen, 1998, S.82).

Die zahlreichen empirischen Befunde wurden auf unterschiedlichste Weise kritisiert. So bemängeln Amelang und Bartussek (1981) die Zusammenfassung von Ergebnissen über mehrere Berufsgruppen hinweg, da für komplexere Berufe höhere Zusammenhänge zu erwarten seien. Zudem gebe es innerhalb einer Berufsgruppe starke Varianzeinschränkungen, da das Berufsniveau, wie bereits vor langem gezeigt (Harrel & Harrel, 1945), mit der Intelligenz zusammenhänge, auch wenn intelligente Menschen in jeder Berufsgruppe vorzufinden seien (Asendorpf, 1999). Hunter und Hunter (1984) gehen im Gegensatz zu McClelland (s.o.) davon aus, dass die ermittelten Zusammenhänge auf Grund der Varianzeinschränkung und der Unreliabilität systematisch unterschätzt werden. Sie zeigten dies anhand einer Reanalyse der gefundenen Validitäten von Ghiselli (1966). Obgleich diese Ansicht und ihre Korrektur-

formeln weitgehend akzeptiert sind, weist Sieglen (1998, S.85) zu Recht darauf hin, dass metaanalytische Techniken und Parameterschätzungen immer auch die Gefahr in sich bergen, die empirische Basis zu verlassen. Zudem werden positive Ergebnisse häufiger zitiert und die meistens als Kriterium herangezogenen Vorgesetztenurteile durch Selbstdarstellungsfähigkeiten des Beurteilten (die wiederum intelligenzabhängig sind) beeinflusst.

Spezifische kognitive Fähigkeiten, die sich inhaltlich allgemein anerkannt in *verbale*, *numerische* und *figurale/räumliche* Komponenten einteilen lassen, zeigen im Vergleich zu *g* nur eine sehr geringe inkrementelle Varianzaufklärung am Berufserfolg (Gottfredson, 2002; Ree & Earles, 1992; Ree & Earles, 1994; Ree, et al., 1994). Thorndike (1986) postulierte hingegen höhere Validitäten spezifischer kognitiver Fähigkeiten bei großen Stichproben. Murphy (1996) empfiehlt deshalb den Einsatz des Prädiktors *g* als „sparsame“ Variante. Die besten Vorhersagen mit spezifischen Fähigkeiten erzielt man dabei durch je nach Arbeitsstelle unterschiedliche Gewichtungen dieser (z.B. Ree & Earles, 1991). Schmidt (2002) weist jedoch darauf hin, dass in solchen Fällen größtenteils auch nichts anderes als *g* gemessen wird, da spezifische kognitive Fähigkeiten immer auch *g* beinhalten würden. Ree und Earles (1994) merken dazu an, dass bei den meisten Faktorisierungen in den rotierten Faktoren *g* enthalten ist. Zudem unterscheiden sich nach Gottfredson (2002) Tests, die hoch, von jenen, die niedrig auf *g* laden, lediglich anhand der Komplexität ihrer Aufgaben. Die allgemeine Intelligenz *g* wird dabei bei jedem kognitiven Test miterfasst und lässt sich gar nicht umgehen.

Lubinski (2004; Achter, Lubinski, Benbow & Eftekhari-Sanjani, 1999; Shea, Lubinski & Benbow, 2001; Webb, Lubinski & Benbow, 2002) entdeckte allerdings im Rahmen seiner Hochbegabungsstudien, dass die sehr unterschiedlichen Profile der Probanden auf den Dimensionen *verbal*, *numerisch* und *räumlich* erhebliche Auswirkungen auf den späteren universitären und beruflichen Werdegang haben. In seinen Studien zeigt sich auch die verbale Überlegenheit der weiblichen im Vergleich zur mathematisch-figuralen Überlegenheit der männlichen Teilnehmer (gekippte Profile). Dieses Muster wurde bereits in zahlreichen anderen Untersuchungen festgestellt (z.B. Geary, 1996, 1998; Halpern, 2000; Humphreys, Lubinski & Yao, 1993; Kimura, 1999; Loehlin, 2000) und teilweise auf gesellschaftliche Stereotypen und zugewiesene Rollen zurückgeführt (Hyde, Fennema & Lamon, 1990). Frauen, da eher verbal orientiert, scheinen zudem den beruflichen Umgang mit Menschen zu präferieren, während Männer die Auseinandersetzung mit Sachen vorziehen (Webb, et al., 2002). Die berufliche Entwicklung wird dabei zusätzlich von den Werten und Interessen der Person beeinflusst. Eine Kongruenz dieser mit den Anforderungen und der Ausgestaltung

des Arbeitsplatzes erhöht die Wahrscheinlichkeit von Höchstleistungen (Lubinski, Benbow, Shea, Eftekhari-Sanjani & Halvorson, 2001).

Flankierend zu den angesprochenen Schwerpunkten ist in den Vereinigten Staaten immer wieder die Diskussion um *Diskriminierungsaspekte* bei der Anwendung von *g* als Prädiktor aufgekommen, da bestimmte Minderheiten (z.B. Farbige und Bürger lateinamerikanischer Abstammung) in kognitiven Tests signifikant schlechter abschneiden (Gottfredson, 2002). Schmidt (2002) betont allerdings, dass alle Personen mit niedriger *g*-Ausprägung, unabhängig von ihrer Herkunft oder Hautfarbe, auch schlechtere Berufsleistungen erbringen würden. Somit sei das Vorgehen fair. Zudem warnt er vor verheerenden ökonomischen und sozialen Folgen, die mit einem Festhalten an einer, bezüglich der Repräsentativität aller Personengruppen quotenorientierten Politik einhergehen kann.

Zusammenfassend kann festgestellt werden, dass auf Grund der sehr komplexen und wenig routinierten Beratungstätigkeit, sich *g* hervorragend als Prädiktor des Berufserfolgs eignen dürfte. Zudem müsste *g* auch im vorliegenden Auswahlverfahren Relevanz besitzen, da sich seine Erfassung, wie beschrieben, wohl nicht umgehen lässt. Betrachtet man spezifischere Dimensionen der Intelligenz, so müssten sich in den Prädiktorendaten unterschiedliche Profile bei den Bewerbern zeigen, die sich auf den späteren Berufserfolg auswirken dürften.

2.5.2.2 Persönlichkeit

Vor dem Aufkommen und der verbreiteten Anerkennung der *Big 5*-Taxonomie, sind für Persönlichkeitsdimensionen, im Vergleich zu kognitiven Fähigkeiten, eher geringe bis moderate prognostische Validitäten für den Berufserfolg gefunden worden (Hunter & Hirsh, 1987; Schmitt, et al., 1984). In den letzten zwei Jahrzehnten hat jedoch das Interesse an Persönlichkeitsvariablen und ihrer Validität hinsichtlich der Vorhersage von Berufserfolg stark zugenommen (Barrick & Mount, 1991; Hough, Eaton, Dunette, Kamp & McCloy, 1990; Mount & Barrick, 1995; Mount & Barrick, 1998; Mount, Barrick & Stewart, 1998; Robertson & Kinder, 1993; Tett, et al., 1991; Tett, Jackson, Rothstein & Reddon, 1994). Der Aufschwung wurde von empirischen Befunden begünstigt, die zu kognitiven Fähigkeiten inkrementelle Varianzaufklärungen am Berufserfolg entdeckten (z.B. McHenry, Hough, Toquam, Hanson & Asworth, 1990). Darüber hinaus entstanden auch neuere Berufserfolgsmodelle, die neben der eigentlichen beruflichen Aufgabenerledigung (*task performance*) auch unterstützende Tätigkeiten (*contextual performance*) betrachten (Borman & Motowidlo, 1997; siehe auch 2.2.2). Barrick, Mount und Strauss (1993) unterscheiden diesbezüglich eine

fähigkeitsabhängige „can do"- und eine motivationsabhängige „will do"-Komponente. Mit der zweiten hängen scheinbar Persönlichkeitsfaktoren zusammen. McHenry et. al. (1990) zeigten beispielsweise im Rahmen ihrer Forschungsarbeit bei der amerikanischen Armee (genannt *Projekt A*) überzeugende Vorhersageleistungen von Persönlichkeitsdimensionen auf den Kriterien *Unterstützung von Kameraden* und *Bereitschaft zu zusätzlichen Anstrengungen*. Zu gleichen Ergebnissen kamen sie in Kombination mit kognitiven Fähigkeitsmaßen für die Kriterien *Anstrengung* und *Führung*. Borman und Hanson (1997) sowie McManus und Kelly (1999) berichten vergleichbare Befunde.

Verschiedene Forscher äußerten jedoch auch Zweifel, ob sich Persönlichkeitsvariablen auf Grund ihrer komplexen und auf einem hohen Abstraktionsniveau angesiedelten Struktur (Funder, 1991; Hough, 1992) für Vorhersagen im Personalbereich eignen. Die Thematik, auch als *Bandbreiten-Genauigkeits-Dilemma* bekannt, spiegelt die im Rahmen der Validitätsanalyse angesprochene Asymmetrie und damit verbundenen unfairen Validitätstests wider (siehe 2.2.4). Ones und Viswesvaran (1996) vertreten die Ansicht, dass breite Persönlichkeitsdimensionen bessere Vorhersagen auf Grund ihrer höheren Reliabilität und der ebenso komplexen (und somit symmetrischen) Dimensionierung des Berufserfolgskriteriums ermöglichen. Barrick und Mount (1991) führten eine Metaanalyse über Zusammenhänge von den *Big 5* und verschiedenen Leistungskriterien bei unterschiedlichen Berufen durch. Sie fanden Korrelationen in Höhe von r=.22 für die Dimension *Gewissenhaftigkeit* und etwas geringere Werte für *Extraversion* (r=.18 und r=.15) bei den Berufsgruppen Manager und Verkäufer. Beide Dimensionen ergaben in einer späteren Untersuchung außerdem bessere Vorhersageleistung bei hoher Autonomie am Arbeitsplatz (Mount & Barrick, 1995). Für *Emotionale Stabilität* hingegen war die Vorhersageleistung sehr gering. *Verträglichkeit* wurde von den Autoren gar als prognostisch nicht tauglich für den Berufserfolg eingestuft. Im Gegensatz dazu fanden Tett et. al. (1991; 1994) gerade für *Verträglichkeit* die höchsten Vorhersagewerte (r=.33), gefolgt von *Offenheit für Erfahrungen*, *Emotionale Stabilität* und *Gewissenhaftigkeit*. Auch für den europäischen Raum erwiesen sich *Gewissenhaftigkeit* und *Emotionale Stabilität* als valide Prädiktoren über verschiedene Berufskriterien und –gruppen hinweg (Salgado, 1997). Für *Offenheit für Erfahrung* und *Verträglichkeit* gilt dies ebenso beim Kriterium *Ausbildungserfolg*. *Extraversion* und *Verträglichkeit* zeigten sich wiederum valide bei Managern.

Im Gegensatz zu der Überzeugung der prognostischen Güte breiter Persönlichkeitsfaktoren, argumentieren Schneider, Hough und Dunette (1996) sowie Schuler (2000), dass für spezifischere Konstrukte generell höhere Validitäten entdeckt wurden. Demnach sollten

also eher engere „traits“ verwendet werden. Vinchur, et. al. (1998) fanden z.B. metaanalytisch bei der Vorhersage von Berufserfolg bei Verkäufern (Vorgesetztenurteil und Verkaufszahlen) für spezifischere Unterdimensionen der *Big 5* (*Potency, Achievement)* höhere Korrelationen als für die fünf Oberdimensionen. Hough (1992) identifizierte *Achievement* (Leistungswille) als validen Prädiktor für eine Reihe von Kriterien. Mit dem Gesamtleistungskriterium korrelierten außerdem *Angepasstheit* und *Selbstkontrolle*. Zusätzlich zeigten sich für die Berufsgruppe der Verkäufer höhere Validitäten für jeweils eine der Subskalen der *Big 5* (*Potency* und *Achievement*), als für *Extraversion* und *Gewissenhaftigkeit*. Bessere Vorhersagen mit Hilfe der Subskalen für *Gewissenhaftigkeit* fanden Mount und Barrick (1995) für globale und spezifische Kriterien.

Auch augenscheinlich valide Eigenschaften wie *Integrität* oder *Kreativität* wurden untersucht. Gerade Erstere erwies sich als guter Prädiktor für die Vorhersage kontraproduktiven Verhaltens (Schmidt & Hunter, 1998). Dessen kriterienspezifische Festlegung ist allerdings auf Grund möglicher Varianzeinschränkungen und Moderatoreffekte problembehaftet (siehe dazu Marcus, Funke & Schuler, 1997). Ones, Viswesvaran und Schmidt (1993) berichten in ihrer Metaanalyse von einer prädiktiven Validität für *Integrität* in Höhe von r=.40 bei Vorhersage des Berufserfolges über alle beruflichen Komplexitätsniveaus hinweg. Das Integritätskonstrukt scheint dabei weitgehend vom Prädiktor kognitive Fähigkeiten unabhängig zu sein und mit den *Big 5*-Faktoren *Emotionale Stabilität*, *Verträglichkeit* und *Gewissenhaftigkeit* moderat zusammenzuhängen. Besonders bei *Ausbildungserfolg* scheint es inkrementelle Validität zur Intelligenz zu besitzen (Schmidt & Hunter, 1998). Torrance (1988) untersuchte in einer Längsschnittstudie den Zusammenhang zwischen Kreativitätstestwerten und späteren beruflichen kreativen Leistungen und fand Korrelationen in Höhe von bis zu r=.58. Allerdings sind die Ergebnisse auf Grund der inhomogenen Kriterienmaße hinsichtlich ihrer Vergleichbarkeit eingeschränkt (Sieglen, 1998).

Die beiden geschilderten Positionen sind nur scheinbar widersprüchlich und lassen sich integrieren. So betonen Borman, et. al. (1997), dass für eine Vorhersage globaler Kriterien wie Berufserfolg als komplexes multidimensionales Konstrukt, eher breitere, bei der Vorhersage spezifischer Kriterien jedoch bevorzugt engere Persönlichkeitseigenschaften zum Einsatz kommen sollten. Diesem „Symmetriegedanken“ folgen auch Hogan, Hogan und Roberts (1996) sowie Schneider und Hough (1995). Auch Guion (1998) plädiert für eine Berücksichtigung mehrerer (spezifischer und genereller) Ebenen bei der Auswahl von Prädiktoren für die Personalauswahl.

Als Fazit lässt sich festhalten, dass es Anhaltspunkte für die prädiktive Validität von Persönlichkeit als Prädiktor für beruflichen Erfolg gibt. Schuler, et. al. (1995) nennen *Emotionale Stabilität*, *Leistungsmotivation*, *soziale Kompetenz*, *Selbstvertrauen*, *Extraversion*, *Offenheit für Erfahrung*, *Gewissenhaftigkeit*, *Verträglichkeit*, „*locus of control*“, *Maskulinität* (*Potency*) und *Dominanz* für den Berufs- bzw. Führungserfolg valide. Zu bemängeln sind allerdings empirisch gut fundierten Modelle zur Kausalität mit Berufserfolg. Zudem scheint es einen Wildwuchs an Persönlichkeitsmerkmalen zu geben, die ein einheitliches Verständnis scheinbar gleicher Konstrukte und eine Interpretation gefundener Ergebnisse erheblich erschweren. Einen möglichen Ausweg könnte nach Guion (1998) die Fokussierung auf typische Verhaltensweisen bestimmter Persönlichkeiten und eine Abkehr von den eigenschaftsorientierten Beschreibungen bieten.

Für die Untersuchung dieser Arbeit bedeutet dies, dass eher generelle Persönlichkeitsdimensionen im Zusammenhang mit dem Beratungserfolg betrachtet werden sollten, um die Symmetrie von Prädiktor und Kriterium sicherzustellen. Dabei ist zu berücksichtigen, dass der Beruf des Beraters eine hohe Affinität zu dem des Verkäufers und des Managers besitzt. Folglich ist, bei Betrachtung der *Big 5*, von einer Validität insbesondere der Persönlichkeitsfaktoren *Gewissenhaftigkeit*, *Extraversion* und *Verträglichkeit* auszugehen.

2.5.2.3 Motivation

Zahlreiche Studien im Zusammenhang mit Motivation stammen aus der Führungsforschung. Schon McClelland (1961) hatte früh Belege gesammelt, dass Führungskräfte gehäuft hoch *leistungsmotiviert* sind. Diese Vermutung gilt allerdings nur eingeschränkt für kleine überschaubare Unternehmen und nicht-technische Führungskräfte auf niedriger Ebene. (Heckhausen, 1989). Die Bedeutung motivationaler Aspekte für den Berufserfolg ist insbesondere auf Grund der problematischen Operationalisierung, speziell im Rahmen der Auswahlsituation, schwierig festzulegen (Schuler, 2000). Motivationspsychologische Instrumente sind häufig motivspezifisch, also an eng umrissene Einzelmotive (z.B. Machtmotiv) gekoppelt, die keine Gesamtmotivation erfassen und alleine selten prädiktiven Wert haben (McClelland, 1985). Zusätzlich führt die interaktionistische Sichtweise zu der Erkenntnis, dass rein personenzentrierte Ansätze erfolglos bleiben dürften (Nachreiner & Müller, 1995). So zeigte bereits Andrews (1967) die Bedeutung des Organisationsklimas für das Wirksamwerden individueller Motive. Litwin und Stringer (1968) kommen zu dem Schluss, dass ein für den wirtschaftlichen Erfolg optimales Organisationsklima entsteht, wenn die leitenden Führungskräfte hohe *Macht*- und *Leistungsmotive* gepaart mit einem niedrigen *Anschlussmo-*

tiv haben. Während McClelland und Boyatzis (1982) sowie Wotschke (1995) zu ähnlichen Ergebnissen kamen, zeigte bei Wainer und Rubin (1971), allerdings bei kleinen Firmen, das *Machtmotiv* keine Relevanz. Kock (1965; 1974, zitiert nach Heckhausen, 1989) identifizierte die Bedeutung des *Leistungsmotivs* nicht als singulär, sondern nur in Interaktion mit dem *Macht*- und dem *Anschlussmotiv*. Verschiedentlich ist allerdings auch die Vermutung geäußert worden, *Leistungsmotivation* habe die höchste Varianzaufklärung am Berufserfolg, wenn man neben der *allgemeinen* auch die *tätigkeitsfeldspezifische* Leistungsmotivation messen könne (Atkinson, 1978; Eckardt & Schuler, 1992). Die Korrelation zwischen Leistungsmotivationswerten und beruflicher Leistung nimmt scheinbar außerdem mit zunehmender zeitlicher Differenz ihrer Messung zu (Helmreich, L. & Carsrud, 1986). Schuler und Prochaska (2000) zeigten empirisch mit ihrem sehr breit aufgestellten, aus vielen Einzeldimensionen bestehenden Leistungsmotivationstest *LMI*, dass hinter diesem faktoranalytisch die Unterdimensionen *Ehrgeiz*, *Unabhängigkeit* und *aufgabenbezogene Motivation* liegen. Sie entdeckten zudem enge Beziehungen von *Leistungsmotivation* zu den *Big 5*-Dimensionen *Gewissenhaftigkeit* und *Neurotizismus*. Mäßige Zusammenhänge ergaben sich für die Einzeldimensionen *Dominanz* mit den beruflichen Kriterien *Positionsniveau* und *Gehalt.*

Zusammenfassend lässt sich festhalten, dass die Untersuchungen zur Motivation und Berufserfolg bisher zu keinen eindeutigen Schlüssen führen. Aus diesem Grund können ebenso wenig Schlussfolgerungen für die vorliegende Untersuchung gezogen werden, zumal Motivationskonstrukte nicht explizit erfasst werden (siehe 2.5.1.4).

2.5.2.4 Biographie

Biographische Daten sind „Informationen, die über die Lebensgeschichte eines Individuums, in Ausschnitten oder insgesamt, Auskunft geben (Jäger, 1995, S.351). Sie können objektive Lebenslaufinformation wie subjektive Einstellungen und individuelle Erfahrungen beinhalten. Die erfassten Daten unterscheiden sich nicht von denen in Interviews oder Persönlichkeitsfragebögen, sind allerdings faktisch nachprüfbar, da sie sich auf tatsächliches Verhalten beziehen. Der dabei zugrunde liegende Gedanke ist, dass der beste Prädiktor zukünftigen Verhaltens das vergangene Verhalten ist (Owens, 1976, S.625). Biographische Daten sind aber bisher ohne theoretisches Fundament geblieben. Hilfsweise können allerdings Sozialisations- und Lerntheorien als Erklärungshintergrund herangezogen werden (Jäger, 1995). Zahlreiche Studien konnten bisher biographischen Daten gute Vorhersagewerte für berufliche Kriterien bescheinigen. Schon Kirkpatrick (1960) fand sehr hohe Zusammenhänge von biographischen Fragebögen bei Managern mit Vorgesetztenurteilen der

Managementleistung (r=.81) und mit Gehalt (r=.44). Metaanalytische Studien kommen, je nach betrachteter Berufsgruppe, zu Validitäten zwischen r=.39 und r=.52 für das Vorgesetztenurteil (Bliesener, 1991; 1992; Reilly & Chao, 1982; Vinchur, et al., 1998). Dabei ist von stark moderierenden Effekten auszugehen (Bliesener, 1996). Schuler und Funke (1989) nennen einen mittleren Validitätskoeffizienten in Höhe von r=.37 für Berufserfolg. Funke, Krauss, Schuler und Stapf (1987) zeigten in ihrer Studie sogar die Überlegenheit im Forschungs- und Entwicklungsbereich gegenüber kognitiven Verfahren. Stehle (1986) kam in seiner Analyse gar zu dem Schluss, dass sie, neben ACs, der valideste Prädiktor für Berufserfolg sind (r=.40 bis r=.70). Mount, Witt und Barrick (2000) fanden inkrementelle Validitäten empirisch hergeleiteter Biographieskalen zur *Intelligenz* und *Persönlichkeit*.

Die Validitätskoeffizienten sind zudem kriterienspezifisch zu differenzieren. Für objektive Kriterien werden höhere Validitäten angenommen (Cook, 1998). So berichten Schmitt, et. al. (1984) von Validitäten in Höhe von r=.52 für *Gehalt* und geringeren r=.23 für *Ausbildungsleistung*, letztere sicherlich auch bedingt durch geringe Verhaltensstabilität von Personen in jungen Jahren (Funke, 1986). Zusätzlich können der Zeitpunkt der Erhebung, sowie die untersuchte Zielgruppe moderierende Einflüsse auf die Validität haben (Schuler, 2000). Höhere Validitäten kommen außerdem verhaltens- und persönlichkeitsbezogenen Informationen sowie für das Unternehmen spezifisch entwickelten Inhalten zu (Cook, 1998). Neben den Validitätsstudien sind auch Untersuchungen zur *Konstruktvalidierung* vorgenommen worden, die ergaben, dass biographischen Inventaren stabile, interpretierbare Faktoren zugrunde liegen (z.B. Baehr & Williams, 1967; Cassens, 1966; Mumford & Owens, 1987). Die theoretische Überlappung der dabei identifizierten Strukturen mit Intelligenz und Persönlichkeitsfaktoren zeigen sich in korrelativen Zusammenhängen mit diesen (Schmidt, 1988; Tenopyr, 1994). Durch biographische Daten gemessene Persönlichkeitskonstrukte haben dabei höhere prognostische Validitäten als jene, die mit Persönlichkeitstests identifiziert wurden (Mumford, Snell & Reiter-Palmon, 1994).

Eine ebenso aus der Biographie ermittelbare Information sind die *Schul-* und *Studienleistungen* des Bewerbers (gängigerweise in Notenform). Sie erfassen einen großen Ausschnitt des Verhaltensspektrums der Person, besitzen allerdings ungünstige messtechnische Eigenschaften (Sieglen, 1998). Dabei gelten gute Schulnoten als ein generelles Indiz für Engagement und Anpassungsbereitschaft (Berthel, 1995). Baron-Boldt, Funke und Schuler (1989) zeigten hohe prädiktive Validitäten von Schulnoten für Ausbildungsleistung (r=.41) und Studienleistungen (r=.46), wobei der valideste Prädiktor die Mathematiknote war. Zur Vorhersage von Berufserfolg scheinen diese allerdings nur begrenzt von Nutzen zu sein

(r=.15; Schuler, 2000). Ihre Validität nimmt mit zunehmenden Zeitabstand zur Kriteriumsmessung zudem ab. Die Validität von *Examensnoten* ist wiederum durch die stark unterschiedlichen Notenmaßstäbe der ausbildenden Institutionen eingeschränkt. O'Leary (1980) berichtet von durchschnittlichen Validitäten von r=.17, Reilly und Chao (1982) von etwas höheren r=.20. Die höhere prädiktive Validität von Noten für *Ausbildungsleistungen* im Vergleich zur Berufsleistung liegen nach Sieglen (1998) in ihren unterschiedlichen intellektuellen Anforderungen. So sind Ausbildungsaufgaben vorgegeben und haben wenig Bezug zu Bedürfnissen, Motiven, Zielen, Plänen oder der Lebenserfahrung des Aufgabenlösers. Zudem sind sie eindeutig, exakt definiert, mit in der Regel nur einer Lösung, die aus der vorhandenen Information erschlossen werden kann. Ebenso eingeschränkt ist auf Grund z.B. rechtlicher Vorgaben oder dem Phänomen des Weglobens die Interpretation von Arbeitszeugnissen (Schuler, 2000). Das *Einholen von Referenzen* schließlich, hat nach Hunter und Hunter (1984) eine prädiktive Validität auf den beruflichen Erfolg von r=.26. Ähnlich hohe Werte (r=.20) berichten Moser und Rhyssen (2001) in einer neueren Studie. Weinert (1992) weist auf die fragliche Aussagekraft, bedingt durch Gefälligkeitsaussagen und der Schwierigkeit, daraus auf Verhalten im zukünftigen Arbeitsumfeld zu schließen, hin.

Die geschilderten Befunde aus den vorherigen Abschnitten führen zu der Erwartung, dass auch die in der Untersuchung vorliegenden biographischen Bewerberinformationen valide sein dürften. Auf Grund der gesamten gesichteten Literatur dürften die Werte für einzelne Variablen in einem Bereich von r=.39 bis r=.52 liegen. Die prognostische Validität von Noten müsste den Befunden nach allerdings eher gering sein, da Untersuchungsgegenstand der Berufserfolg und nicht die Ausbildungsleistung ist.

2.5.2.5 Wissen

Der Einfluss von Wissen auf den beruflichen Erfolg ist Gegenstand zahlreicher Untersuchungen gewesen. Sujan, Sujan und Bettman (1988; 1991) z.B. zeigten in ihren Studien mit Verkaufsangestellten, dass erfolgreiche Verkäufer umfangreichere vernetzte Wissensstrukturen über ihre Kunden besitzen und über effektivere Verkaufstrategien verfügen als weniger erfolgreiche Kollegen. McCloy, et. al. (1994) entdeckten bei der amerikanischen Armee, wie von ihnen postuliert, dass *allgemeine kognitive Fähigkeiten* auf *deklaratives* und *prozedurales Wissen* Einfluss haben. Hunter (1986) berichtet von hohen Korrelationen zwischen Fähigkeiten, Berufswissen und Berufsleistung. Schmidt, et al. (1986) zeigten pfadanalytisch die Bedeutung von Berufswissen, das durch *Berufserfahrung* und *kognitive Leistungsfähigkeit* beeinflusst wird, für *Berufsleistung* bei Berufen mittlerer Komplexität. Dabei

war der Effekt bei dem Kriterium *Vorgesetztenurteil* stärker als bei der *Arbeitsprobe* und bei Konstanthaltung der *Berufserfahrung* erhöhte sich der direkte Effekt auf die *berufliche Wissensaneignung*. Dye, Reck und McDaniel (1993) untersuchten metaanalytisch den Zusammenhang von Berufswissen und Berufsleistung und kamen zu der Erkenntnis, dass die Validität von Berufswissenstests generalisierbar ist und *Inhaltsgleichheit* und *Komplexität* der Berufe einen moderierenden Effekt ausüben. Für „job knowledge", gemessen mit für den Arbeitsplatz spezifisch entwickelten Wissenstests, errechneten sie eine durchschnittliche Validität von r=.45, bzw. r=.62 bei starker Nähe des Wissens zum Arbeitsplatz. In ihren Untersuchungen zu *tacit knowledge* bei Managern berichten Sternberg, et. al (2000) Korrelationen mit simulierten Leistungen von r=.61 und, im Vergleich zu Maßen der Intelligenz, doppelt so hohe Validitäten bei beruflichen Kriterienmaßen. Gottfredson (2003a; 2003b) kritisiert allerdings die selektive Studienwahl bei der Ergebnisdarstellung und zeigte genau gegenteilige Ergebnismuster. Obwohl also Fachkenntnisse bzw. Wissen offenbar beträchtliche prognostische Validität besitzt, werden sie als Prädiktor in der Praxis der Personalauswahl eher selten eingesetzt. Gründe hierfür liegen in ihrem Ruf, rasch überholt zu sein oder in ihrer guten Verfügbarkeit durch geschickte Informationszugriffe (Schuler, 2000).

Den geschilderten Befunden nach müssten sich auch in der Untersuchung des vorliegenden Auswahlverfahrens Zusammenhänge zwischen den gemessenen Wissensaspekten und späterer beruflicher Leistung aufdecken lassen. Außerdem dürfte die Berufserfahrung, die aus den biographischen Informationen der Bewerber zu entnehmen ist (bisherige Praktika und Anstellungen) ebenfalls mit dem Berufserfolg als Berater einhergehen.

2.5.3 Überblick ausgewählter eignungsdiagnostischer Verfahren

Die vorherigen Abschnitte haben gezeigt, dass es eine Vielzahl von empirischen Befunden im Bereich der Personalauswahl, mit höchst unterschiedlichen Validitätsergebnissen auch für ein und denselben Prädiktor, gibt. Schuler (2000, S.165) nennt es deshalb problematisch, von *der* Validität eines Verfahrens zu sprechen, da diese immer neben Stichprobenfehlern auch von systematischen Größen beeinflusst würde. Außerdem sei der Einfluss der Methodenvarianz auf das Ergebnis bei den Kriterien nicht geringer als bei den Prädiktoren. Schuler, et al. (1995) stellten außerdem fest, dass die Qualität der Kriterien erheblich die Validität beeinflussen kann. Auch metaanalytisch zeigten Schmitt, Gooding, Noe und Kirsch (1984), dass, je nach Kriterienwahl, die prognostischen Validitätskoeffizienten gleichbleibender Prädiktoren variieren. Als weitere Moderatoren nennt Schuler die *Zielgruppe* und die *Kompetenz des Durchführenden*.

Tabelle 14: Durchschnittliche Validitätskoeffizienten ausgewählter Auswahlverfahren zur Vorhersage von Berufserfolg (Berufsleistung) und inkrementelle Validitäten zu kognitiven Leistungstest (aus Schmidt & Hunter, 1998, S.265)

Auswahlverfahren	Prognosegüte[a]	Multiples R[b]	Inkrementelle Validität zu Intelligenztests
Allgemeine Intelligenztests	.51		
Integritätstests	.41	.65	.14
Test zur Erfassung berufsrelevanten Wissens	.48	.58	.07
Arbeitsproben	.54	.63	.12
Interview strukturiert	.51	.63	.12
Interview unstrukturiert	.38	.55	.04
Assessment Center	.37	.53	.02
Biographische Fragebogen	.35	.52	.01
Referenzen	.26	.57	.06
Berufserfahrung	.18	.54	.03
Interessen	.10	.52	.01
Graphologie	.02	.51	.00

[a] korrigiert um Varianzeinschränkung und Kriteriumsunreliabilität
[b] Validitätskoeffizient bei Hinzunahme des Instruments zusätzlich zu kognitiven Leistungstests im Rahmen einer multiplen Regression

Alles in allem erfordert die Vielzahl der Ergebnisse Möglichkeiten, um der Informationsüberflutung zu bewältigen. Eine dieser Möglichkeiten bieten die bereits mehrfach angesprochenen Verfahren der Validitätsgeneralisierung (Burke, Raju & Pearlman, 1986). Diese integrieren die statistischen Befunde einzelner Untersuchungen quantitativ zu einem Ergebnis. Die populärste metaanalytische Technik ist die von Hunter und Schmidt (Hunter, Schmidt & Jackson, 1982; Hunter & Schmidt, 1989; 1990). Dabei werden die durchschnittlichen Effektkenngrößen der Einzelstudien um Artefakte (Stichprobenfehler, Messfehler, Varianzeinschränkungen) bereinigt und, wenn weiter Differenzen zwischen den Kenngrößen bestehen, nach möglichen Moderatorvariablen gesucht. Mit dieser Technik konnten Schmidt und Hunter die bis Ende der Siebziger populäre *Situational Specifity Theory*, die Validitätsunterschiede auf Arbeitsstellencharakteristika zurückführen wollte, weitgehend entkräften. Für die Praxis der Personalauswahl sind Metaanalysen von herausragender Bedeutung, da sie als Anhaltspunkt für die Auswahl geeigneter Prädiktoren und zugehöriger Instrumente sowie die Identifikation möglicher Moderatorvariablen dienen können. Für die Studie in dieser Arbeit sollen sie vor allem als Anhaltspunkt für mögliche Verbesserungen des vorliegenden

Verfahrens dienen. Sie sind somit auch Basis für die Berechnung möglicher Nutzeneffekte durch Modifizierungen des bestehenden Vorgehens.

Tabelle 14 bildet die Validitäten von Auswahlverfahren auf Basis der Metaanalyse von Schmidt und Hunter (1998) ab, die wiederum eine Zusammenfassung zahlreicher anderer wesentlicher Metaanalysen darstellt. Die ersten drei genannten Verfahren sind standardisierte Methoden; die restlichen werden zu den unstandardisierten Methoden gezählt (Kanning & Holling, 2002). Die Ergebnisse zeigen, dass der beste Einzelprädiktor die *Arbeitsprobe* ist. Diese ist allerdings im Vergleich zum ähnlich validen *Intelligenztest* kostspielig und nur bei erfahrenen Bewerbern anwendbar (Schmidt & Hunter, 1998, S.264).

Intelligenztests sind somit der bedeutendste und fundierteste Prädiktor von Berufserfolg, berufsrelevantem Lernen und Ausbildungserfolg. Bei der Kombination von Intelligenztest und einem anderen Verfahren zeigen sich die höchsten Validitäten (über r=.60) in Kombination mit *Arbeitsproben*, *Integritätstests* oder *strukturierten Interviews*. Teilweise sogar höhere Validitäten berichten Schmidt und Hunter für die Vorhersage von Ausbildungserfolg[11], was insofern nicht verwundert, da *g* ja als Fähigkeit zum Lernen (s.o.) aufgefasst wird. Die große Bedeutung kognitiver Leistungstests konnte von Salgado, Anderson, Moscoso, Bertua und de Fruyt (2003) in einer Metaanalyse, in der eine mittlere Prognosegüte für Berufserfolg von r=.62 ermittelt wurde, auch für den europäischen Raum bestätigt werden. Schuler und Schmitt (1987, S.262) berichten wesentlich geringere Koeffizienten. Diese sind allerdings nicht um die Unreliabilität und die Varianzeinschränkung korrigiert, so dass ihre Aussagefähigkeit eingeschränkt ist. Zudem ist zu beachten, dass die dargestellte Prognosegüte für kognitive Leistungstests in der Metaanalyse von Schmidt und Hunter (1998) für mittel komplexe Berufe gilt. Für den Beruf des Unternehmensberaters dürften wohl höhere Validitäten, vergleichbar zu der Gruppe der Manager, in Höhe von um die r=.58 (Hunter & Hunter, 1984) zu erwarten sein. Salgado et. al. (2003) zeigten in einer weiteren europäischen Metaanalyse für kognitive Leistungstests sogar Validitäten von r=.64 bei komplexen Berufen. Judge, Higgins, Thoresen und Barrick (1999) lieferten in einer Längsschnittstudie, die den Zusammenhang von *Big 5*, *g* und *Berufserfolg* untersuchte, erste Hinweise auf inkrementelle Validitäten von mehr als zwei Prädiktoren. Sie differenzierten dabei Berufserfolg in eine extrinsische (Gehalt, Position im Unternehmen) und eine intrinsische Komponente (Arbeitszufriedenheit). Die *Big 5* und *g* korrelierten zu r=.64 mit dem extrinsischen Berufserfolg. Schmidt und Hunter weisen zudem darauf hin, dass die hohen

prognostischen Validitäten von *Intelligenztests*, *Berufserfahrung* und *Integritätstests* auf einem indirekten Effekt über *Berufswissen* beruhen.

Tabelle 15: Validität, Praktikabilität und Akzeptanz eignungsdiagnostischer Verfahren für externe Personalauswahl in der Einschätzung der Verwender (adaptiert aus Schuler, Frier & Kaufmann, 1991 und Schuler, et al., 1993)

Auswahlverfahren	Validität[a]	Praktikabilität[b]	Akzeptanz[c]	Anwendungshäufigkeit [%]
Analyse der Bewerbungsunterlagen	1.63	1.33	1.13	98
Referenzen	1.99	2.08	1.82	71
Interview strukturiert mit Personalabteilung	1.40	1.74	1.26	70
Interview unstrukturiert mit Personalabteilung	2.00	1.33	1.48	57
Interview strukturiert mit Fachabteilung	1.45	1.90	1.39	49
Interview unstrukturiert mit Fachabteilung	1.99	1.43	1.48	69
Gruppendiskussion	1.63	2.53	1.71	51
Persönlichkeitstests	2.20	2.36	2.38	21
Leistungstests	1.75	2.12	1.94	47
Intelligenztests	1.96	2.15	2.08	34
Arbeitsproben	1.32	2.11	1.53	44
Biographische Fragebögen	2.00	1.98	2.25	21
Assessment Center	1.33	2.89	1.72	39
Graphologisches Gutachten	2.90	2.25	2.78	9

[a] prozentual gewichtet nach Nennung mit gut=1, mittel=2 und schlecht=3; [b] prozentual gewichtet nach Nennung mit einfach=1, mittel=2 und aufwendig=3; [c] prozentual gewichtet nach Nennung mit voll akzeptiert=1, eingeschränkt=2 und nicht akzeptiert=3

Neben der Frage der Validität, ist die *Praktikabilität* und *Akzeptanz* der einzelnen Verfahren entscheidend für die Implementierung eines Auswahlverfahrens in einem Unternehmen. Schuler, Frier und Kaufmann (1991) untersuchten diese in einer empirischen Erhebung für den deutschen Raum (alte Bundesländer). Die Ergebnisse sind in *Tabelle 15* dargestellt. Dabei zeigt sich, dass die Anwender die höchste Validität den *Arbeitsproben*, dem *Assessment Center* und dem *strukturierten Einstellungsinterview* zuschreiben. Als praktikabel werden das *unstrukturierte Einstellungsinterview* und die *Analyse der Bewerbungsunterlagen* eingestuft. Die höchste Akzeptanz findet sich bei der *Analyse der Bewerbungs-*

[11] Problematisch ist hierbei, dass es in den U.S.A. keine vergleichbaren Standards im Sinne einer deutschen Ausbildung gibt; vielmehr wird hier der Trainingserfolg bei Weiterbildungsmaßnahmen erfasst

unterlagen, dem *strukturierten Einstellungsinterview*, der *Arbeitsprobe* und dem *unstrukturierten Interview*. *Assessment Center*, *Gruppendiskussionen*, und Tests, wie z.B. *Intelligenztests*, scheinen u.a. aus Akzeptanz- und Praktikabilitätsgründen in der Praxis auf Widerstände zu stoßen. Ähnliche Akzeptanzwerte auf Bewerberseite fanden Schuler, et. al. (1995) auch in einer neueren Untersuchung.

Im Folgenden soll abschließend ein Abriss der diagnostischen Instrumente gegeben werden, die auch in der vorliegenden Untersuchung zum Einsatz kommen. Eine ausführliche Übersicht dieser und weiterer diagnostischer Verfahren nach unterschiedlichen Unterteilungskriterien findet der interessierte Leser in Amelang und Zielinski (2002), Kanning und Holling (2002), Sarges und Wottawa (2001), Schuler (2000; 2001b) sowie Schuler und Funke (1995).

2.5.3.1 Analyse der Bewerbungsunterlagen

Die *Analyse der Bewerbungsunterlagen* wird im Bereich der Personalauswahl universell als Vorauswahlinstrument verwendet (Schuler, et al., 1993) und ist das am häufigsten eingesetzte Verfahren (Seibt & Kleinmann, 1991, siehe *Tabelle 15*). Neben der Überprüfung der Einhaltung formaler Aspekte und Voraussetzungen beinhaltet sie im Wesentlichen die Beurteilung des Anschreibens (Stil, Inhalt, Struktur), sowie die Bewertung der biographischen Informationen des Lebenslaufes und der beigelegten Arbeits- und Ausbildungszeugnisse (Schuler, 2000). Problematisch ist die bisher mangelnde Kenntnis über die optimale Verwertung der Information. Praktiker gehen dabei offenbar in einem iterativen Wechsel von Detailanalyse und gestalthafter Gesamtbetrachtung vor und kommen so zu einem Urteil (Schuler, 2000). Das Vorgehen besteht dabei in einer Art „Schlechtestenauslese“ (Hossiep, 1995; Machwirth, Schuler & Moser, 1996). Dies ist problematisch, da, wie Kanning (2002) anmerkt, an *Bewerbungsunterlagen* die gleichen testtheoretischen Qualitätskriterien wie bei Testverfahren angelegt werden sollten. Die von den Praktikern zur Selektion herangezogenen Hypothesen sind allerdings selten wissenschaftlich überprüft (Rastetter, 1999). Dies gilt vor allem für die Informationen aus dem Lebenslauf (z.B. Berufserfahrung, Auslandsaufenthalte). Für deren Tauglichkeit als Indikatoren künftiger Verhaltensweisen gibt es zwar Anhaltspunkte (Hollmann & Reitzig, 1995), diese sind jedoch bisher nicht statistisch abgesichert worden. Auch Arbeitszeugnisse sind, besonders auf Grund ihrer sprachlichen Interpretationsschwierigkeit, einer fehlenden Standardisierung sowie gesetzlicher Bestimmungen, in ihrer Aussagefähigkeit und prognostischen Validität eingeschränkt.

In dem untersuchten Auswahlverfahren dieser Arbeit wird ebenfalls eine Vorauswahl der Bewerber anhand der Bewerbungsunterlagen vorgenommen. Dabei wird auf Grund mangelnder Kenntnis über das Verwertungsoptimum der Informationen ebenfalls im Sinne einer „Schlechtestenauslese“ vorgegangen. Die biographischen Informationen des Lebenslaufes der Bewerber sind allerdings Bestandteil der statistischen Analysen des empirischen Teils. Es besteht deshalb die Hoffnung, dass im Verlauf der Untersuchung, statistisch abgesicherte Ergebnisse Anhaltspunkte liefern können, welche Information wie berücksichtigt werden sollte. Damit könnte das geschilderte Problem der mangelnden testtheoretischen Qualitätskriterien abgeschwächt werden.

2.5.3.2 Interviews

Interviews (zur Definition siehe Wiesner & Cronshaw, 1988) werden bei der Personalauswahl nahezu immer durchgeführt. Sie gelten als die in der psychologische Praxis am häufigsten eingesetzte Methode zur Informationsgewinnung über menschliches Verhalten (Westhoff, 2000). Die dabei gewonnene Information kann sehr umfangreich sein und z.B. Aspekte des Lebenslaufes, persönliche Einstellungen, Interessen, Motive oder Berufsvorstellungen betreffen. Quellen zur Urteilsbildung sind dabei verbale (z.B. Inhalt, Rhetorik) sowie nonverbale (Aussehen, Auftreten usf.) Informationen des Bewerbers (vgl. Burnett & Motowidlo, 1998). Im Gegensatz zu anderen Auswahlinstrumenten wird der Interviewprozess durch eine Vielzahl von Determinanten beeinflusst (siehe Lopez, 1965) und ist von einer starken Dynamik gekennzeichnet. Diese zeigt sich auch in einem hohen Ausmaß der *Reaktivität*, d.h. einer wechsel- und gegenseitigen Beeinflussung der Interviewteilnehmer (Blum & Naylor, 1968). Darüber hinaus wird das Interviewverhalten durch das Rollenverständnis und die -interaktion der Teilnehmer bedingt. Auf Grund der hohen multiplen Anforderungen (Prozessbeteiligter und Beobachter) ist dabei eine hohe Handlungskompetenz des Interviewers erforderlich (Kompa, 1989. S.165ff.). Interviews unterliegen zahlreichen *Quellen der Verzerrung*, die mit Hilfe der *Sozialen Urteilsbildungstheorie* (siehe auch 2.2.2.3) erklärt werden können. Die wichtigste auf Seiten des Bewerbers ist das *impression management*, d.h. die Formung und Kontrolle des Eindruckes über die eigene Person, die andere gewinnen sollen (siehe Barrick, Patton & Haughland, 2000). Zudem können Merkmale der Interviewers (Alter, Geschlecht, Status) das Verhalten des Bewerbers bedingen sowie sich aus der Interaktion ergebende Einflüsse (z.B. *„self-fulfilling prophecy“*; siehe Kompa, 1989. S.189) entstehen. Schließlich können verzerrte Wahrnehmungen (z.B. *Stereotype*, *Vorurteile*) und Interpretationen des Interviewers (z.B. implizite Persönlichkeitstheorien), u.U. bedingt durch

Vorabinformationen und Erwartungen, das Urteil beeinträchtigen (zusammenfassend siehe Schuler, 2002).

Interviews lassen sich in *strukturierter* und *unstrukturierter* Form durchführen (Huffcut, Roth & McDaniel, 1996; McDaniel, Whetzel, Schmidt & Maurer, 1994). Letztere Variante beinhaltet weder vorgefertigte Fragen noch eine vorgegebene Systematik zum Beurteilen des Interviewten. Wenn überhaupt, dann wird ein *klinisches Urteil*, im Sinne eines Gesamteindrucks über den Bewerber, vielleicht auch in Form von Ratings auf unterschiedlichen Dimensionen abgegeben. Im Gegensatz hierzu sind beim *strukturierten Interview* die Fragenstruktur, Anforderungen und Beurteilungskriterien basierend auf einer AA vorgegeben und der Ablauf und die Beurteilungsskalen (idealerweise verhaltensverankert) standardisiert. Dies gilt auch für die Regeln, wie man zu einem Urteil kommt. Dabei ist der Einsatz mehrerer Beurteiler und Interviewerschulungen empfehlenswert (Kanning, 2002). Eine Übersicht von (teilweise allerdings etwas praxisfernen) Empfehlungen für die Ausgestaltung strukturierter Interviews liefern Campion, Palmer und Campion (1997, siehe auch *Tabelle 68*). Ein klassisches Beispiel für ein gut strukturiertes Interview ist das *Multimodale Interview* (*MMI*) nach Schuler (1992a). Es integriert dabei die diagnostischen Ideen des trimodalen Ansatzes der Berufseignungsdiagnostik aus *Abbildung 7*, indem es sowohl *eigenschafts*-, als auch *simulations*- und *biographieorientierte* Fragestellungen kombiniert. Beim „*Behavioural Description Interview"* (*BDI*; Janz, 1982; Janz, Hellervik & Gilmore, 1986) und dem „*Structured Behavioural Interview"* (*SBI*, Motowidlo, et al., 1992) wird der Bewerber nach erlebten kritischen Ereignissen gefragt, während das „*Situational Interview"* (*SI,* Latham, Saari, Pursell & Campion, 1980) hypothetische zukünftige Problemstellungen zum Inhalt hat. Eine weitere hochstrukturierte Interviewvariante ist das „*Comprehensive Structured Interview"* (*CSI,* Pursell, Campion & Gaylord, 1980). Das *Entscheidungsorientierte Gespräch* (*EOG*; Westhoff & Kluck, 2003) schließlich ist eine Technologie, die Regelsysteme zur Planung, Durchführung und Auswertung voll strukturierter Interviews zur Verfügung stellt. Insgesamt haben sich die beschriebenen Formate in der Praxis allerdings nicht durchgesetzt und sind nur bei den jeweiligen Erfindern Bestandteil von Forschungsarbeiten gewesen (Lochner-Hannen, 2002).

Interviews sind seit längerem, insbesondere in den USA., Gegenstand intensiver Forschungen. Das Gütekriterium der *Interrater-Reliabilität*, also der Beurteilerübereinstimmung, eigentlich ein Maß der Objektivität, variiert dabei je nach betrachteter Dimension (Wahrnehmungs-, Bewertungs-, Prognose-, Entscheidungsleistung und Erschließung latenter Konstrukte). Zudem hängt es von der Komplexität der Beurteilungsleistung ab (Kompa, 1989;

Trost, 1996). Außerdem variieren die Werte je nach zu beurteilendem Merkmal (Mayfield, 1964; Wagner, 1949), wobei sich die *strukturierte* Interviewvariante und die Beurteilung der *Intelligenz* als Merkmal am reliabelsten erweisen. Eine große Anzahl von Untersuchungen kommt auf Werte um r=.50 oder leicht darüber, während bei freien Auswahlgesprächen lediglich Objektivitätswerte von um r=.30 erzielt werden (Schuler, 2002). Hohe Objektivitätsmaße von um r=.80 konnten hingegen für das *MMI* festgestellt werden (Schuler, et al., 1995). *Intrarater-Reliabilitäten,* also die *Stabilität*, sind demgegenüber, auf Grund des Wesens des Interviews und der einhergehenden Unmöglichkeit, parallele unabhängige Situationen zu schaffen, schwer ermittelbar. Trost (1996) berichtet eine Stabilität von Interviewwerten von r=.60, wobei das *SI* eine Höhe von nur r=.04 erreicht.

Zur *prognostischen Validität* von Interviews[12] sind ebenfalls zahlreiche metaanalytische Untersuchungen durchgeführt worden, die z.T. auf Grund methodischer Mängel zu stark divergierenden Ergebnissen kommen, die in ihrer Vergleichbarkeit eingeschränkt sind (siehe dazu Lochner-Hannen, 2002, S.70ff. und Robertson & Smith, 2001). Reilly und Chao (1982) und Hunter und Hunter (1984) kommen in ihren Untersuchungen zu einer durchschnittlichen Validität von Interviews von lediglich r=.19 bzw. r=14. Eine Metaanalyse von Wright, Lichtenfels und Pursell (1989) ergab eine höhere korrigierte durchschnittliche Validität *strukturierter Interviews* von r=.37. Hinweise, dass diese eine höhere prognostische Validität aufweisen als ihre *unstrukturierte* Variante, ergaben bereits in frühen Jahren in einigen Studien (z.B. Mayfield, 1964; Ulrich & Trumbo, 1965; Wagner, 1949). Auch McDaniel et. al. (1994) berichten korrigierte Validitäten von r=.37 (gesamt), r=.44 (strukturiert) und r=.33 (unstrukturiert). Wiesner und Cronshaw (1988) kommen dabei zu vergleichbaren Ergebnissen wie Schmidt und Hunter (1998), mit korrigierten Werten von r=.47 (insgesamt), r=.62 (strukturiert) und r=.31 (unstrukturiert). Ähnliche Ergebnisse berichten Huffcut und Arthur (1994) sowie Campion, Pursell und Brown (1988). Cortina, Goldstein, Payne, Davison und Gilliland (2000) zeigten metaanalytisch am Berufserfolg inkrementelle Validitäten hochstrukturierter Interviews zur *Intelligenz* und *Gewissenhaftigkeit* ($R_{inkrementell}$=.123). Für das *MMI* wurden in verschiedenen Studien Validitäten zwischen r=.30 und r=.50 ermittelt (Schuler, 2002, S.206ff.). Conway, Jako und Goodman (1995) berechneten metaanalytisch durch mangelnde Objektivität bedingte Validitätsobergrenzen von r=.34 für gering strukturierte, r=.56 für mittel strukturierte und r=.67 für hochstrukturierte Interviews. Nach Hossiep (1995) liegt die geringere Validität unstrukturierter Interviews insbesondere an der frühen

[12] Kompa (1989) weist darauf hin, dass das eigentliche Messinstrument der Interviewer durch seine Urteilsbildung ist und man deshalb besser von der *Validität des Interviewers* sprechen sollte

Urteilsbildung, die mit einem höheren Gewicht in die Informationsverarbeitung eingeht und den weiteren Gesprächsverlauf beeinflusst, sowie an den klassischen Urteilsfehlern (siehe 2.2.2.3). Kataoka, Latham und Whyte (1997) fanden z.B. eine höhere Resistenz von SI, die genaue Beurteilungsanleitungen enthalten, im Vergleich zu den weniger strukturierten BDI und „klassischen" strukturierten Interviews gegenüber *Anchoring*-Effekten. Bei gleicher Beurteilungsanleitung wurden hingegen auch gleich hohe prognostische Validitäten berichtet (Day & Carroll, 2003).

Bezüglich neben der Strukturiertheit von Interviews weiterer, *differenzieller Validitäten* liegen unterschiedlichen Befunde vor. Hunter und Hunter (1984) fanden z.B. für das Vorgesetztenurteil als Kriterium die höchste Validität. Schmidt und Rader (1999) berichten hingegen gleich hohe Validitäten für Vorgesetztenurteil und objektive Leistungskriterien. Auch McDaniel et. al. (1994) fanden kriterienspezifisch (r=.37 für Berufserfolg, r=.36 für Trainingserfolg, r=.20 für Verweildauer) sowie inhaltsspezifisch variierende korrigierte Validitäten (r=.50 für situative, r=.39 für tätigkeitsbezogene und r=.29 für psychologische Inhalte). Campion, Campion und Hudson (1994) zeigten in einer Einzelstudie zudem höhere korrigierte Validitäten biographischer gegenüber zukunftsbezogener, situativer Fragen (r=.51 vs. r=.39). Zu der gleichen Erkenntnis kommen Motowidlo und Burnett (1995) und Pulakos und Schmitt (1995). Die Validität scheint ebenso abhängig von der zu beurteilenden Dimension zu sein. Schmitt (1976) und Ulrich und Trumbo (1965) empfehlen deshalb ihren Einsatz auf die Beurteilung von *Motivation* und *sozialen Komponenten* zu beschränken. Vereinzelt wurden auch „*board"*- oder „*panel-interviews"*, bei denen der Bewerber durch mehrere Personen gleichzeitig interviewt wird, höhere Validitäten zugeschrieben (z.B. Arvey & Campion, 1982). Dies steht aber im Widerspruch zu neueren Untersuchungen, die keine oder genau gegenteilige Effekte entdeckten (Conway, et al., 1995; Marchese & Muchinsky, 1993; McDaniel, et al., 1994; Wiesner & Cronshaw, 1988). In diesem Zusammenhang wurde auch von einer Überlegenheit von Konsensurteilen gegenüber gemittelten individuellen Urteilen berichtet (z.B. McDaniel, et al., 1994; Pulakos, Schmitt, Whitney & Smith, 1996). Schmidt und Zimmermann (2004) berichten von der Ebenbürtigkeit 3-4 gemittelter unabhängiger Urteile aus unstrukturierten Interviews zu einem strukturierten Einzelurteil. In einer Einzelstudie kamen Tullar, Mullins und Caldwell (1979) außerdem zu dem Schluss, dass sich bei einer vorab vereinbarten kurzen *Interviewdauer* (15 Min.) die Interviewer schnell entscheiden und eine längere Interviewdauer bei komplexen Tätigkeiten zu empfehlen ist.

Ergebnisse aus Untersuchungen zur *Konstruktvalidität* von Interviews sind noch nicht abschließend beurteilbar (Sarges, 1995a). Huffcut, Conway, Roth und Stone (2001) fanden

metaanalytisch sieben Konstruktkategorien, die mithilfe von Interviews in der Praxis gemessen werden. Zu ihnen gehören die der Persönlichkeitskategorie zugeordneten *Gewissenhaftigkeit*, *Emotionale Stabilität* und *Extraversion* als die häufigsten. Schmidt und Rader (1999) vermuten, dass Interviews Konstrukte generalisierbarer Validität wie *Gewissenhaftigkeit* oder *allgemeine Intelligenz* messen. Die Zusammenhänge zu *allgemeiner Intelligenz* und nicht-kognitiven *Temperament-Variablen* zeigten sich auch in anderen Studien (Campion, et al., 1994; Huffcut, et al., 1996; McDaniel, et al., 1994; Salgado & Moscoso, 2002; Schuler & Funke, 1989). Salgado und Moscoso (2002) kommen zu dem Schluss, dass verhaltensbezogene strukturierte Interviews *Fachkenntnisse*, *Berufserfahrung* und *soziale Kompetenz*, konventionelle Varianten dagegen neben *sozialen Fähigkeiten*, *Intelligenz* und *globale Persönlichkeitsmerkmale* erfassen (siehe auch Huffcut, Conway, et al., 2001). Barrick, et. al. (2000) zeigten in einer Einzelstudie allerdings, dass dies gerade für die berufserfolgsrelevanten Dimension *Emotionale Stabilität* und *Gewissenhaftigkeit* nicht gilt, wobei sich keine moderierenden Effekte der Struktur oder der Inhalte des Interviews zeigten. Conway, et. al. (1995) berichten zudem von höheren Homogenitäten *unstrukturierter* Interviews im Vergleich zu *strukturierten* Interviews, die zu differenzierteren Beurteilungen führen. Schuler (2002) spricht in diesem Zusammenhang von einem Globalkonstrukt „Eignung für eine bestimmte Position“. *MMIs* schließlich ermöglichen die Erfassung vorher intendierter Konstrukte in zufrieden stellendem Maße (Schuler, 2002). Die Konstruktvalidität scheint also nicht abschließend geklärt zu sein. Verschiedene Untersuchungen mit *MTMM-Methoden* kamen zudem beim Vergleich von Interviewformen auf höhere oder gleich hohe *diskriminante* im Vergleich zu *konvergenten* Validitäten (Conway & Peneno, 1999; Facteau, Facteau, Jackson & Becton, 2000; Huffcut, Weekley, Wiesner, DeGroot & Jones, 2001). Das Fehlen von *Konstruktvalidität* bei vorhandener *Inhalts-* und *Vorhersagevalidität* wurde bei vergleichbaren Befunden mit *ACs* auch als *Validitätsparadoxon* (Arthur, Woehr & Maldegen, 2000) bezeichnet und wird auch bei strukturierten Interviews vermutet (Van Iddekinge, Raymark, Eidson & Attenweiler, 2004).

Strukturierte Interviews sind zwar prognostisch valider, sie sind aber in ihrer Entwicklung auch teuerer als die unstrukturierte Form oder käuflich erwerbbare Tests. Zudem ist bei einem zu hohen Strukturierungsgrad die Abgrenzung zu einem Test nicht mehr so genau gegeben (Interviews als „mündliche“ Tests, Campion, et al., 1994). In einem solchen Fall könnte man gleich auf die ökonomische Anwendung von Fragebögen übergehen. Zusätzlich kann eine „Überstrukturierung“ den Auswahlcharakter der Situation im Vergleich zur Anwerbung zu stark betonen und negativ auf den Bewerber wirken (Goodale, 1989).

Eine moderate Strukturierung, die Nachfragen oder die Klärung von Inkonsistenzen ermöglicht, scheint auch im Rahmen der Beurteilung von Persönlichkeitseigenschaften sinnvoll (Binning, LeBreton & Adorno, 1999). Dies gilt auch, wenn das Interview selbst als zwischenmenschliche Verhaltensstichprobe angesehen wird (Lanyon & Goodstein, 1997). Andererseits haben angemessen strukturierte Interviews den Bewerbern gegenüber eine höhere *Fairness* als ihre unstrukturierte Variante (Hough & Oswald, 2000). Zudem haben Interviews allgemein eine höhere Augenscheinvalidität als Tests. Zusätzlich sind sie für kleinere Unternehmen leichter zu entwickeln, einfacher als eine Validierung von Tests vorzunehmen, und haben eine gute Akzeptanz beim Management, da dieses sich stärker in das Auswahlverfahren eingebunden fühlt (Campion, et al., 1988). Auch in Studien zur *sozialen Validität*, d.h. inwiefern Bewerber ein Instrument als akzeptable soziale Situation erleben (Schuler & Stehle, 1983), zeigte sich das Interview den anderen Verfahren als überlegen (Schuler, 2002). Insgesamt kann deshalb die Applikation von *Intelligenztests* und *strukturierten Interviews* bei der Personalauswahl als eine interessante, hoch valide und bezüglich ihrer Gesamtkosten ökonomische Kombination angesehen werden (Schmidt & Hunter, 1998, S.267).

In dem untersuchten Auswahlverfahren dieser Arbeit kommen ebenfalls Interviews zur Anwendung. Die dabei gewonnene Information deckt die gesamte Bandbreite dessen ab, was eingangs dieses Kapitels aufgezählt wurde. Die eingesetzten Interviewer sind Berater des Unternehmens, die in Interviewtechniken geschult wurden. Es ist deshalb davon auszugehen, dass sie eine ausreichende Handlungskompetenz zur Durchführung des Interviews besitzen. Berater dürften sich zudem wegen ihrer durch Berufserfahrung angereicherten kognitiven Struktur und ihren analytischen Fähigkeiten gut als Interviewer eignen (Guion, 1998, S.621f.). Allerdings ist dabei davon auszugehen, dass Vorabinformationen über den Bewerber, die jedem Interviewer vorliegen, Erwartungen generieren, die das Urteil beeinträchtigen können. Zudem handelt es sich bei den untersuchten Interviews lediglich um semi-strukturierte Varianten (siehe auch McDaniel, et al., 1994), bei denen der Ablauf nicht standardisiert ist. Lediglich die vorliegenden Beurteilungsskalen, auf denen die Bewerber eingestuft werden, sind einheitlich mit teilweise verhaltensverankerten Dimensionen. Es dürften deshalb die in diesem Kapitel erläuterten Nachteile unstrukturierter Interviews zum Ausdruck kommen.

2.5.3.3 Assessment Center

Assessment Center (*AC*) sind multiple Verfahrenstechniken, die aus mehreren, teilweise oben genannten eignungsdiagnostischen Instrumenten bestehen, und können somit als

integratives Vorgehen verstanden werden. Bestandteile von *ACs* sind *situative Verfahren* (Hossiep, 1995), d.h. Verfahren, die wichtige Erfolgsfaktoren aus dem sozial-interaktiven Bereich situativ, nach dem Prinzip der Verhaltenssimulation, erfassen wollen. Dazu gehören beispielsweise die führerlose Gruppendiskussion, Vorträge, Präsentationen, Rollenspiele und Fallstudien. Ferner können sie weitere Einzelverfahren wie Fähigkeitstests, Persönlichkeitstests, biographische Fragebögen und Einstellungsinterviews beinhalten (vgl. Fisseni & Fennekels, 1995; Jeserich, 1991; Kleinmann, 1997; Sarges, 1996, Schuler, 1992). Charakteristisches Merkmal ist dabei, dass an ihnen sowohl mehrere Bewerber als auch mehrere unabhängige Beurteiler teilnehmen (Schuler, 2000). *ACs* sollten auf einer AA aufbauen (Kanning, 2002) und erfreuen sich einer großen Beliebtheit im Rahmen der Führungskräfte- und Hochschulabsolventenauswahl. Vorbild einer Vielzahl von *ACs* ist die *AT&T*-Management-Studie, bei der Validitäten mit Berufserfolg (Beförderungskriterium) in Höhe von r=.46 festgestellt wurden (Bray, et al., 1974). Hunter und Hunter (1984) und Schmitt (1984) berichten ähnlich hohe korrigierte Validitäten für das Vorgesetztenurteil (r=.43) als Kriterium. Neuere *ACs* fokussieren hauptsächlich auf *Arbeitsproben* und *Tätigkeitssimulationen*, da diese eine höhere Transparenz und Partizipation aller Beteiligten fördern. Es kann allerdings, wie Schuler (2000) anmerkt, auch an der mangelnden Kompetenz der meisten Durchführenden bezüglich des geeigneten Einsatzes von Testverfahren liegen, zumal es bisher keine wissenschaftlich fundierten Möglichkeiten der Auswahl und Konstruktion der einzusetzenden Verfahren gibt.

Die Leistungen in *ACs* werden üblicherweise pro Dimension anhand mehrerer Aufgaben von (im Idealfall geschulten) Beurteilern mit Hilfe vorgegebener Skalen eingeschätzt. Sie orientieren sich somit am *Multi-Trait-Multi-Method-Ansatz* (Campbell & Fiske, 1959). Sie beinhalten neben Verhaltensbeschreibungen auch Interpretationen durch den Beurteiler (Hossiep, 1995) und unterliegen damit den gleichen potenziellen Fehlerquellen wie die Leistungsbeurteilungen (siehe 2.2.2). Alternativen zu den einfachen Skalen sind *Checklisten* und *verhaltensverankerte Skalen*. Die Verlässlichkeit der Beurteilungen kann dabei durch eine *Anforderungs-Aufgaben-Matrix* erhöht werden (Schuler, 1991a). Zudem ist der Einsatz von *Rotationsplänen*, d.h. Beobachter beurteilen wechselnde Bewerber auf abwechselnden Dimensionen, sinnvoll (Kanning, 2002).

Untersuchungen zur *Konstruktvalidität* haben gezeigt, dass die Korrelationen zwischen den einzelnen Beurteilerdimensionen von *ACs* hoch und somit die *diskriminante Validität* gering ist (Maukisch, 1989; Sackett & Dreher, 1982). Gleichzeitig zeigen sich geringe Korrelationen gleicher Dimensionen von verschiedenen Beurteilern, also hohe konvergente

Validitäten (Schuler, 2000). Obgleich also die Konstruktvalidität bei Auswertung von ACs mit der MTMM-Matrix also gering ausfällt, geht sie in der Regel mir einer hohen prognostischen Validität einher. Dieses *Validitätsparadox* ist schon seit längerem Gegenstand kontroverser Diskussionen in der AC-Forschung. Nach Arthur et al. (2000) kann dies an den stark messfehlerbehafteten Beobachterurteilen liegen, die ein Nachweis der Konstruktvalidität verhindern. Dies ist sicherlich teilweise auf die im Rahmen der *sozialen Urteilsbildung* bereits angesprochene, sich an globalen Kriterien orientierende Urteilsbildung der Beurteiler und der fehlenden Trennung von Beobachtung und Bewertung zurückzuführen. Schuler (2000) betont jedoch, dass die Schlussfolgerung auf eine mangelnde Konstruktvalidität daraus nicht zulässig ist, da gerade die multimodale Messung der Eignungsdiagnostik die Repräsentativität der AC-Tätigkeiten erhöhen soll und die Urteile immer eine Kombination zwischen Verhalten, Eigenschaft und Messmethode darstellen. Zudem wird durch eine *Aggregation* der Urteile der Beurteiler eine Erhöhung der Reliabilität erreicht (siehe 2.2.4), wobei eine Diskussion der Einzelurteile nicht die prognostische Validität der Verfahren erhöht (Schuler, 2000, S.127). Sie ist jedoch im Rechtfertigungskontext gegenüber dem Bewerber, sowie aus der Beurteilertrainingsperspektive sinnvoll. Zusätzlich sind anforderungs- und verfahrensbedingte Mängel, sowie Mängel bei der Erfassung relevanter Verhaltensweisen als Grund für mangelnde Konstruktvalidität ins Feld geführt worden (Höft & Funke, 2001; Kleinmann, 1997). Scholz und Schuler (1993) kommen zu der Erkenntnis, dass das Hauptkonstrukt, dass mit Hilfe von ACs gemessen wird, die allgemeine Intelligenz (*g*) ist. Untersuchungen zur Konstruktvalidität von *ACs* mit Außenkriterien zeigen außerdem Korrelationen zu Persönlichkeitsmerkmalen (z.B. Extraversion, soziale Kompetenz und Leistungsmotivation) sowie mittlere Korrelationen zu Intelligenztestwerten (Diemand & Schuler, 1991; Scholz & Schuler, 1993; Zimmermann & Schuler, 1991). Zudem scheint sich die Beteiligung von Psychologen positiv auf die Konstruktvalidität auszuwirken (Sagie & Magnezy, 1997).

Die *prognostische Validität* scheint bei *ACs* einer großen Schwankungsbreite zu unterliegen. Thornton, Gaugler, Rosenthal und Bentson (1987) berichten metaanalytisch von Werten zwischen r=.11 und r=.63 (95% Konfidenzintervall) mit einer durchschnittlichen Validität von r=.37. Als die Validität erhöhende Moderatorvariablen identifizierten sie die Beteiligung von Psychologen, eine hohe Anzahl unterschiedlicher Aufgaben sowie gegenseitige Beurteilungen der Bewerber. Schuler (2000) rät zudem zur Verwendung psychologischer Tests, strukturierter Interviews, einem hohen Standardisierungsgrad sowie einer konstrukt- und zeitbezogenen Optimierung. Neben der guten Prognosekraft (unter der Voraus-

setzung der Einhaltung angemessener Standards) und einer guten Akzeptanz (Kanning, 2002), liegen die Vorteile von *ACs* bei der Personalauswahl auch in der Möglichkeit, Bewerber mit Führungstätigkeiten vertraut zu machen, ihre Selbsteinschätzung zu verbessern und einen sozialen Vergleich zu ermöglichen (Schuler, 2000). Zudem sind die hohe Realitätsnähe, die Berücksichtigung dynamischer Arbeitsplatzaspekte und die Möglichkeit, eine Vielzahl von Fähigkeiten, Neigungen und sonstigen Eigenschaften durch mehrere Beurteiler angemessen feststellen zu können, vorteilhaft (Hossiep, 1995). Es gibt allerdings auch einige Kritikpunkte. Dazu gehören der hohe Zeit- und der besonders entwicklungstechnische Kostenaufwand (Hossiep, 1995) sowie die bereits angesprochene mangelnde *Konstruktvalidität*. Zudem herrscht wenig Klarheit darüber, welche Aufgaben maximales und welche Aufgaben typisches Verhalten hervorrufen (Schuler, 2000). Darüber hinaus ist auch auf *Reaktivitätseffekte*, d.h. das Zeigen erwünschten Verhaltens durch den Bewerber bei Kenntnis der aufgabenspezifischen Erwartung, hingewiesen worden (Bungard, 1987). Schuler (2000, S.132) vertritt jedoch die Ansicht, dass diese Fähigkeit ebenfalls einen prognostischen Wert beinhalten dürfte und somit als „veritabler Teil des Wirkungszusammenhangs“ aufgefasst werden sollte.

Das in dieser Arbeit untersuchte Auswahlverfahren kann ebenfalls als AC eingestuft werden. Neben den genannten Interviews beinhaltet dieses allerdings lediglich eine Gruppenarbeit (nur für Hochschulabsolventen) und eine Fallpräsentation als weitere Übungen. Dabei nehmen sowohl zahlreiche Bewerber als auch unterschiedliche Beurteiler teil. Die Beurteiler stufen, wie bereits angesprochen, dabei die Bewerber auf Dimensionen mit Hilfe einer Skala ein. Gemäß der Untersuchungen zur Konstruktvalidität dürfte deshalb auch im vorliegenden Fall geringe diskriminante und hohe konvergente Validitäten, zum Ausdruck kommen. Geht man davon aus, dass das Verfahren eine durchschnittliche prognostische Güte aufweist, so dürfte diese um den Wert r=.37 liegen.

2.5.4 Zusammenfassung

Nunmehr 100 Jahre anhaltender Forschung haben zu der Erkenntnis geführt, dass die geeigneten Modelle der Intelligenz hierarchisch, mit einem Generalfaktor (*g*) auf oberster Ebene, aufgebaut sind. Im Bereich der Persönlichkeitsforschung hat sich das *Fünf-Faktoren-Modell* der Persönlichkeit als Leitmodell durchgesetzt, allerdings stehen hier, im Gegensatz zur Intelligenzforschung, plausible und empirisch fundierte Modelle zum Zusammenhang der Faktoren mit Berufserfolg bisher aus. Bei der Interessenforschung ist die auf Holland (1996) zurückgehende *RIASEC-Taxonomie* im Personalbereich am weitesten verbreitet. Als

die im Zusammenhang mit Berufserfolg bedeutendsten Motive gelten das *Leistungs*- und das *Machtmotiv*. Schließlich wird beim Zusammenhang von Wissen und Berufserfolg häufig auf die Unterscheidung von *prozeduralem* und *deklarativem* Wissen Wert gelegt, sowie *Berufswissen* konzeptionell unterschieden. Das Konzept der *praktischen Intelligenz* hingegen konnte sich bisher als gesichertes Konstrukt nicht durchsetzen.

Das Konstrukt g ist auf Grund einer umfassenden intensiven Forschung das im Vergleich zu anderen Prädiktoren am besten untersucht und theoretisch fundiert (Schmidt & Hunter, 1998, S.246). Die Bedeutung dürfte sogar mit wachsender Komplexität der Umwelt (und damit des Lebens allgemein) weiter zunehmen (Lubinski, 2004). Unterschiede auf den spezifischen kognitiven Fähigkeitsprofilen können unterschiedliche Präferenzen der Informationsverarbeitung (Datenart, Prozesse) bedingen und somit ebenso wertvolle Informationen im Auswahlkontext liefern. Varianzaufklärungen gemeinsam mit *g* von bis zu R^2=.50 an der Berufsleistung sind dabei nicht ungewöhnlich (Lubinski, 2004, S.106).

Bei den Persönlichkeitskonstrukten sind hingegen bisher eher moderate prognostische Validitäten vorgefunden worden. Je nach betrachtetem Kriterium scheinen, unter Einbehaltung symmetrischer Konstellationen, sowohl spezifische als auch generelle Konstrukte prognostisch wertvoll zu sein. Die höchste Validität scheint dabei *Integrität*, die im Wesentlichen *Gewissenhaftigkeit* beinhaltet, zu haben. Auch von Zusammenhängen zwischen Motivation und Berufs- bzw. Führungserfolg ist auszugehen. Die Ergebnisse deuten bei Managern auf die Vorteilhaftigkeit von hoher *Machtmotivation*, *Leistungsmotivation* und *Selbstkontrolle* sowie einer geringen *Anschlussmotivausprägung* hin. Genaue Befunde über die Höhe dieser Zusammenhänge müssen jedoch in weiteren Validitätsstudien erforscht werden.

In der Praxis hat sich die Analyse der Bewerbungsunterlagen als Vorauswahlinstrument etabliert, allerdings bedarf die prognostische Tauglichkeit der dabei gewonnen Informationen weiterer Forschung. Sehr häufig findet ebenso das *unstrukturierte Interview* Anwendung, obwohl die in ihrer Entwicklung aufwändigere strukturierte Variante sich psychometrisch als weitaus bessere Alternative erwiesen hat. Ein *Assessment Center*, wie es auch in der vorliegenden Studie verwendet wird, ist ein mehrere Instrumente integrierendes Vorgehen, das eine gute prognostische Validität und geringe Konstruktvalidität besitzt. Bezüglich eignungsdiagnostischer Instrumente lässt sich allerdings festhalten, dass *Intelligenztests* aus Validitätsaspekten die Methode der Wahl bei Personalauswahl für Arbeitsplätze, die Informationsverarbeitung erfordern, sein sollten. Andere Verfahren sollten als Ergänzung zu *Intelligenztests* verstanden werden, die, teilweise bedeutende, inkrementelle Varianzaufklärung am Kriterium ermöglichen (Schmidt & Hunter, 1998).

3 Empirischer Teil

„Es ist nicht genug zu wissen, man muss auch anwenden;
es ist nicht genug zu wollen, man muss auch tun.“

Johann Wolfgang von Goethe (1749-1832)

Der Einsatz eignungsdiagnostischer Instrumente ist in der Praxis vornehmlich durch augenscheinvalide, aber oftmals ungeeignete Vorgehensweisen gekennzeichnet, obwohl, wie im letzten großen Abschnitt berichtet, in zahlreichen empirischen Studien wertvolle Erkenntnisse über Einsatzmöglichkeiten, Validität und Akzeptanz der Verfahren gewonnen wurden. Zudem geben die im Kapitel 2.3 erläuterten Nutzenmodelle den Personalverantwortlichen das Instrumentarium zur Hand, um auch in der betriebswirtschaftlichen Diskussion um die Verteilung geplanter Investitionen in den Unternehmen bestehen zu können. Insofern ist auf der „Wissensseite“ aus wissenschaftlicher Perspektive bereits einiges geleistet worden. Die Praxisseite, also die Verantwortlichen der Personalabteilungen in den Unternehmen, sollte nun den zahlreichen Erkenntnissen durch ihre Anwendung auch breiten Zugang zu jener Welt ermöglichen, für die sie auch entwickelt wurden. Die Unternehmensberatung, die Untersuchungsgegenstand der vorliegenden Arbeit ist, hat dies erkannt und ihre Datenbasis für eine umfangreiche Analyse zur Verfügung gestellt. Bei dieser kommen einige der beschriebenen Modelle zur Anwendung und einzelne empirische Befunde werden erneut durchleuchtet. Nachfolgend sollen aber zunächst die sich aus dem Theorieteil ergebenden, zentralen Fragestellungen der Untersuchung erläutert werden, bevor das Vorgehen der Untersuchung dargestellt wird.

3.1 Zentrale Fragestellungen der Untersuchung

Betrachtet man das angewandte Auswahlverfahren isoliert, also lediglich die Prädiktorenseite aus *Abbildung 14*, so sind folgenden Fragestellungen von Interesse:

1. *Lassen sich die geschilderten empirischen Befunde zur Reliabilität auch bei den Instrumenten des vorliegenden AC replizieren?*

Unter 2.2.2.3 und 2.2.5.1 wurde erläutert, welche potenziellen Fehlerquellen mit einer Leistungsbeurteilung, die im *AC* von den Beobachtern und Interviewern durchgeführt wird, einhergehen. Es ist deshalb anzunehmen, dass diese sich negativ auf die Reliabilität der Urteile sowohl der Interviews als auch der Beobachtungen des untersuchten Auswahlverfahrens auswirken. Demnach dürfte die Konsistenz der Urteile auf Grund des semi-strukturierten Charakters der Interviews etwas höher als die von Schuler (2002) berichteten Objektivitäts-

werte offener Interviews (r=.30) liegen. Für die Beobachter sind demgegenüber etwas bessere Werte zu erwarten, da sie die Bewerber im Rahmen gleicher Situationen beurteilen und somit die durch die Situation bedingte Varianz ($\sigma^2{}_{ZwS}{}^{PR}$; $\sigma^2{}_{PxS}{}^{PR}$) verringert wird. Hier wird vermutet, dass sie etwa die gleiche Höhe wie bei Vorgesetztenurteilen (etwa r=.50) haben, da sie vergleichbaren Fehlerquellen unterliegen.

2. Gibt es auf den Dimensionen des AC charakteristische Profile?

Das vorliegende kompensatorische Verfahren (siehe auch 2.2.6) gibt Anlass zu der Vermutung, dass akzeptierte Bewerber stark divergierende Profile auf den Dimensionen des *AC* aufweisen. Dabei dürfte sich zeigen, dass niedrige Werte auf einer Dimension mit hohen Werten auf anderen Dimensionen ausgeglichen werden. Von Interesse ist zudem, wie die unterschiedlichen Ausprägungen im Zusammenhang mit der Biographie, also den Informationen aus dem Lebenslauf, in Verbindung stehen. Entsprechend der geschilderten empirischen Befunde dürften beispielsweise Frauen eher Stärken im *verbalen*, *kreativen* und *sozialen* Bereich und weniger im analytischen Bereich aufweisen (Lubinski, 2004; Achter, et al., 1999; Shea, et al., 2001; Webb, et al., 2002). Zudem soll festgestellt werden, wie sich der Ausbildungsstand auf die Leistungen im *AC* auswirkt. Ein höherer Bildungsstand müsste mit einem höheren Wissen einhergehen und somit zu besseren Werten auf den wissensrelevanten Dimensionen führen. Die Kenntnis des Ausbildungsstandes dürfte zudem die Bildung von Stereotypen bei den Beurteilern begünstigen (siehe auch 2.2.5.1) und ebenfalls zu besseren Einstufungen führen.

3. Treffen die Erkenntnisse zur Konstruktvalidität von ACs auch im vorliegenden Fall zu?

Unter 2.5.3.3 wurden empirische Befunde erläutert, die geringe *diskriminante* und hohe *konvergente* Validitäten zwischen den Dimensionen bei *ACs* feststellten. Dies dürfte in einem Hauptkonstrukt, das durch das AC gemessen wird, begründet liegen. Dabei kann es sich um ein Globalurteil der Eignung oder, wie von Scholz und Schuler (1993) postuliert, um die allgemeine Intelligenz (*g*) handeln. Entsprechend wird erwartet, dass, und soll geprüft werden, ob auch bei dem vorliegenden Verfahren ein solches Konstrukt gemessen wird.

4. Sind die intendierten Prädiktoren des ACs angemessen operationalisiert worden?

Unter 2.5 wurde angesprochen, welche Konstrukte in welcher Form gängigerweise operationalisiert werden. Darüber hinaus wurde aufgezeigt, welche Gütekriterien einzelne Ausgestaltungsformen der auch in dem vorliegenden Verfahren angewandten Instrumente

aufweisen. Es soll deshalb dargestellt werden, wie das vorliegende Vorgehen mit dem Ziel, seine Güte zu optimieren, modifiziert werden kann.

Analog zu der Prädiktorenseite sind auf der Kriterienseite folgende Fragestellungen relevant:

5. Lassen sich bei den Mitarbeitern auf den Leistungsdimensionen charakteristische Profile erkennen?

Die in der zweiten Fragestellung thematisierten Profilunterschiede der Bewerber auf der Prädiktorenseite müssten sich auch im Berufserfolg zeigen. Die geschilderten empirischen Befunde zum *Berufswissen* (2.5.2.5) geben zudem Grund zu der Annahme, dass Bewerber mit einem entsprechenden beratungsspezifischen Vorwissen oder beruflicher Erfahrung, insgesamt bessere Leistung bei den Projekten erbringen dürften. Entsprechend soll untersucht werden, ob dies auch im vorliegenden Fall zutrifft. Zusätzlich sollen diese Leistungsunterschiede inhaltlich interpretiert werden.

6. Welche Konstrukte liegen hinter den erfassten Dimensionen des Kriteriums?

Die Extraktion von einem *Generalfaktor*, der einen Großteil der Varianz aufklärt, ist bei Leistungsbeurteilungen durch Vorgesetzte, wie unter 2.2.2.3 geschildert, metaanalytisch bestätigt worden (Viswesvaran, et al., 2005). Er wird neben den klassischen Urteilsfehlern (insb. *Halo-Effekt*) darauf zurückgeführt, dass symmetrisch sehr generell angesiedelte Faktoren (wie *g* oder die Persönlichkeitseigenschaft *Gewissenhaftigkeit*) den Berufserfolg auf nahezu allen Dimensionen beeinflussen und somit zu hohen Interkorrelationen führen (Feldmann, 1981; Motowidlo, et al., 1997). Entsprechend soll untersucht werden, ob sich auch bei den als Kriterium verwendeten Vorgesetztenurteilen der hier durchgeführten Evaluationsstudie, ein derartiges generelles Konstrukt wieder findet.

7. Sind die Kriterien angemessen operationalisiert worden?

Unter 2.2.2 wurden mögliche Operationalisierungen der Kriterien bei Evaluationsstudien dargestellt. Im Rahmen der vorliegenden Arbeit sollen die aktuell verwendeten Kriterien diesbezüglich durchleuchtet und praktische Verbesserungsvorschläge zur Operationalisierung abgeleitet werden.

Bezüglich des Zusammenhangs von Prädiktoren und Kriterien, also der *Validität* des Auswahlverfahrens, stehen folgende Fragestellungen im Betrachtungsfokus:

8. Entspricht die prognostische Validität des angewandten ACs den empirischen Befunden?

Die metaanalytischen Befunde von Schmidt und Hunter (*Tabelle 14*) führen zu der Erkenntnis, dass die durchschnittliche prognostische Validität von *ACs* bei r=.37 liegt. Diese kann aber, wie unter 2.5.3.3 diskutiert, in Abhängigkeit von der Ausgestaltung, erheblich schwanken. Im vorliegenden Fall soll festgestellt werden, wie hoch die prognostische Validität des *AC* ist. Zudem soll darauf eingegangen werden, welche Ursachen für eine mögliche Validitätsminderung verantwortlich sein können. Zusätzlich sollen mögliche Zusammenhänge zwischen den einzelnen Dimensionen des *AC* und den Kriteriendimensionen des Vorgesetztenurteils inhaltlich interpretiert werden.

9. Wie hoch ist die prognostische Validität biographischer Lebenslaufinformationen?

Neben den *AC*-Dimensionen sind ebenfalls die biographischen Lebenslaufinformationen von Interesse. Unter 2.5.2.4 wurde dargestellt, dass diese bemerkenswerte prognostische Validitäten für den Berufserfolg aufweisen können. Es soll deshalb am vorliegenden Fall untersucht werden, ob dies auch für die Auswahl von Unternehmensberatern gilt. Zudem soll überprüft werden, inwieweit sich die besprochene geringe prognostische Validität von Schulleistungen für den Berufserfolg replizieren lässt.

10. Wie können die Prädiktoren optimal ausgewählt und kombiniert werden?

Unter 2.2.5 wurde ausführlich erläutert, dass eine geeignete Auswahl, Kombination und unterschiedliche Gewichtung der Prädiktorendaten zu einem prognostischen Validitätsgewinn führen kann. Es wird deshalb ebenfalls untersucht, inwiefern dies im vorliegenden Fall zutrifft. Zudem sollen die Variablen, die inkrementelle Varianzen am Beratungserfolg aufklären, inhaltlich interpretiert werden.

11. Sind die Symmetrieprinzipien zwischen Prädiktor und Kriterium eingehalten worden?

Unter 2.2.4 wurde thematisiert, dass zur Durchführung fairer Validitätstests die Einhaltung von *Symmetrieprinzipien* auf Prädiktor- und Kriterienseite erforderlich ist. Wittmann (1985; 1990) zeigte, dass Kriterium und Prädiktor hinsichtlich ihrer Breite und inhaltlichem Bezug übereinstimmen sollten, da es ansonsten zu erheblichen Validitätseinbußen kommen kann. Im Rahmen der Studie soll untersucht werden, ob dies im vorliegenden Fall zutrifft und die Berücksichtigung solcher Symmetrieprinzipien bei der Validitätsstudie zu neuen Erkenntnissen führen kann. Dabei soll dargestellt werden, welche Indizien es für eine Neuordnung der Symmetrieebenen der durchgeführten Validitätsstudie gibt und welcher möglicherweise daraus entstehende Validitätsgewinn damit einhergeht.

Schließlich beschäftigt sich die Untersuchung mit den Folgen der diagnostischen Entscheidung und den Implikationen für die Praxis:

12. Welche Folgen hat die Urteilsbildung auf die Verteilung diagnostischer Entscheidungen?

Analog 2.2.6 sollen auch die Folgen der angewandten diagnostischen Entscheidung untersucht werden. Dabei sollen die Verteilung der Bewerber gemäß *Abbildung 21* ermittelt und die praktischen Konsequenzen exemplarisch aufgezeigt werden.

13. Welchen monetären Nutzen erbringt das angewandte Auswahlverfahren?

Im Abschnitt 2.3 wurden die Möglichkeiten der Nutzenberechnung bei Unternehmensberatungen erläutert und eine Vorgehensweise zur Berechnung erarbeitet. Diese soll auch auf den vorliegenden Fall angewandt und somit der monetäre Nutzen quantifiziert werden. Darüber hinaus soll berechnet werden, welche monetären Auswirkungen eine Optimierung der angewandten Vorgehensweise mit sich bringt.

14. Welche praktischen Verbesserungsvorschläge ergeben sich aus der Untersuchung?

Die abschließende, für den Auftraggeber dieser Arbeit entscheidende Fragestellung betrifft die Implikationen der Ergebnisse für die Praxis. Dabei werden Gestaltungsvorschläge für eine Verbesserung des angewandten Auswahlverfahrens diskutiert und Handlungsempfehlungen abgeleitet.

3.2 Untersuchungsvorgehen

Nachfolgend soll ein Überblick über den Inhalt und Ablauf der Untersuchung gegeben werden. Dieser beinhaltet neben der Beschreibung des Untersuchungsgegenstandes und des Ablaufs des Auswahlverfahrens, die Beschreibung der Stichprobe. Anschließend wird auf die Operationalisierung von Prädiktoren und Kriterien eingegangen, bevor die angewandten Methoden detailliert werden.

3.2.1 Untersuchungsgegenstand

Gegenstand der vorliegenden Untersuchung ist eine internationale Strategieberatung (im Folgenden ISB genannt) mit weltweit 33 Büros und 1.700 Angestellten (Stand 2002), die in Form einer Partnerschaft organisiert ist. Zu den inhaltlichen Beratungsschwerpunkten gehören Wachstums- und Internationalisierungsstrategien, Portfolio- und Wertemanagement, Post-Merger-Integration, Restrukturierung und Marketing. Damit wurde 2002 ein Umsatz von über 500 Millionen Euro generiert. Das Unternehmen orientiert sich bei der Zusammen-

setzung der Beratungsteams an der Form einer Matrixorganisation, an deren Achsen einzelne funktionale und branchenspezifische Organisationseinheiten (*Competence Center*) angesiedelt sind (*Abbildung 32*).

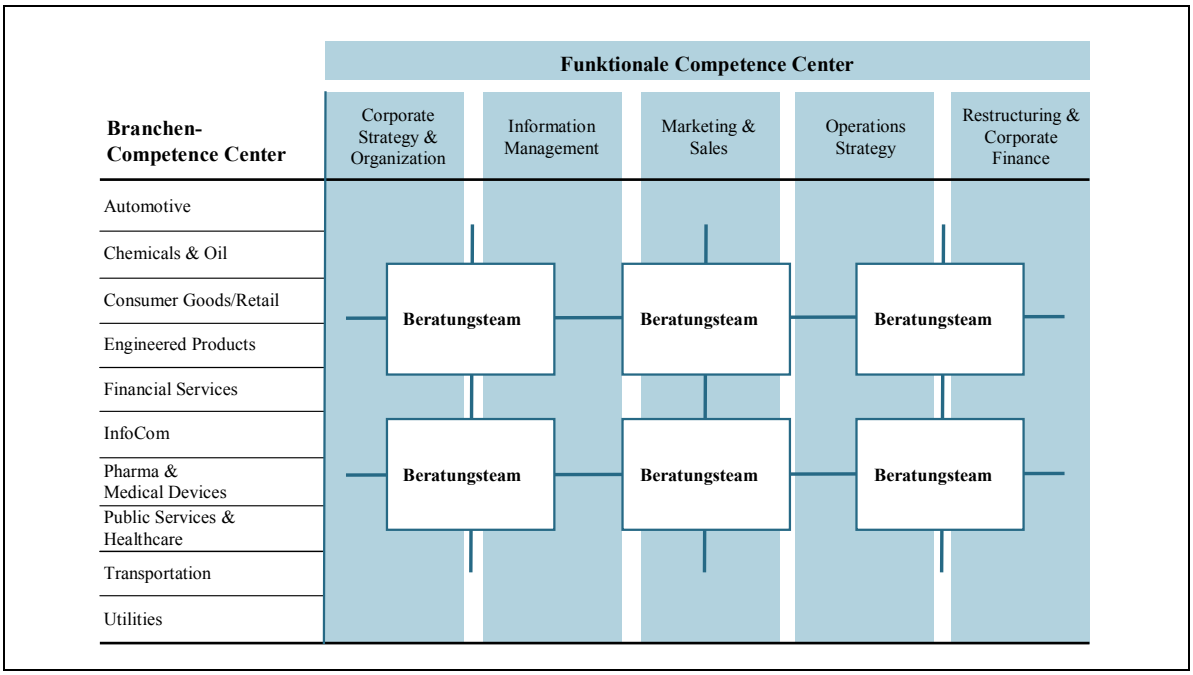

Abbildung 32: Zusammensetzung der Beratungsteams nach *Competence Center*

Innerhalb der operativen Beratungsteams gibt es dabei unterschiedliche Funktionsstufen. Berufseinsteiger mit einem Universitätsabschluss fangen i.d.R. als *Junior-Consultant* (*JC*) an und werden dann, gängigerweise im Jahresrhythmus, sukzessive auf folgende Funktionsstufen befördert: *Consultant* (*C*), *Senior Consultant 1* (*SC1*), *Senior Consultant 2* (*SC2*), *Projektmanager 1* (*PM1*) und *Projektmanager 2* (*PM2*). Quereinstiege auf einer höheren Funktionsstufe als *JC* sind für promovierte und berufserfahrene Bewerber ebenso möglich. Über die operativen Beratungsteams hinaus gibt es die höchste Funktionsstufe der Partner (*Associate Partner* (*AP*) und *Partner* (*P*)), die an dem Unternehmen beteiligt sind und mehrere Projekte gleichzeitig verantworten.

3.2.2 Beschreibung des Auswahlverfahrens

Das Auswahlverfahren der ISB beinhaltet insgesamt fünf Schritte (*Abbildung 33*). Dabei findet zunächst eine Sichtung der Bewerbungsunterlagen durch die Personalabteilung statt. In Abhängigkeit des Gesamtbildes, das die Bewerberunterlagen hinterlassen (im Sinne eines *klinischen Urteils*), wird der Bewerber zu dem eigentlichen Auswahlverfahren eingeladen, das sich aus *Interviews* und der Bearbeitung einer *Fallstudie* zusammensetzt. Der Bewerber führt i.d.R. zwei 60-minütige *Fachinterviews* mit jeweils einem Berater. Die Inhalte

Abbildung 33: Ablauf des Auswahlverfahrens der ISB

der Gespräche sind nicht vorgegeben, sondern bleiben dem Interviewer überlassen. Üblicherweise beziehen sie sich auf betriebswirtschaftliches Wissen, Problemstellungen aus der Praxis, Rollenspiele sowie Marktabschätzungsaufgaben. Darüber hinaus wird ein ebenso langes *Persönlichkeitsinterview* geführt. Dieses hat vorzugsweise Gespräche über Biographie, Vorlieben, Interessen, Motivation und Selbsteinschätzungen des Bewerbers zum Inhalt. Zudem werden oftmals kleinere allgemeine Denkaufgaben („*brain-teaser*“) gestellt, um die analytischen Fähigkeiten des Bewerbers einschätzen zu können. Schließlich werden Teile des Gesprächs als Stressinterview geführt, d.h. der Bewerber wird z.B. mit aggressiven Fragen gezielt unter Druck gesetzt, um seine Stressresistenz zu prüfen. *Persönlichkeits*- und *Fachinterviews* lassen sich nicht exakt abgrenzen. So kommen in beiden Interviewformen auch Inhalte des jeweils anderen Interviews vor. Die Bezeichnungen *Fach*- und *Persönlichkeitsinterview* beziehen sich also eher auf die Schwerpunktsetzung der behandelten Themen und weniger auf die Reinform des Interviews. Im Folgenden wird der Ausdruck *Interview* benutzt, wenn Bezug auf beide Formen genommen wird.

Nach jedem *Interview* füllt der Berater einen Beurteilungsbogen aus, mit dem er den Bewerber auf den in *Abbildung 33* dargestellten Dimensionen auf einer Skala von eins bis sechs einstuft. Zudem gibt er ein Urteil darüber ab, ob der Bewerber eingestellt werden soll oder nicht. Die einzelnen zu bewertenden Dimensionen sind dabei auf den Beurteilungsbögen stichwortartig detailliert. Beispielhafte Bögen (allerdings mit einer jüngst eingeführten Fünf-Punkte-Skala) sind in *Anhang V* dargestellt. Da das Interview nicht einem strukturierten Ablauf folgt, durch die Skalierung die Bewertung aber einen standardisierten Charakter hat, handelt es sich dabei nicht um eine reine unstrukturierte Form. Vielmehr ist es als semistrukturiert einzustufen (siehe auch McDaniel, et al., 1994). Zwischen den Interviews haben die Berater etwa fünf Minuten Zeit, um sich über den Bewerber auszutauschen und für bestimmte Dinge zu sensibilisieren. Nach den drei Interviews folgt eine einstündige Mittagspause, in der die Berater ausführlicher über die Kandidaten und die Erfahrungen aus dem Vormittag diskutieren. Der Nachmittag besteht aus der Bearbeitung einer Fallstudie, die sich in eine Stunde *Vorbereitungszeit* (Analysieren des Problems und Vorbereitung der Präsentation) und eine Stunde *Präsentation* mit Diskussion der Ergebnisse aufteilt. Berufseinsteiger, die sich für eine Stelle als *JC* bewerben, bereiten den Fall in einer Gruppe von durchschnittlich fünf Bewerbern gemeinsam vor und werden dabei i.d.R. von drei Beratern beobachtet. Die Präsentation der Ergebnisse sowie die anschließende Diskussion finden auch gemeinsam in der Gruppe statt. Bewerber für höhere Funktionsstufen bereiten sich hingegen eine Stunde lang unbeobachtet und alleine vor. Bei der anschließenden Präsentation und Diskussion müs-

sen sie sich ebenfalls alleine den Fragen von drei Beratern (die in der Präsentation die Klientenrolle übernehmen) stellen. Im Anschluss an die jeweiligen Beobachtungen (Vorbereitung bei *JC-Anwärtern* und Präsentation bei allen) werden die Kandidaten erneut auf den in *Abbildung 33* genannten Dimensionen eingestuft. Im Folgenden wird der Ausdruck *Beobachtung* gewählt, wenn Bezug auf die am Nachmittag durchgeführte Fallvorbereitung und -präsentation genommen wird. In einer abschließenden Sitzung werden, unter Moderation eines Mitarbeiters aus dem Personalbereich, alle Beurteilungen (*Interview* und *Beobachtung*) am Flipchart dargestellt und der Gesamtdurchschnittswert der Beurteilungen berechnet. Dieser bildet die Ausgangsbasis für die Gruppendiskussion in der ein abschließendes Konsensurteil gebildet wird. Üblicherweise wird ein Kandidat bei Ablehnung durch mehr als einen der zahlreichen Beobachter und Interviewer (ca. fünf bis sieben) oder einem Durchschnittswert kleiner 4.0 nicht eingestellt. Nach dem Fällen der Entscheidung überbringt jeweils ein Berater dem Interviewten diese und gibt bei Bedarf ein ausführliches Feedback.

Die Berater, die die Interviews und Beobachtungen durchführen, müssen eine Mindestbetriebszugehörigkeit von einem Jahr haben und werden in einem zweitägigen Seminar zum Thema *Interviewtechniken* geschult. Im Rahmen des Seminars wird ihnen der Recruitingprozess erklärt und werden die Ziele und Phasen von *Interviews* besprochen. Zudem werden die Teilnehmer für Fehler bei der Perzeption und Urteilsbildung sensibilisiert. Darüber hinaus werden unterschiedliche Frage- und Gesprächstechniken genauso erlernt wie die professionelle Handhabung von Absagen und Feedback.

3.2.3 Beschreibung der Stichprobe

Die Stichprobe besteht aus insgesamt 209 akzeptierten Bewerbern aus den Jahren 2000 bis 2003. Die Mehrheit der Bewerber wurde in den Jahren 2001 (54%) und 2002 (24%) eingestellt. Teilweise sind die Informationen zur Stichprobe lückenhaft. So war bei 39 Bewerbern aus dem Datensatz nicht feststellbar, auf welcher Funktionsstufe sie in das Unternehmen eingestiegen sind. Die Gründe liegen bei 29 von ihnen in einem forcierten Ausscheiden nach sehr kurzer Zeit, so dass bei der Personalabteilung die notwendigen Informationen nicht verfügbar waren. Von den restlichen zehn Mitarbeitern ist nur einer zum Ende des Betrachtungszeitraums in dem Unternehmen noch beschäftigt. Der Großteil der verbleibenden Mitarbeiter (N=170) ist als *JC* in das Unternehmen eingestiegen (58%). Rund 19% sind als *C*, die restlichen 23% auf höheren Funktionsstufen (*SC* oder *PM*) eingestiegen (*Tabelle 16*).

Tabelle 16: Einstiegsstufe nach Einstiegsjahr der Untersuchungsstichprobe (Anzahl)

Einstiegsstufe	2000	2001	2002	2003	Gesamt
Junior Consultant	20	38	37	3	98
Consultant	10	20	1	1	32
Senior Consultant 1	4	17	5	1	27
Senior Consultant 2	2	2	0	0	4
Projekt Manager	0	8	1	0	9
Gesamt	36	85	44	5	170

Die akzeptierten Bewerber sind zu einem Großteil männlichen Geschlechts (82%). Bei drei Mitarbeitern liegt keine Geschlechtsinformation vor. Das durchschnittliche Eintrittsalter beträgt 29 Jahre, wobei der jüngste Einsteiger 23 und der älteste 40 Jahre alt ist. Alle Bewerber der Stichprobe besitzen einen Hochschulabschluss, davon 49 von einer Privatuniversität. Die Mehrzahl hat ein betriebswirtschaftliches Studium absolviert (62%), von fünf Personen ist der Studienabschluss nicht bekannt. 67 Personen besitzen zwei, 11 Personen drei Hochschulabschlüsse. Von allen Personen haben 32 (15%) zum Zeitpunkt des Einstiegs einen Dr.-Titel und 13 (11%) einen MBA-Abschluss. Rund die Hälfte der Personen der Untersuchungsstichprobe (49%) ist im Betrachtungszeitraum ausgeschieden, 52 (28%) von ihnen forciert.

3.2.4 Operationalisierung der Prädiktoren

Die Prädiktorendatenbasis aus dem AC wurde aus den vorliegenden Bögen von der Personalabteilung elektronisch erfasst, anonymisiert und dem Autor zur Verfügung gestellt. Es liegen dabei allerdings nur für 205 Mitarbeiter Beurteilungen vor und dabei auch nicht auf allen Dimensionen. Dies liegt darin begründet, dass nur solche Dimensionen von den jeweiligen Beratern beurteilt wurden, zu deren Beurteilung sie im Rahmen des Verfahrens auch Gelegenheit hatten. Zudem haben sich in dem Betrachtungszeitraum die Beurteilungsdimensionen teilweise verändert. Der Autor hat in Absprache mit der Personalabteilung des Unternehmens die unterschiedlichen alten und neuen Dimensionen verglichen und bei semantischer Ähnlichkeit zusammengefasst. Jeder Bewerber der Stichprobe wurde durchschnittlich im Rahmen der Interviews 3.4 und bei den Beobachtungen 3.3 mal beurteilt. Insgesamt sind Beobachtungswerte für 199 und Interviewwerte für 201 Bewerber vorhanden. Die meisten Bewerber sind dabei von zwei Interviewern und drei Beobachtern beurteilt worden (*Abbildung 34*).

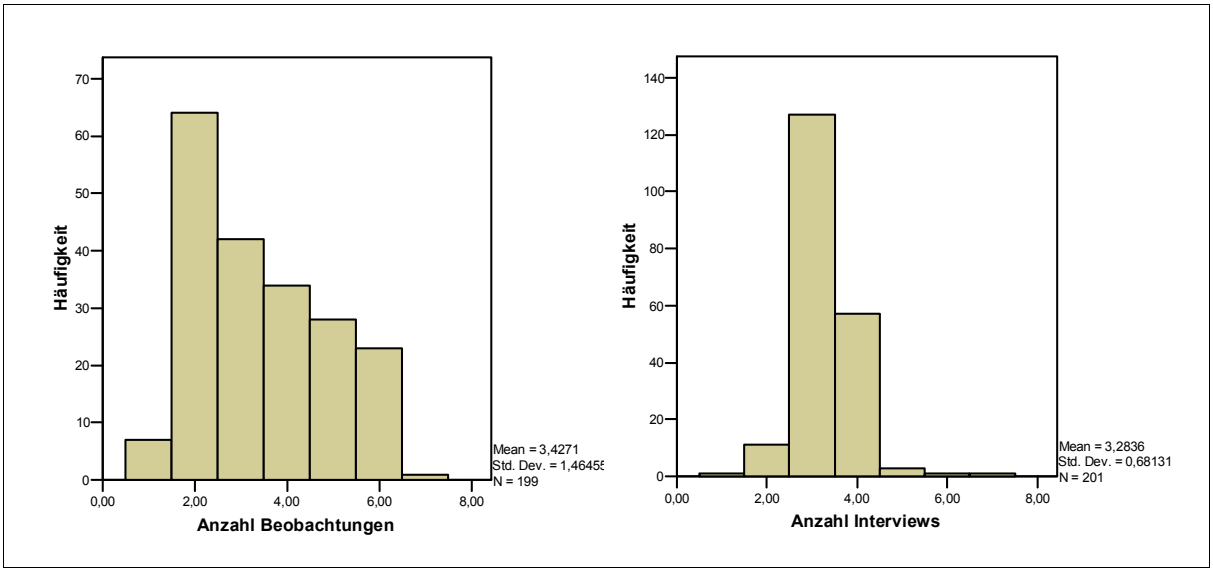

Abbildung 34: Häufigkeiten der Beobachtungen und Interviews je Bewerber

Die einzelnen Daten wurden vom Autor über alle Interviewer und Beobachter getrennt, sowie über beide gemeinsam aggregiert. Die Anzahl der aggregierten Urteile schwankt entsprechend dimensions- und personenspezifisch. Dieses Vorgehen ist aus Sicht des Autors sinnvoll, da eine interviewer- oder beobachterspezifische Betrachtung auf Grund fehlender Werte und der Tatsache, dass nicht jeder Interviewer bzw. Beobachter alle Bewerber bearbeitet hat, unmöglich ist. Zudem sind die Einzelurteile durch die Interaktion der Berater mit den Meinungen der anderen Berater kontaminiert. Außerdem bildet das aggregierte Urteil von Beobachtung und Interview auch die Basis für die Enddiskussion des Auswahlverfahrens, an dessen Ende die finale Einstellungsentscheidung getroffen wird. Es ist somit der geeignete Prädiktor für die spätere Validitätsanalyse.

Die Technik der *Aggregation* der Urteile führt zu Informationsverlusten, die aber aus Reliabilitäts- und Validitätsgründen von dem Autor in Kauf genommen werden (siehe Schmidt, et al., 2000). Dreher, Ash und Hancock (1988) haben genau dieses Vorgehen bei Validierungsstudien kritisiert und plädieren für eine separate Berücksichtigung der systematischen Urteilstendenzen der einzelnen Beurteiler (Disaggregationsmethode). Moser und Reuter (2001) konnten allerdings mit Hilfe von *Montecarlo-Analysen* zeigen, dass beide Vorgehensweisen zu nahezu gleichen Ergebnissen führen, weshalb auch nachfolgend an der *Aggregation* festgehalten wird. *Tabelle 17* zeigt die Inhalte und die Variablenbezeichnung der Prädiktoren aus dem *AC* (im Folgenden *Recruitingvariablen* genannt) sowie die Anzahl der Personen, für die Werte vorhanden sind. Zudem sind aus dem theoretischen Teil erwähnte Konstrukte dargestellt, die semantische Ähnlichkeit oder Überschneidungen mit den je-

weiligen Dimensionen haben. Eine detaillierte Beschreibung der einzelnen Dimensionen findet sich auf den Beurteilungsbögen in *Anhang V-1*. Zudem wird in *Anhang V-2* detailliert, welche Dimensionen über die Instrumente aggregiert wurden.

Tabelle 17: Beurteilte Dimensionen des AC (Recruitingvariablen)

Variable	Dimension	N	Verwandte Konstrukte
Team	Team-/Integrationsfähigkeit	203	Verträglichkeit
Konfl	Konflikt- und Kritikfähigkeit	87	Emotionale Stabilität
Komm	Kommunikationsstärke/Eloquenz	35	Extraversion
Charis	Ausstrahlung/Charisma	35	-
VWL	VWL-Kenntnisse	203	Akademisches Wissen, kristalline Intelligenz
Allg	Allgemeinwissen	86	Kristalline Intelligenz
Pragma	Pragmatismus/Ergebnisorientierung	202	Gewissenhaftigkeit
Unt	Unternehmerisches Denken	204	Berufswissen, Enterprising
LogAn	strukturiertes/logisch-analytisches Denken	200	Fluide Intelligenz
Meth	Methoden-Kenntnisse	204	Berufswissen, kristalline Intelligenz, prozedurales Wissen
Kreat	Kreativität/ Flexibilität	204	Offenheit für Erfahrung
Fach	Fachkenntnisse	204	Berufswissen, kristalline Intelligenz
BWL	BWL-Kenntnisse	203	Akademisches Wissen, kristalline Intelligenz, deklaratives Wissen
Intell	Intellektuelle Neugierde	203	Typisches Intellektuelles Engagement
Frust	Frustrationstoleranz	204	Emotionale Stabilität
Engl	Englischkenntnisse	43	Akademisches Wissen, kristalline Intelligenz, deklaratives Wissen
Fuehr	Durchsetzungskraft/Führungsfähigkeit	201	Extraversion
AllBil	Allgemeinbildung	32	Kristalline Intelligenz
SelOrg	Selbstorganisation, Engagement, Biss	202	Gewissenhaftigkeit
Leader	Leadership	196	Extraversion
Zielor	Zielorientierung	196	Gewissenhaftigkeit
Fit_ISB	Fit ISB-Team	147	Verträglichkeit, Person-Organisation-Fit
Exot	Exotenwissen	145	Berufswissen , kristalline Intelligenz, deklaratives Wissen
Decis	Entscheidung für Einstellung (j/n)	204	-
MeanRecr	Durchschnittswert aus Recruiting	204	-

Die Interviewer und Beobachter, die die vorliegende Bewerberstichprobe beurteilten, sind Berater unterschiedlicher Funktionsstufen des Unternehmens. Dabei haben zahlreiche von ihnen an mehreren Recruitings teilgenommen und hatten zu den Zeitpunkten teilweise unterschiedliche Funktionsstufen, da sie im Betrachtungszeitraum befördert wurden. *Abbildung 35* zeigt die Anzahl der Urteile aus Interview und Beobachtung in Abhängigkeit der Funktionsstufe des Beurteilers.

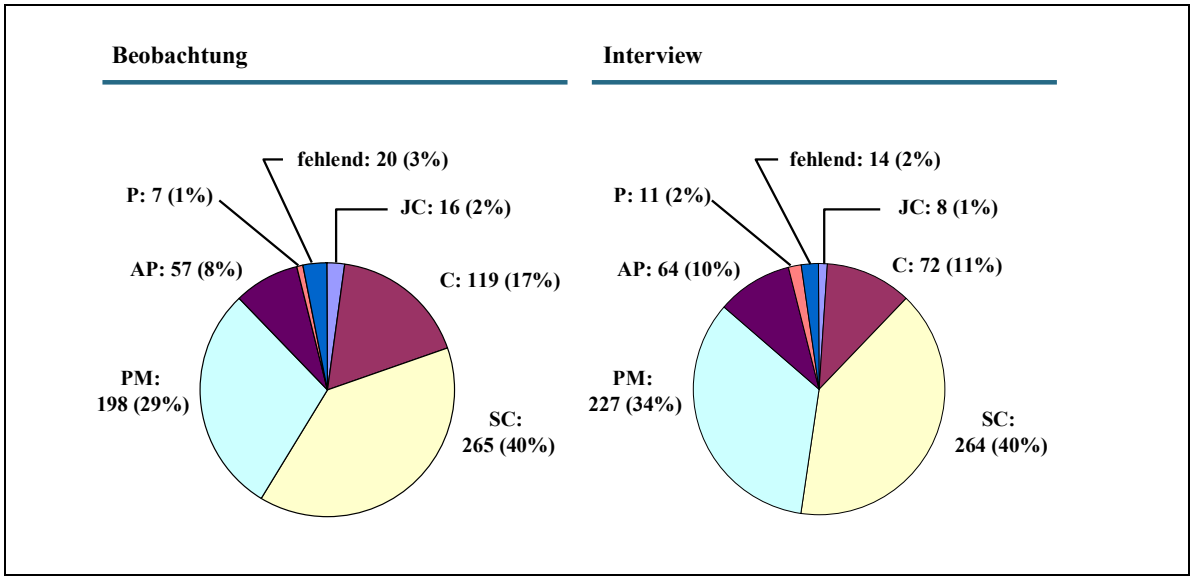

Abbildung 35: Häufigkeiten der Urteile aus Interviews und Beobachtung nach Funktionsstufe des Beurteilers

Neben den Beurteilungen aus dem Recruiting wurden als weitere Prädiktoren *biographische* Informationen aus den Bewerbungsunterlagen erfasst (im folgenden *CV-Variablen* genannt). Diese beinhalteten demographische Variablen, Ausbildungs-, sowie Praktika- und Berufsinformationen (Dauer, Anzahl, Beratungsrelevanz). Die dabei verwendete Kodierung wurde vom Autor wegen teilweise fehlender Informationen auf Basis plausibler Annahmen vorgenommen. So wurden z.B. die Zeiträume studienbegleitender Tätigkeiten und Werkstudententätigkeiten mit dem Faktor 0.5 angesetzt, da von der Annahme ausgegangen wurde, dass die Hälfte des angegebenen Zeitraumes tatsächlich gearbeitet wurde. Die Kodierung der einzelnen Daten kann beim Autor bei Bedarf angefordert werden, und wird, wenn zum besseren Verständnis nötig, bei den Analysen detaillierter erklärt. Vereinzelt waren Informationen aus den vorliegenden Lebensläufen nicht zu entnehmen oder einzelne Lebensläufe nicht mehr vorhanden, so dass es zu fehlenden Werten kommt. Dadurch hat wiederum nicht jeder Mitarbeiter Werte auf allen Variablen. Zudem war es nicht möglich, vergleichbare Werte für Studienabschlussnoten zu bekommen, so dass diese aus der Betrachtung ausgeschlossen werden mussten. *Tabelle 18* fasst die erfassten Variablen zusammen.

Tabelle 18: Erfasste CV-Variablen der Bewerberstichprobe

Variable	Inhalt (Kodierung)	N
Eintrittsalter	Alter des Bewerbers in Jahren bei Eintritt in das Unternehmen	206
Sex	Geschlecht des Bewerbers (0=weiblich; 1=männlich)	206
Status	Familienstand (0=ledig; 1=verheiratet)	205
Abinote	Abiturnote des Bewerbers (Skala 1-6)	195
FSP	Anzahl der analytischen Fächerschwerpunkte im Abitur[a]	167
Note_FSP	Durchschnittsnote der Fächerschwerpunkte im Abitur (Skala 1-6)	159
Note_D	Deutschnote im Abitur (Skala 1-6)	151
Note_M	Mathematiknote im Abitur (Skala 1-6)	104
Note_E	Englischnote im Abitur (Skala 1-6)	98
Note_Aus	Abschlussnote der Berufsausbildung (Skala 1-6)	44
BWLStud	Absolvent eines betriebswirtschaftlichen Studiums (0=nein; 1=ja)	204
Uni_Art	Universitätsart auf der studiert wurde (0=öffentlich; 1=privat)	204
Anz_Absch_tot	Anzahl der Hochschulabschlüsse gesamt	204
Dr	Inhaber eines Dr.-Titels (0=nein; 1=ja)	205
MBA	Inhaber eines *Master of Business Administration* (0=nein; 1=ja)	204
Pra_N	Anzahl absolvierter Praktika des Bewerbers	195
Pra_t	Zeit in Monaten, die der Bewerber insgesamt an Praktika absolviert hat	195
Pra_UB	Ehemaliger Praktikant einer Unternehmensberatung (0=nein; 1=ja)	195
Pr_N_UB	Anzahl absolvierter Unternehmensberatungspraktika	195
Pra_ISB	Ehemaliger Praktikant des Unternehmens (0=nein; 1=ja)	205
FG	Mitglied der Fördergruppe des Unternehmens für Studenten (0=nein; 1=ja)	205
Job	Berufserfahrung (0=nein; 1=ja)	204

[a] Folgende Fächer des Abiturleistungskurses wurden als analytisch betrachtet: Mathematik, Wirtschaftswissenschaften (BWL, VWL), Biologie, Technik, Physik und Chemie

3.2.5 Operationalisierung der Kriterien

Als Kriterium für die Evaluationsstudie wurden Projektbeurteilungen (auch Projektevaluationen genannt) der eingestellten Mitarbeiter aus den Jahren 2000 bis 2003 herangezogen. Sie sind auf Englisch verfasst (da sie international eingesetzt werden) und wurden anonymisiert von der Personalabteilung aus einer Datenbank entnommen.

Tabelle 19: Kriteriendimensionen der Projektbeurteilungen

Variablenname[a]	Dimension	Kurzbeschreibung der Inhalte
uSKH_MEAN	Specialist Know-how	Funktionales, branchenspezifisches, methodisches und EDV-spezifisches Wissen des Mitarbeiters
uPS_MEAN	Problem Solution	Problemlöseverhalten bezüglich Datensammlung, Analysen, Schlussfolgerungen und Maßnahmenableitung
uQW_MEAN	Quality of Work	Generelle Qualität der erbrachten Leistung (Strukturierung der Arbeit, Zuverlässigkeit, Zeitrahmen, Kosten und Güte der Dokumentation)
uBUP_MEAN	Behaviour under Pressure	Ausgeglichenheit, Rationalität, Motivation und Hilfsbereitschaft unter Stressbedingungen, sowie konstante Qualität der erbrachten Leistung
uCA_MEAN	Client Acceptance	Akzeptanz bei Mitarbeitern und Geschäftsführung des Klienten, Arbeitsverhältnis sowie Einfühlungsvermögen in die Probleme des Klienten
uCLIECO_ MEAN	Client Communication	Ausdrucksfähigkeit und Sprache gegenüber dem Klienten, Qualität geleiteter Arbeitskreise und durchgeführter Interviews
uKHM_MEAN	Know-how-Management	Nutzung intern vorliegender Informationsressourcen, strukturierte Dokumentation neuer Erkenntnisse und Informationsversorgung des restlichen Teams
uTW_MEAN	Teamwork	Kooperationsbereitschaft gegenüber Kollegen, Kritikfähigkeit und Akzeptanz aufgezeigter Verbesserungspotenziale, Beitrag zu Ideenfindung und Akzeptanz durch Kollegen auf persönlicher wie professioneller Ebene
uMeanEvaltot	Durchschnittswert über alle Variablen	Mittelwert der oben beurteilten Dimensionen. Grundlage für Personalentscheidungen (Beförderung, Bonizahlung)

[a] Das Kürzel *u* vor dem Variablennamen kennzeichnet, dass es sich um die ungewichtete Aggregationsvariante handelt (siehe Text) Die ab dem SC-Level zusätzlich eingeführte Dimension „Projektmanagement“ wurde wegen zu kleiner Fallzahlen (N=26) aus der Betrachtung ausgeschlossen

Die Beurteilungen haben den Charakter einer Vorgesetztenbeurteilung und beinhalten eine Einstufung auf einer Fünf-Punkte-Bewertungsskala für zahlreiche Variablen. Üblicherweise wird diese von dem Projektmanager nach Abschluss des Projektes oder bei langen Projekten auch während Projektes vorgenommen. Es gibt dabei zwei sich auf Untervariablenebene unterscheidende Bögen, wovon einer ab der und der andere bis zu der *SC-Funktionsstufe* gilt. Die Dimensionen beider Bögen und die Untervariablen sind in *Anhang VI* beschrieben. Im Laufe des Betrachtungszeitraums haben sich zahlreiche Beurteilungsdimensionen geändert. Der Autor hat deshalb nach semantischer Ähnlichkeit in Absprache mit der Personalabteilung verschiedene Variablen zusammengefasst bzw. nach dem aktuellen Beurteilungsbogen neu gruppiert. Es liegen dabei Beurteilungen für die Funktionsstufen *JC*, *C* und *SC1* (in Folgenden *SC* genannt) vor. *Tabelle 19* zeigt die Oberdimensionen, ihre Variablenbezeichnung und eine Kurzbeschreibung ihrer Inhalte (Details siehe *Anhang VI*). Für

insgesamt 160 Mitarbeiter der Untersuchungsstichprobe liegen Beurteilungen vor. Dabei haben einige dieser Mitarbeiter Beurteilungen auf nur einer, andere wiederum, die in dem Zeitraum befördert wurden, Beurteilungen auf mehr als einer Funktionsstufe. Die einzelnen Kohorten (*JC*, *C* und *SC*) sind demnach nicht als unabhängig voneinander zu betrachten, da die Stichproben teilweise die gleichen Personen beinhalten. Bei dem Vergleich der Ergebnisse der Funktionsstufen untereinander muss dieser Aspekt berücksichtigt werden. *Abbildung 36* zeigt die Zusammensetzung der Beurteilungen nach Funktionsstufen.

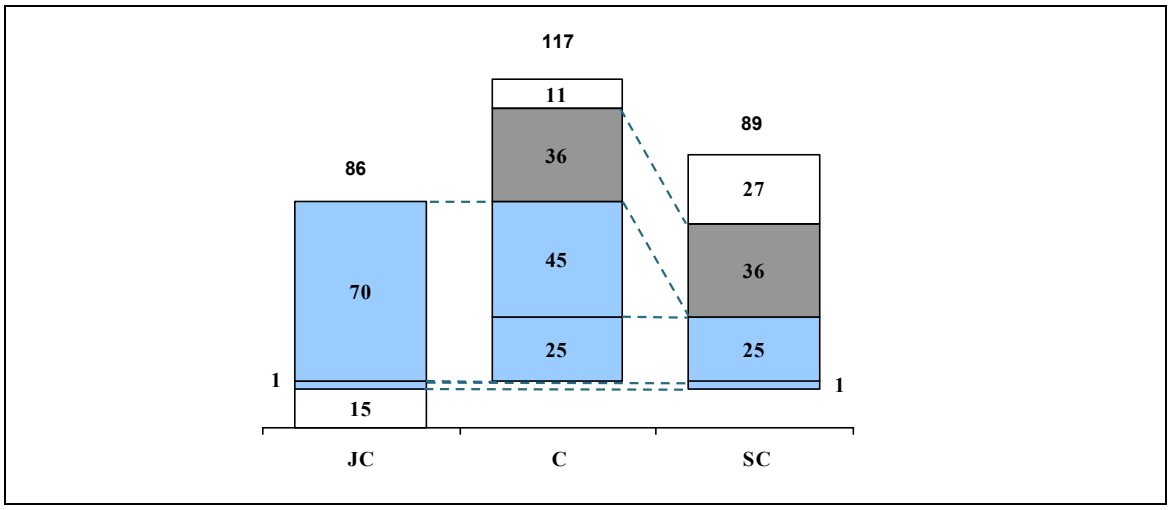

Abbildung 36: Zusammensetzung Projektbeurteilungen nach Funktionsstufe

Dabei zeigt sich, dass bei 86 Personen Projektbeurteilungen auf der *JC-Stufe* vorhanden sind. Von diesen 86 haben 70 ebenfalls Projektbeurteilungen auf der *C-Stufe* und 25 auf der SC-Stufe vorliegen. Weitere 36 Personen wurden nur auf der *C*- und der *SC -Stufe* beurteilt. Von 53 Personen gibt es lediglich auf einer Funktionsstufe Beurteilungen (15 *JC*, 11 *C* und 27 *SC*).

Zur Berechnung der Ausprägungen auf den Dimensionen wurde wie folgt vorgegangen: Der Durchschnitt der Untervariablen (aus *Anhang VI*) wurde nach Oberdimensionen gebildet (Bezeichnung _MEAN) und diese wieder zusammen zu einem Gesamtwert aggregiert. Dabei wurden sowohl die Unterdimensionen je Oberdimension als auch die Oberdimensionen bei der Aggregation gleichgewichtet. Die gebildeten Oberdimensionen wurden über die Anzahl der Beurteilungen innerhalb der Funktionsstufen (*JC*, *C* und *SC*), auf die sie sich beziehen, für jeden Mitarbeiter aggregiert. Dabei wurden die Beurteilungen, unabhängig von der Länge des Beurteilungszeitraumes, ebenfalls gleich gewichtet. Da dieses Vorgehen auch bei der Jahresendbeurteilung der Mitarbeiter angewandt wird, scheint es nach Meinung des Autors auch geeignet für die Verwendung als Kriterium. Zudem ergab eine Gewichtung

der Projektbeurteilungen nach Projektlaufzeiten bei der Aggregation nur marginale Unterschiede zu der ungewichteten Variante. Aus diesem Grund hat sich der Autor für das auch in der Realität angewandte Kriterium (also die ungewichtete Variante) entschieden. *Abbildung 37* visualisiert funktionsstufenspezifisch, bei wie vielen Personen wie viele Beurteilungen aggregiert wurden. Dabei zeigt sich, dass in jeder Kohorte durchschnittlich annähernd drei Beurteilungen zu einem Gesamtwert aggregiert wurden.

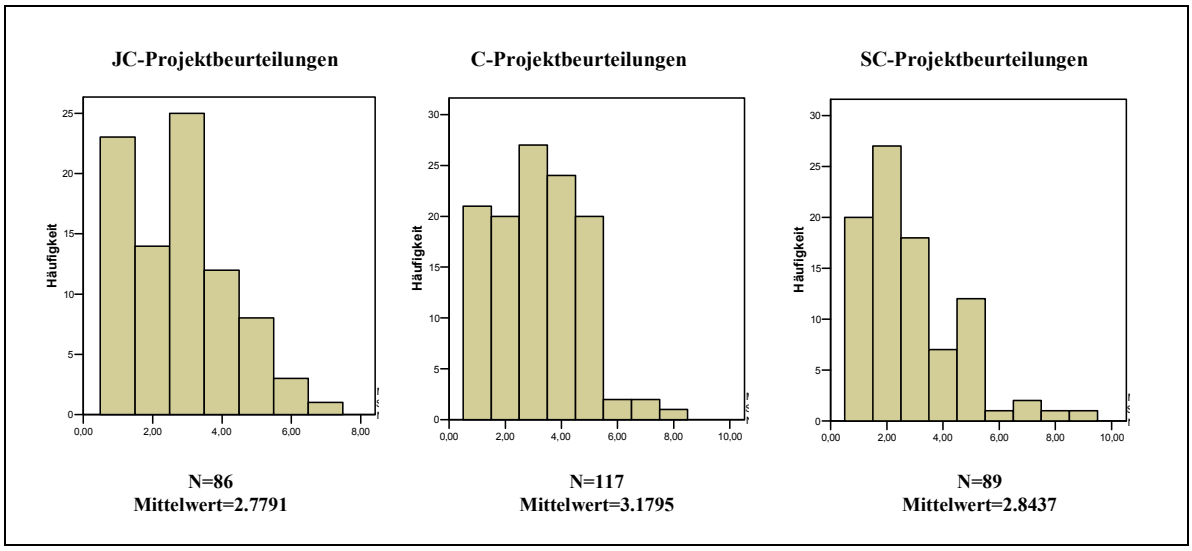

Abbildung 37: Personenanzahl nach Menge aggregierter Urteile je Funktionsstufe

Die bisher dargestellten Kriterien erfordern im Rahmen der späteren Validitätsanalyse eine nach Kohorten (*JC*, *C*, *SC*) getrennte Betrachtung. Die Unmöglichkeit, aus den einzelnen Projektbeurteilungen ein kohortenübergreifendes Gesamtkriterium zu bilden, führt dazu, dass, je nach untersuchtem Validitätszusammenhang, immer nur ein Teil der gesamten Stichprobe herangezogen werden kann. Neben den Dimensionen aus Projektbeurteilungen wurden deshalb zwei weitere Kriterien erfasst. Die Variable *Forced Exit* (0=nein; 1=ja) wurde als dichotomes Surrogat für Berufserfolg verwandt. Sie gibt an, ob der Mitarbeiter im betrachteten Zeitraum forciert ausgeschieden ist. Als weitere dichotome Variable wurde in die Betrachtung das *Angestelltenverhältnis* (0=angestellt; 1=ausgeschieden) einbezogen. Ausgeschiedene Mitarbeiter sind dabei nicht zwangsläufig erfolglos, ein freiwilliges Ausscheiden aus der Firma ist allerdings aus Sicht des Arbeitgebers i.d.R. nicht wünschenswert. Die Verwendung der Variable als Kriterium ist somit gerade aus praktischer Perspektive von besonderem Interesse.

3.2.6 Methode

Abbildung 38 visualisiert nochmals die Datenstruktur, die die Ausgangsbasis für die nachfolgenden Analysen ist.

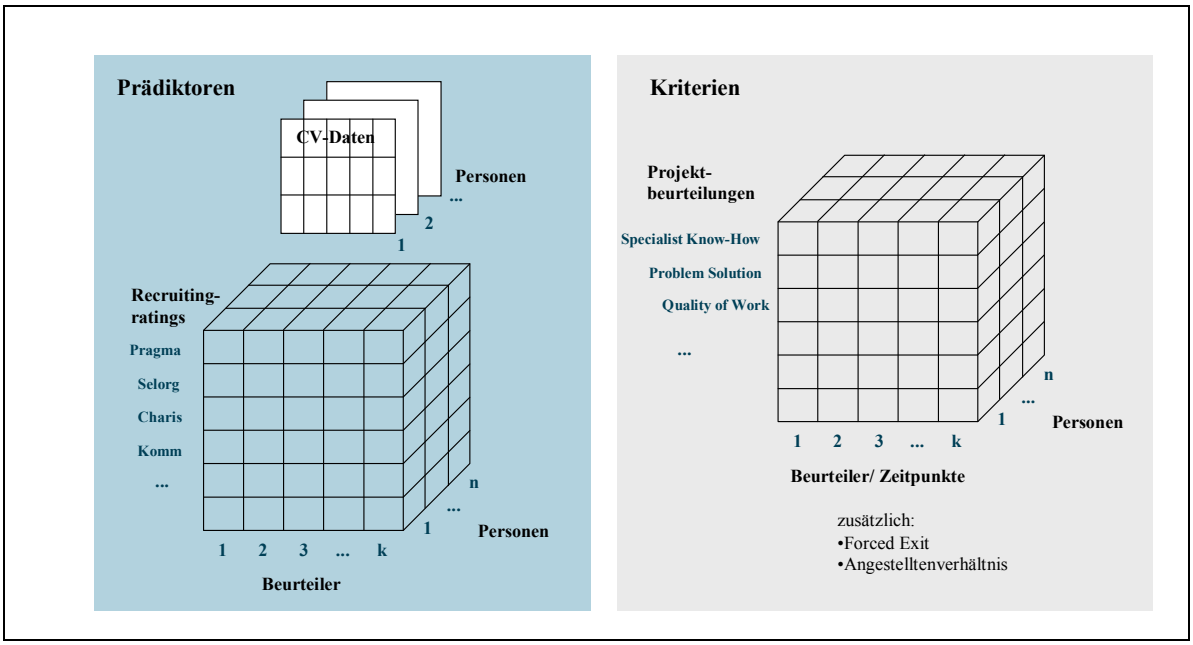

Abbildung 38: Datenstruktur der Studie

Auf Grund dieser Struktur wurde die Evaluationsstudie analog *Abbildung 1* durchgeführt. Demnach wurden zunächst für die gesamte Stichprobe die Prädiktorendaten analysiert (kohortenübergreifende Prädiktorenanalyse). Im Anschluss standen die Zusammenhänge zwischen Prädiktoren und Kriterien im Mittelpunkt. Da die Kriteriendaten funktionsspezifisch vorliegen und je nach Funktionslevel unterschiedlich hohe Anforderungen gelten, war dabei die Bildung eines Gesamtkriteriumswertes nicht möglich. Aus diesem Grund wurde die weitere Untersuchung in unterschiedliche Abschnitte, die sich mit Leistungen als *JC*, *C* oder *SC* beschäftigen, eingeteilt. Zusätzlich wurde in einem separaten Abschnitt auf erweiterte Fragestellungen (Operationalisierung von Kriterium und Prädiktor, Urteilsbildung und Verteilung diagnostischer Entscheidung) kritisch durchleuchtet. Schließlich standen im letzten Abschnitt die Nutzenaspekte im Vordergrund. Es ergeben sich somit insgesamt sechs große Analyseabschnitte: *Kohortenübergreifende Prädiktorenanalyse*, *Analyse der Kohorte JC*, *Analyse der Kohorte C*, *Analyse der Kohorte SC, erweiterte Fragestellungen* und *Nutzenanalyse*.

Am Anfang der ersten vier Abschnitte wurde dabei zunächst eine deskriptive Analyse der Prädiktoren- und bei der Kohortenbetrachtung auch der Kriteriendaten vorgenommen. Dabei wurden für jede Variable deskriptive Kennwerte errechnet (Mittelwert, Median, Mini-

mum, Maximum, Standardabweichung). Zudem wurde eine visuelle Inspektion des Datenmaterials mit Hilfe von Boxplots, Stem-and-Leaf-Plots und Histogrammen vorgenommen. Schließlich wurden die Variablen auf Normalverteilung überprüft, da diese für einige der weiteren Analysen Voraussetzung ist. Zu Überprüfung der Normalverteilung wurde der *Kolgomorov-Smirnov-Anpassungstest* angewandt. Dieser testet die Abweichung von einer Normalverteilung. Bei einer Wahrscheinlichkeit von $p<.05$ wird dabei davon ausgegangen, dass die Verteilung der Variablen signifikant von der Normalverteilung abweicht. Bei solchen Variabeln wurden im weiteren Verlauf entsprechend nichtparametrische Tests angewendet. Mit der deskriptiven Analyse sollte Transparenz über das vorliegende Datenmaterial gewonnen werden. Sie diente der Eliminierung von Ausreißern und potenziell fehlerbehafteter Daten. Die somit bereinigte Datenbasis wurde für alle weiteren Analysen verwendet. Welche Methoden dabei für welche Fragestellungen aus 3.1 in den einzelnen Abschnitten zur Anwendung kamen, soll nachfolgend skizziert werden.

3.2.6.1 Kohortenübergreifende Prädiktorenanalyse

Im Rahmen dieser *kohortenübergreifenden Prädiktorenanalyse* standen die Fragestellungen (1) bis (3) und (10) im Vordergrund.

Bei der *Reliabilitätsanalyse* (*Fragestellung (1)*) wurde auf Basis der Einzelurteile in den Interviews und den Beobachtungen als Übereinstimmungsmaß der *Intra-Class-Correlation-Koeffizient* (*ICC*) mit Einweg-Zufallseffekten berechnet, um die Urteilsübereinstimmung je Dimension festzustellen. Der Koeffizient beruht auf der Annahme, dass der Mittelwert der Population für alle Beobachtungen gleich und die Differenz zwischen den Urteilen, die auf die Beurteiler zurückgehen, zufällig und nicht systematischer Natur ist. Dies traf auch im vorliegenden Fall zu, da die jeweils untersuchten Stichproben der Bewerber, innerhalb der Beobachtungen und der Interviews gleich, die Zusammensetzung der Urteiler pro Bewerber allerdings unterschiedlich war. Im Folgenden wird der Begriff *Reliabilität* verwendet, da durch diese methodische Restriktion keine „reinen" Rater-Reliabilitäten berechnet werden können. Basis der Untersuchung waren mindestens drei *Beobachtungen* bzw. *Interviews* je Person (N=47). Es konnten deshalb die Variablen *Selorg, Leader, Zielor, Fuehr, Fit_ISB, Exot, VWL, Allg, Engl* und *AllBil* nicht berücksichtigt werden. Darüber hinaus wurden auch Übereinstimmungsmaße für Interviews und Beobachtung zusammen, einmal mit gleicher Stichprobe (K=6) über alle Dimensionen und dann mit maximal möglicher Personenanzahl (je nach Variable $73<N<154$) pro Dimension, gerechnet. Im Anschluss wurden *Whisker-Plots* erstellt, um einen Anhaltspunkt dafür zu bekommen, ob Unterschiede in

den Urteilen auf das *CC* oder die *Funktionsstufe* des Urteilers zurückzuführen sind. Basis der *Whisker-Plots* ist der Mittelwert des Gesamtwertes aus dem Recruiting, einmal nach *CC* und einmal nach *Funktionsstufe* des Beurteilers.

Die Untersuchung charakteristischer Profile (*Fragestellung (2)*) wurde getrennt nach *Einstiegslevel als JC* (ja/nein) betrachtet. Dies liegt in der Tatsache begründet, dass für JC-Anwärter andere Maßstäbe gelten als für Bewerber auf höheren Stufen. Eine gemeinsame Betrachtung beider Gruppen würde dazu führen, dass Personen mit identischer Leistung im Recruiting unterschiedliche Einstufungen bekämen. Somit wären die anhand biographischer Informationen vorgenommenen Gruppenvergleiche nicht mehr aussagekräftig. Da nicht genügend hohe Fallzahlen für eine Differenzierung der Bewerber für höhere Stufen vorlag (C und SC), wurden diese zusammengefasst. Basis für die Untersuchung waren alle Variablen mit N>150. Ferner wurden alle Bewerber aus dem *Jahr 2000* aus der Untersuchungsgruppe eliminiert, da diese auf Grund einer weniger restriktiven Einstellungspolitik im Recruiting signifikant besser beurteilt wurden (siehe dazu 3.3.2). Die zwei Bewerberstichproben wurden anhand folgender *CV-Variablen* gruppiert: *Sex*, *BWL_Stud*, *MBA*, *Dr*, und *Uni_Art*. Darüber hinaus wurden die Gruppen nach den Kriterien *Forced Exit* und *Angestelltenstatus* aufgeteilt. Zur Profildarstellung wurden jeweils die Mittelwerte der verglichenen Gruppen in einem Liniendiagramm gegenübergestellt. Zusätzlich wurden die ausgewählten Recruitingdimensionen *Kommunikationsfähigkeit* und *logisch-analytische Denken* als Ersatzgröße für quantitative Fähigkeiten bzw. verbale Fähigkeiten herangezogen. Hier wurden Mittelwerte der Personen, deren Werteausprägung auf den 25%-Randquartilen lagen, ebenfalls mit Hilfe eines Liniendiagramms gegenübergestellt. Ferner wurde ein t-Test für unabhängige Stichproben gerechnet, um festzustellen, ob die Mittelwerte der zwei Dimensionen innerhalb des oberen Quartils der einen Dimension signifikant voneinander abweichen. Abschließend wurden zwei zufällig ausgewählte Personen mit gleicher Ausprägung auf dem Gesamtwert aus Recruiting auf den einzelnen Dimensionen hinsichtlich ihrer Ausprägung verglichen. Auch hier diente der Visualisierung ein Liniendiagramm. Ziel der Profilanalysen war es, verschiedene Muster zu identifizieren, die bei unterschiedlichen Personengruppen auftreten und eventuell Aufschluss über Zusammenhänge mit der späteren Leistung liefern. Zudem können die Profile unterschiedliche gruppenspezifische Entwicklungspotenziale aufzeigen.

Zur Untersuchung der Konstruktvalidität (Fragestellung (3)) wurden Faktorenanalysen mit Recruitingvariablen für die Kohorte „*Einstiegslevel als JC*“ durchgeführt. Die getrennte Betrachtung wurde vorgenommen, da bei der späteren *Analyse der JC-Kohorte* auf die JC-spezifischen Faktoren zurückgegriffen wird. Auf eine Betrachtung der Kohorte „*Ein-*

stieg auf einem höheren Level" wurde auf Grund zu geringer Fallzahlen (N=49) verzichtet, da sich daraus keine robuste Faktorenstruktur ergeben dürfte. Basis der Untersuchung sind auch hier Personen, die nach 2000 in das Unternehmen eingestiegen sind, da in diesem Jahr, im Vergleich zu den Folgejahren, andere Recruitingziele im Vordergrund standen (*War for Talents*). Es ist davon auszugehen, dass diese das Urteil beeinflussen (siehe auch Murphy, Cleveland, Skattebo & Kinney, 2004). Wegen zu kleiner Fallzahlen wurden dabei die Variablen *Exot*, *Engl*, *AllBil*, *Fit_ISB*, *Fuehr*, *Selorg*, *Leader* und *Ziel* aus der Berechnung ausgeschlossen. Zunächst wurde eine explorative Hauptkomponentenanalyse (VARIMAX mit anschließender Rotation) der Prädiktorenseite mit listenweisem Fallausschluss vorgenommen (Kaiser-Guttmann Kriterium, Eigenwerte >1, $N_{JC\text{-}Einstieg}$=70). Im Anschluss wurden die gleichen Variablen *unrotiert* faktorisiert (ebenfalls Hauptkomponentenanalyse, Kaiser-Guttmann Kriterium, Eigenwerte>1). Darauf folgend wurden die extrahierten Faktoren hinsichtlich ihrer Variablenstruktur inhaltlich interpretiert. Anschließend wurde der korrelative Zusammenhang der Faktorwerte mit den dichotomen Variablen *Angestelltenstatus* und *Forced Exit* untersucht (Spearman-Rho-Korrelationskoeffizient; $N_{JC\text{-}Einstieg}$=70). Ziel dabei war es, Anhaltspunkte darüber zu bekommen, ob die identifizierten Konstrukte Relevanz für den späteren Berufserfolg besitzen.

Im letzten Analyseschritt der kohortenübergreifenden Prädiktorenanalyse sollte der Zusammenhang zwischen allen verfügbaren Prädiktorendaten und dem Berufserfolg aufgedeckt werden (*Fragestellung (10)*). Da kein einheitliches Projektbeurteilungskriterium vorlag (s.o.), wurden ersatzweise die genannten Kriterien *Angestelltenstatus* und *Forced Exit* verwendet. Um die geeignete Kombinationsmöglichkeit der Prädiktoren zu identifizieren, wurde das Verfahren der binären logistischen Regression angewandt. Mit dem Verfahren wird die Abhängigkeit einer dichotomen Variablen von anderen unabhängigen Variablen eines beliebigen Skalenniveaus untersucht. Im vorliegenden Fall wurde folglich dieses Verfahren angewandt, um mögliche Prädiktoren für Mitarbeiter zu identifizieren, deren Abgang erzwungen wurde, oder die das Unternehmen verlassen haben (FP). Ziel ist dabei, diese Anteile durch eine stärkere Berücksichtigung dieser Prädiktoren bei der Auswahl, zu reduzieren. Dabei wurden explorativ logistische Regressionen (schrittweise) mit ausgewählten *CV*- und den *Recruitingvariablen* gemeinsam durchgeführt. Die abhängigen dichotomen Variablen

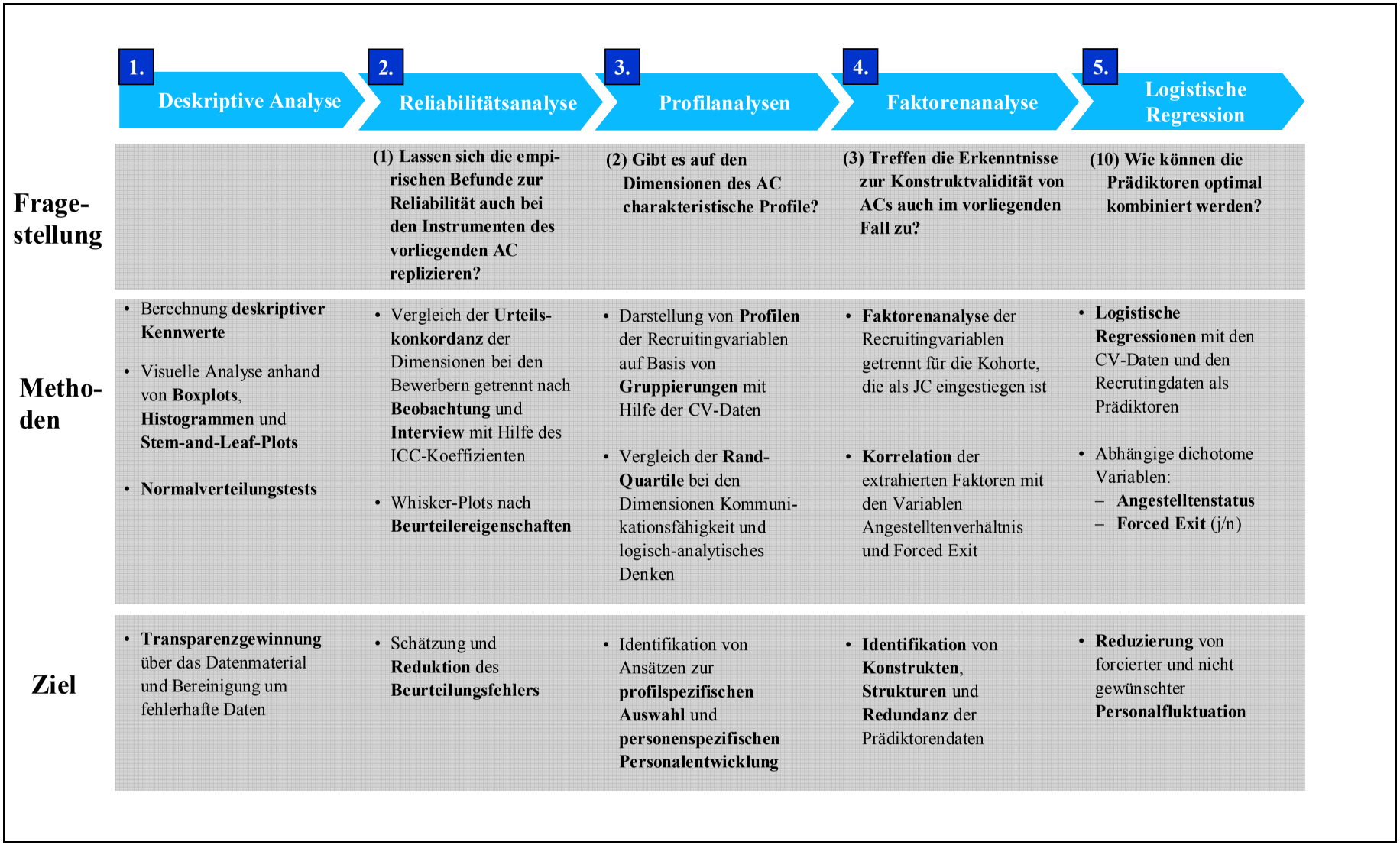

Abbildung 39: Vorgehen der kohortenübergreifenden Prädiktorenanalyse

waren *Forced Exit* (j/n) und *Angestelltenverhältnis* (ausgeschieden vs. angestellt). Auf Grund der hohe Sensitivität der Regressionen gegenüber fehlenden Werten wurden lediglich jene Bewerber in die Regressionsrechnung einbezogen, die Werte auf allen untersuchten Variablen hatten (listenweiser Fallausschluss, N=142). Folgende Variablen wurden für die Regressionsrechnung verwendet: *Pragma*, *Team*, *Konfl*, *Komm*, *Fach*, *LogAn*, *Intell*, *Charis*, *Frust*, *Kreat*, *Meth*, *Unt*, *BWL*, *VWL*, *Allg*, *Einstieg*, *Eintrittsalter*, *Sex*, *Abinote*, *BWLStud*, *Dr*, *Uni_Art*, *Pra_ISB*, *Pra _t*, *Pra_N*, *Pra_UB*, *FG*, *MBA* und *Job*. Bei der abhängigen Variable *Angestelltenverhältnis* wurde dabei die Variable *Eintrittsjahr* als erster Block in die Regression aufgenommen, um die Varianzaufklärung über natürliche Fluktuation, in Abhängigkeit der Verweildauer, zu berücksichtigen. *Abbildung 39* fasst das Vorgehen und die Zielsetzungen der kohortenübergreifenden Prädiktorenanalyse abschließend zusammen.

3.2.6.2 Kohortenspezifische Analysen

Auf Grund ihres ähnlichen Charakters sollen nachfolgend die in den Analysen der drei Kohorten *JC*, *C* und *SC* angewandten Methoden gemeinsam in einem Abschnitt dargestellt werden. Ziel der Analysen innerhalb der Kohorten ist dabei die Beantwortung der Fragestellungen (5), (6) und (8) bis (11).

Zur Untersuchung charakteristischer Profile auf den Kriteriendimensionen (*Fragestellung (5)*) wurden erneut die Bewerber anhand ausgewählter dichotomer Variablen gruppiert. Ziel dabei war es, Anhaltspunkte für eine Optimierung der Vorauswahl der Bewerber anhand biographischer Informationen zu gewinnen. Die Gruppierung erfolgte nach jenen Variablen, für die eine genügend hohe Differenzierung vorlag. So konnte beispielsweise bei den *JCs* die Variable *Familienstand* nicht zur Gruppierung herangezogen werden, da lediglich eine Person zum Zeitpunkt des Einstiegs als *JC* verheiratet war. Die Gruppierung erfolgte nicht auf Basis einer theoretischen Fundierung, sondern lediglich mit dem Zweck, explorativ Unterschiede zwischen den Gruppen festzustellen. Aus diesem Grund wurden explorative zweiseitige t-Tests für unabhängige Stichproben auf Mittelwertunterschiede auf den Projektbeurteilungsdimensionen gerechnet. Dabei besteht die Gefahr der α-Fehler-Inflation, da aus Anregungsgründen der Hypothesengenerierung auf eine Adjustierung nach Bonferroni verzichtet wurde. *Tabelle 20* zeigt die Gruppierungsvariablen für die einzelnen betrachteten Kohorten und die Anzahl dahinter liegender Fälle. Aus den Mittelwertunterschieden wurden in einem weiteren Schritt Effektstärken errechnet, um eine dimensionsübergreifende Vergleichbarkeit der Differenzen zu ermöglichen. Darauf folgend wurden erneut Gruppen anhand biographischer dichotomer Variablen gebildet, von denen Annahmen

über einen Einfluss auf die berufliche Leistung vorlagen. Die Forschergruppe um Lubinski (2004; Achter, et al., 1999; Shea, et al., 2001; Webb, et al., 2002) hat im Rahmen ihrer Hochbegabungsstudien gezeigt, dass bereits in jungen Jahren Präferenzen für eher *numerische* oder eher *verbale* Tätigkeitsschwerpunkte, in Abhängigkeit von der eigenen Fähigkeit, gezeigt werden. Dies dürfte sich folglich auch in den von den Bewerbern zur Schulzeit gewählten Fächerschwerpunkten abbilden. Aus diesem Grund wurde die Gruppe der *JCs* und *der Cs* ausgewählt, die im Rahmen der Erlangung der Allgemeinen Hochschulreife mindestens einen analytischen Fächerschwerpunkt (FSP) hatten und diese den restlichen *JCs* bzw. *Cs* gegenübergestellt. Da die Tätigkeit der Beratung gleichermaßen *verbale* wie *analytische* Fähigkeiten erfordert, dürften sich, so die Annahme, keine nennenswerten Unterschiede zwischen den zwei Gruppen zeigen. Die geschilderten empirischen Befunde zum *Berufswissen* (2.5.2.5) geben zudem Grund zu der Annahme, dass Bewerber mit einem entsprechenden beratungsspezifischen Vorwissen oder beruflicher Erfahrung insgesamt bessere Leistung in den Projekten erbringen dürften. Entsprechend sollten diese signifikant höhere Werte bei den Projektbeurteilungen haben.

Tabelle 20: Gruppierungssystematik explorativer t-Tests nach Kohorten

		Anzahl der Fälle		
Gruppierungvariable	**Verglichene Ausprägungen**	**JC**	**C**	**SC**
Eintrittsjahr	Eintrittsjahr =2000	15	-	24
	Eintrittsjahr >2000	68	-	62
Geschlecht	Weiblich	16	-	-
	Männlich	67	-	-
Universitätsart	Privat	23	23	16
	Öffentlich	60	90	69
Forced Exit	Ja	10	12	12
	Nein	73	102	74
Angestelltenverhältnis	Angestellt	56	78	32
	Ausgeschieden	27	36	54
Einstiegsstufe	Einstieg als JC	-	85	-
	Einstieg als C	-	32	-
	Einstieg als SC	-	-	27
	Kein Einstieg als SC	-	-	59
Familienstand	Ledig	-	105	78
	Verheiratet	-	7	8

So ist davon auszugehen, dass Mitarbeiter mit einem BWL-Studium solchen Kollegen in der Leistung überlegen sind, die kein solches Studium absolviert haben. Ebenso müss-

ten Mitarbeiter mit einem *MBA-Abschluss* sowie Mitarbeiter, die im Rahmen von Praktika Beratungserfahrung sammeln konnten, signifikant besser abschneiden. Analog dürften Bewerber mit einem Dr.-Titel auf Grund ihres höheren Bildungsstandes insbesondere auf der Dimension *Specialist-Know-how* besser beurteilt werden. *Tabelle 21* zeigt die Eckpunkte der Gruppenbildung anhand dieser biographischen Variablen. Auch hier wurden zweiseitige t-Tests für unabhängige Stichproben auf Mittelwertunterschiede auf den Projektbeurteilungsdimensionen (ohne Signifikanzniveau-Adjustierung) gerechnet. Zu Vergleichbarkeitszwecken wurden die Mittelwertunterschiede wiederum in Effektstärken transformiert.

Tabelle 21: Gruppierungssystematik für t-Tests nach Kohorten

		Anzahl Fälle		
Gruppierungvariable	**Verglichene Ausprägungen**	**JC**	**C**	**SC**
Analytischer FSP	Mindestens ein analytischer FSP	48	70	-
	Kein analytischer FSP	28	31	-
Studium der BWL	Ja	59	73	54
	Nein	24	40	31
MBA-Abschluss	Ja	9	15	11
	Nein	74	98	74
Dr.-Titel	Ja	-	14	14
	Nein	-	99	71
Praktikum bei einer UB	Ja	61	-	-
	Nein	22	-	-
Externe Berufserfahrung	Ja	-	-	35
	Nein	-	-	50

FSP= Fächerschwerpunkt; UB= Unternehmensberatung

Die Ergebnisse aller Mittelwertvergleiche wurden ferner zur Interpretationserleichterung als Profile mit Hilfe von Liniendiagrammen visualisiert. Abschließend wurde untersucht, inwieweit sich die Leistung auf den Projekten mit zunehmender Stufenzugehörigkeit der Beurteilten verändert. Zu diesem Zweck wurden die Mittelwerte der Gesamtwerte aus den Projektbeurteilungen im Zeitverlauf dargestellt und über eine einfaktorielle Varianzanalyse mit Messwiederholung auf Unterschiede getestet, um mögliche Lerneffekte zu visualisieren. Es wurden dabei die ersten drei Beurteilungen auf jeder Stufe herangezogen. Dabei ist allerdings zu berücksichtigen, dass es sich nicht um identische Messzeitpunkte, sondern um unterschiedliche Messzeitpunkte je Person handelt, da die Beurteilungen i.d.R. am Ende der jeweiligen Projekte, die eine unterschiedliche Dauer haben, erfolgen. Es handelt sich also lediglich um in ihrer Rangreihe sukzessive Beurteilungen der *JCs*. Da es zahlreiche Personen

gibt, für die nicht mindestens drei Beurteilungen vorliegen, wurden die jeweiligen Stichprobengrößen dadurch reduziert (N_{JC}=49 ; N_C=76; N_{SC}=41). Die Mittelwertunterschiede der Messzeitpunkte wurden wiederum zur besseren Interpretation in Effektstärken transformiert (Glass'sche Deltas).

Zur Ermittlung der prognostischen Validität der Prädiktor- und Kriterienvariablen (Fragestellungen (8) und (9)) wurden Korrelationsanalysen durchgeführt. In Abhängigkeit der betrachteten Kohorte wurde dabei das Vorgehen variiert.

Bei der *JC-Kohorte* wurden zunächst für ausgewählte *CV-Variablen* Rangkorrelationen (Spearman-Rho) mit den Projektbeurteilungsdimensionen, unter Einbezug der gesamten Stichprobe, berechnet. Ziel dabei war es, Anhaltspunkte über mögliche Zusammenhänge zwischen biographischen Daten und späterem Berufserfolg zu erlangen. Da der Autor von der Annahme positiver Zusammenhänge der ausgewählten Variablen ausging, wurden folgende Variablen einseitig gestestet: *Eintrittsalter, Note_FSP, Note_D, Note_M, Note_E, Abinote, Anz_Absch_tot* und *Pra_t*. Je nach betrachteter Variable variierte dabei auf Grund fehlender Werte die Stichprobengröße (42< N_{JC}<84). Für die weitere Analyse des Zusammenhangs aus Recruiting und Projektbeurteilungen wurden danach 15 Mitarbeiter, die 2000 eingestellt wurden, wegen potenzieller Verfälschungen aus der Untersuchungsgruppe ausgeschlossen. T-Test-Mittelwertvergleiche hatten hier zu der Erkenntnis geführt, dass diese Mitarbeiter bei den Projektbeurteilungen signifikant schlechter, im Recruiting hingegen signifikant besser beurteilt wurden als der Rest der Stichprobe (siehe dazu *3.3.2.3*). Im Anschluss wurden für die bereinigte Gruppe *Korrelationsmatrizen* und *Influence-Plots* der *Recruitingvariablen* mit den *Projektbeurteilungen* erstellt. Mithilfe der *Influence Plots* aus den Kombinationen der Gesamtwerte untereinander und den Gesamtwerten mit den jeweiligen Unterdimensionen des anderen Gesamtwertes wurden sieben Personen identifiziert, die einen Einfluss auf die Korrelation von mehr als .04 hatten. Diese Personen haben zwar außer ihrem großen Einfluss auf die Korrelationen keine weiteren erkennbaren Gemeinsamkeiten, sie können aber die Ergebnisse potenziell verfälschen. Die Personen (im Folgenden Ausreißer genannt) wurden deshalb aus der Stichprobe eliminiert. Die Stichprobengrößen für die Korrelationsanalysen mit der Recruitingdaten wurden dabei entsprechend kleiner (N_{JC}=59). Bei den im Anschluss berechneten Korrelationen wurde sukzessive von der Gesamtwertebene auf die Unterdimensionen übergegangen und so alle Kombinationen miteinander korreliert. Bei den *Gesamtwerten aus Recruiting* und *Projektbeurteilungen* wurde die Pearson-Korrelation (einseitig, N_{JC}=59) berechnet. Darüber hinaus wurde in einer separaten Berechnung eine Korrektur des Selektionseffektes vorgenommen. Hierzu wurde die Standardab-

weichung auf dem Gesamtrecruitingwert, der für die gesamte Bewerberpopulation vorlag (SD=.344) herangezogen. Für die Unterebenen wurden Rangkorrelationen (Spearman-Rho, einseitig, je nach Vergleich $18< N_{JC}<60$) angewandt, um eine potenzielle Gefährdung der Ergebnisse durch Verletzung des Intervallskalenniveaus auszuschließen.

Darüber hinaus wurden, erneut unter Ausschluss der 2000er Werte, Korrelationen der *Abiturnote* als einzige nicht-dichotome normalverteilte *CV-Variable* mit den Dimensionen der Projektbeurteilungen berechnet (Pearson-Korrelation, einseitig ohne die genannten Ausreißerwerte; $N_{JC}=59$). Zusätzlich wurde exemplarisch ein regressionsanalytisches Ersetzen fehlender Werte bei den Recruitingvariablen vorgenommen und die Korrelationen mit dem Gesamtwert aus Projektbeurteilungen des Datensatzes mit ersetzten Werten, dem Ursprungsdatensatz gegenübergestellt. Dabei zeigten sich starke Veränderungen der Signifikanzniveaus bei Variablen, die geringe Fallzahlen in der Ursprungsstichprobe aufweisen (z.B. Englischkenntnisse und Führungsfähigkeit) sowie teilweise nennenswerte Veränderungen in der Höhe der Korrelationen. Der Autor hat deshalb beschlossen, auf ein Ersetzen fehlender Werte in dieser Arbeit zu verzichten und auf tatsächlich vorhandene Werte zurückzugreifen.

Bei der C-Kohorte wurde ähnlich vorgegangen. Auch hier wurden zunächst für folgende *CV-Variablen* Rangkorrelationen (Spearman-Rho, einseitig) mit den Projektbeurteilungsdimensionen, unter Einbezug der gesamten Stichprobe, berechnet: *Eintrittsalter, Note_FSP, Note_D, Note_M, Note_E, Abinote, Job_t, Pra_t*. Je nach betrachteter Variable variierte dabei die Stichprobengröße ($57< N_C<115$). Auf Grund der im Vergleich zu den *JCs* größeren Stichprobe wurden im Anschluss nur jene *Cs* betrachtet, bei denen auf der Kriterienseite mindestens drei Projektbeurteilungen vorlagen, um eine Mindestaggregation sicherzustellen. Mithilfe des gleichen Vorgehens wie bei der JC-Kohorte wurde zusätzlich ein Ausreißer identifiziert und aus der Stichprobe eliminiert ($N_C=70$). Auf Grund der größeren Stichprobe war hier der gewählte Grenzwert ein Einfluss von .06 auf die Korrelation. Auf eine Bereinigung der Daten um jene Personen, die 2000 eingestiegen waren, wurde verzichtet, da nur zwei Personen dieser Gruppe als C eingestiegen sind. Bei den im Anschluss berechneten Korrelationen wurde analog sukzessive von der Gesamtwertebene auf die Unterdimensionen übergegangen und so alle Kombinationen miteinander korreliert. Bei den *Gesamtwerten aus Recruiting* und *Projektbeurteilungen* wurde die Pearson-Korrelation (einseitig, $N_C=70$) berechnet und ebenfalls eine Korrektur des Selektionseffektes (mithilfe der Populationsstandardabweichung auf dem Recruiting-Gesamtwert SD=.344) vorgenommen. Für die Unterebenen wurden Rangkorrelationen (Spearman-Rho, einseitig, je nach Vergleich

$9< N_C<71$) angewandt, um eine potenzielle Gefährdung der Ergebnisse durch Verletzung des Intervallskalenniveaus auszuschließen. Darüber hinaus wurden Korrelationen der *Abiturnote* als einzige nicht-dichotome normalverteilte *CV-Variable* mit den Dimensionen der Projektbeurteilungen berechnet (Pearson-Korrelation ohne die genannten Ausreißerwerte, einseitig, $N_C=71$) und grafisch für den Gesamtwert aus Projektbeurteilung dargestellt.

Bei der *SC-Kohorte* wurde für die gleichen *CV-Variablen* wie bei der C-Kohorte Rangkorrelationen (Spearman-Rho, einseitig, je verglichener Dimension $34<N_{SC}<85$) mit den Projektbeurteilungsdimensionen unter Einbezug der gesamten Stichprobe berechnet. Da in der *SC*-Stichprobe zwar 24 Personen 2000 eingestiegen sind, keiner davon jedoch auf der *SC*-Stufe, wurde auf eine Bereinigung dieser Personen verzichtet. Mithilfe des gleichen Vorgehens wie bei der *JC-Kohorte* wurden zusätzlich sechs Ausreißer identifiziert und aus der Stichprobe eliminiert ($N_{SC}=79$). Anschließend wurde erneut sukzessive von der Gesamtwertebene auf die Unterdimensionen übergegangen und so alle Kombinationen miteinander korreliert. Bei den *Gesamtwerten aus Recruiting* und *Projektbeurteilungen* wurde die Pearson-Korrelation (einseitig, $N_{SC}=79$) berechnet. Zudem wurde auf Grund mangelnder Signifikanz auf eine Berechnung des Selektionseffektes verzichtet und Korrelationen für die Eintrittsjahre 2000 vs. Rest nach Einstiegskohorten mit und ohne Ausreißer gegenübergestellt. Ziel war dabei festzustellen, inwieweit die Validität sich in den so gruppierten Stichproben unterscheidet. Für die Unterebenen wurden Rangkorrelationen (Spearman-Rho, einseitig, je nach Vergleich $4<N_{SC}<80$) errechnet. Darüber hinaus wurden Korrelationen der *Abiturnote* als einzige nicht-dichotome normalverteilte *CV-Variable* mit den Dimensionen der Projektbeurteilungen berechnet (Pearson-Korrelation ohne die genannten Ausreißerwerte, einseitig; $N_{SC}=78$) und grafisch für den Gesamtwert aus Projektbeurteilung dargestellt.

Die optimale Kombination der Recruitingvariablen (*Fragestellung (10)*) wurde mit verschiedenen multiplen Regressionsanalysen untersucht. Auch hier variiert das Vorgehen je Kohorte. Gegenstand der Regressionen waren auf der Prädiktorenseite die Recruitingvariablen sowie die *Abiturnote* als ausgewählte CV-Variable.

Bei der *JC-Kohorte* wurde die um Ausreißerwerte korrigierte Personengruppe, die *nach 2000* eingestellt wurde, als Stichprobe herangezogen. Darüber hinaus wurden für die Regressionsrechnungen auf Grund der hohen Sensitivität gegenüber fehlenden Werten nur jene Personen einbezogen, bei denen alle Werte für die verwendeten Variablen vorhanden waren (listenweiser Fallausschluss; $N_{JC}=52$). In einem ersten Schritt wurden erneut die *Recruitingvariablen* einem Normalverteilungstest unterzogen, um jene Variablen auszuschlie-

ßen, die nicht die Bedingungen multipler Regression erfüllen und bei nicht genügend großer Fallzahlen nicht in die Regression einbezogen werden sollten (siehe dazu Bortz, 1993, S.417). Folgende Recruitingvariablen wurden aus der weiteren Betrachtung ausgeschlossen: *Engl, Fuehr, AllBil, Selorg, Leader, Zielor, Fit_ISB, Exot*. Im Anschluss wurden zunächst Regressionen der einzelnen Dimensionen der Projektbeurteilungen aus dem *Gesamtwert des Recruitings* und der *Abiturnote* separat sowie hierarchisch mit beiden zusammen vorhergesagt. Danach wurde zunächst jede *Projektbeurteilungsdimension* explorativ über eine schrittweise Regression vorhergesagt, mit und ohne *Abiturnote* als vorgestelltem Block, um maximale Varianz aufzuklären. Wegen hoher *Multikolinearität* wurden zusätzlich alle Variablen unter Hinzunahme der *Abiturnote* auch als *Set* von Indikatoren der Unterdimensionen der Projektbeurteilungen in eine *blockweise Regression* mit der *Abiturnote* als erstem Block eingebracht. Dabei wurde auf Grund der Unmöglichkeit, eine inhaltlich sinnvolle Reihenfolge der Sets vorzugeben, explorativ maximierend vorgegangen. Die *Sets* wurden sachlogisch abgeleitet und über Korrelationsmatrizen visualisiert. Die Definitionen der einzelnen *Set*s finden sich in *Tabelle 22*.

Tabelle 22: Set-Bildung der Recruitingvariablen bei der JC-Kohorte

Set	**Inhaltlich zugeordnete Variablen**
Soziale Fähigkeiten	Team, Konfl, Komm, Charis
Prozessuales Wissen	LogAn, Meth, Kreat
Fokussierung ("Adlerblick")	Pragma, Unt
Ausdauer	Intell, Frust
Spezialwissen	Fach, BWL
Allgemeinwissen	VWL, Allg

Bei allen *Sets* zeigten sich hohen Interkorrelationen zwischen den Variablen. Bei der Vorhersage mit den *Sets* wurde der Beitrag der einzelnen Variablen nicht spezifiziert, sondern nur der inkrementelle Vorhersagebeitrag des Sets interpretiert. Alle Regressionsmodelle wurden hinsichtlich weiterer Prämissenverletzungen (*Nichtlinearität*, *Heteroskedastizität* und *Normalverteilung* der Residuen) überprüft und keine nennenswerten Anomalien entdeckt.

Bei der *C-Kohorte* wurde ebenfalls zunächst die Recruitingvariablen einem Normalverteilungstest unterzogen, um jene Variablen auszuschließen, die nicht den Bedingungen multipler Regression entsprechen. Ausgangsbasis der Analysen war dabei die um den Aus-

reißerwert korrigierte Personengruppe mit mindestens drei Beurteilungen auf Kriterienseite (s.o.). Es wurden die gleichen acht Recruitingvariablen aus der weiteren Betrachtung ausgeschlossen wie bei der *JC-Kohorte*. Darüber hinaus wurden in die Regressionsrechnungen, auf Grund der hohen Sensitivität gegenüber fehlenden Werten nur jene Personen einbezogen, bei denen alle Werte für *Abiturnote* und die verwendeten Variablen vorhanden waren (listenweiser Fallausschluss; N_C=67). Zunächst wurden Regressionen der einzelnen Dimensionen der Projektbeurteilungen aus dem *Gesamtwert des Recruitings* und der *Abiturnote* separat vorhergesagt. Anstelle der Set-Bildung wurde bei der C-Kohorte eine andere Regressionsvariante gewählt. Diese beinhaltete die inhaltlich-sachlogische Ableitung von Prädiktoren, bei denen ein Zusammenhang mit dem jeweiligen Kriterium vermutet wurde.

Tabelle 23: Inhaltlich sachlogisch zugeordnete Variablen je Kriteriendimension (Cs)

Kriterium	**Inhaltlich-sachlogisch zugeordnete Variablen**
Durchschnittlicher Gesamtwert	Alle
Specialist-Know-how	Fach, Intell, Meth, BWL,VWL
Problem Solution	Pragma, LogAn, Fach, Intell, Frust, Kreat, Meth, Unt, BWL
Quality of Work	Pragma, Fach, LogAn, Kreat, Meth, Unt, BWL, VWL
Behaviour under Pressure	Pragma, Team, Konfl, Komm, Frust, Kreat, Unt
Client Acceptance	Alle
Client Communication	Alle
Know-how-Management	Fach, Meth, BWL, Pragma, Team
Teamwork	Team, Konfl, Komm, Charis, LogAn, Kreat

Die vom Autor vorgenommene Zuordnung ist in *Tabelle 23* dargestellt. Dabei wurden auf Grund der Unmöglichkeit, eine inhaltlich sinnvolle Reihenfolge dieser Variablen vorzugeben, diese explorativ maximierend in eine schrittweise Regression eingegeben. Im Anschluss wurde jede *Projektbeurteilungsdimension* explorativ über eine schrittweise Regression durch alle Recruitingvariablen vorhergesagt, um maximale Varianz aufzuklären. Alle Regressionsmodelle wurden hinsichtlich weiterer Prämissenverletzungen (*Nichtlinearität*, *Heteroskedastizität* und *Normalverteilung der Residuen*) überprüft und keine Anomalien festgestellt.

Wegen der ernüchternden korrelativen Zusammenhänge, die sich aus der Korrelationsanalyse ergaben, wurden bei der *SC-Kohorte* lediglich drei Regressionsvarianten gerechnet. Ausgangsbasis der Analysen war dabei wiederum die um die Ausreißerwerte korrigierte

Personengruppe, ohne Berücksichtigung der *Recruitingvariablen*, die in den anderen beiden Kohorten ebenfalls ausgeschlossen wurden. Auch hier wurden nur jene Personen einbezogen, bei denen alle Werte für *Abiturnote* und die verwendeten Variablen vorhanden waren (listenweiser Fallausschluss; N_{SC}=70). Zunächst wurden Regressionen der einzelnen Dimensionen der Projektbeurteilungen aus dem *Gesamtwert des Recruitings* und der *Abiturnote* separat vorhergesagt. Im Anschluss wurde jede *Projektbeurteilungsdimension* lediglich explorativ über eine schrittweise Regression durch alle *Recruitingvariablen* vorhergesagt, um maximale Varianz aufzuklären. Alle Regressionsmodelle wurden auch hier hinsichtlich weiterer Prämissenverletzungen (*Nichtlinearität, Heteroskedastizität und Normalverteilung der Residuen*) überprüft und keine Anomalien festgestellt.

In dem letzten großen Analyseblock sollten die Konstruktvalidität des Kriteriums sowie die Einhaltung der Symmetrieprinzipien zwischen Prädiktor und Kriterium untersucht werden (*Fragestellungen (6) und (11))*.

Bei der *JC-Kohorte* wurde zu diesem Zweck zunächst eine Faktorenanalyse der Kriterienseite (Hauptkomponentenanalyse) mit den *Projektbeurteilungsdimensionen* durchgeführt (VARIMAX mit anschließender Rotation; Extraktion nach Scree; listenweiser Fallausschluss; N_{JC}=61) und die extrahierten, unrotierten und rotierten Faktoren interpretiert. Basis der Faktorenanalyse waren die Werte aller Mitarbeiter, *die nach 2000 eingestellt* wurden, bereinigt um Ausreißer (s.o.). Ziel dabei war es, Aufschluss über die hinter den Kriterien liegenden Konstrukte zu erlangen und zu prüfen, ob im vorliegenden Fall ebenfalls die Extraktion von einem *Generalfaktor* bestätigt wird (siehe Viswesvaran, et al., 2005). Die Faktoren auf der Prädiktorenseite wurden aus der kohortenübergreifenden Prädiktorenanalyse übernommen, da hier eine wesentlich größere Stichprobe vorlag (rotierte und unrotierte Lösungen; $N_{JC\text{-}Einstieg}$=70). Die extrahierten Faktoren wurden in einem weiteren Schritt auf beiden Seiten hinsichtlich korrelativer Zusammenhänge untereinander und mit den *Gesamtwerten* aus *Projektbeurteilungen* und *Recruiting* untersucht (Spearman-Rho; je nach Variablenkombination $51<N_{JC}<55$). Die Korrelationen sollten Indizien für die Generalitätsebene der betrachteten *Prädiktor*- und *Kriteriendimensionen* liefern. Im Anschluss wurden schrittweise Regressionen der *Kriterienfaktoren* aus den *Prädiktorfaktoren* für die unrotierten und die rotierten Varianten gerechnet (listenweiser Fallausschluss; N_{JC}=54), um festzustellen, welche Konstrukte prognostische oder inkrementelle Varianzaufklärung aufweisen. Zusätzlich wurde eine zweite Faktorenanalyse aus den Erkenntnissen der korrelativen und regressionsanalytischen Untersuchung heraus gerechnet, in der die unterschiedliche vermutete Generalitätsebene auf Kriterienseite im Sinne der *Brunswik´schen Symmetrieprinzipien*

berücksichtigt wurde. Die problematischen klientenzentrierten Dimensionen *Client Communication* und *Client Acceptance*, die vermutlich bei *JCs* nicht beurteilt werden können (siehe 3.3.2), wurden dabei vernachlässigt. Bei der Faktorisierung wurden auf der Kriterienseite die Dimensionen *Specialist-Know-how, Know-how-Management, Teamwork* und *Problem Solution* gemeinsam faktorisiert (Hauptkomponentenanalyse; VARIMAX mit anschließender Rotation; listenweiser Fallausschluss; N_{JC}=61). Der dabei entstehende Faktor *Allgemeines Arbeitsverhalten* wurde erneut mit den auf seiner Generalitätsebene vermuteten Dimensionen *Quality of Work* und *Behaviour under Pressure* faktorisiert (Hauptkomponentenanalyse, VARIMAX mit anschließender Rotation, Extraktion von nur einem Faktor; listenweiser Fallausschluss; N_{JC}=61). Der extrahierte Generalfaktor wurde anschließend hinsichtlich korrelativer Zusammenhänge mit der *Abiturnote* und den in der Korrelationsanalyse signifikanten Variablen *Charis*, *Komm*, *Intell* und *VWL* dem ursprünglich extrahierten ersten Faktor und den Gesamtwerten aus Projektbeurteilungen gegenübergestellt (Spearman-Rho; paarweise Fallausschluss; N_{JC}=52).

Bei der *C-Kohorte* wurde analog vorgegangen. Die Faktoren auf der Prädiktorenseite wurden aus der Prädiktorenanalyse übernommen (rotierte und unrotierte Lösungen, $N_{JC\text{-}Einstieg}$=70). Im Anschluss wurde eine Faktorenanalyse der Kriterienseite (Hauptkomponentenanalyse) mit den *Projektbeurteilungsdimensionen* durchgeführt (Extraktion nach Scree; listenweiser Fallausschluss; N_C=75) und die extrahierten (VARIMAX mit anschließender Rotation) Faktoren interpretiert. Basis der Faktorenanalyse waren die Werte aller Mitarbeiter, die mindestens drei Projektbeurteilung hatten, bereinigt um Ausreißer. Die extrahierten Faktoren wurden auf *Prädiktor-* und *Kriterienseite* hinsichtlich korrelativer Zusammenhänge untereinander und mit den *Gesamtwerten* aus *Projektbeurteilung* und *Recruiting* untersucht (Spearman-Rho). Für die Kombinationen mit den Recruitingfaktoren aus der kohortenübergreifenden Prädiktorenanalyse ergab sich eine geringere Fallzahl ($N_{JC\text{-}Einstieg}$=44). Für die Kombination zwischen den *Projektbeurteilungsfaktoren* und den *Recruitingvariablen* war hingegen eine größere Fallzahl vorhanden (N_C=70). Im Anschluss wurden erneut schrittweise Regressionen mit den Faktoren gerechnet (listenweiser Fallausschluss; N_C=44). Dabei wurden lediglich jene Mitarbeiter betrachtet, die als *JC* eingestiegen sind. Darüber hinaus wurde, wie bei der *JC-Kohorte*, eine zweite Faktorenanalyse aus den Erkenntnissen der korrelativen und regressionsanalytischen Untersuchung heraus gerechnet, in der erneut die vermuteten Generalitätsebenen der Kriterienvariablen im Sinne der *Brunswik´schen Symmetrieprinzipien* berücksichtigt wurde. Der Faktor *Allgemeines Arbeitsverhalten* wurde analog zur *JC-Kohorte* faktorisiert (Extraktion nach Scree; listenweiser Fallausschluss;

N_C=75). Auf oberster Ebene wurden hingegen die Dimensionen *Client Communication* und *Client Acceptance* in die Faktorisierung einbezogen. In einem weiteren Schritt wurde der so neu gebildete *Beratungserfolgsfaktor* hinsichtlich korrelativer Zusammenhänge mit den in der Korrelationsanalyse signifikanten Variabeln *Pragma*, *Team*, *Komm*, *LogAn*, *Intell*, *Unt*, *BWL*, und *dem Gesamtwert aus Recruiting* dem ursprünglich extrahierten unrotierten ersten Faktor und den *Gesamtwerten aus Projektbeurteilungen* gegenübergestellt (Spearman-Rho; paarweise Fallausschluss; je nach Variable $53<N_{JC}<56$). Zusätzlich wurde in einem explorativen Pfadmodell die maximale Varianzaufklärung bei dem *Beratungserfolgsfaktor* dargestellt. Bei der Modellrechnung wurden die signifikanten Prädiktorenvariablen *Kommunikationsfähigkeit* und *logisch-analytisches Denken*, die verbale und numerische Fähigkeiten beinhalten dürften, zur Vorhersage herangezogen (N=75).

Bei der *SC-Kohorte* wurden ebenfalls zunächst die Faktoren auf der Prädiktorenseite aus der Prädiktorenanalyse übernommen (rotierte und unrotierte Lösungen, $N_{JC\text{-}Einstieg}$=70)). Im Anschluss wurde eine Faktorenanalyse der Kriterienseite (Hauptkomponentenanalyse) mit den *Projektbeurteilungsdimensionen* durchgeführt (VARIMAX mit anschließender Rotation; Extraktion nach Scree; listenweiser Fallausschluss; N_{SC}=82) und die extrahierten Faktoren interpretiert. Basis der Faktorenanalyse waren die Werte aller Mitarbeiter, bereinigt um Ausreißer. Die extrahierten Faktoren wurden auf *Prädiktor*- und *Kriterienseite* hinsichtlich korrelativer Zusammenhänge untereinander und mit den *Gesamtwerten* aus *Projektbeurteilung* und *Recruiting* untersucht (Spearman-Rho). Für die Kombinationen mit den Recruitingfaktoren aus der kohortenübergreifenden Prädiktorenanalyse ergaben sich geringe Fallzahlen ($N_{JC\text{-}Einstieg}$=17). Für die Kombination zwischen den *Projektbeurteilungsfaktoren* und den Recruitingvariablen waren hingegen größere Fallzahlen vorhanden ($51>N_{SC}>57$). Auf eine anschließende Regressionsrechnung und Pfadanalysen wurde auf Grund mangelnder Validität verzichtet.

3.2.6.3 Erweiterte Fragestellungen

In den vorherigen zwei Abschnitten wurde erläutert, mit welchen Methoden einige der zentralen Fragestellungen dieser Arbeit angegangen wurden. Neben den reinen statistischen Analysen soll im Rahmen dieser Arbeit im Kapitel 3.4 aber auch die Auswahl und Operationalisierung der Prädiktoren und Kriterien kritisch hinterfragt werden (*Fragestellungen (4)* und *(7)*). Dabei lässt das vorliegende Datenmaterial keine umfassenden statistischen Analysen zu. Vielmehr sollen die im Rahmen des Theorieteils geschilderten Erkenntnisse mit dem vorliegenden Vorgehen abgeglichen werden. *Abbildung 5* zeigte eingangs dieser

Arbeit, dass eine umfassende Evaluation zahlreiche, den Rahmen dieser Arbeit sprengende Fragestellungen betrifft. Dies trifft auch auf diese beiden Fragestellungen zu, so dass hier vereinzelte Punkte aufgegriffen werden. Dabei wird zunächst auf der Kriterienseite überprüft, ob das verwendete Projektbeurteilungssystem den gesamten Bereich dessen abdeckt, was man als Beratungserfolg verstehen kann. Der Autor vergleicht an dieser Stelle die Dimensionen der Projektbeurteilung mit dem entwickelten *Fünf-Säulen-Modell* erfolgreicher Beratung. Darüber hinaus sollen auf der Prädiktorenseite Hinweise gewonnen werden, wie die derzeitige Prädiktorenauswahl und Operationalisierung verbessert und gegebenenfalls durch Alternativen ersetzt oder ergänzt werden kann.

Neben der Operationalisierungsthematik stehen im Kapitel 3.4.2 die Urteilsbildung und die Folgen der diagnostischen Entscheidung im Betrachtungsfokus (*Fragestellung (12)*). Hier soll am vorliegenden Beispiel erläutert werden, wie die Urteilsbildung verbessert werden kann. Zudem wird dargestellt, welche Verteilung der Bewerber in die vier Klassen FP, FN, RP und RN das vorliegende Vorgehen generiert.

Unter 3.5 wird der Nutzen des Verfahrens behandelt (*Fragestellung (13)*). Neben der ökonomischen Bewertung der aktuellen, werden dabei auch die ökonomische Nutzeneffekte alternativer Vorgehensweisen thematisiert. Basis der Berechnung ist dabei das Boudreau-Modell aus *Gleichung 19*. Kapitel 3 endet schließlich mit einer Darstellung der aus allen Ergebnissen resultierenden Implikationen für die Praxis (*Fragestellung (14)*).

3.3 Ergebnisse

Bei der Durchführung der in den nachfolgenden Kapiteln dargestellten Analysen wurden die Programme *Excel* (zur Datenaufbereitung), *SPSS 12.0* und *SYSTAT 10* verwendet. Zur Pfadanalyse wurde *AMOS 4* benutzt. Detaillierte Berechnungsergebnisse finden sich im Anhang. Im Folgenden wird aus Komplexitätsgründen nur auf die im Rahmen der Interpretation wesentlichen Tabellen und Grafiken eingegangen.

3.3.1 Kohortenübergreifende Prädiktorenanalyse

In einem ersten Analyseschritt wurden die Prädiktorendaten aus dem Recruiting separat betrachtet, um die Güte, insbesondere die *Objektivität* und die *Konstruktvalidität* des verwendeten Auswahl- bzw. Ratingverfahrens, unabhängig von dessen prognostischer Validität einschätzen zu können. Nach einer deskriptiven Analyse, um Transparenz über das vorliegende Datenmaterial zu erlangen, wurden die Reliabilitäten untersucht. Hier dürften, analog zu den geschilderten empirischen Befunden aus 2.5.3.2, auf Grund der semi-

strukturierten Interviewform Werte geringer als r=.50 zu erwarten sein. Bei den Beobachtungen in der Gruppenarbeit oder den Präsentationen könnten die Werte etwas höher liegen, da alle Beurteiler die gleiche Situation beobachten und somit situationsspezifische Abweichungen und Schwankungen weitgehend unterbleiben. Darüber hinaus soll untersucht werden, welche Profile die ausgewählten Bewerber auf den einzuschätzenden Dimensionen haben. Das geschilderte Auswahlverfahren, das einen kompensatorischen Charakter aufweist, dürfte die Entstehung unterschiedlicher Profile begünstigen, da ein Ausgleich schwacher Werte auf einer Dimension durch stärkere Werte auf einer anderen Dimension möglich ist. Die Profilunterschiede müssten sich besonders auf den von der Forschergruppe um Lubinski (2004; Achter, et al., 1999; Shea, et al., 2001; Webb, et al., 2002) untersuchten Gegenpolen *verbaler* und *numerischer* Fähigkeiten zeigen (figurale Komponenten wurden bei dem Auswahlverfahren nicht erfasst). Zusätzlich sollte festgestellt werden, ob bei der Personalauswahl bereits Anhaltspunkte für ein Ausscheiden des Mitarbeiters aus dem Unternehmen (forciert vs. nicht forciert) gewonnen werden können. Während ein freiwilliges Ausscheiden bei dem Unternehmen auf Grund des Wesens des Beratungsgeschäftes nach einer gewissen Verweildauer sowieso zu erwarten ist, ist die Reduktion von forciert ausscheidenden Mitarbeitern von besonderem Interesse. Mit ihr dürfte nicht nur eine Einsparung von ineffektiven Personalentwicklungskosten, sondern auch eine Erhöhung der durchschnittlichen Beratungsqualität einhergehen. Zu diesem Zweck wurden mehrere logistische Regressionen explorativen Charakters zur Vorhersage des Ausscheidens gerechnet.

Schließlich soll eine explorative Faktorenanalyse Anhaltspunkte darüber liefern, welche Konstrukte durch das Auswahlverfahren erfasst werden. Gemäß den Untersuchungen zur Konstruktvalidität von *ACs* (siehe 2.5.3.3) dürften auf Grund des *Validitätsparadoxons* insgesamt geringe *diskriminante* Validitäten und die Extraktion von einem Generalfaktor, der den Großteil der Varianz aufklärt, zu erwarten sein.

3.3.1.1 Methodische Restriktionen

Die Ergebnisse sind unter zahlreichen methodischen Restriktionen zu interpretieren. So ist der *Stichprobenumfang* einzelner Variablen und deren Kategorien sehr klein und damit die Effekte schwer zu generalisieren. Kleine Stichproben führen zwangsläufig zu geringen Teststärken, so dass anzunehmen ist, dass mögliche wahre Effekte in den Analysen nicht aufgedeckt werden konnten. Dadurch, dass zusätzlich in die Analysen (z.B. bei der binär logistischen Regression) sehr viele Variablen eingeflossen sind, ist eine α-Fehler-Inflation, die zu zufällig signifikanten Effekten führt, ebenfalls nicht auszuschließen. Problematisch ist

ebenso die Gefahr des *Selektionsfehlers* und eine einhergehende *Varianzreduktion*, die zu einer Unterschätzung der Effekte führen kann. Außerdem besteht eine hohe Anfälligkeit der Analysen, z.B. bei der *Profilbildung*, für *Ausreißerwerte*. Es ist deshalb von einer Instabilität der berechneten Koeffizienten auszugehen. Zusätzlich können sich durch *paarweise fehlende Werte,* je nach Vergleichsdimensionen, die Effekte stark verändern. Zudem sind die hohe *Multikolinearität* und damit konfundierte Varianzanteile zwischen den Variablen problematisch, da sie zu einer Verletzung der Prämissen der *Regressionen* und einer Ineffizienz der β-Koeffizienten führen. Schließlich geht die vorgenommene Faktorisierung der Variablen mit einem *Varianzverlust* einher, der zu einer Effektverzerrung führen kann, da die extrahierten Faktoren nicht den in der Realität verwendeten Prädiktoren entsprechen. Somit ist die Interpretierbarkeit der Ergebnisse insgesamt eingeschränkt.

3.3.1.2 Deskriptive Analyse

Tabelle 24 bildet die deskriptiven Kennwerte der *Recruitingvariablen* ab. Sie zeigt durchweg hohe Mittelwerte und geringe Standardabweichungen. Zudem haben einzelne Variablen kleine Fallzahlen, da nicht alle Dimensionen bei allen Personen beurteilt wurden und sich die Beurteilungsdimensionen im Zeitverlauf geändert haben. Dies betrifft vor allem die Dimensionen *Leadership*, *Zielorientierung*, *Fit ISB-Team* und *Exotenwissen*. Ein insgesamt relativ hoher Durchschnittsgesamtwert aus Recruiting mit 4.82 und ein Minimalwert von 3.76 deuten zusätzlich auf einen restriktiven Einstellungs-*Cutoff* hin.

Die Betrachtung der Verteilungseigenschaften (Histogramme, Kolgomorov-Smirnov-Anpassungstest) lieferte folgende Erkenntnisse: Die Verteilungen der Recruitingvariablen sind höchst unterschiedlich. Während einige Variablen signifikant von der Normalverteilung abweichen (z.B. *Pragmatismus*, *Selbstorganisation* und *Teamfähigkeit*), trifft dies auf andere (z.B. *Leadership*, *Gesamtmittelwert*) nicht zu. Die visuelle Inspektion deutet darauf hin, dass die Abweichungen hauptsächlich in einer schmalgipfligen, rechtslastigen Verteilung um den Wert 5.0 liegen. Da die Stichprobe nur aus Personen besteht, die eine Zusage bekommen haben, ist insgesamt nicht mehr von einer Normalverteilung auszugehen. Stattdessen liegt eine Verteilungsform vor, bei der der linke Teil der Normalverteilung „abgeschnitten“ und diese somit steiler wird. Dies führt zu einer Varianzreduktion der Recruitingdaten. Diese fällt insgesamt relativ gering aus (von .344 auf .332 beim Gesamtmittelwert), da zum Auswahlverfahren nur eine bereits ausgewählte Gruppe von Personen zugelassen wird, die bestimmte Bedingungen erfüllt. Zudem zeigt sich, dass die angewandte Skalierung nicht in ausreichendem Maße ausgeschöpft und hinreichend präzise eingesetzt wird, da die Beurteilungen sich

in einer sehr engen Spannweite bewegen. Zusätzlich können die angesprochenen Urteilsfehler (insb. *Halo-Effekt*, *zentrale Tendenz* und *Mildeffekte*, siehe auch Guion, 1998, S.577ff.), denen die *Interviewer* und *Beobachter* unterliegen, ebenso zu einer Varianzreduktion beitragen. Schließlich führt die Technik der *Aggregation* der Urteile zusätzlich zu einem Informationsverlust, der aber aus Reliabilitäts- und Validitätsgründen von dem Autor in Kauf genommen wird (siehe 3.2.4).

Tabelle 24: Deskriptive Kennwerte der Recruitingvariablen

Recruitingdimension	N	MW	Med.	Min.	Max.	SD	NV (j/n)
Pragmatismus/Ergebnisorientierung	203	4.84	4.75	3.50	5.75	.46	n
Selbstorganisation, Engagement, Biss	87	4.86	4.75	3.50	6.00	.60	n
Leadership	35	4.77	4.75	3.50	6.00	.71	j
Zielorientierung	35	4.73	4.75	3.50	6.00	.58	n
Team-/Integrationsfähigkeit	203	5.06	5.06	3.50	6.00	.50	n
Durchsetzungskraft/Führungsfähigkeit	86	4.74	4.75	3.50	6.00	.63	j
Konflikt- und Kritikfähigkeit	202	4.94	4.75	3.50	6.00	.47	n
Kommunikationsstärke/Eloquenz	204	4.93	5.00	3.29	6.00	.45	n
Fachkenntnisse	200	4.58	4.75	2.56	6.00	.64	n
strukturiertes/ logisch-analytisches Denken	204	4.77	4.75	2.88	6.00	.48	n
Intellektuelle Neugier	204	4.82	4.91	3.50	6.00	.43	n
Ausstrahlung/Charisma	204	4.90	4.96	3.50	5.75	.47	n
Frustrationstoleranz	203	4.89	4.75	3.19	6.00	.50	n
Kreativität/ Flexibilität	203	4.82	4.75	2.88	6.00	.58	n
Methoden-Kenntnisse	204	4.55	4.54	2.88	6.00	.50	n
Fit ISB-Team	43	4.96	4.75	3.50	6.00	.72	n
Unternehmerisches Denken	201	4.81	4.75	2.25	6.00	.63	n
Exotenwissen	32	4.93	4.75	3.50	6.00	.82	j
BWL-Kenntnisse	202	4.63	4.75	2.88	6.00	.68	n
VWL-Kenntnisse	196	4.70	4.75	2.25	6.00	.60	n
Allgemeinwissen	196	4.74	4.75	3.08	6.00	.61	n
Englischkenntnisse	147	4.93	4.75	2.25	6.00	.74	n
Allgemeinbildung	145	4.85	4.75	3.50	6.00	.61	n
Entscheidung für Einstellung (j/n)	204	.94	1.00	.43	1.00	.11	n
Gesamtmittelwert aus Recruiting	204	4.82	4.79	3.76	5.73	.33	j

N= Stichprobenumfang; MW=Mittelwert; Med.=Median; Min.=Minimum; Max.=Maximum; SD=Standardabweichung; NV=Normalverteilungsannahme

Bei den nicht-dichotomen *CV-Daten* zeigt sich ein differenzierteres Bild (*Tabelle 25*). Die Zeit in Monaten, die die Bewerber insgesamt an Praktika absolviert haben (*Pra_t*),

schwankt beispielsweise stark einzelfallbedingt. Hier finden sich entsprechend höhere Standardabweichungen und Maximalwerte. Auf der anderen Seite differenzieren andere Variablen weniger zwischen den Personen. So hat die *Abiturnote,* ebenfalls selektionsbedingt, eine Schwankungsbreite zwischen 1.0 und 3.2. Zudem waren bei einzelnen Variablen wie *Mathematik-*, *Englisch-*, *Ausbildungs-* und *Fächerschwerpunktnote* nur geringe Fallzahlen vorhanden. Bei den *CV-Variablen* dürfte zudem die Vorauswahl durch die Personalabteilung zusätzlich zu Varianzeinschränkungen führen. Lediglich bei der *Abiturnote* ist von der Normalverteilungsannahme auszugehen. Alle anderen Variablen weichen signifikant von dieser ab, z.T. bedingt durch wenige Kategorien (z.B. Anzahl der Universitätsabschlüsse), so dass hier im weiteren Verlauf nonparametrische Tests ihre Anwendung finden.

Tabelle 25: Deskriptive Kennwerte nicht-dichotomer CV-Variablen

CV-Variablen	N	MW	Med.	Min.	Max.	SD	NV (j/n)
Eintrittsalter	206	29	28	23	40	3	n
Abinote	195	1.9	1.9	1.0	3.2	.5	j
Note_FSP	159	2.0	2.0	1.0	4.0	.7	n
Note_D	151	2.4	2.0	1.0	4.0	.8	n
Note_M	104	2.2	2.0	1.0	5.0	.9	n
Note_E	98	2.2	2.0	1.0	5.0	.9	n
Note_Aus	44	1.7	1.9	1.0	2.8	.5	n
Anz_Absch_tot	204	1	1	1	3	1	n
Pra_N	195	5	5	1	12	2	n
Pra_t	195	20.4	17.0	.0	63.0	12.7	n
Pra_N_UB	195	1	1	0	4	1	n

N= Stichprobenumfang; MW=Mittelwert; Med.=Median; Min.=Minimum; Max.=Maximum; SD=Standardabweichung; NV=Normalverteilungsannahme

Die Boxplots und Stem-and-Leaf-Plots verdeutlichen sowohl bei den *CV-* als auch bei den *Recruitingvariablen* eine große Anzahl von Ausreißerwerten, die eine Eliminierung der Fälle wenig sinnvoll erscheinen lässt, da diese, je nach Variable, unterschiedliche Personen betreffen. Neben dem relativ kleinen Stichprobenumfang einzelner Variablen, kann hier auch die Neigung zu Extremurteilen bei den betreffenden Beratern ursächlich sein.

3.3.1.3 Reliabilitäts- und Beurteileranalyse

Bei der Reliabilitätsanalyse wurde auf Basis der Einzelurteile in den Interviews und den Beobachtungen als Übereinstimmungsmaß der *Intra-Class-Correlation-Koeffizient* (*ICC*) mit Einweg-Zufallseffekten berechnet. Ziel war dabei, die Urteilsübereinstimmung je Dimension festzustellen.

Tabelle 26: ICC-Koeffizienten mit Einweg-Zufallseffekten nach Recruitingvariablen

Untersuchte Übungen	Beobachtung[a]			Interview[a]			Overall[a]			Overall		
Recruitingvariablen	ICC	k	N	ICC	k	N	ICC	k	N	ICC	k	N
Konflikt- und Kritikfähigkeit	.293**	3	47	-.-	-.-	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Kommunikationsstärke/ Eloquenz	.407**	3	47	.304**	3	31	.147*	6	16	.188**	6	153
Fachkenntnisse	.335**	3	47	-.-	-.-		-.-	-.-	-.-	-.-	-.-	-.-
Strukturiertes/ logisch-analytisches Denken	.383**	3	47	.134	3	31	.154*	6	16	.156**	6	99
Intellektuelle Neugier	.286**	3	47	.049	3	31	.107	6	16	.133**	6	74
Ausstrahlung/Charisma	.347**	3	47	.346**	3	31	.214**	6	16	.169**	6	111
Methoden-Kenntnisse	.411**	3	47	-.-	-.-		.144*	6	16	.193**	6	36
Team-/ Integrationsfähigkeit	-.-	-.-	-.-	.258**	3	31	-.-	-.-	-.-	-.-	-.-	-.-
Pragmatismus/ Ergebnisorientierung	-.-	-.-	-.-	-.91	3	31	-.-	-.-	-.-	-.-	-.-	-.-
Frustrationstoleranz	-.-	-.-	-.-	.236*	3	31	-.-	-.-	-.-	-.-	-.-	-.-
Kreativität/ Flexibilität	-.-	-.-	-.-	.089	3	31	-.-	-.-	-.-	-.-	-.-	-.-
Unternehmerisches Denken	-.-	-.-	-.-	.199*	3	31	-.-	-.-	-.-	-.-	-.-	-.-
BWL-Kenntnisse	-.-	-.-	-.-	.352**	3	31	-.-	-.-	-.-	-.-	-.-	-.-

N= Anzahl beurteilter Personen; k= Anzahl der Beurteiler
[a] gleich bleibende Stichprobe über alle Dimensionen
* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant

Tabelle 26 zeigt die Ergebnisse der Reliabilitätsuntersuchung nach Recruitingvariablen. Insgesamt sind die Konsistenzmaße der Urteile bei *Beobachtung* und *Interview* ernüchternd und erreichen in keinem Fall Effekte >.50. Dabei zeigt sich, dass die Urteilsübereinstimmung bei der *Beobachtung* durchgängig besser ist als bei den *Interviews.* Sie haben eine mittlere Höhe von r=.30 bis r=.40. Bei *Kommunikationsfähigkeit* und *methodischen Kenntnissen* zeigen sich bei den genannten Übungen, in denen die Bewerber Fallstudien bearbeiten und präsentieren, die höchsten Übereinstimmungen. Ebenso relativ hoch ist diese für *Charisma* und *intellektuelle Neugier* sowie *logisch-analytisches Denken*. Schließlich gibt es auch eine, relativ gesehen, mittlere Übereinstimmung der Urteile über das *fachliche Wissen*. Der höchste Übereinstimmungswert zeigt sich in den *Interviews* bei *BWL-Kenntnissen*. Bei der *Kommunikationsfähigkeit* gibt es auch mittlere Übereinstimmungen in den Einzelinterviews. Die *Teamfähigkeit* wird gängigerweise in dem persönlichkeitsbezoge-

nen Teil des Interviews beurteilt (klassische Überlegung: „Könnten Sie sich vorstellen, mit dem Bewerber 24h unter Hochdruck zusammenzuarbeiten?“) und zeigt eher schwache Übereinstimmungen. Keine einheitlichen Urteile finden sich zu den Urteilsdimensionen *logisch-analytisches Denken*, *intellektuelle Neugier* und *Kreativität*. Bei der Betrachtung der Dimensionen, die mit beiden Methoden beurteilt werden (*Interview* und *Beobachtung*) zeigt sich, dass die Übereinstimmung der Urteile weiterhin sinkt.

Insgesamt entsprechen die Ergebnisse zur Reliabilität der angewandten Instrumente den Erwartungen aus *Fragestellung (1)*. Die beobachteten Übereinstimmungsmaße bei den Einschätzungen auf verschiedenen Dimensionen müssen allerdings vor dem Hintergrund zahlreicher weiterer Prämissen interpretiert werden: Grundlage des Übereinstimmungswertes pro Urteilsdimension sind über die Instrumente (Beobachtung bzw. Interview) jeweils unterschiedliche Bewerber und Bewerberanzahlen, keine gleichen Interviewer bzw. Beobachter für jeden Bewerber und keine einheitlichen standardisierten Aufgaben. So kommt es zu einer Konfundierung der Effekte, die auf die Person des Bewerbers, den Beurteiler, die Situation und die durchgeführte Aufgabe bzw. Das Interview sowie deren Interaktionen zurückzuführen sind (siehe auch McGraw & Wong, 1996). Dennoch sind die Maße in ihrer Höhe innerhalb eines Instrumentes interpretierbar, da für die Interpretation der Genauigkeit des Urteils, die Höhe der einzelnen potenziellen Fehlerquellen (z.B. bezogen auf die Person des *Raters* σ^2_{ZwS}, σ^2_{ZwP} und σ^2_{PxS}) zunächst irrelevant ist.

Die Werte für das *Interview* liegen mit r=.20 bis r=.30 in dem Bereich der berichteten Objektivitätswerte *offener Interviews* (Schuler, 2002). Auf Grund des teilstrukturierten Charakters der Interviews wären leicht höhere Werte (zwischen r=.30 und r=.50) zu erwarten gewesen. Das Fehlen einer kompletten Interviewer-Bewerber-Matrix mit den damit verbundenen methodischen Restriktionen lässt die etwas geringeren Werte allerdings plausibel erscheinen. Zudem kann die *hohe Komplexität* der beurteilenden Leistung (keine einfachen Tätigkeiten) ursächlich für die niedrigen Werte sein (Kompa, 1989; Trost, 1996). Des Weiteren haben die Interviewer auch die Gelegenheit, sich einen Allgemeineindruck über den Bewerbers zu bilden, was sich ungünstig auf die Reliabilität von zu beurteilenden Einzeldimensionen auswirken dürfte (Kiker & Motowidlo, 1998). Vor dem Hintergrund, dass, ähnlich einer mündlichen Prüfungssituation, im *Interview* tatsächlich *BWL-Kenntnisse* in Form von betriebswirtschaftlichen Problemstellungen abgeprüft werden, ist die höchste Ausprägung auf dieser Variablen (r=.352) beim *Interview* erklärbar. Offensichtlich reicht dies aber nur aus, um sehr schwache Übereinstimmungen bei der Urteilsfindung zum *unternehmerischen Denken* (r=.199) zu finden, das allerdings eine uneindeutig definierte Dimension darstellt.

Im Gespräch findet, oftmals unter Stressbedingungen, ein mündlicher *Informationsaustausch* statt, der den Interviewern eine Möglichkeit bietet, die *kommunikativen Fähigkeiten* (r=.304) und die *Frustrationstoleranz* (r=.236) des Gegenübers zu bewerten. Höhere Übereinstimmungen sind, da diese Fähigkeiten gut beobachtbar sind, deshalb nicht ungewöhnlich (siehe Barrick, et al., 2000). Die Dimension *Charisma* hingegen wird, wie weiter unten gezeigt, vermutlich als allgemeines Konstrukt oder *Urteilsaggregat*, über die gesamte Interviewzeit (im Sinne eines Gesamteindruckes) gebildet (r=.346). Obwohl keine *Teamarbeit* direkt beobachtet wird, scheint es trotzdem schwache Übereinstimmungen zwischen den Interviewern zu geben (r=.258), die zumindest teilweise ein einheitliches Teamfähigkeits-Verständnis im Unternehmen vermuten lassen. Die fehlende Übereinstimmung bei den Urteilsdimensionen *logisch-analytisches Denken* und *intellektuelle Neugier* (beide nicht signifikant), ist insofern verwunderlich, da in den *Interviews* die angesprochenen "brain-teaser" angewandt werden, zu deren Lösung gerade diese Fähigkeiten erforderlich sind. Zudem werden intelligenzaffinen Dimensionen die besten Reliabilitäten in Interviews zugesprochen (Mayfield, 1964; Wagner, 1949). Die Vermutung liegt nahe, dass der Charakter der Aufgaben sich interviewerspezifisch erheblich unterscheidet und kein einheitliches *Konstruktverständnis* bei den Interviewern vorliegt. Ein gleichartiges Phänomen findet sich vermutlich bei dem Begriff der *Kreativität*. Auch hier gibt es keine Übereinstimmung der Urteile.

Auch die Beobachtungswerte erreichen geringe *Reliabilitäten*. Die insgesamt zum Interview relativ höheren Werte liegen vermutlich in dem, im Vergleich zum *Interview standardisierten* Prozess begründet und entsprechen den Erwartungen. Der Ablauf ist zeitlich und inhaltlich exakt getaktet und die vorgegebenen Fallstudien haben alle einen ähnlichen Charakter. Die Interaktion mit dem Beurteiler findet in nur geringem Maße statt und alle Beobachter beurteilen die exakt gleiche Situation bei jedem Bewerber. Deswegen ist eine relativ höhere Übereinstimmung plausibel, da die mögliche ungewünschte Varianz (σ^2_{ZwS}) weitaus geringer ist als beim *Einzelinterview*. Die relativ hohen Werte von *Kommunikationsstärke* (r=.407) und *Methodenkenntnissen* (r=.411) können darin begründet sein, dass bei der *Fallvorbereitung* und der *Präsentation* gerade diese am besten zum Ausdruck kommen. Ebenso gilt dies für *Charisma* (r=.347) und *intellektuelle Neugier* (r=.286) sowie *logisch-analytisches Denken* (r=.383), die sich bei der Fallbearbeitung zeigen dürften. Hier haben die Beobachter die Gelegenheit, den Bewerber bei der Suche nach Lösungsansätzen zu bewerten. Schließlich gibt es auch eine mittlere Übereinstimmung der Urteile über das *fachliche Wissen* (r=.335), das Voraussetzung ist, um die *Problemstellung* zu bearbeiten und zu lösen. Schwächen dürften hier deshalb leichter zu erkennen und zu beurteilen sein.

Die geringe Höhe signifikanter Übereinstimmungen bei der Betrachtung der Dimensionen, die mit beiden Methoden beurteilt werden (*Interview* und *Beobachtung*) kann am Zufallsfehler und der geringen Teststärke durch die kleinere Stichprobe liegen (N=16). Die Berechnung der *Overall-Konsistenz* mit maximale Fallzahl zeigt aber auch hier keine nennenswerten Konsistenzerhöhungen. Allerdings ist hier die eingangs erklärte *Effektkonfundierung* wirksam, da es sich nicht um die gleiche Stichprobe handelt. Der Schluss liegt nahe, dass im Gegensatz zur allgemeinen Erkenntnis, dass *MTMM-Ansätze* die Validität erhöhen, hier durch Messung mit Instrumenten, die eine geringe *Konstruktvalidität* aufweisen, die *Reliabilität* der gemessenen Werte möglicherweise unterschätzt wird.

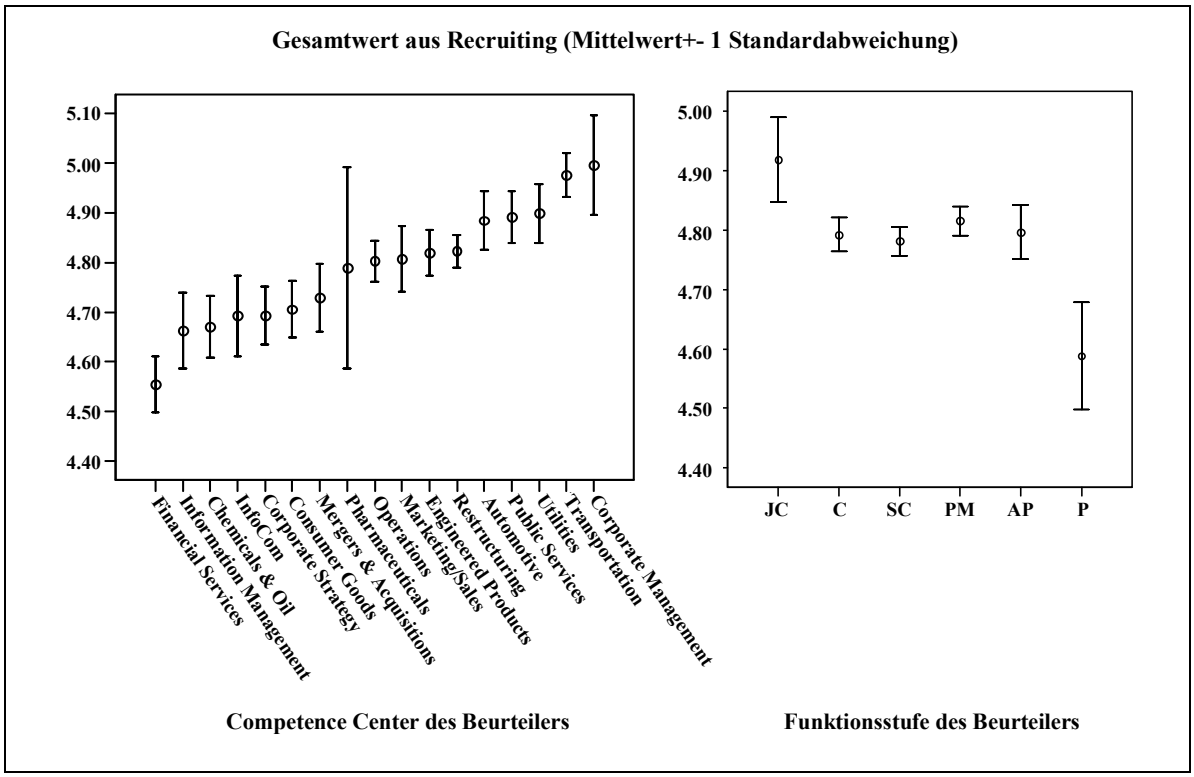

Abbildung 40: Urteilsstrenge nach Funktionsstufe und Competence Center

Der Vergleich der *Beurteilerstrenge* nach Stufenzugehörigkeit des Beurteilers ergibt, dass *JCs* am besten und *Ps* am strengsten im Recruiting beurteilen (*Abbildung 40*). Recruiter mit einer höheren Funktionsstufe sind u.U. besser in der Lage Entwicklungsmöglichkeiten des Gegenübers zu beurteilen und sind dadurch eher bereit, strengere Urteile zu fällen. Allerdings findet hier ebenfalls eine Konfundierung von Effekten statt, da die Varianzanteile, die auf die beurteilte Person, auf den Beurteiler und auf dessen Stufe zurückzuführen sind, sich nicht trennen lassen, weil nicht für alle Personen von allen Beurteilern auf allen Stufen Urteile vorhanden sind. Außerdem ist von mehreren Beurteilern mehr als ein Urteil in dem

Datensatz enthalten. Zudem sind für die *JC-Stufe* und die *P-Stufe* nur sehr geringe Fallzahlen vorhanden, da diese am Recruitingprozess nur selten teilnehmen. Dennoch lassen sich Indizien für unterschiedliche Beurteilungsmaßstäbe gewinnen, da die Standardfehler der Mittelwerte teilweise überschneidungsfrei sind.

Außerdem scheint es CC-spezifische Bewertungsmaßstäbe zu geben. Betrachtet man den Gesamtwert, so sind die schlechtesten Gesamtbeurteilungen bei den *CCs Financial Services*, *Information Management*, *Corporate Strategy* und *Consumer Goods* vorzufinden.

3.3.1.4 Profilanalyse

Bei den Profilanalysen wurden die Personengruppen getrennt nach *Einstiegslevel als JC* (ja/nein) anhand von *CV-Daten* gruppiert und ihre Ausprägungen auf den Recruitingdaten über Profildiagramme visualisiert. Insgesamt wurden Profilvergleiche in Abhängigkeit von sieben Variablen durchgeführt. *Tabelle 23* zeigt ihre Häufigkeiten und *Abbildung 41* und *42* exemplarische Profilvergleiche.

Die Profilvergleiche ergaben, dass es unterscheidbare Muster bei der Einschätzung auf den Variablen bei den auf Basis der *CV-Daten* gebildeten Gruppen gibt. Somit konnten die erwarteten charakteristischen Profile aus *Fragestellung (2)* teilweise auch in der vorliegenden Untersuchung aufgezeigt werden. Bei dem Vergleich der *Geschlechterprofile* zeigen sich Unterschiede in einem sehr geringen Ausmaß. So haben weibliche Bewerber beim *JC-Recruiting* ein tendenziell schlechteres *BWL-Know-how*. Besser als ihre männlichen Bewerberkollegen schneiden sie auf den Dimensionen *Frustrationstoleranz, Kreativität* und *Charisma* ab. Da die Quote der wirtschaftswissenschaftlichen Absolventen unter Frauen und Männern gleich ist, kann bei *BWL-Kenntnissen* ein Studiengangeffekt ausgeschlossen werden. Allerdings muss hier der kleine Frauenanteil der Stichprobe (N=12) berücksichtigt werden. Beim *SC-Recruiting* gibt es geschlechtsspezifische Unterschiede bei *Teamfähigkeit*. Diese ist bei den Frauen ebenso wie die *Kommunikationsfähigkeit* und *Kreativität* stärker ausgeprägt. Hier werden also tendenziell die Ergebnisse aus der *JC-Einstiegskohorte* repliziert. Bei den *analytischen Fähigkeiten* sind Frauen allerdings leicht besser als ihre männlichen Kollegen, obwohl, analog zu den Erkenntnissen von Lubinski (Lubinski, Benbow, et al., 2001; Lubinski, Webb, Morelock & Benbow, 2001) eher bessere Werte bei den männlichen Bewerbern zu erwarten gewesen wären. Durch den Selektionseffekt ist anzunehmen, dass sich überwiegend solche Personen bei einer Unternehmensberatung bewerben, die sehr analytisch geprägt sind, und somit, auch auf Grund des kleineren Frauenanteils (N=36), in der Stichprobe keine genügend hohe Differenzierung für eine Generalisierung vorliegt. Darüber

hinaus stehen die Ergebnisse im Einklang mit den genannten wissenschaftlichen Untersuchungen, die Stärken der Frauen eher im *verbalen*, *kreativen* und *sozialen* Bereich und weniger im analytischen Bereich attestieren (2004; Achter, et al., 1999; Shea, et al., 2001; Webb, et al., 2002).

Tabelle 27: Variablen der Profilvergleiche und ihre Häufigkeiten

Gruppierungsvariablen	Ausprägung	Einstieg als JC	Einstieg als C oder höher
Geschlecht	Weiblich	12	8
	Männlich	66	48
Erzwungenes Ausscheiden	Nein	69	47
	Ja	9	9
Angestelltenstatus	Angestellt	56	29
	Ausgeschieden	22	27
MBA-Abschluss	Nein	71	50
	Ja	6	5
Dr.-Titel	Nein	75	38
	Ja	2	17
BWL- Studium	Nein	24	20
	Ja	53	35
Universitätsart	Privat	56	47
	Öffentlich	21	8

Personen, deren Abgang aus dem Unternehmen im Laufe der Tätigkeit *forciert* wurde, haben im *JC-Recruiting* geringere *Frustrationstoleranz*- und *Charisma*-Werte (*Abbildung 41*). Dies ist plausibel, da die Beratertätigkeit *Stressresistenz*, *Ausdauer* und somit hohe *Frustrationstoleranz* erfordert. Eine Kreuzvalidierung könnte zeigen, ob die angewandte Personalentwicklung u.U. keine Verbesserung auf diesen Dimensionen erreicht. Beim Begriff *Charisma* liegt es, wie angesprochen, nahe, dass jeder Beurteiler darunter ein Aggregat an positiven Eigenschaften (*Kommunikationsstärke*, *Sympathie*, *gutes Auftreten* u.a.) versteht und dieses mit einem forcierten Ausscheiden zusammenhängt. Darüber hinaus scheinen im *JC-Recruiting forciert* ausgeschiedene Mitarbeiter ihre fachlichen und persönlichen Schwächen durch *Kreativität* auszugleichen, ein Aspekt, der durch das kompensatorische Vorgehen begünstigt wird. Dass gerade diese Personen als besonders *kreativ* eingeschätzt wurden, ist dennoch zunächst überraschend. Die *JC-Tätigkeit* ist mit einer Art Ausbildungsphase zu vergleichen, in der das Aufgabenspektrum nur begrenzt Raum für kreative Entfaltung zulässt. Eine Erklärung könnte deshalb sein, dass kreative *JCs* in ihrer Tätigkeit als *JC* nicht in ausreichendem Maße kreativ gefördert werden und diese somit schnell frustriert sind. Dies kann

wiederum zu Konflikten und letztlich zum Ausscheiden aus der Firma führen. Beim *SC-Recruiting*, bei dem das spätere Aufgabenspektrum sehr wohl *Kreativität* abverlangt, scheinen sich beide Gruppen (*Forced Exit* vs. *kein Forced Exit*) nur noch auf wissensrelevanten Variablen zu unterscheiden. Diese Schwächen scheinen insbesondere durch soziale Aspekte (*Teamfähigkeit*, *Konfliktfähigkeit* und *Kommunikationsfähigkeit*) ausgeglichen zu werden.

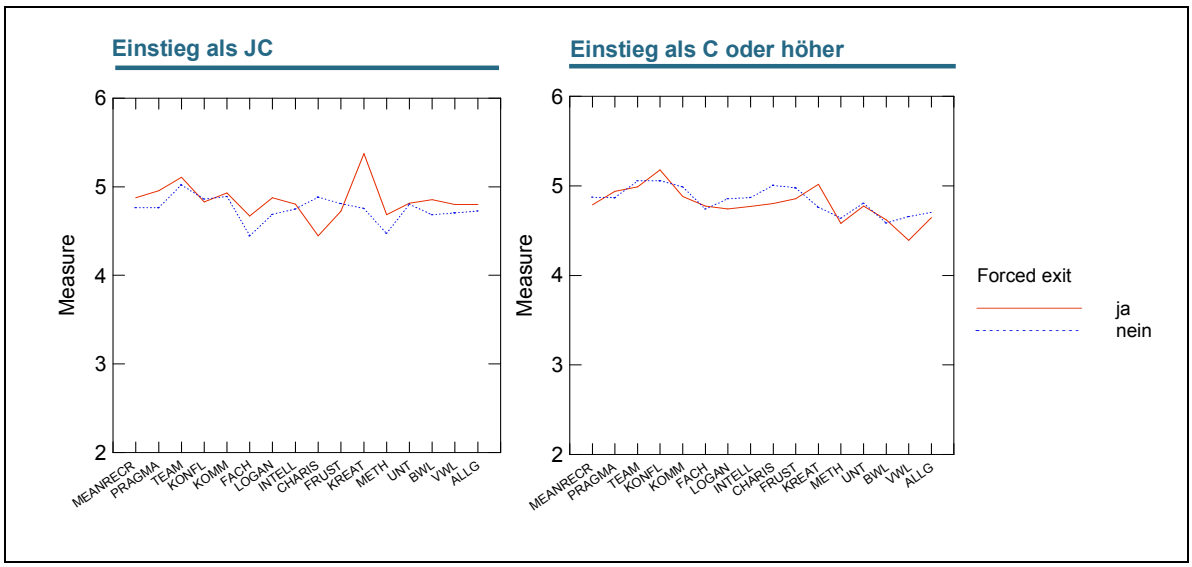

Abbildung 41: Profilvergleiche am Beispiel *Forced Exit*

MBA-Absolventen werden insgesamt genauso besser eingestuft wie Personen mit einem Dr.-Titel, was mit ihrem höheren Ausbildungsstand einhergeht (*Abbildung 42*). Da das *MBA-Studium* sehr teamarbeits- und fallstudienorientiert ist, sind auch die höheren Einstufungen auf den *sozialen Dimensionen* (*Teamfähigkeit, Konfliktfähigkeit* und *Kommunikationsfähigkeit*) und *Pragmatismus* zu erklären. Das geringere *Fach-Know-how* kann daran liegen, dass die meisten Master-Studiengänge im Ausland absolviert wurden und die Erwartungshaltung des Interviewers von deutschen Fachkenntnissen (z.B. Bilanzierungsrichtlinien) ausgeht. Bei promovierten Bewerbern wäre es möglich, dass der Dr.-Titel einen Beurteilerfehler auslöst (selektive Wahrnehmung, *illusory correlation*), der dazu führt, dass mehr Informationen aufgenommen werden, die Stereotype bestätigen, als ihnen zu widersprechen (z.B. „Wenn jemand promoviert hat, muss er logisch-analytisch und methodisch gut sein“). Beim *JC-Recruiting* zeigen sich beim *unternehmerischen Denken* und den *BWL-Kenntnissen* Unterschiede in Abhängigkeit des Studiums (BWL (ja/nein)). Diese sind in ihrer Ausprägung allerdings kleiner als erwartet. Zum Einen sind die Anforderungen im Recruiting bezüglich detailliertem Wissen nicht so hoch, zum Zweiten wird vermutlich gerade bei den Nicht-BWLern darauf geachtet, dass ein Basiswissen vorhanden ist. Folglich sind die Werte

in der Selektionsstichprobe nicht stark genug differenziert. Beim *SC-Recruiting* werden Bewerber mit einem *BWL-Studium*, bei *BWL-* und *VWL-Kenntnissen* ebenso besser bewertet, eine Erkenntnis, die nicht überraschend ist. *Nicht-BWLer* gleichen ihr fehlendes Wissen mit ihrer Persönlichkeit aus (höhere *Charisma*-Werte). Keine nennenswerten Unterschiede gibt es bei den Profilen *ausgeschiedener* im Vergleich zu *nicht ausgeschiedenen* Mitarbeitern sowie bei der *Universitätsart,* in der das Studium absolviert wurde (*privat* vs. *öffentlich).*

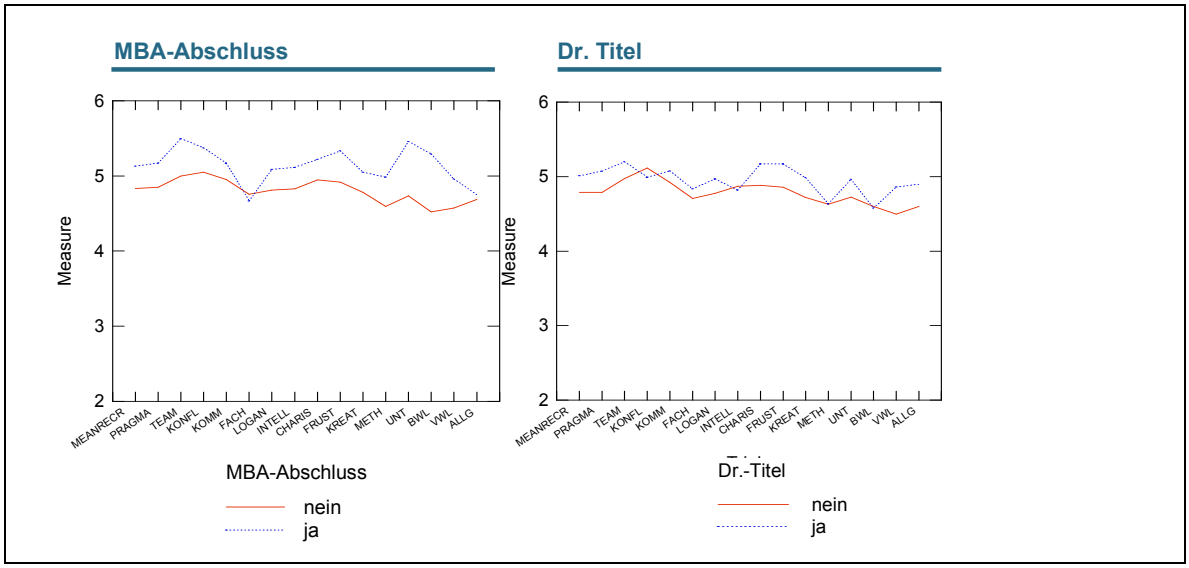

Abbildung 42: Profilvergleiche am Beispiel *MBA-Abschluss* und *Dr.-Titel*

Darüber hinaus wurde in einem Vergleich von verbalen und analytischen Fähigkeiten festgestellt, dass Personen mit hohen *analytischen Fähigkeiten* (als Kenngröße für „*quantitative-numerical abilities*“) oder mit hohen *kommunikativen Fähigkeiten* (als Kenngröße für „*verbal abilities*“) auf der jeweils anderen Dimension signifikant schwächer sind (*tilted profiles*). Sie gleichen somit ihre Defizite aus. *Abbildung 43* verdeutlicht die Vergleiche der jeweils unteren und oberen Quartile der *analytischen* und *kommunikativen Fähigkeiten*. Ein anschließender *t-Test* für abhängige Stichproben zum Vergleich der beiden Dimensionen in den oberen Quartilen ergab zudem hoch signifikante Mittelwertunterschiede (*Tabelle 28*). Damit konnten die erwarteten gekippten Profile (*tilted profiles*) bei den Bewerbern repliziert werden. Bei der Interpretation ist allerdings zu berücksichtigen, dass die Unterschiede teilweise auch durch Regressionseffekte bedingt sein können.

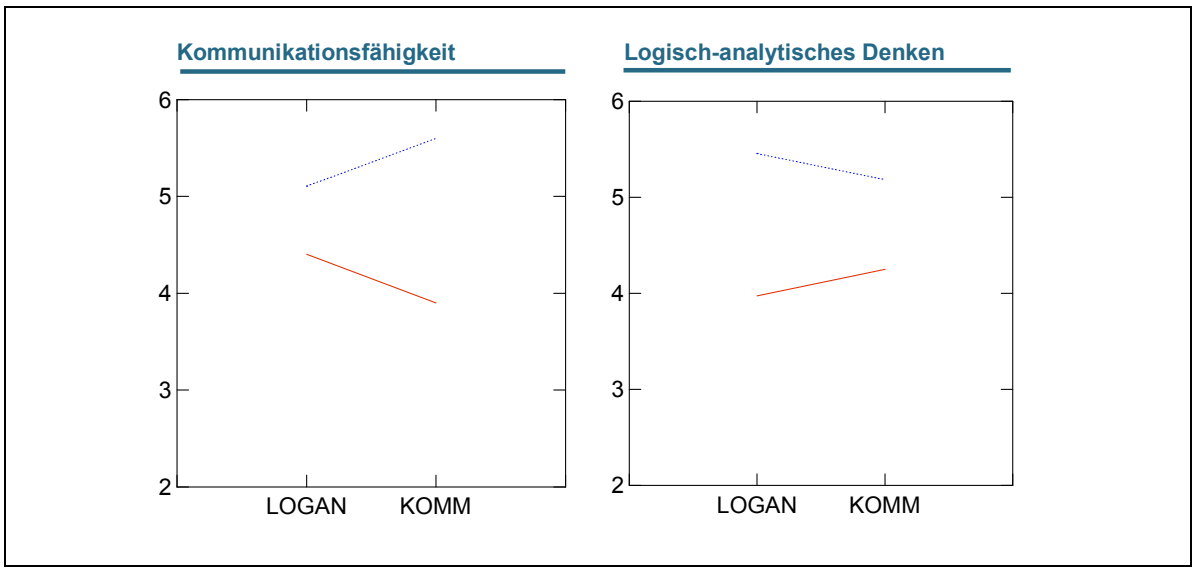

Abbildung 43: 25%-Randquartile kommunikativer und logisch-analytischer Fähigkeiten (N=26)

Tabelle 28: t-Test für abhängige Stichproben zum Vergleich der oberen Quartilwerte auf den Variablen Kommunikationsfähigkeit und logisch-analytisches Denken (N=26)

	Variable des betrachteten obersten Quartils	
	Komm	LogAn
MW Komm	5.589	5.244
MW Logan	5.199	5.515
MW Diff.	.390	-.291
95% Konfidenzintervall	.248 bis .531	-.435 bis -.148
df	25	25
p	<.001	<.001

N= Anzahl Personen im obersten Quartil der Variablen; MW=Mittelwert; df= Freiheitsgrade; p= Signifikanz des t-Tests (zweiseitig)

Des Weiteren zeigen sich im exemplarischen Profilvergleich zweier Bewerber mit identischem *Recruiting-Gesamtdurchschnitt* die Tücken des *kompensatorischen* Verfahrens. Bewerber mit schwachen Werten auf einigen Dimensionen können diese mit besseren Werten auf anderen ausgleichen (*Abbildung 44*).

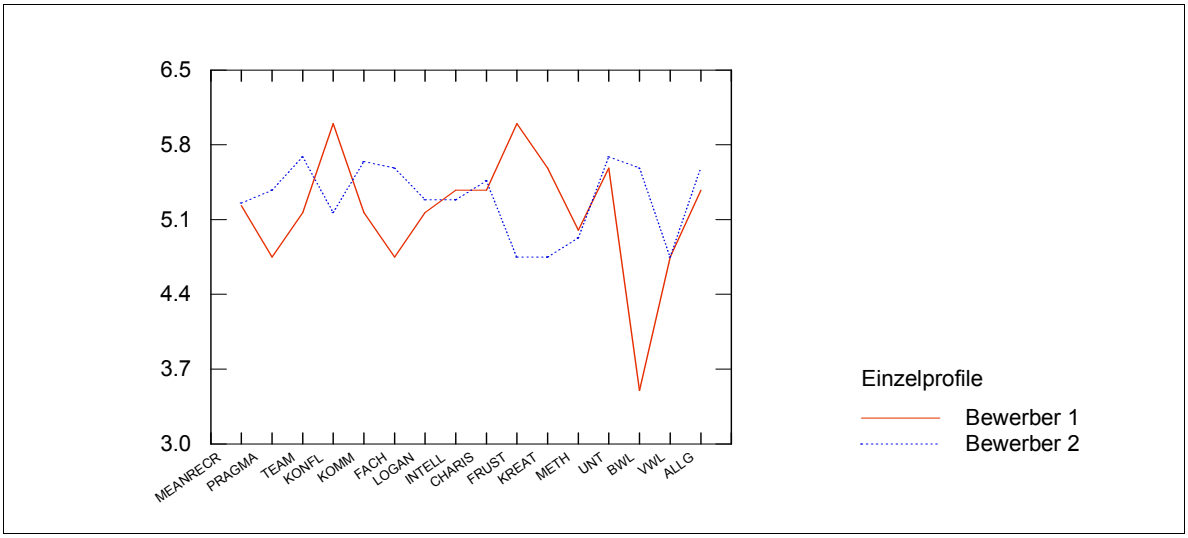

Abbildung 44: Profilvergleich von zwei Bewerbern bei identischem Recruiting-Gesamtwert

3.3.1.5 Faktorenanalyse

Im Rahmen dieses Analyseabschnitts wurde zunächst eine Faktorenanalyse der Prädiktorenseite (Hauptkomponentenanalyse; VARIMAX mit anschließender Rotation) mit den Recruitingvariablen für die Kohorten „Einstieg als JC“ durchgeführt. Auf eine faktoranalytische Untersuchung der Kohorte „kein Einstieg als JC“ wurde verzichtet, da auf Grund der kleinen Stichprobe (N=49) nicht von einer robusten Faktorenstruktur auszugehen war (siehe auch Gorsuch, 1983). Darauf folgend wurden die extrahierten Faktoren hinsichtlich ihrer Variablenstruktur inhaltlich interpretiert und ihr Zusammenhang mit der Ausscheidungsart und *forciertem* Abgang untersucht. Die faktoranalytischen Ergebnisse zeigt *Tabelle 29*.

Dabei gibt es einzelne Variablen, die auf mehreren Faktoren hoch laden, ohne dass eine eindeutige Beziehung zu erkennen ist. Dennoch lassen sich Tendenzen bei der Faktorenbildung interpretieren. So gibt es sowohl bei der *rotierten* als auch bei der *unrotierten* Lösung, einen *Allgemeinfaktor,* auf dem nahezu alle Variablen relativ hoch laden. Der rotierte *Allgemeinfaktor* beeinflusst bei *JC-Recruiting-Teilnehmern* am stärksten die *Kreativität*, die *BWL*-, die *VWL-Kenntnisse* und das *Allgemeinwissen*. Die Bedeutsamkeit des allgemeinen Faktors bei der Konstruktvalidität von ACs (allgemeine Intelligenz, Globalurteil, *Fragestellung (3)*) konnten folglich faktoranalytisch deutlich repliziert werden. Auch Scholz und Schuler (1993) kommen zu dem Schluss, dass das Hauptkonstrukt, das mit Hilfe von *ACs* gemessen wird, die allgemeine Intelligenz (*g*) ist. Das bedeutet, dass Personen mit hohen Werten auf diesem Faktor, auf den hoch ladenden Variablen besser abschneiden als der Rest. Bei den generellen Dimensionen des *Allgemeinwissens* und *VWL-Kenntnissen* als ebenso

allgemeinem wirtschaftlichem Know-how, scheint dies auf Anhieb plausibel. Bei *BWL-Kenntnissen* drängt sich die Vermutung auf, dass, wie bei den Profilvergleichen bereits angesprochen, im *JC-Recruiting* die Anforderungen an die *BWL-Kenntnisse* durch allgemeine Intelligenz größtenteils erfüllt werden können. Spezielles Wissen ist nur in einem geringerem Maße erforderlich. Dass im *JC-Recruiting* die BWL-Problemstellungen ein eher allgemeines Niveau haben, steht mit diesem Ergebnis im Einklang. Bei der Variablen *Kreativität* wurde bereits im Rahmen der *Reliabilitätsanalyse* angesprochen, dass dieser Begriff wenig einheitlich gehandhabt wird. Auf Grund der schwierigen Definition ist auch hier anzunehmen, dass die Beurteiler sich bei der Beurteilung des *Kreativitätsaspektes* auf Grund mangelnder Messinstrumente und mangelndem Verständnis eher an allgemeinen Eigenschaften wie der *Intelligenz* bzw. einem allgemeinen Eignungsfaktor orientieren.

Tabelle 29: Rotierte und unrotierte Komponentenmatrizen der Recruitingvariablen

	Rotierte Komponentenmatrizen[a]			**Unrotierte Komponentenmatrizen[b]**		
(N=70)	1	2	3	1	2	3
Pragma	.569		.353	.654		
Team	.504	.600		.615		.419
Konfl		.680	.351	.470	.569	
Komm	.445	.714	.228	.786	.347	
Fach			.850	.469		-.711
LogAn	.412		.691	.777		
Intell	.369	.495	.473	.747		
Charis		.833		.483	.675	
Frust	.483	.433		.586		
Kreat	.713			.601		
Meth	.454		.755	.790		
Unt	.563		.325	.611		
BWL	.711		.401	.740	-.351	
VWL	.670			.726		
Allg	.746			.568	-.340	
VA[c]	25.7	43.8	61.4	42.4	52.3	61.4

Faktor unrotiert	1	2	3	4	5
Eigenwert	6.361	1.485	1.359	.925	.871

a Kohorte „Einstieg als JC“; Extraktionsmethode: Hauptkomponentenanalyse. Rotationsmethode: Varimax mit Kaiser-Normalisierung. Die Rotation ist in 6 Iterationen konvergiert.
b Extraktionsmethode: Hauptkomponentenanalyse.
c Varianzaufklärung, kumulativ in Prozent dargestellt.
Dargestellt sind nur Faktorladungen mit Betrag >.300.

Der zweite rotierte Faktor beinhaltet bei der *JC-Recruiting-Stichprobe* größtenteils *soziale Komponenten*. Es ist anzunehmen, dass Personen mit einer hohen *Sozialkompetenz*

kommunikativer, *teamfähiger* und *charismatischer* sind. Darüber hinaus können sie auf Grund ihrer höheren *Sozialkompetenz* auch besser mit Konflikten umgehen. Entsprechend sind die Faktorladungen auf diesen Variablen am höchsten. Hier zeigt sich auch ein schwacher Zusammenhang zu der Variablen *Forced Exit* (r=-.202) (*Tabelle 30*). Demnach trennt sich das Unternehmen *forciert* tendenziell von solchen Mitarbeitern, die beim *JC-Recruiting* eine niedrigere Sozialkompetenz aufweisen. Der dritte *Problemlösungsfähigkeits*-Faktor deutet darauf hin, dass Bewerber, die hier hoch eingestuft sind, gute *Fach*- und *Methodenkenntnisse* und gute *logisch-analytische Fähigkeiten* besitzen. Diese Kombination dürfte die Lösung gestellter Aufgaben und Probleme erheblich erleichtern.

Die unrotierte Faktorlösung zeigt, dass der erste Faktor das *Gesamtlevel* des Recruitingergebnisses abbildet. Das Ergebnis spiegelt, durch die Extraktion eines Allgemeinfaktors, die in zahlreichen Studien gefundene geringe diskriminante und damit beklagte mangelnde *Konstruktvalidität* von *ACs* wider (z.B. Sackett & Dreher, 1982). Dies kann allerdings mit dem vorliegenden Datensatz nicht mit einer *MTMM-Matrix* verifiziert werden. Beurteiler und Beobachter füllen im vorliegenden Auswahlverfahren ihre Ratingbögen *nach* den Interviews und Beobachtungen aus, was eine Orientierung an einem *Globalurteil* und somit in der angewandten Faktorenanalyse die Extraktion eines *Globalfaktors* begünstigt. Russel (1985) kam in einer explorativen Faktorenanalyse von *AC-Ratings* zu vergleichbaren Ergebnissen. Vermutlich werden außerdem im vorliegenden Fall eher Leistungen und weniger Eigenschaften beurteilt (siehe Cook, 1998, S.186ff.). Dies ist oft dann der Fall, wenn die Assessoren entweder zu viele Dimensionen oder zu viele Bewerber zur Einstufung bekommen. Dies führt zu einem Informationsüberfluss, der die Urteiler zwingt, sich beim Urteil an der *Gesamtleistung* des Bewerbers zu orientieren (Kleinmann & Koller, 1997). Die Vermutung liegt nahe, dass dies, bei der Vielzahl zu beurteilender Dimensionen, auch auf den vorliegenden Fall zutrifft. Zudem ist die Bezeichnung einiger Dimensionen (z.B. *Kreativität* oder *Charisma*) sehr generell und nicht direkt in den Übungen beobachtbar, so dass auch hier die Bildung von Globalurteilen begünstigt wird (Kleinmann, Exler, Kuptsch & Koller, 1995). Die zweite unrotierte Komponente der *JC-Einsteiger* gibt die *Profilstruktur* auf den Dimensionen (*verbal-numerisches* Kontinuum) wieder. Hier zeigt sich auch ein schwacher Zusammenhang zu der Variablen *Forced Exit* (*Tabelle 30*). Demnach trennt sich das Unternehmen *forciert* tendenziell von solchen Mitarbeitern, die beim JC-Recruiting eine niedrigere *Sozialkompetenz* aufweisen (r=-.264). Es wird der mit den Faktorwerten der *rotierten Lösung* gefundene Zusammenhang erwartungsgemäß repliziert. Der dritte unrotierte *JC-Recruiting-Faktor* beinhaltet Problemlösungsfähigkeitskomponenten.

Tabelle 30: Korrelation der Faktoren mit *Angestelltenverhältnis* und *Forced Exit*

Kohorte „Einstieg als JC“ (N=70)			**r** (Spearman-Rho)	
rotiert (j/n)	Faktor	Faktorbezeichnung	Angestelltenverhältnis	Forced Exit
j	1	Allgemeinfaktor	-.059	.120
j	2	Sozialkompetenz	-.072	-.202*
j	3	Problemlösungsfähigkeit	.135	.122
n	1	Levelfaktor	-.044	.071
n	2	Profilfaktor (verbal-numerisch)	-.039	-.264*

* Die Korrelation ist auf dem 0.05 Niveau signifikant (einseitig)
Bei den unrotierten Faktoren sind lediglich die ersten zwei inhaltlich interpretierbaren Faktoren dargestellt

3.3.1.6 Logistische Regressionsanalyse

Im vorliegenden Fall wurde die binäre logistische Regression explorativ angewandt, um mögliche Prädiktoren für Mitarbeiter zu identifizieren, deren Abgang erzwungen wurde, oder die das Unternehmen verlassen haben (FP). Zur Beantwortung der Frage nach einer optimalen Prädiktorenkombination (*Fragestellung (7)*) konnten dabei einige Indizien gewonnen werden. Die Ergebnisse sind in den *Tabellen 31* und *32* dargestellt.

Tabelle 31: Logistische Regression der abhängigen Variable *Angestelltenverhältnis* (N=142)

Schritt	**Aufgenommene Variablen**	**B-Gewicht**	**Nagelkerkes R^2**	**R^2inkrementell**
0	Eintrittsjahr	-.872**	.116**	-.-
1	Eintrittsjahr	-.851**		
	Einstiegsstufe	.383*	.160*	.044
2	Eintrittsjahr	-.974**		
	Einstiegsstufe	.556**		
	Praktikumsanzahl	.333**	.231**	.071
3	Eintrittsjahr	-1.105**		
	Einstiegsstufe	.570*		
	Praktikumsanzahl	.396**		
	VWL	-871*	.282**	.051

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
Kodierung Angestelltenstatus: 0=angestellt; 1=ausgeschieden; listenweiser Fallausschluss

Die Ergebnisse zeigen, dass mit abnehmendem *Einstiegsjahr* die *Ausstiegswahrscheinlichkeit* steigt. Dies ist insofern nicht verwunderlich, da der Beruf des Unternehmensberaters traditionell mit einer hohen Fluktuationsrate verbunden ist und die Anzahl der Mitarbeiter, die das Unternehmen verlassen, ab einer gewissen Verweildauer zunimmt. Da-

bei zeigte sich zunächst, dass in der Untersuchungsstichprobe das *Eintrittsjahr* Varianz bezüglich des Ausscheidens von Mitarbeitern aufklärt.

Darüber hinaus scheint die *Einstiegsstufe* ebenfalls mit dem Ausscheiden in Verbindung zu stehen. Zum Einen kann dies genau die gleichen Gründe haben wie beim *Einstiegsjahr*: Personen, die auf einer höheren Stufe einsteigen, kommen früher in Bereiche, in denen die Fluktuation sowieso höher ist, und scheiden dann im betrachteten Zeitraum vorher aus. Eine weitere Erklärung könnte jedoch sein, dass Mitarbeiter, die bereits Berufserfahrung besitzen, größere Schwierigkeiten haben, als Berater zu arbeiten, da sie andere Arbeitsgewohnheiten haben. Zudem ist anzunehmen, dass sie auf Grund früherer Berufserfahrung mehr Alternativen im Berufsleben sehen und kritischer gegenüber dem Beraterarbeitsalltag eingestellt sind als Berufsanfänger. Dies könnte ein Ausscheiden aus der Firma begünstigen. Das gleiche gilt für Mitarbeiter, die schon eine Vielzahl von *Praktika* absolviert haben und somit sehr viel Einblick in unterschiedliche berufliche Tätigkeiten bekommen haben. Um eine Konfundierung der Effekte auszuschließen, wurde zusätzlich ein zweiseitiger t-Test zum Vergleich der Ausscheidungswahrscheinlichkeit aller *Cs* im Jahr 2001 in Abhängigkeit ihrer Eintrittsstufe (*JC* oder *C*) gerechnet. Dabei bestätigt sich die Vermutung, dass Personen mit einer höheren Einstiegsstufe signifikant früher das Unternehmen verlassen ($N_{JC\text{-}Einstieg}$=23; $N_{C\text{-}Einstieg}$=25; df=46, MW_{Diff}=-.379; p=.028).

Des Weiteren scheinen Personen mit hohen *volkswirtschaftlichen Kenntnissen* länger im Unternehmen zu bleiben. Da die Variable *VWL* per Definition einen eher allgemeinen Charakter aufweist und, vom Verständnis beim Rating im Recruiting her gesehen, das Verstehen und das Interesse für gesamtwirtschaftliche Zusammenhänge beinhaltet, ist dieses Ergebnis plausibel. Wirtschaftlich interessierte Mitarbeiter sind sicherlich erfüllter durch eine abwechslungsreiche, wirtschaftlich orientierte Tätigkeit wie die des Beraters, die Einblicke in eine Vielzahl von Bereichen und Branchen bietet.

Nach Berücksichtigung des *Eintrittsjahres* (d.h. der Tatsache, dass im Laufe der Zeit sowieso Mitarbeiter ausscheiden) kann man durch die logistische Regression 62% der Bewerber bei der Auswahl richtig klassifizieren. Durch die Hinzunahme der Prädiktoren *Einstiegslevel*, *VWL-Kenntnisse* und *Praktikumsanzahl* erhöht sich diese Rate auf 72.5%, d.h. die Anzahl jener Bewerber, die man richtig bezüglich ihres Ausscheidens vorhersagen kann, erhöht sich um über 10%. Betrachtet man lediglich den Anteil, der tatsächlich ausscheidet, so erhöht sich hier die korrekte Klassifikation von 27.6% auf 58.6% um über 30%.

Bei der Vorhersage jener Mitarbeiter, die das Unternehmen *forciert* verlassen haben, stellen sich die Ergebnisse etwas anders dar (*Tabelle 32*). So ist ein valider Prädiktor die *Geschlechtszugehörigkeit*. Die Wahrscheinlichkeit, dass bei Personen weiblichen Geschlechts das Ausscheiden forciert wird, ist dabei größer als bei männlichen Kollegen. Die Gründe können vielfältiger Natur sein. Da in den späteren Leistungsvergleichen keine signifikanten Unterschiede zwischen den Geschlechtern zu erkennen sind, ist davon auszugehen, dass es sich um einen Effekt in Kombination mit den anderen Prädiktoren handelt. Dabei fällt auf, dass der Anteil an weiblichen Neueinstellungen im Jahr 2000 (34%) höher war als in den darauf folgenden Jahren (17%). Spätere Untersuchungen zeigen, dass die im Jahr 2000 eingestellten Mitarbeiter auf allen Dimensionen schlechter abschneiden als Mitarbeiter, die zu einem späteren Zeitpunkt eingestellt wurden. Im Jahr 2000 gab es ein Marktungleichgewicht zwischen dem Angebot an hoch qualifizierten Mitarbeitern und der unternehmensinternen Nachfrage (*War for Talents*, *New Economy-Boom*). Deshalb ist davon auszugehen, dass aus Wachstumsgründen die Einstellungspolitik weniger restriktiv gehandhabt wurde. Folglich wurde qualitativ weniger gutes Personal eingekauft. Dies gilt gleichermaßen für Frauen und Männer. Da aber der Anteil an Frauen in der Untersuchungsstichprobe *im Jahr 2000* größer ist als in den anderen Jahren, werden die weniger guten Mitarbeiterinnen aus 2000 in der Gesamtstichprobe stärker gewichtet und damit der Prädiktor *Geschlechtszugehörigkeit* signifikant.

Tabelle 32: Logistische Regression der abhängigen Variable *Forced Exit* (N=142)

Schritt	Aufgenommene Variablen	B-Gewicht	Nagelkerkes R^2	R^2inkrementell
1	Geschlecht	-1.234*	.063*	-.-
2	Charisma	-1.102*		
	Geschlecht	-1,290*	.124**	.061
3	Charisma	-1.436**		
	Geschlecht	-1.393*		
	Dr.-Titel	1.469*	.180**	.056
4	Charisma	-1.904**		
	Geschlecht	-1.519**		
	Dr.-Titel	1.704**		
	Konfliktfähigkeit	1.076**	.223**	.043

Kodierung Forced Exit: 0=nein; 1=ja; listenweiser Fallausschluss

Des Weiteren wird bei im Recruiting als weniger *charismatisch* eingestuften Mitarbeitern ein Ausscheiden eher *forciert*. In späteren Analysen zeigt sich, dass *Charisma* signi-

fikant mit Erfolgskomponenten (Projektbeurteilungsvariablen) zusammenhängt. Dabei ist der hohe Generalitätsgrad von Bedeutung, da der Begriff *Charisma* ein sehr allgemeines Konstrukt darstellt, das viele Unterkomponenten beinhalten kann. Auch ist von einer einheitlichen Definition unter den Beurteilern im Recruiting nicht auszugehen (s.o.).

Darüber hinaus zeigt sich, dass promovierte Mitarbeiter häufiger aus dem Unternehmen forciert ausscheiden als nicht-promovierte Kollegen. Unter 3.3.3 wird gezeigt, dass promovierte *Cs* durchgängig schlechtere Beurteilungen aufweisen, d.h. der forcierte Ausstieg liegt auch hier vermutlich an weniger guten Leistungen. Eine mögliche Erklärung kann darin liegen, dass promovierte Mitarbeiter überwiegend direkt als *C* oder höher einsteigen. Damit werden sie mit Neueinsteigerproblemen und höheren Anforderungen konfrontiert (s.o.). Dies trifft auf vergleichbare Kollegen mit *JC-Erfahrung* nicht zu (indirekter Effekt). Zudem könnte die Gewohnheit des wissenschaftlichen Arbeitens und Denkens für die sehr praktisch orientierte Beratungstätigkeit kontraproduktiv wirken. Die Variable *Konfliktfähigkeit* wirkt dabei als *Suppressor*, der irrelevante Varianzanteile bei *Charisma* unterdrückt und vermutlich auf ein rein methodisches Artefakt zurückzuführen ist. Durch die vier Prädiktoren lässt sich der Anteil der insgesamt richtig klassifizierten Mitarbeiter insgesamt nicht erhöhen. Allerdings gelingt es auf Kosten der RP, die FP besser vorherzusagen. Bei ihnen erhöht sich die richtige Klassifikation um 13%.

Die gefundenen Zusammenhänge müssen immer vor dem Hintergrund der methodischen Restriktionen beurteilt werden. Sie sollten Anhaltspunkte bieten, um bei den Bewerbungen für charakteristische Muster sensibilisiert zu werden. Es ist dabei klar, dass die Schlussfolgerung nicht sein kann, in Zukunft beispielsweise ausschließlich Mitarbeiter einzustellen, die wenige Praktika absolviert haben oder nur solche männlichen Geschlechts. Hier spielen, wie man bei den späteren Analysen der Zusammenhänge mit der erbrachten Leistung in den Kohorten erkennen wird, zahlreiche andere Faktoren eine Rolle, die Berücksichtigung finden sollten. Zudem ist es empfehlenswert, die gefundenen Zusammenhänge mit Hilfe einer weiteren Stichprobe zu replizieren.

3.3.2 Analyse der Kohorte Junior-Consultants

Ausgangspunkt des weiteren Vorgehens in diesem Abschnitt ist die deskriptive Analyse der Prädiktor- und Kriterienvariablen. Zunächst wurden basierend auf ausgewählten *CV-Variablen* t-Tests für unabhängige Stichproben auf Mittelwertunterschiede und deren Effektstärken auf allen Projektbeurteilungsdimensionen gerechnet. Ziel dabei war es, signifikante Profilunterschiede zwischen den untersuchten Gruppen festzustellen und somit An-

haltspunkte für eine Verbesserung der Vorauswahl von Bewerbern zu gewinnen. Zudem soll eine Varianzanalyse mit Messwiederholung Auskunft darüber geben, inwieweit sich die Bewerber mit zunehmender Funktionsstufenzugehörigkeit in den Projektbeurteilungen verbessern. In einem weiteren Schritt wurde über Korrelationsanalysen die Validität des Auswahlverfahrens und der *CV-Variablen* für den Berufserfolg ermittelt. Die anschließenden Regressionsanalysen sollen aufdecken, im welcher Form die Prädiktoren aus dem Recruiting optimal kombiniert werden sollten. Ferner werden im Rahmen einer Faktorenanalyse, Anhaltspunkte für die Konstruktvalidität des gemessenen Kriteriums gewonnen. Darüber hinaus sollen sie Erkenntnisse bei der Untersuchung der Generalitätsebenen von Prädiktor und Kriterium liefern.

3.3.2.1 Methodische Restriktionen

Nachfolgend werden die Ergebnisse der Analysen dargestellt und inhaltlich interpretiert. Dabei ist allerdings eine Vielzahl, teilweise bereits erwähnter, methodischer Restriktionen zu berücksichtigen. So ist der *Stichprobenumfang* einzelner Variablen und deren Kategorien sehr klein und damit die Effekte schwer zu generalisieren. Kleine Stichproben führen zwangsläufig zu geringen Teststärken. So können beispielsweise t-Tests, bei einer Stichprobengröße von N~80, mit einer hinreichend großen Teststärke von 80%, nur eingeschränkt *wahre Effekte* von mindestens d=0.5 auch tatsächlich bei einem 5% Signifikanzniveau aufdecken (Cohen, 1988). Es ist deshalb auch anzunehmen, dass mögliche wahre Effekte in den Analysen nicht entdeckt werden konnten. Problematisch ist ebenso die Gefahr des *Selektionsfehlers* und die einhergehende Varianzreduktion, die zu einer Unterschätzung der Effekte führen kann. Außerdem besteht eine hohe Anfälligkeit der Analysen für *Ausreißerwerte*. Es ist deshalb von einer Instabilität der berechneten Koeffizienten, die sich auch teilweise in den Konfidenzintervallen widerspiegeln (Vorzeichenwechsel) auszugehen. Die teilweise vorgenommene Eliminierung der Ausreißerwerte erschwert zudem die Generalisierbarkeit der Befunde. Ähnlich problematisch ist die große Anzahl fehlender Werte. Ein vom Autor vorgenommenes regressionsanalytisches Ersetzen dieser ergab bei den Korrelationsanalysen, dass die Berechnungen sich in ihrer Höhe und Signifikanz stark verändern. Der Autor verzichtet deshalb auf eine Datenanpassung mittels eines solchen Vorgehens und zieht die tatsächlichen Werte zur Analyse heran. Zusätzlich kann es zu einer *α-Fehlerinflation* bei fehlender Testhierarchie kommen, da es bei einer großen Anzahl von Analysen und Testdurchläufen zwangsläufig per Zufall zu Effekten kommt, die nicht auf „wahren“ Gegebenheiten beruhen. Zudem sind die hohe *Multikolinearität* und damit konfundierten Varianzanteile zwischen den Variablen problematisch, da sie zu einer Verletzung der Prä-

missen der multiplen Regression und einer Ineffizienz der β-Koeffizienten führen. Schließlich gehen die vorgenommenen Faktorisierungen der Variablen mit einem *Varianzverlust* einher, der zu einer Effektverzerrung beim Vergleich mit dem realen Vorgehen führen kann. Somit ist die Interpretierbarkeit der Ergebnisse eingeschränkt.

3.3.2.2 Deskriptive Analyse

Bei den *Recruitingvariablen* und den *CV-Variablen* wurden die Erkenntnisse aus der kohortenübergreifenden Prädiktorenanalyse weitgehend repliziert. Die deskriptiven Befunde sollen deshalb nur kurz skizziert werden. So sind die *Recruitingvariablen* auch bei der *JC-Stichprobe* durch hohe Mittelwerte und geringe Standardabweichungen bei allen Dimensionen gekennzeichnet. Zudem gibt es auf einigen Dimensionen eine geringe Anzahl an Beurteilungen. Der *Gesamtmittelwert* aus *Recruiting* von 4.78 und der hohe Minimalwert von 3.88 deuten auch hier auf einen restriktiven *Einstellungs-Cutoff* hin. Ebenso gibt es eine Vielzahl an Ausreißerwerten, die je nach Variable immer unterschiedliche Personen betreffen. Während einige Recruitingvariablen signifikant von der Normalverteilung abweichen (z.B. *Pragmatismus*, *Selbstorganisation* und *Teamfähigkeit*), trifft dies auf andere (z.B. *Charisma* oder *Gesamtmittelwert*) nicht zu.

Die visuelle Inspektion der Histogramme deutet wie bei der kohortenübergreifenden Analyse darauf hin, dass die Abweichungen von der Normalverteilung hauptsächlich in einer schmalgipfligen rechtslastigen Verteilung um den Wert 5.0 liegen. Die Gründe für die geringe Varianz und die Verteilungseigenschaften der *Prädiktorvariablen* wurden bereits unter 3.3.1.2 beschrieben. Sie werden deshalb an dieser Stelle nicht erneut dargestellt.

Bei den *CV-Daten* wurde auf jene Variablen fokussiert, die bei den *JCs* variieren. Hier zeigt sich ein differenzierteres Bild. Auf vielen Variablen gibt es, wie bei der kohortenübergreifenden Analyse, höchst unterschiedliche Werte, teilweise einzelfallbedingt. Auf einzelnen Variablen finden sich entsprechend hohe Standardabweichungen. Auf der anderen Seite differenzieren einige Variablen weniger zwischen den Personen. Zudem waren bei einzelnen Variablen wie *Mathematiknote*, *Englischnote* oder *Fächerschwerpunktnote* nur geringe Fallzahlen vorhanden. Bei den *CV-Variablen* ist lediglich für die *Abiturnote*, dem *Eintrittsalter* und der *Ausbildungsnote* von der Normalverteilungsannahme auszugehen. Alle anderen Variablen weichen signifikant von dieser ab, z.T. bedingt durch wenige Kategorien, so dass hier im weiteren Verlauf nonparametrische Tests ihre Anwendung finden.

Der *Gesamtmittelwert aus Projektbeurteilungen* (*Tabelle 33*) ist mit 3.54 weit über dem theoretisch zu erwartenden Wert von 3.0, bei dem laut Definition alle Erwartungen

durch den *JC* voll und ganz erfüllt werden. Dass nicht davon ausgegangen werden kann, dass die Mehrheit der *JCs* die Erwartungen übertreffen bzw. im Umkehrschluss die Erwartungshaltung der Beurteiler niedrig ist, führt zu der Vermutung, dass die Handhabung des Beurteilungsinstrumentes wenig streng ist. *Behaviour under Pressure* und *Teamwork* sind dabei die durchschnittlich am besten beurteilten Dimensionen. Alle Dimensionen haben durchweg geringe Standardabweichungen. Die geringe Differenzierung bei den evaluierten Mitarbeitern ist vermutlich auf die gleichen klassischen Urteilsfehler, wie auf der Prädiktorenseite beschrieben, und auf die Aggregation zurückzuführen. Dabei halten sich die Ausreißer in Grenzen. Zudem sind innerhalb der vorhandenen Spannweite, die Werte relativ gut differenziert. Vereinzelte Ausreißer finden sich auf den Dimensionen *Behaviour under Pressure*, *Quality of Work* und *Know-how-Management.* Von nahezu allen Mitarbeitern sind Urteile auf allen Dimensionen vorhanden. Die *Projektbeurteilungsdaten* weichen nicht signifikant von der Normalverteilung ab.

Tabelle 33: Deskriptive Kennwerte der Kriterien (Projektbeurteilungen) bei JCs

Kriterienvariablen	N	MW	Med.	Min.	Max.	SD	NV (j/n)
Specialist-Know-how	86	3.48	3.45	2.50	4.17	.32	j
Problem Solution	86	3.51	3.50	2.75	4.25	.31	j
Quality of Work	86	3.52	3.53	2.75	4.25	.29	j
Behaviour under Pressure	85	3.70	3.67	2.67	5.00	.39	j
Client Acceptance	85	3.55	3.56	3.00	4.20	.33	j
Client Communication	85	3.50	3.50	2.88	4.25	.30	j
Know-how-Management	86	3.45	3.42	2.83	4.33	.30	j
Teamwork	86	3.64	3.67	3.00	4.50	.30	j
Durchschnittlicher Gesamtwert	86	3.54	3.56	2.90	4.00	.24	j

N= Stichprobenumfang; MW=Mittelwert; Med.=Median; Min.=Minimum; Max.=Maximum; SD=Standardabweichung; NV=Normalverteilungsannahme

3.3.2.3 Mittelwert- und Profilvergleiche

Basierend auf ausgewählten *CV-Variablen* wurden zunächst *explorative* t-Tests auf Mittelwertunterschiede und deren Effektstärken auf allen Projektbeurteilungsdimensionen gerechnet (*Tabelle* 34*)*. Die Ergebnisse können einige Anhaltspunkte für eine Optimierung der Vorauswahl der Bewerber anhand biographischer Informationen (*Fragestellung (5)*) aufzeigen. Die in den Mittelwertvergleichen aufgedeckten Unterschiede sind allerdings nicht als kausal zu interpretieren, sondern bieten Anhaltspunkte für mögliche Zusammenhänge und

Hintergründe unterschiedlicher Leistungen bei den eingestellten Mitarbeitern. Obgleich sie sich in einem sehr kleinen absoluten Umfang bewegen, sind sie in ihrer Ausprägung und Richtung größtenteils plausibel und interpretierbar, mit teilweise beachtlichen Effektstärken.

Tabelle 34: Explorative t-Tests ausgewählter CV-Kategorien für JCs (Effektstärken)

	Gruppen				
Kriterienvariablen	Eintrittsjahr 2000 (ja vs. nein)	Beschäftigungs-verhältnis (angestellt vs. ausgeschieden)	Forced Exit (ja vs. nein)	Universitätsart (öffentlich vs. privat)	Geschlecht (w vs. m)
Specialist-Know-how	-.953**	.588*	-.698*	-.235	-.427
Problem Solution	-.444**	.625**	-1.145**	-.053	-.161
Quality of Work	-.444	.676**	-.863**	-.212	-.044
Behaviour under Pressure	-.599*	.525*	-.843*	-.008	-.533
Client Acceptance	-.752**	.724**	-1.089**	-.198	-.362
Client Communication	-.335	.427	-.792*	-.079	.072
Know-how-Management	-.108	.363	-.757*	.056	.013
Teamwork	-.821*	.336	-1.156*	-.478	-.481
Durchschnittlicher Gesamtwert	-.738**	.730**	-1.226**	-.199	-.324

Basis der Effektgrößen d ist der Mittelwertunterschied = 1.Gruppe – 2. Gruppe
* Der t-Test ist auf dem 0.05 Niveau signifikant (zweiseitig)
** Der t-Test ist auf dem 0.01 Niveau signifikant (zweiseitig)
Zur Anzahl der Fälle siehe Tabelle 20

Die Ergebnisse zeigen, dass die im Jahr 2000 eingestellten Mitarbeiter auf allen Dimensionen schlechter abschneiden, als Mitarbeiter, die zu einem späteren Zeitpunkt eingestellt wurden. Dabei handelt es sich größtenteils um signifikante, starke Effekte. Da in jenem Jahr ein Marktungleichgewicht zwischen dem Angebot an hoch qualifizierten Mitarbeitern und der unternehmensinternen Nachfrage bestand (*War for Talents, New-Economy-Boom*) ist davon auszugehen, dass aus Wachstumsgründen, die Einstellungspolitik weniger restriktiv gehandhabt wurde. Somit kam es vermutlich zur Einstellung qualitativ weniger guten Personals. Darüber hinaus sind die ausgeschiedenen Mitarbeiter (*Unforced* und *Forced Exits*) in zahlreichen Leistungen signifikant schlechter als die restlichen Mitarbeiter (mittlere bis starke Effekte). Dies deutet auf eine effektive Personalpolitik des Unternehmens, wegen der die tendenziell schlechteren Mitarbeiter das Unternehmen verlassen müssen, hin. Absolventen von *Privatuniversitäten* sind hingegen auf allen Dimensionen leicht besser, besonders bei *Teamwork*. Dies lässt sich möglicherweise durch das bessere Lernumfeld (Quote Studenten/Professoren) und die gezielte Förderung von *Teamarbeit* in den meisten privaten Univer-

sitäten erklären. Die insgesamt schwachen Effektgrößen beruhen allerdings genauso wenig auf signifikanten Mittelwertunterschieden wie die Unterschiede beim geschlechtspezifischen Vergleich.

Tabelle 35: Gruppenvergleich (t-Test) biographischer Kategorien für JCs (Effektstärken)

	Gruppen			
Kriterienvariablen	Mind. ein analyt. FSP (ja vs. nein)	BWL-Studium (ja vs. nein)	MBA (ja vs. nein)	Praktikum bei UB (ja vs. nein)
Specialist-Know-how	-.022	-.043	.362	-.214
Problem Solution	-.105	.069	.493	.046
Quality of Work	.072	.195	.836*	.092
Behaviour under Pressure	.005	.150	-.089	-.066
Client Acceptance	-.396	-.442	.139	-.068
Client Communication	-.339	-.368	-.342	-.605*
Know-how-Management	-.285	-.222	.396	-.167
Teamwork	-.191	.000	.316	.076
Durchschnittlicher Gesamtwert	-.199	-.075	.344	-.141

Basis der Effektgrößen d ist der Mittelwertunterschied = 1.Gruppe – 2. Gruppe
* Der t-Test ist auf dem 0.05 Niveau signifikant (zweiseitig)
** Der t-Test ist auf dem 0.01 Niveau signifikant (zweiseitig)
Zur Anzahl der Fälle siehe Tabelle 21

In einem weiteren Schritt wurden Zusammenhänge weiterer *biographischer Informationen* und *beruflicher Leistung getestet*. *Tabelle 35* zeigt die Effektstärken der genannten *biographischen* Gruppenvergleiche. Keine Unterschiede zeigten sich in Abhängigkeit des *analytischen FSP*. Auf Grund der unterschiedlichen Projektinhalte, die mal mehr analytische und mal mehr kreative Schwerpunkte haben, und der Bedeutung der sprachlichen Ausdrucksfähigkeit im Beruf des Beraters war dieses Ergebnis zu erwarten. Das Vorhandensein eines BWL-Abschlusses zeigt ebenso keine signifikanten Unterschiede. Absolventen, die in einer Unternehmensberatung anfangen, beschäftigen sich oftmals mit Spezialthemen, die sie in einem BWL-Studium nicht gelernt haben. Es findet ein „learning on the job“ statt. Somit ist das Wissen aus dem Studium vermutlich nur begrenzt von Nutzen, da es nur wenig berufsrelevantes (*job knowledge*, siehe Hunter & Hunter, 1984) oder praktisches Wissen (*tacit knowledge*, siehe Sternberg, et al., 2000) enthält. Vielmehr ist die Fähigkeit, sich schnell in neue, praxisrelevante Themen einzuarbeiten, gefragt. Im Sinne Ackermans dürfte dabei der fluiden Intelligenz (siehe 2.5.1.1) eine bedeutende Rolle zukommen. *MBA-Absolventen* haben hingegen durchgängig bessere Beurteilungen. Einen signifikanten Unterschied gibt es

dabei auf der Dimension *Quality of Work* (sehr starker Effekt). Dies liegt vermutlich daran, dass durch das *MBA-Studium* eine stärkere Spezialisierung stattfindet. Außerdem weist die Arbeit anhand von Fallstudien einen hohen Praxischarakter und eine hohe Affinität zur Beratungstätigkeit auf. Die vermittelten und gelernten Inhalte dürfte deshalb höhere Überschneidungen mit dem notwendigen *Berufswissen* aufweisen.

Schließlich sind Personen, die ein Praktikum bei einer Unternehmensberatung absolviert haben, signifikant schlechter in der Klientenkommunikation (mittlere Effektstärke). Obwohl eine Inspektion des Streudiagrammes keinen nennenswerten Ausreißerwert ergab, ist dieses Ergebnis dennoch verwunderlich, da tendenziell bessere Beurteilungen auf Grund der Gelegenheit zum Aufbau von Berufswissen erwartet wurden. Gründe könnten u.a. in einem zu selbstsicheren Auftreten gegenüber den Kunden liegen. Zudem kann es sein, dass gerade beratungserfahrene *JCs* eher mit Klienten kommunizieren und deshalb eine realistischere Beurteilung dieser Dimension möglich ist, während andere *JCs* wenig oder gar keinen Kundenkontakt haben. Die dargestellten Ergebnisse aus den *Tabellen 34* und *35* wurden ferner zur Interpretationserleichterung als Profilvergleiche visualisiert. *Abbildung 45* zeigt drei ausgewählte Beispiele.

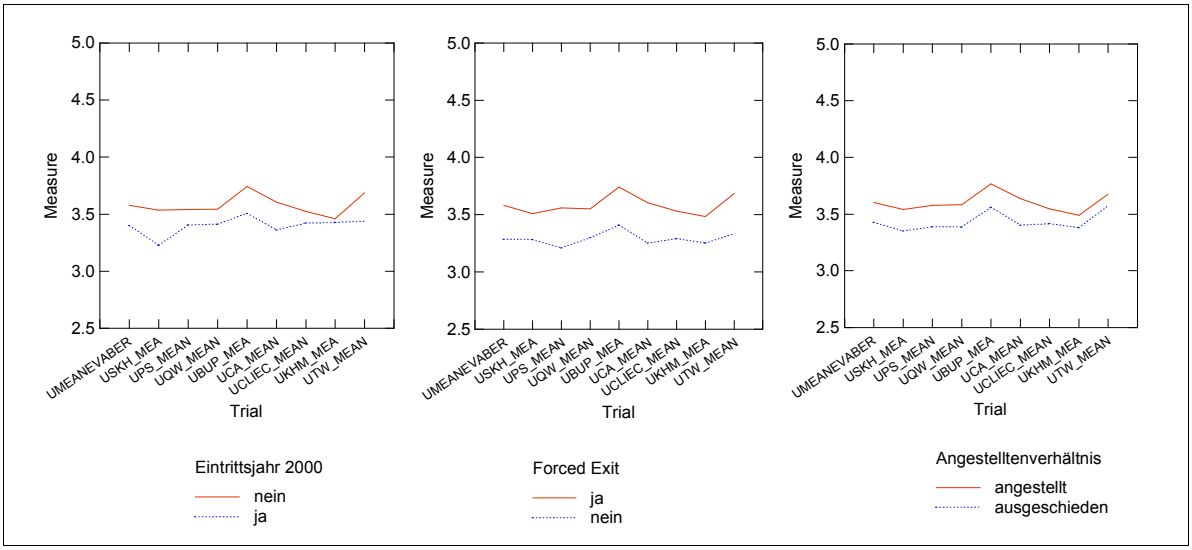

Abbildung 45: Beispielhafte Profilvergleiche der Projektbeurteilungen

Abschließend wurden die Projektbeurteilungsmittelwerte im Zeitverlauf untersucht und über eine einfaktorielle Varianzanalyse mit Messwiederholung auf Unterschiede getestet, um möglich Lerneffekte zu visualisieren. *Tabelle 36* zeigt die Effektstärken (Glass'sche Deltas) der Mittelwertunterschiede. Sie zeigen, dass sich der Durchschnittswert der Projektbeurteilungen im Laufe der Zeit über alle Personen mit fortschreitendem Messzeitpunkt signifikant verbessert. Auf den Unterdimensionen lässt sich dieses Ergebnis nur teilweise

replizieren. So ist auf den Dimensionen *Behaviour under Pressure* und *Client Acceptance* von keinem Unterschied auszugehen. Dennoch gehen, wenn auch nicht signifikant, die Effekte in die erwartete Richtung. Auf die Schwierigkeit der Beurteilung klientenbezogener Skalen bei JCs wurde bereits hingewiesen. Das Verhalten unter Stresssituation hingegen ist möglicherweise an unveränderbare Persönlichkeitseigenschaften wie *Emotionale Stabilität* gekoppelt.

Tabelle 36: Übersicht Mittelwertunterschiede im Zeitverlauf (Glass'sche Deltas) (N=49)

	Messzeitpunkte		
Kriterienvariablen	**1 vs. 2**	**1 vs. 3**	**2 vs. 3**
Specialist-Know-how	-.083	-.425*	-.337*
Problem Solution	-.055	-.514*	-.400*
Quality of Work	-.277*	-.548*	-.283*
Behaviour under Pressure	-.083	-.250	-.136
Client Acceptance	-.103	-.245	-.128
Client Communication	-.054	-.325*	-.286
Know-how-Management	.205	-.403*	-.484*
Teamwork	-.174	-.577*	-.409*
Durchschnittlicher Gesamtwert	-.110	-.522*	-.382*

* Signifikanter Effekt auf 0.05 Niveau (Paarvergleiche der ANOVA)

3.3.2.4 Korrelationsanalyse

In einem ersten Schritt wurden für ausgewählte *CV-Variablen* Rangkorrelationen mit den Projektbeurteilungsdaten unter Einbezug der gesamten Stichprobe berechnet (*Tabelle 37*). Erwartungsgemäß ergaben sich für zahlreiche biographische Informationen nennenswerte prognostische Validitäten (Fragestellungen (9)). Die signifikanten Korrelationen von *Eintrittsalter* mit *Client Acceptance* (r=.206) und *Know-how-Management* (r=.213) deuten darauf hin, dass ältere *JCs* eher vom Klienten akzeptiert werden. Dies steht im Einklang mit dem Trend zur Bevorzugung seniorer Berater durch die Klienten (Lechner, et al., 2005). Zudem sind sie besser im Wissensmanagement, was mit der höheren Erfahrung zu tun haben kann.

Bei den Noten zeigen sich breite Zusammenhänge auf den Variablen. Dabei hängen alle, mit Ausnahme der *Mathematiknote,* signifikant mit *Specialist-Know-how* und *Problem Solution* zusammen. Dass bessere Noten in der Ausbildung mit einem höheren Wissensstand einhergehen ist ein weniger überraschendes Ergebnis. Darüber hinaus korrelieren die

Deutsch- und die *Englischnote* signifikant mit der Gesamtleistung (r=-.222 und r=-.283). Die Validität ist damit höher als die von Schuler (2000) für Schulnoten berichtete prognostische Validität bei der Vorhersage von Berufserfolg von (r=.15). Der Erfolg bei Beratungstätigkeiten hat zu einem großen Teil mit der sprachlichen Ausdrucksfähigkeit des Beraters gegenüber dem Klienten, die teilweise in der Note enthalten sein dürfte, zu tun. Insbesondere bei der Dimension *Client Acceptance* (r=-.331) scheint die Korrelation deshalb plausibel. Die Korrelation der *Mathematiknote* mit der Dimension *Teamwork* ist dadurch zu erklären, dass *Teamwork* auf den Einzelvariablen auch die Komponente des Beitrags des Einzelnen zu Problemlösungen im Team, also eine analytische Komponente, beinhaltet (siehe Anhang VI).

Tabelle 37: Korrelationen Projektbeurteilungsdimensionen mit ausgewählten CV-Variablen (JCs)

Korrelationen (Spearman-Rho)								
	Eintritts-alter	Note_FSP	Note_D	Note_M	Note_E	Abinote	Anz_Absch_tot	Pra_t
uMeanEvaltot	.141	-.175	-.222*	-.025	-.283*	-.153	.103	.283**
uSKH_MEAN	.027	-.250*	-.264*	-.156	-.339*	-.212*	.076	.160
uPS_MEAN	.172	-.122	-.221*	-.029	-.379**	-.313	.193*	.268**
uQW_MEAN	.061	-.125	-.066	-.094	-.275*	-.243*	.088	.282**
uBUP_MEAN	.063	-.105	-.132	.052	-.120	-.067	.081	.078
uCA_MEAN	.206*	-.102	-.219*	.132	-.331*	-.035	.009	.258**
uCLIECO_MEAN	.018	-.105	-.047	.078	.043	-.022	-.057	.222*
uKHM_MEAN	.213*	-.078	-.292*	.186	-.255*	.117	.122	.201*
uTW_MEAN	.149	-.073	-.190	-.251*	-.120	.099	.169	.217*
N	83[a]	73	63	48[a]	44[a]	80[a]	83[a]	83[a]

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).
** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).
[a] Für die Dimension uCLIECO_MEAN gilt Stichprobengröße=N-1.

Zudem hängt *Problem Solution* mit der Anzahl der Hochschulabschlüsse zusammen. Hier könnte die Fähigkeit, „über den Tellerrand hinauszuschauen“, die bei Personen mit mehr als einem Studium ausgeprägter sein müsste, eine Rolle spielen. Schließlich gibt es bei den *CV-Daten* einen nahe liegenden, selbsterklärenden Zusammenhang zwischen der *Praxiserfahrung* (Dauer von Praktika) und fast allen *Dimensionen der Projektbeurteilung*, auf den schon im Rahmen des *Berufswissens* eingegangen wurde (siehe 3.3.2.3). Die Korrelationen sind in ihrer Höhe (.20<r<.30) geringer als die von Dye, Reck und McDaniel (1993) berichteten prädiktiven Validitäten für Berufswissenstests (r=.45 und r=.62). Vor dem Hin-

tergrund, dass die innerhalb von Praktika gesammelte Erfahrung nur teilweise mit Berufserfahrung gleichzusetzen ist und bei der Wissensaneignung zudem die kognitive Leistungsfähigkeit moderiert, ist die Höhe nachvollziehbar. Sie liegt sogar leicht über den metaanalytisch berichteten Validitäten für Berufserfahrung von Schmidt und Hunter (1998) in Höhe von r=.18.

Unter der Annahme gleicher Symmetrieebenen wurde der Zusammenhang zwischen den Gesamtwerten aus Recruiting und Projektbeurteilung betrachtet (*Abbildung 46*). Nach Eliminierung der Ausreißerwerte gibt es einen signifikanten Zusammenhang zwischen dem *Recruiting*- und dem *Projektbeurteilungsgesamtwert* von r=.276. Diese mittlere Effektgröße erhöht sich durch Rückrechnung des *Selektionsfehlers* auf die Grundgesamtheit auf r=.301[13] und liegt zwar unterhalb, aber dennoch in der Nähe der metaanalytisch berichteten Validität für *unstrukturierte* Interviews (r=.380) und Assessment Center (r=.370). Auf die *Attenuationskorrektur* des Prädiktors wurde dabei verzichtet, da die *operationale* Validität des verwendeten Instrumentes von Interesse war (siehe auch Sackett, Laczo & Arvey, 2002). Auch eine *Attenuationskorrektur* des Kriteriums wurde nicht vorgenommen, da auf Grund der Datenstruktur die Interrater-Reliabilitäten nicht bekannt waren. Zudem hat die Verwendung aggregierter Kriteriendaten zu einer nicht bezifferbaren Reliabilitätserhöhung geführt. Die Attenuationskorrektur mit der Interrater-Reliabilität würde somit zu einer Erhöhung der korrigierten Validität des Instrumentes führen, die höher als die „wahre“ Validität ist. Bei durchschnittlich 2.78 Beurteilungen pro *JC* und einer angenommenen Interrater-Reliabilität von r=.52 für Vorgesetztenurteile (Viswesvaran, et al., 1996) ergibt sich beispielsweise nach der *Spearman-Brown-Formel* (*Gleichung (8)*) eine Reliabilität von r=.750. Attenuationskorrigiert führt dies zu einer Validität von r=.347, die noch näher an der metaanalytisch berichteten Validität von *ACs* liegt und den Erwartungen entspricht. Dieses Vorgehen ist allerdings nur unter der Annahme paralleler Tests, also unter der Voraussetzung, dass einzelne Beurteiler auch das exakt gleiche Konstrukt beurteilen, sinnvoll. Dies ist beim vorliegenden Verfahren fraglich (siehe auch Guion, 1998). Die Verwendung der Interraterübereinstimmung als Reliabilitätsmaß für Vorgesetztenurteile über die Berufsleistung ist zudem Gegenstand zahlreicher kontroverser Diskussionen gewesen (siehe dazu Murphy & De Shon, 2000a, 2000b; Schmidt, et al., 2000). Die Korrelation von r=.301 wird folglich als konservative Schätzung der Validität des betrachteten *ACs* angesehen und in den weiteren Berechnungen zum Nutzen beibehalten.

Die *Abiturnote* (*Abbildung 46*) korreliert mittelstark mit dem *Gesamtwert der Projektbeurteilung* (r=-.381) und den meisten Unterdimensionen (*Anhang VIII-I*). Lediglich für die Dimensionen *Client Communication* und *Know-how-Management* sind diese nicht signifikant. Auf Kriteriumsseite attenuationskorrigiert erhöht sich die Korrelation mit dem *Gesamtwert der Projektbeurteilung* sogar auf r=-.439 und liegt somit in einem Bereich, der eigentlich eher bei der Vorhersage von Ausbildungsleistungen und nicht von Berufserfolg angenommen wird (r=.41; Baron-Boldt, et al., 1989). Die Tätigkeit als *JC* ist allerdings mit einer Art Ausbildungsphase vergleichbar und durch Zuarbeiten, langsames Hinführen zu der Thematik und Unterstützung durch erfahrene Kollegen gekennzeichnet. Sie hat damit andere intellektuelle Anforderungen als eine Beratungstätigkeit auf einer höheren Ebene (siehe dazu Sieglen, 1998). Eine erfolgreiche Bewältigung dieser Tätigkeiten hängt nicht unwesentlich von Gewissenhaftigkeit, Zuverlässigkeit, Fleiß, Ausdauer und Intelligenz ab, alles Eigenschaften, die auch durch die *Abiturnote* abgebildet werden.

Beim Betrachten der Überkreuzeffekte (unfaire Validitätstests, *Anhang VIII-1*) zeigt sich, dass der *Gesamtwert aus Recruiting* mit den Dimensionen *Specialist-Know-how*, *Quality of Work*, *Behaviour under Pressure*, *Know-how-Management* und *Teamwork* signifikant korreliert (mittlere Effekte). Keine Indizien gibt es bezogen auf das Verhalten gegenüber dem Klienten (*Client Acceptance*, *Client Communication*). Das angewandte Rollenspiel (im Rahmen des AC), mit dem dieses abgeprüft werden soll und bei dem der Interviewer die Rolle des Klienten übernimmt, scheint deshalb die Realität nicht in ausreichendem Maße widerzuspiegeln ($R_{PR}=0$). Sie könnte auch in den Bewertungsdimensionen des Recruiting nicht ausreichend vertreten sein. Zudem kann hier die bei den Mittelwertvergleichen bereits erwähnte Schwierigkeit herrschen, diese Dimensionen auf Projektbeurteilungsseite zu erfassen, wenn *JCs* wenig oder gar keinen Kundenkontakt haben ($R_{KR}=0$). Gleiches gilt für *Problem Solution*, wo sich ebenfalls keine signifikante Korrelation zeigt. Die tägliche Arbeit von *JCs* ist von kleineren Aufgaben, administrativen Tätigkeiten und Zuarbeiten gekennzeichnet (s.o.). Oftmals ist hier keine Gelegenheit für komplexes Problemlösen ($R_{KR}=0$). Dies wird meistens von erfahrenen Kollegen übernommen. Die Analyse der Prädiktordaten hat zudem gezeigt, dass die Beurteilung logisch-analytischer Fähigkeiten vermutlich nicht ausreichend reliabel ist.

---

[13] Die Berechnung des „restriction of range"-Effektes wurde mit der von Paul Barrett entwickelten Software *Correlation Attenuation Corrections v.2.1* vorgenommen. Basis ist die Standardabweichung des Gesamtwertes der Bewerberpopulation SDx von 1999-2001, die mit der Skalierung 1-6 beurteilt wurden. Die Werte insgesamt sind auf Grund einer variierenden Berechnungssystematik etwas geringer, d.h. die Berechnung ist konservativ.

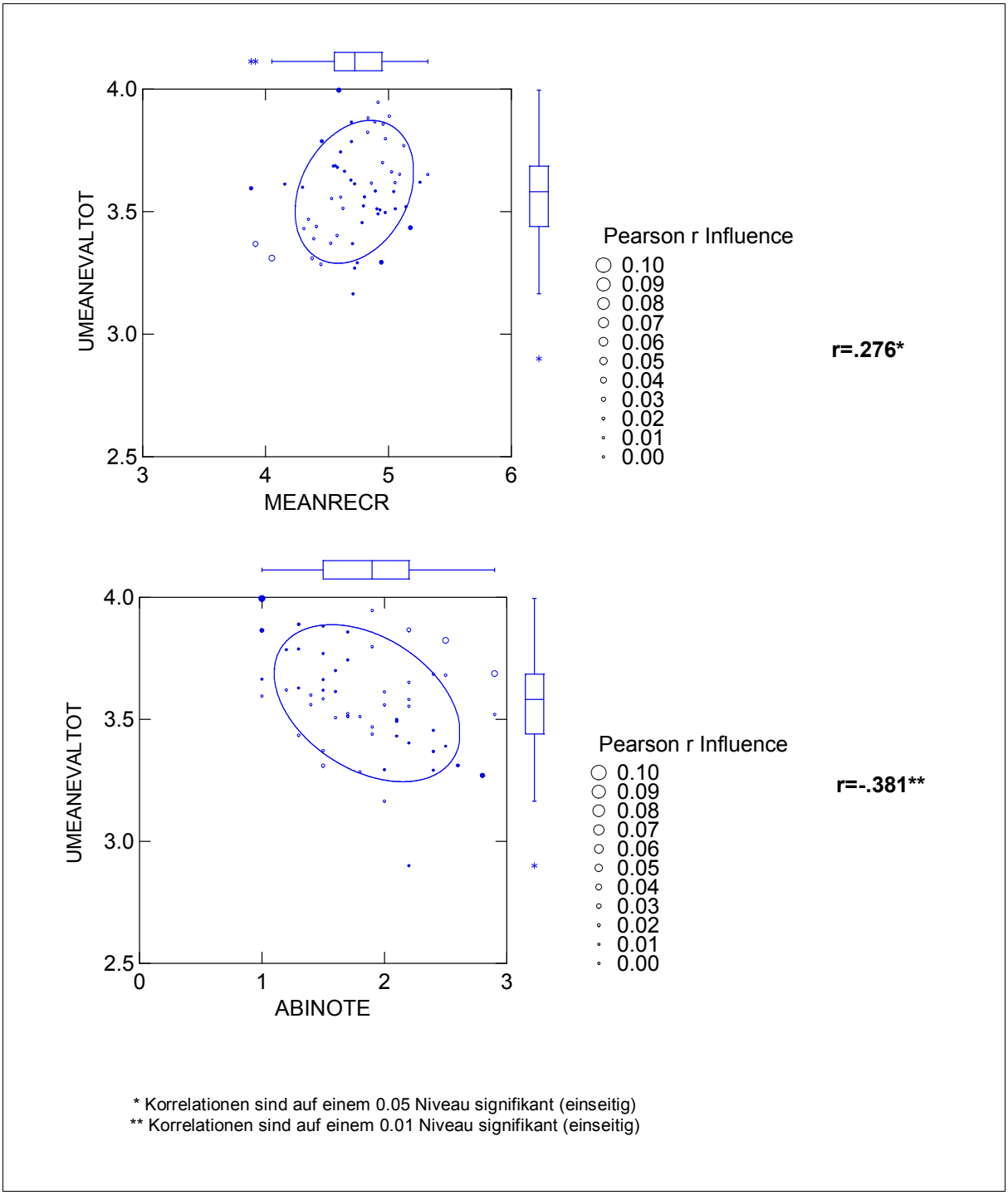

Abbildung 46: Korrelationen Abiturnote und Recruiting-Gesamtwert mit Gesamtwert aus Projektbeurteilungen (JC) (N=59)

Auf der Variablenebene des Recruitings zeigen sich kleinere bis mittlere signifikante Effektgrößen bei der *Kommunikationsfähigkeit* (r=.248), dem *Charisma* (r=.335), intellektueller Neugier (r=.237) und *VWL-Kenntnissen* (r=.305) mit dem Gesamtwert der Projektbeurteilung. Obwohl auch hier zunächst von einer Verletzung der Symmetrieprinzipien auszugehen war, hat *Charisma* einen stärkeren Zusammenhang mit der späteren *Gesamtleis-*

tung, als der *Gesamtwert des Recruiting*. Offensichtlich werden *charismatische* Persönlichkeiten und solche, die *kommunikative* Fähigkeiten besitzen, insgesamt besser beurteilt. Gründe können zunächst entweder in ihrer tatsächlich besseren Leistung auf bestimmten Dimensionen oder in einem Generaleffekt auf oberster Symmetrieebene liegen. Bei *Charisma*, das vermutlich Überschneidungen mit den Big 5-Dimensionen *Extraversion*, *Emotionale Stabilität* und *Verträglichkeit* hat, und *Kommunikationsfähigkeit,* das Überlappungen mit *Extraversion* aufweisen dürfte, handelt es sich um eher generelle Persönlichkeitseigenschaften. Dass die Leistung des Beraters eine sehr komplexe Dimensionierung aufweist, kommt diesen breit gefassten Persönlichkeitseigenschaften entgegen. Eine weitere Erklärung könnte darin liegen, dass *charismatische*, *kommunikative* Mitarbeiter ihre Ideen besser „an den Mann" bringen können, also mehr Gehör bei ihren Mitmenschen finden. Da, besonders im Projektgeschäft, Erfolge meistens nur mit dem Kunden erreicht werden können (siehe 2.4), ist dieser Zusammenhang nachvollziehbar. Im Einklang mit den Ergebnissen steht, dass Extraversion in Metaanalysen als valider Prädiktor für berufliche Leistung bei Verkäufern und Managern mit Korrelationen um die r=.18 bestätigt werden konnte (Barrick & Mount, 1991). Dies gilt insbesondere bei großer Autonomie am Arbeitsplatz, die bei Unternehmensberatern (auch bei *JCs*) zweifelsohne vorhanden ist. Ähnliche Ergebnisse finden sich in europäischen Studien für Verträglichkeit (Salgado, 1997). Ebenso wurden für mündliche *Kommunikationsfähigkeit* als Persönlichkeitsdimension bei Managementberufen positive Zusammenhänge festgestellt (Sarges, 1994). Es ist allerdings auch eine Beeinflussung der Beurteiler durch *charismatische* und *kommunikative* Personen denkbar. Das würde bedeuten, dass tendenziell bessere Ratings, unabhängig von der tatsächlich erbrachten Leistung, gegeben werden. Eine Schwierigkeit besteht zudem in diesem Zusammenhang in der Abgrenzung des Begriffs *Charisma* und einem einheitlichen (Konstrukt-)Verständnis unter den Beurteilern ($R_{PR}<1$), so dass sie sich durch ein Globalurteil leiten lassen. Die analoge Argumentation (Generaleffekt vs. spezifisch) kann für die *VWL-Kenntnisse* (die als Verständnis allgemeiner wirtschaftlicher Zusammenhänge zu sehen sind) angewandt werden.

Die Korrelationsanalyse der Variablenebene aus Recruiting und Projektbeurteilung (faire Validitätstest) zeigt wiederum zahlreiche signifikante Zusammenhänge (*Anhang VIII-1*). Die Dimension *Specialist-Know-how* korreliert dabei auf Variablenebene mit *Fachkenntnissen* (r=.252), *VWL-* (r=.313) und *BWL-Kenntnissen* (r=.243) signifikant. Da diese Variablen Spezialwissen sind, entspricht dies den Erwartungen. Der Zusammenhang mit *Intellektueller Neugier* (r=.290) lässt sich dadurch erklären, dass die *Specialist-Know-how-Untervariablen* inhaltlich auch die *Entwicklung* von Spezialwissen beinhalten und diese von

einer höheren *intellektuellen Neugier* (im Sinne des *TIE*; Rolfhus & Ackerman, 1999) begünstigt wird. Interessant ist hier auch der fast signifikante hohe Zusammenhang mit *Exotenwissen (*r=.342), der die Vermutungen unterstreicht.

Bei *Quality of Work,* das mit einer Vielzahl an Variablen korreliert, handelt es sich um einen sehr breit gefassten Begriff, der eine sehr große Anzahl an Einflussfaktoren beinhaltet. Aus diesem Grund sind die zahlreichen Korrelationen nicht überraschend. Das breite Begriffsverständnis spiegelt sich auch in der untersten Erklärungsebene von *Quality of Work* wider, bei der u.a. die wenig präzisierten Begriffe *Effizienz* und *Zuverlässigkeit* der Arbeit genannt sind (siehe *Anhang VI*). Auf eine Interpretation der einzelnen Variablen wird aus diesem Grund verzichtet. Die Korrelationen von *Behaviour under Pressure* mit *Pragmatismus* (r=.355) und *Unternehmerischem Denken* (r=.367) scheinen plausibel, da in einer Stresssituation gerade solche Eigenschaften (pragmatische Herangehensweise und Blick auf das Wesentliche) das Entstehen guter Ergebnisse begünstigen. Gute *Methoden-* (r=.290) und *BWL-Kenntnisse* (r=.305) sind dabei sicher Voraussetzung für gute Leistungen auf den beiden o.g. Dimensionen.

Neben der *Selbstorganisation* und der *Teamorientierung* sind *kommunikative Fähigkeiten*, *Charisma* und *VWL-Kenntnisse* mit der Dimension *Know-how-Management* signifikant korreliert. Bei den letzteren drei Variablen sei wiederum auf deren vermutlichen Allgemeincharakter verwiesen. Dass zu einem guten *Wissensmanagement* auch die *Fähigkeit der Selbstorganisation* gehört, ist selbsterklärend. Die Korrelation von *Teamfähigkeit* (r=.245) lässt sich durch eine der Untervariablen bei *Know-how-Management*, die den Informationsaustausch mit Teammitgliedern beinhaltet, deuten. Mit der Dimension *Teamwork* korrelieren neben den allgemeinen *Variablen VWL, Charisma* und *Kommunikationsfähigkeit*, auch *Teamfähigkeit* (selbsterklärend). Oftmals sind hier Personen gefragt, die in kritischen Situationen die Verantwortung und Führung übernehmen, womit sich auch der Zusammenhang mit *Führungsfähigkeit* (r=.370) und *Leadership* (r=.345) deuten lässt. Eine Untervariable von *Teamwork* beinhaltet die aktive Suche nach neuen Lösungen, was den Zusammenhang mit *Intellektueller Neugier* (r=.349) sinnvoll erscheinen lässt. Allerdings könnte *Intellektuelle Neugier* auch eine höhere Symmetrieebene haben, da es auch mit dem *Gesamtwert der Projektbeurteilung* korreliert. Zudem beinhaltet *Teamwork* auch die Zusammenarbeit mit dem Service-Bereich, eine Komponente, die mit *Selbstorganisation* zu tun hat (wo besorgt man welche Informationen).

3.3.2.5 Regressionsanalyse

Zunächst wurden Regressionen der einzelnen Dimensionen der Projektbeurteilungen aus dem *Gesamtwert des Recruitings* und der *Abiturnote* separat sowie hierarchisch mit beiden zusammen vorhergesagt. Danach wurde zunächst jede *Projektbeurteilungsdimension* explorativ über eine schrittweise Regression vorhergesagt, mit und ohne *Abiturnote* als vorgestelltem Block, um maximale Varianz aufzuklären. Auf Grund hoher *Multikolinearität* wurden zusätzlich alle Variablen unter Hinzunahme der *Abiturnote* auch als *Set* von Indikatoren der Unterdimensionen der Projektbeurteilung in eine *blockweise Regression*, mit der *Abiturnote* als erstem Block, eingebracht. Bei der Vorhersage mit den *Sets* wurde der Beitrag der einzelnen Variablen nicht spezifiziert, sondern nur der inkrementelle Vorhersagebeitrag interpretiert. *Tabelle 38* zeigt die Ergebnisse im Überblick. In *Anhang VIII-2,3* und *4* finden sich die Detaillierungen zu den einzelnen Regressionsrechnungen. Die bivariaten regressionsanalytischen Befunde variieren dabei auf Grund der etwas kleineren Stichprobe (listenweiser Fallausschluss) leicht im Vergleich zu den bivariaten Korrelationen aus dem vorherigen Kapitel. Sie spiegeln aber im Wesentlichen die bereits erläuterten Ergebnisse wider, weshalb im Folgenden nur auf einzelne Aspekte nochmals eingegangen wird.

Die Regressionen zeigen erwartungsgemäß, dass der mit Abstand beste Prädiktor für die erbrachte Leistung insgesamt (R^2=.145) und auch auf den unteren *Projektbeurteilungsebenen*, die *Abiturnote* ist. Bei den Unterdimensionen klärt die *Abiturnote*, mit Ausnahme von *Client Communication* und *Know-how-Management*, auf allen Dimensionen Varianz zwischen rd. R^2=.07 und R^2=.17 auf. Lediglich bei *Behaviour under Pressure* ist die Varianzaufklärung des *Gesamtwertes aus Recruiting* überlegen (R^2=.17). Die regressionsanalytischen Zusammenhänge des *Gesamtwertes aus Recruiting* deuten darauf hin, dass das Auswahlverfahren sehr zentriert auf die drei Dimensionen *Quality of Work*, *Behaviour under Pressure* und *Teamwork* (fast signifikant) ist. Die Interviews werden weitgehend unter „Stressbedingungen“ durchgeführt, die dem Beurteiler eine gute Gelegenheit bieten, das Verhalten in so einer Situation einzuschätzen. Darüber hinaus können die Beobachter in der Gruppenarbeit und -diskussion offensichtlich die *Teamarbeit* beurteilen. Auch die Qualität der Arbeit, beispielsweise durch die Fallpräsentation, kann, wenn auch zu einem geringeren Maße, vorhergesagt werden. Zudem wurde dieser Dimension ein eher hohes Generalitätsniveau zugesprochen. Bei diesen Dimensionen hat der Gesamtwert entsprechend auch die höchsten Regressionsgewichte von β=.30 bis β=.40. Bei der Vorhersage von *Specialist-Know-how* und *Know-how-Management* sind diese etwas geringer (um β=.20) und relativ nahe an der 5% Signifikanzschwelle.

Bei der Vorhersage des *Gesamtwertes aus Projektbeurteilung* durch die *Abiturnote* und den *Gesamtwert aus Recruiting* ist das Gesamtmodell mit einem R^2 =.194 sehr signifikant. Lediglich das β-Gewicht des *Gesamtwertes aus Recruiting* verfehlt knapp die Signifikanzschwelle. Auch hier ist also vor dem Hintergrund der o.g. Restriktionen ein zusätzlicher wahrer Effekt nicht unwahrscheinlich. Das gleiche gilt wie oben für die Dimensionen *Specialist-Know-how* und *Know-how-Management.* Inkrementelle Varianz zur *Abiturnote* klärt der *Gesamtwert aus Recruiting* bei den Dimensionen *Quality of Work* und *Behaviour under Pressure* auf.

Tabelle 38: Gesamtübersicht der Regressionsergebnisse der JC-Kohorte (N=52)

Prädiktoren	**Abinote**	**Gesamtwert Recruiting**	**Abinote + Gesamtwert Recruiting**	**Recruitingvariablen (schrittweise)**	**Abinote + Recruitingvariablen (schrittweise)**	**Abiturnote + Sets (hierarchisch)**
Kriterienvariablen	R^2_{korr}	R^2_{korr}	R^2_{korr}	R^2_{korr}	R^2_{korr}	R^2_{korr} [a]
Durchschnittlicher Gesamtwert	.145**	.058**	.145**	.078*	.350**	.330**
Specialist-Know-how	.139**	.034	.139**	.059*	.201**	.371**
Problem Solution	.101*	-.007	.101*	-,-	.157**	.123*
Quality of Work	.100*	.078*	.168**	.115**	.239**	.192*
Behaviour under Pressure	.074*	.172**	.235**	.249**	.389**	.359**
Client Acceptance	.067*	-.020	.067*	.060*	.187**	.150**
Client Communication	.038	-.019	.019	-,-	.038	-,-
Know-how-Management	.025	.037	.057	.074*	.098	-,-
Teamwork	.071*	.053	.071*	.075*	.148**	.135*

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} :Schrumpfungskorrigierter Determinationskoeffizient
[a] Es sind lediglich die Werte signifikanter Blöcke dargestellt.

Bei dem explorativen Versuch der maximalen Varianzaufklärung durch schrittweise Regression mit den Recruitingvariablen zeigen sich auf allen Dimensionen leicht höhere signifikante Varianzaufklärungen als durch den *Gesamtwert aus Recruiting*. Wird jedoch die *Abiturnote* hinzugezogen, so steigt die Varianzaufklärung vereinzelt auf über R^2=.30 (beim *Gesamtwert* und *Behaviour under Pressure*, siehe *Tabelle 38*).

Die Erhöhung beruht größtenteils auf *Supressoreffekten* (z.B. *Konfliktfähigkeit* und *Frustrationstoleranz*), die teilweise das Signifikanzniveau einzelner Variablen verschieben sowie den Variablen *Charisma* und *VWL* mit einem vermutlich höheren Generalitätsniveau. In ihrer Höhe bemerkenswert sind diese Zuwächse bei *Quality of Work* und *Behavior under Pressure* von über 5%. Dies gilt ebenso bei Einbezug der *Abiturnote* in die schrittweise Reg-

ression. Die genannten *Suppressorkonstellationen* des Datensatzes führen zu keinen erheblichen Varianzverschiebungen in den Analysen, was die Vermutung erhärtet, dass sie *stichprobenspezifisch* sind oder durch einzelne Beurteilungen bewirkt wurden. Deshalb wird auf eine Interpretation der *Suppressoren* verzichtet. Ein weiteres, bereits angesprochenes Phänomen ist die sehr hohe *Multikolinearität* zwischen zahlreichen Variablen. Diese verletzt die Prämissen der multiplen Regression, führt zu einer Ineffizienz der β-Koeffizienten und einer eingeschränkten Interpretierbarkeit der Ergebnisse. Dem wurde versucht über *blockweise Regression* inhaltlich geclusterter Variablen (Sets) entgegenzutreten. Interpretiert wurden dabei lediglich die inkrementellen Varianzen unter dem Gesichtspunkt der Signifikanz des F-Wertes für den gesamten Block (*Anhang VIII–6*).

Die hierarchische explorative Regression mit Hilfe der Sets zeigt ebenfalls relativ hohe Varianzaufklärungen bei dem *Gesamtwert*, bei *Behaviour under Pressure* und bei *Specialist-Know-how* mit Werten um ebenfalls R^2=.30. Keine nennenswerten Effekte finden sich bei *Client Communication* und *Know-how-Management.* Allerdings ergibt sich bei *Specialist-Know-how* als einziger Dimension ein Validitätsgewinn mit den inhaltlich erwarteten aufgenommenen Blöcken *Abinote*, *Spezialwissen*, *Prozessuales Wissen* und *Soziale Fähigkeit*. Den höchsten inkrementellen Anteil leistet dabei das Set *Soziale Fähigkeit*. Offensichtlich spielen bei der Aneignung und Entwicklung von *Spezialwissen* nicht nur das vorhandene Know-how sondern auch die *sozialen Fähigkeiten* gegenüber anderen Mitarbeitern eine Rolle. Dies lässt sich vor dem Hintergrund der Teamarbeit auf den Projekten sowie dem Wissenstransfer von Berater zu Berater (Erfahrungsaustausch) als gängigste Quelle des Wissensaufbaus erklären. Wenig kommunikative, verschlossene Mitarbeiter haben hier größere Schwierigkeiten an die Information zu gelangen und bauen entsprechend weniger Wissen auf.

Zusammenfassend lässt sich feststellen, dass die gefundenen Zusammenhänge in ihrer Richtung zwar interpretierbar, in ihrer Ausprägung jedoch unter den Erwartungen geblieben sind. Die *Abiturnote* ist durchweg der beste Prädiktor und kann nur durch ein explorative Vorhersage (schrittweise Regression) übertroffen werden. Dabei werden aber nur sehr wenige Variablen aufgenommen. Diese sind zudem schwer zu interpretieren. Zusätzlich verwischen eine Vielzahl ebenso wenig interpretierbarer *Supressoreffekte* die eigentlichen Zusammenhänge. Neben der *Stichprobengröße* ist vor allem die hohe *Multikolinearität* sowohl auf der Prädiktoren- als auch auf der Kriterienseite problematisch. Durch sie sind die Personen auf den einzelnen Dimensionen wenig unterscheidbar und die Varianz reduziert. Dies kann an der verwendeten Operationalisierung der Konstrukte liegen, die im vorliegenden Fall sehr

anfällig gegenüber Beurteilungsfehlern ist (siehe auch 2.2.2.3). Zudem ist von einer *Verletzung der Symmetrieprinzipien* auszugehen, da auf Kriterien- wie Prädiktorenseite allgemeine und spezifische Variablen aggregiert werden und dies zu unfairen Validitätstests führt. Mit Hilfe der Faktorenanalyse sollten deshalb weitere Zusammenhänge aufgedeckt werden.

3.3.2.6 Faktorenanalyse

Auf die detaillierte Interpretation der Faktorenanalyse der *Prädiktorendaten* wird an dieser Stelle verzichtet und auf 3.3.1.5 verwiesen. *Tabelle 39* zeigt die korrelativen Zusammenhänge der *Recruitingfaktoren* mit den *Projektbeurteilungsdimensionen*.

Tabelle 39: Korrelationsmatrix der Faktorwerte der Recruitingfaktoren mit Projektbeurteilungsdimensionen (Spearman-Rho; N=54)

	Extrahierte rotierte Faktoren			**Extrahierte unrotierte Faktoren[a]**	
Kriterienvariablen	Allgemeinfaktor	Sozial-kompetenz	Problemlösungsfähigkeit	Levelfaktor	Profilfaktor
Durchschnittlicher Gesamtwert	.258*	.214	.029	.272*	.038
Specialist-Know-how	.198	.037	.164	.226	-.085
Problem Solution	-.042	.182	.170	.104	.132
Quality of Work	.258*	.320**	.039	.341**	.139
Behaviour under Pressure	.318**	.145	.128	.362**	-.059
Client Acceptance	.031	.079	-.101	-.050	.030
Client Communication	.067	-.021	-.022	.022	-.065
Know-how-Management	.303*	.138	-.063	.239*	-.048
Teamwork	.294*	.306*	-.061	.301*	.163

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant (einseitig)
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant (einseitig)
[a] Von den 3 extrahierten Komponenten sind die ersten 2 dargestellt.

Signifikante Korrelationen zeigen sich beim Allgemeinfaktor (rotiert) und dem Levelfaktor (unrotiert) erwartungsgemäß mit dem *Gesamtwert aus Projektbeurteilung* und den generellen Dimensionen *Quality of Work* und *Behaviour under Pressure*. Zudem sind die Korrelationen zu *Know-how-Management* und *Teamwork* signifikant. Inhaltlich kann dies bedeuten, dass eine allgemeine Komponente (u.U. allgemeine Intelligenz *g* oder eine *Globalurteilstendenz*) eine gute Leistung auf diesen Dimensionen begünstigt. Keine Zusammenhänge gibt es mit *Specialist-Know-how* und *Problem Solution,* von denen eher anzunehmen ist, dass sie von *Fähigkeiten* wie *fluider* und *kristalliner Intelligenz* beeinflusst werden. Dar-

über hinaus zeigen sich ebenfalls keine Zusammenhänge zu den *klientenzentrierten* Dimensionen (siehe dazu 3.3.2.4). Der rotierte Recruitingfaktor *Sozialkompetenz* korreliert signifikant mit *Teamwork* (r=.306). Da auch eine fast signifikante Korrelation mit dem *Projektbeurteilungsgesamtwert* (r=.214) vorliegt, ist davon auszugehen, dass *Sozialkompetenz* sich auch günstig auf die Gesamtleistung (Level) auswirkt. Dadurch korreliert *Sozialkompetenz* ebenfalls mit dem sehr generellen Konstrukt *Quality of Work* (r=.320). Keine erkennbaren Zusammenhänge gibt es beim rotierten *Problemlösungsfaktor* aus Recruiting (zur Erklärung siehe ebenfalls 3.3.2.4). Darüber hinaus ist die Korrelation zwischen dem *unrotierten Profilfaktor* aus Recruiting und *Teamwork* fast signifikant. Da hier von einem *verbal-numerischen*-Kontinuum auszugehen ist, bedeutet dies, dass hohe Werte auf dem Profilfaktor (hohe *verbale-soziale* Fähigkeiten) mit guter *Teamarbeit* einhergehen.

Tabelle 40: Komponentenmatrizen der Projektbeurteilungsvariablen (N=61)

	Extrahierte unrotierte Faktoren[a]		**Extrahierte rotierte Faktoren[a,b]**	
Kriteriendimensionen	1	2	1	2
Specialist-Know-how	.757		.751	.304
Problem Solution	.794		.410	.726
Quality of Work	.835		.509	.679
Behaviour under Pressure	.598	.585		.835
Client Acceptance	.729	-.464	.850	
Client Communication	.671	-.355	.733	
Know-how-Management	.732		.625	.403
Teamwork	.741		.352	.710
VA[c]	54.1%	65.9%	34.6%	65.9%

Faktor unrotiert	1	2	3	4	5
Eigenwert	4.326	.949	.664	.540	.498

Dargestellt sind nur Faktorladungen mit Betrag >.300.
a Extraktionsmethode: Hauptkomponentenanalyse.
b Rotationsmethode: Varimax mit Kaiser-Normalisierung. Die Rotation ist in 3 Iterationen konvergiert.
c Varianzaufklärung, kumulativ dargestellt

Die *Faktorenanalyse* der *Projektbeurteilungsdimensionen* (*VARIMAX-Rotation)* führt wiederum zu zwei Faktoren, die nicht eindeutig interpretierbar sind (*Tabelle 40*). Der zweite rotierte Faktor beinhaltet dabei stärker *Problem Solution*, *Teamwork* und *Behaviour under*

Pressure. Dabei korreliert keine der *Recruitingvariablen* mit dem ersten und fast alle Variablen mit dem zweiten Faktor (*Tabelle 41*).

Tabelle 41: Korrelationsmatrix Projektbeurteilungsfaktoren mit Recruitingvariablen (Spearman-Rho; N=52)

	Extrahierte unrotierte Faktoren		**Extrahierte rotierte Faktoren**	
Recruitingvariablen	Levelfaktor Projektbeurteilung	Profilfaktor Projektbeurteilung	F1 rotiert	F2 rotiert
MeanRecr	.283*	.318*	-.073	.440**
Team	.187	.317*	-.105	.368**
Konfl	-.019	.209	-.215	.106
Komm	.224	.265*	-.013	.369**
Charis	.367**	.254*	.054	.438**
VWL	.282*	-.028	.220	.148
Allg	.145	.104	.070	.162
Pragma	.145	.332*	-.159	.334**
Unt	.232*	.177	-.015	.303*
LogAn	.112	.329**	-.144	.307*
Meth	.158	.313*	-.104	.326**
Kreat	.106	.184	-.073	.180
Fach	.047	.100	-.040	.121
BWL	.222	.292*	-.035	.357**
Intell	.233*	.052	.087	.233**
Frust	.103	.180	-.063	.250*
Abinote	-.402**	-.061	-.249*	-.255*

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant (einseitig)
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant (einseitig)

Bei der *unrotierten Faktorenlösung* der *Projektbeurteilungsvariablen* zeigt sich, dass der erste Faktor wie bei den Recruitingvariablen das Gesamtlevel der Leistung widerspiegelt während der zweite Faktor die Profilstruktur abbildet (siehe *Tabelle 40* und *Abbildung 47*). Die Ergebnisse der Faktorenanalyse sprechen für den in *Fragestellung (6)* vermuteten Generalfaktor beim Vorgesetztenurteil. Der unrotierte *Levelfaktor der Projektbeurteilungen* korreliert dabei in nahezu identischer Konstellation mit den *Recruitingvariablen* wie der *Gesamtwert aus der Projektbeurteilung*. Dies unterstreicht seinen Level-Charakter (*Tabelle 41*).

Die Extraktion von einem *Generalfaktor*, der einen Großteil der Varianz aufklärt, ist bei Leistungsbeurteilungen durch Vorgesetzte auch metaanalytisch berichtet worden (Viswesvaran, et al., 2005). Sie wird neben den klassischen Urteilsfehlern (insb. *Halo-Effekt*) darauf zurückgeführt, dass symmetrisch sehr generell angesiedelte Faktoren (wie *g* oder die Persönlichkeitseigenschaft *Gewissenhaftigkeit*) den Berufserfolg auf nahezu allen Dimensionen beeinflussen und somit zu hohen Interkorrelationen führen (Feldmann, 1981; Motowidlo, et al., 1997). Hulin (1982) weist zudem darauf hin, dass die meisten Fähigkeiten positiv zusammenhängen und somit Personen, die Stärken auf einer Fähigkeit haben, automatisch auch in anderen Bereichen stark sind. Diese Vermutungen werden im vorliegenden Fall dadurch erhärtet, dass der Levelfaktor mittlere Korrelationen zur *Abiturnote* hat (r=-.402). Letztere hat sicherlich Überlappungen mit der allgemeinen Intelligenz. Der zweite unrotierte Profilfaktor zeigt hingegen keine korrelativen Zusammenhänge mit der *Abiturnote* aber mittlere signifikante Korrelationen mit zahlreichen Recruitingvariablen (*Tabelle 41*). Zudem werden offenbar durch das Recruiting die Profile besser vorhergesagt als das allgemeine Leistungsniveau, da auch die Korrelationen des unrotierten Profilfaktors (Faktor 2 aus *Tabelle 41*) mit dem *Gesamtwert aus dem Recruiting* höher ausfallen als die des Levelfaktors (r=.318 vs. r=.283). Der *zweite rotierte Faktor* der Projektbeurteilungen scheint am stärksten jene Komponenten abzubilden, die durch das *Recruitingverfahren* vorhergesagt werden können (hohe Korrelation von r=.440 mit Durchschnittswert aus Recruiting). Auch hier gibt es eine signifikante Korrelation mit der *Abiturnote* von r=.249.

Personen unterscheiden sich bei dem Vergleich der Randquartile des Profilfaktors (*Abbildung 47*) hauptsächlich in zwei Bereichen: *Problem Solution*, *Behaviour under Pressure* und *Quality of Work* auf der einen und klienten- und wissenszentrierten Bereichen auf der anderen Seite. Eine gewisse Ähnlichkeit zu den Dimensionen *quantitatv-numerisch-* vs. *verbal-geprägt* ist hier nicht von der Hand zu weisen.

In einem weiteren Schritt wurden Regressionen mit den Faktorwerten gerechnet. Die schrittweise Regressionsanalyse (siehe *Anhang VIII-7*) zeigt entsprechend erhebliche Validitätsgewinne gegenüber den Variablen der *Recruiting-* und *Projektbeurteilungsseite* durch den *Levelfaktor aus dem Recruiting* bei der Vorhersage des zweiten *unrotierten Projektbeurteilungsfaktors*. Bei den restlichen Kombinationen zeigen sich hingegen keine Validitätszugewinne. Die Korrelationsanalysen hatten erste Indizien dafür geliefert, dass es sich bei den gemessenen Variablen um Variablen unterschiedlicher *Symmetrieebenen* handeln könnte. So zeigte beispielsweise *Charisma* eine höhere Korrelation mit dem *Gesamtwert aus Projektbeurteilung* (r=.335) als der *Gesamtwert aus Recruiting* (r=.276). Deshalb wurde eine zweite

Faktorenanalyse gerechnet, in der die unterschiedlichen vermuteten Generalitätsebenen auf der Kriterienseite berücksichtigt und die problematischen klientenzentrierten Dimensionen *Client Communication* und *Client Acceptance* ausgeschlossen wurden.

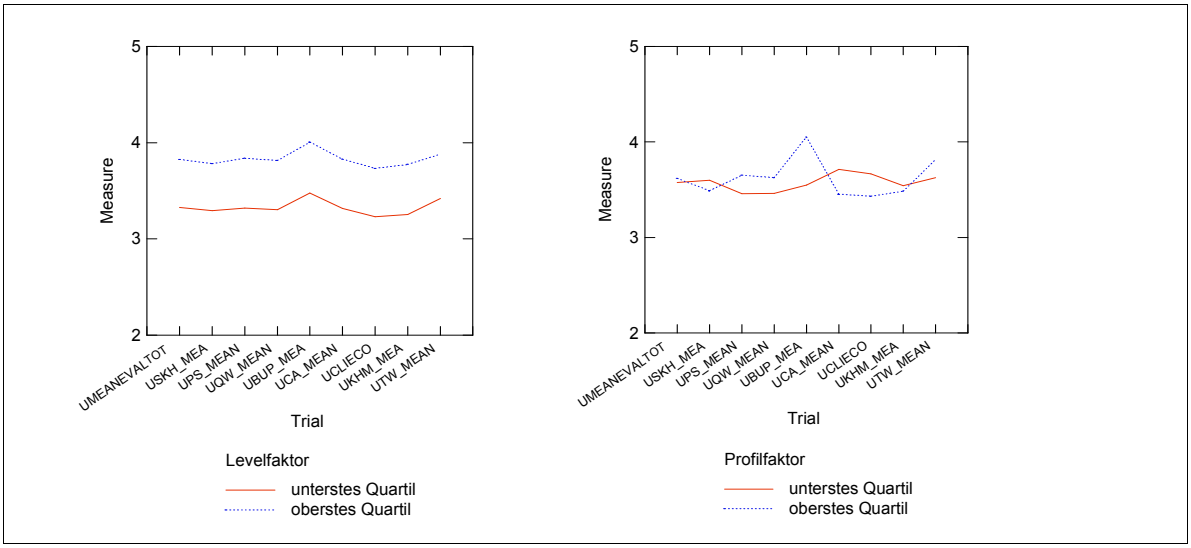

Abbildung 47: Profilvergleich Projektbeurteilung Randquartile Level- und Profilfaktor (unrotiert)

Tabelle 42: Korrelationen ausgewählter Recruitingvariablen mit verschiedenen Kriterien (Spearman-Rho; N=52)

	Unt	Intell	Komm	Charis	VWL
Beratungserfolg-Faktor	.312*	.271*	.324**	.441**	.293*
Unrotierter Levelfaktor	.232*	.233*	.224	.367**	.282*
Gesamtwert Projektbeurteilung	.232*	.235*	.237*	.366**	.305*

Die neu extrahierten Faktoren auf der Kriterienseite zeigten höhere Korrelationen mit den relevanten Prädiktorvariablen als der ursprüngliche *unrotierte Levelfaktor* oder der *Gesamtwert aus der Projektbeurteilung*. Die Neuordnung der Kriterienseite und erneute Faktorisierung (*Berufserfolg-Faktor)* führte folglich zu einem Validitätsgewinn für den die nun passende Symmetrie der betrachteten Variablen ursächlich sein kann.

3.3.3 Analyse der Kohorte Consultants

Ausgangspunkt des Vorgehens in diesem Abschnitt ist erneut die deskriptive Analyse der Prädiktoren- und Kriterienvariablen. Im Anschluss werden basierend auf ausgewählten *CV-Variablen* t-Tests für unabhängige Stichproben auf Mittelwertunterschiede und deren Effektstärken auf allen Projektbeurteilungsdimensionen gerechnet. Ziel dabei ist es, auch bei

der *C-Kohorte* signifikante Profilunterschiede zwischen den untersuchten Gruppen festzustellen und somit Anhaltspunkte für eine Verbesserung der Vorauswahl von Bewerbern zu gewinnen. Zudem wird erneut eine Varianzanalyse mit Messwiederholung durchgeführt. Sie soll Auskunft darüber geben, inwieweit sich die Bewerber mit zunehmender Funktionsstufenzugehörigkeit auf den Projektbeurteilungen verbessern. In einem weiteren Schritt wird über Korrelationsanalysen die Validität des Auswahlverfahrens und der *CV-Variablen* für den Berufserfolg ermittelt. Die anschließenden Regressionsanalysen sollen aufdecken, in welcher Form die Prädiktoren aus dem Recruiting optimal kombiniert werden sollten. Ferner werden im Rahmen von Faktoren- und Pfadanalysen Anhaltspunkte für die Konstruktvalidität des gemessenen Kriteriums gewonnen. Dabei wird die Symmetrie von Prädiktor und Kriterium untersucht, um Erkenntnisse bei der Untersuchung der Generalitätsebenen von Prädiktoren und Kriterien zu erlangen.

3.3.3.1 Methodische Restriktionen

Die Vielzahl der unter 3.3.2.1 bereits erwähnten Restriktionen gelten ebenfalls bei der Interpretation der Ergebnisse der *C-Kohorte* und sollen hier nur nochmals kurz aufgezählt werden: geringer *Stichprobenumfang* und geringe *Teststärken*, Gefahr des *Selektionsfehlers* und einhergehende *Varianzreduktion*, hohe Anfälligkeit der Analysen für *Ausreißerwerte* und *fehlende* Werte, potenzielle *α-Fehlerinflation*, hohe *Multikolinearität* zwischen den Variablen und der mit der Faktorisierung einhergehende *Varianzverlust*. Zusätzlich zu diesen Restriktionen ist von einem Varianzverlust durch *Drop-Out-Effekte* auszugehen. Das heißt, dass die *C-Kohorte* bereits um ausgeschiedene Mitarbeiter, die als *JC* in dem Unternehmen eingestiegen sind und wegen schlechter Leistung das Unternehmen verlassen mussten, bereinigt ist. Die *„überlebenden"* ehemaligen *JCs* dürften damit durchschnittlich besser sein als die Ursprungskohorte und eine geringere Varianz besitzen, was zu einer Effektunterschätzung führen kann.

3.3.3.2 Deskriptive Analyse

Bei den *Recruitingvariablen* und den *CV-Variablen* wurden auch in der *C-Kohorte* die Erkenntnisse aus der kohortenübergreifenden Prädiktorenanalyse weitgehend repliziert. Dies ist nicht verwunderlich, da es sich bei den Stichproben zu einem großen Teil (73%) um identische Personen handelt. Die deskriptiven Kennwerte zeigen bei den *Recruitingvariablen* hohe Mittelwerte und geringe Standardabweichungen sowie, analog zu den *JCs*, vereinzelt kleine Fallzahlen. Die Gründe für die in dem deskriptiven Teil der *C-Kohorte* festgestellte geringe Varianz und die Verteilungseigenschaften der *Prädiktorvariablen* wurden bereits

unter 3.3.1.2 beschrieben und gelten auch hier. Die *CV-Variablen* weisen teilweise hohe Standardabweichungen (durch Ausreißer, *Abbildung 48*) und einzeln geringe Fallzahlen auf.

Der *Gesamtmittelwert aus Projektbeurteilungen* (*Tabelle 43*) ist mit 3.63 leicht höher als bei der *JC-Kohorte* und ebenfalls weit über dem theoretisch zu erwartenden Wert von 3.0, bei dem laut Definition alle Erwartungen durch den *C* voll erfüllt werden. Da bei der Mehrheit der *Cs* ebenso wenig von einem Übertreffen der Erwartungen bzw. im Umkehrschluss von einer niedrigen Erwartungshaltung der Beurteiler auszugehen ist, ist ebenfalls zu vermuten, dass die Handhabung des Beurteilungsinstrumentes auf dieser Stufe wenig streng ist. Zudem zeigt eine geringe Standardabweichung aller Dimensionen die geringe Differenzierung bei den evaluierten Mitarbeitern, die auf die gleichen klassischen Urteilsfehler und auf die Aggregation zurückzuführen ist. Von nahezu allen Mitarbeitern sind auf allen Dimensionen Urteilswerte vorhanden. Bei den Recruitingvariablen gibt es, wie bei den *JCs,* eine Vielzahl an Ausreißerwerten. Bei den Kriterienvariablen halten sich die Ausreißer in Grenzen (*Abbildung 48*).

Tabelle 43: Deskriptive Kennwerte der Kriterien (Projektbeurteilungen) bei *Cs*

Kriterienvariablen	N	MW	Med.	Min.	Max.	SD	NV (j/n)
Specialist-Know-how	117	3.59	3.58	3.00	4.69	.29	j
Problem Solution	117	3.63	3.67	2.96	4.44	.25	j
Quality of Work	117	3.58	3.61	2.75	4.33	.23	j
Behaviour under Pressure	117	3.78	3.80	2.67	4.40	.30	j
Client Acceptance	116	3.61	3.61	3.00	4.50	.32	j
Client Communication	117	3.62	3.63	2.75	4.60	.33	j
Know-how-Management	117	3.51	3.50	2.94	4.13	.23	j
Teamwork	117	3.68	3.70	2.75	4.25	.26	j
Durchschnittlicher Gesamtwert	117	3.63	3.65	2.90	4.28	.20	j

N= Stichprobenumfang; MW=Mittelwert; Med.=Median; Min.=Minimum; Max.=Maximum; SD=Standardabweichung; NV=Normalverteilungsannahme

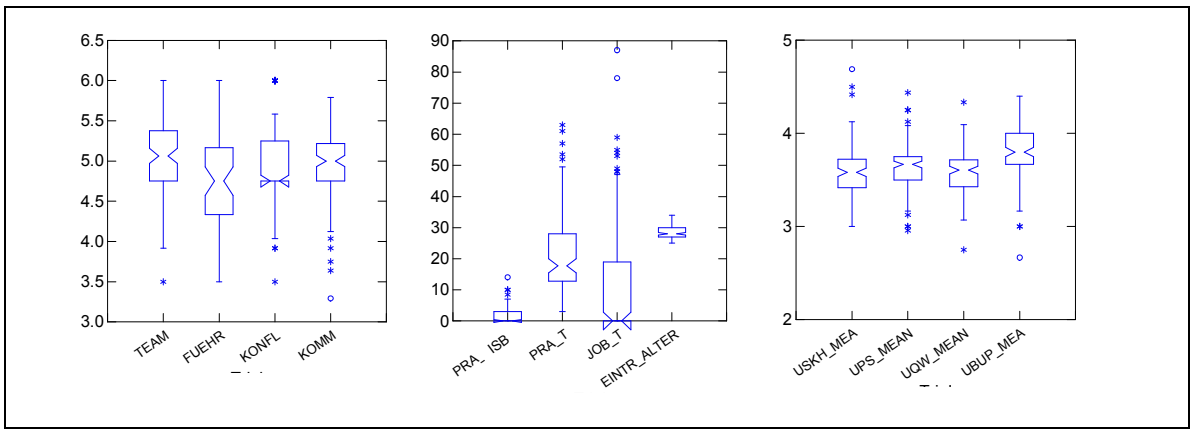

Abbildung 48: Boxplotbeispiele für Recruiting-, CV- und Projektbeurteilungsvariablen

Ebenso weichen einige Recruitingvariablen signifikant von der Normalverteilung ab (z.B. *Pragmatismus*, *Selbstorganisation* und *Teamfähigkeit*). Die visuelle Inspektion der Histogramme zeigt auch hier, dass die Abweichungen von der Normalverteilung hauptsächlich in einer schmalgipfligen, rechtslastigen Verteilung um den Wert 5.0 liegen. Bei den *CV-Variablen* ist lediglich bei der *Abiturnote* und bei der *Ausbildungsnote* von der Normalverteilungsannahme auszugehen. Die *Kriterienvariablen* weichen nicht signifikant von der Normalverteilung ab.

3.3.3.3 Mittelwert- und Profilvergleiche

Die Ergebnisse aus den Mittelwert- und Profilvergleichen ergänzen die Erkenntnisse zur Optimierung der Vorauswahl der Bewerber anhand biographischer Informationen (*Fragestellung (5)*). Einige Befunde der *JC-Kohorte* konnten zudem erhärtet werden, da auf Grund der größeren Stichprobe höhere Teststärken vorliegen. Zudem ist auf der Funktionsstufe der *Cs* wegen des breiteren Aufgabenspektrums eher eine tatsächliche Einschätzung der Leistung auf den Dimensionen möglich. Allerdings ist die Stichprobe, wie oben erwähnt, größtenteils identisch, so dass ähnliche Ergebnisse zu erwarten waren.

Tabelle 44 zeigt die Effektstärken der explorativen t-Tests bei der Analyse der *C-Kohorte* mit ausgewählten *CV-Variablen*. Die dargestellten Ergebnisse wurden zur Interpretationserleichterung als Profilvergleiche visualisiert (siehe *Abbildung 49*). Jene *Cs*, die bereits als *JCs* in das Unternehmen eingestiegen sind, sind auf allen Dimensionen der Projektbeurteilungen besser. Signifikante mittlere Effektgrößen finden sich auf den Dimensionen *Problem Solution* und *Quality of Work*. Dieses Ergebnis ist vor allem vor dem Hintergrund der höheren Erfahrung der *Cs*, die bereits als *JC* gearbeitet haben, zu erklären. Dadurch ist davon auszugehen, dass sie ein höheres Berufswissen (*job knowledge*) besitzen. Direkteinsteiger unter den *Cs* haben gerade anfangs stärker Gewöhnungsschwierigkeiten und

mit der „Akklimatisierung“ zu kämpfen. Sie kennen die Arbeitsstandards sowie die Art und Weise wie Probleme gelöst werden nicht und machen übliche „Anfängerfehler“. Demgegenüber sind ehemalige *JCs* im Vorteil. Deshalb sind von Ihnen höhere *Quality of Work*- (sehr breite Begrifflichkeit) und bessere *Problem Solution*-Werte zu erwarten. Zudem haben sich diese Dimensionen bei den *JCs* als vermutlich bis zu einem gewissen Grad erlernbar herausgestellt, da sie sich im Zeitverlauf verbessern. Zu berücksichtigen ist allerdings dabei, dass die Kohorte jener *Cs*, die als *JC* eingestiegen sind, um jene Personen bereinigt ist, die wegen schlechter Leistung das Unternehmen bereits verlassen mussten (*Drop-Out-Effekt*). Dies bedeutet, dass die „Überlebendengruppe“ tendenziell bessere Leistungen bringen dürfte und somit ein Teil des Effektes darauf zurückzuführen ist.

Tabelle 44: Explorative t-Tests ausgewählter CV-Kategorien für Cs (Effektstärken)

	Gruppen				
Kriterienvariablen	Einstieg als JC (ja vs. nein)	Angestellten-status (angestellt vs. ausgeschieden)	Forced Exit (ja vs. nein)	Universitätsart (öffentlich vs. privat)	Familien-stand (ledig vs. verheir.)
Specialist-Know-how	.132	.170	-.666*	-.460*	-.308
Problem Solution	.470*	.339	-.631*	-.646**	.644
Quality of Work	.428*	.437*	-.740*	-.324	.925*
Behaviour under Pressure	.084	.556**	-.758*	-.134	.601
Client Acceptance	.267	.329	-.184	-.013	.526
Client Communication	.334	.347	-.535	-.203	.425
Know-how-Management	.273	.024	-.203	-.599**	.152
Teamwork	.118	.284	-.413	-.346	1.135*
Durchschnittlicher Gesamtwert	.366	.439*	-.740*	-.459*	.730

Basis der Effektgrößen d ist der Mittelwertunterschied = 1.Gruppe – 2. Gruppe
* Der t-Test ist auf dem 0.05 Niveau signifikant (zweiseitig)
** Der t-Test ist auf dem 0.01 Niveau signifikant (zweiseitig)
Zur Anzahl der Fälle siehe Tabelle 20

Die Erkenntnisse aus der *JC-Kohorte* bezüglich der wesentlich schlechteren Leistung ausgeschiedener Mitarbeiter (*Unforced* und *Forced Exits*) konnten auch in der *C-Kohorte* repliziert werden. Demzufolge haben *ausgeschiedene* Mitarbeiter signifikant schlechtere Beurteilungen als solche, die noch im Angestelltenverhältnis sind (mittlere Effekte). Dies gilt ebenso für die forciert Ausgeschiedenen im Vergleich zu allen anderen Mitarbeitern (starke Effekte). Die Vermutung einer effektiven Personalpolitik, bei der tendenziell schlech-

tere Mitarbeiter das Unternehmen verlassen und die bereits bei der *JC-Kohorte* festgestellt wurde, wird durch dieses Ergebnis gestützt.

Absolventen von *Privatuniversitäten* sind hingegen auf allen Dimensionen leicht besser, besonders bei *Problem Solution*, *Specialist Know-how, Know-how-Management* sowie dem *Gesamtwert aus Projektbeurteilung*. Die teilweisen starken Effektgrößen beruhen im Gegensatz zur *JC-Kohorte* auf signifikanten Mittelwertunterschieden. Dies wurde bereits bei den *JCs* auf das bessere Lernumfeld und die Arbeit mit Fallstudien zurückgeführt. Verheiratete *Cs* sind mit Ausnahme von *Specialist-Know-how* auf allen Dimensionen schlechter als ledige Kollegen, bei *Quality of Work* und *Teamwork* signifikant (starke Effekte). Diese Effekte sind vor dem Hintergrund zu erklären, dass sich der hohe Zeitaufwand, der mit dem Beruf des Beraters einhergeht, schlecht mit einer privaten Beziehung vereinbaren lässt. Entsprechend kann sich dies auf die Qualität der Arbeit auswirken. Die schlechtere Beurteilung im Teamverhalten kann sich auf Freizeitaktivitäten (am Projektstandort) beziehen, die ein Verheirateter vielleicht eher nutzt, um in Kontakt mit seinem Partner zu treten, anstatt sie mit den anderen Teammitgliedern zu verbringen.

Tabelle 45: Gruppenvergleich (t-Test) biographischer Kategorien für Cs (Effektstärken)

	Gruppen			
Kriterienvariablen	Mind. ein analyt. FSP (ja vs. nein)	BWL-Studium (ja vs. nein)	MBA (ja vs. nein)	Dr.-Titel (ja vs. nein)
Specialist-Know-how	-.086	.289	.556*	-.299
Problem Solution	-.088	.454*	.721**	-.366
Quality of Work	-.313	.355	.496	-.653*
Behaviour under Pressure	.017	.199	.446	-.599*
Client Acceptance	-.238	.093	.332	-.176
Client Communication	-.116	.282	.275	-.129
Know-how-Management	-.485*	.312	.436	-.521
Teamwork	-.107	.246	.366	-.357
Durchschnittlicher Gesamtwert	-.235	.381	.616*	-.538

Basis der Effektgrößen d ist der Mittelwertunterschied = 1.Gruppe – 2. Gruppe
* Der t-Test ist auf dem 0.05 Niveau signifikant (zweiseitig)
** Der t-Test ist auf dem 0.01 Niveau signifikant (zweiseitig)
Zur Anzahl der Fälle siehe Tabelle 21

Bei den anschließenden Gruppenvergleichen weiterer *biographischer* Kategorien (*Tabelle 45*) schneiden Personen ohne *analytischen FSP* beim *Know-how-Management* signifi-

kant besser ab (mittlere Effektstärke, siehe auch *Abbildung 49*). Gründe können darin liegen, dass gerade solche Mitarbeiter eher auf externes Know-how angewiesen sind, um sich weiterzuentwickeln und deshalb in dieser Dimension mehr Anstrengungen vornehmen als die eher analytisch geprägten Kollegen. Das Vorhandensein eines *BWL-Abschlusses* führt ebenfalls zu durchweg besseren Beurteilungen, insbesondere bei *Problem Solution.* Es erleichtert offensichtlich die Arbeit als *C*. Dieses Ergebnis war bei der Kohorte der *JCs* nicht vorzufinden und wurde mit der mangelnden Praxisrelevanz der Studiumsinhalte begründet. Für prozedurales Wissen, das sicherlich durch *Problem Solution* teilweise abgedeckt wird, scheint dies offensichtlich nicht zu gelten. Möglich wäre, dass *JCs* kaum Gelegenheit zur Lösung „richtiger" Probleme gegeben wird, da dies, wie erläutert, erfahrene Kollegen übernehmen und damit diese Dimension nicht beurteilbar ist. Folglich zeigt sich bei ihnen auch kein Effekt. Bei den *Cs* hingegen scheint das deklarative Wissen über betriebswirtschaftliche Zusammenhänge bei der Lösung von Problemen bedeutsam zu sein.

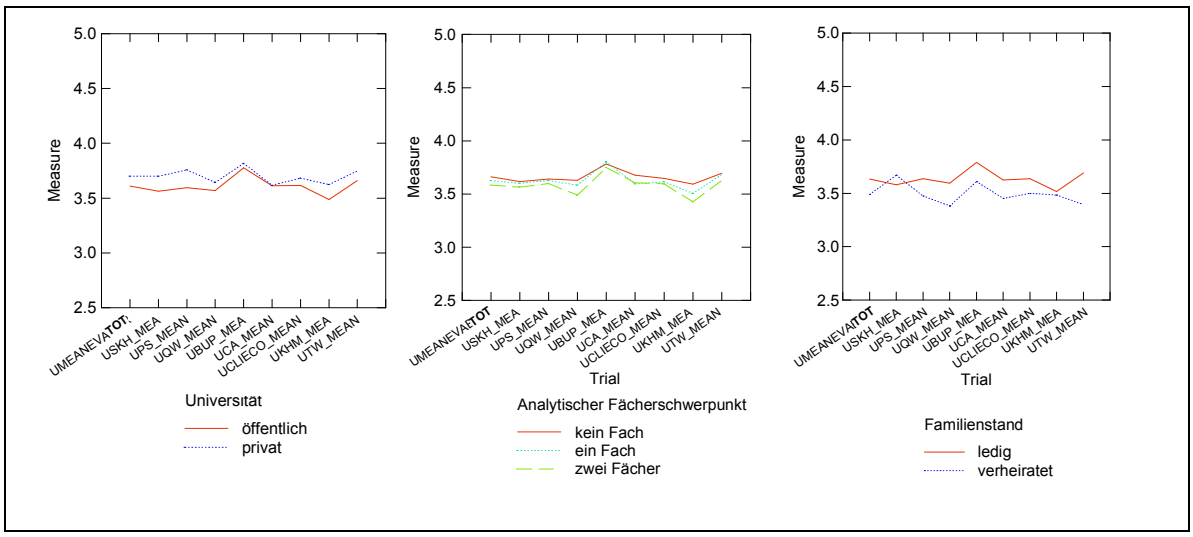

Abbildung 49: Beispielhafte Profilvergleiche der Projektbeurteilungen bei Cs

Erwartungsgemäß haben *MBA-Absolventen* durchgängig bessere Beurteilungen als Nicht-Absolventen. Signifikante Unterschiede gibt es auf den Dimensionen *Specialist-Know-how*, *Problem Solution* und dem *Gesamtwert*. Da beim *MBA-Studium* eine starke Spezialisierung stattfindet und die Arbeit einen hohen Praxischarakter und eine starke Affinität zur Beratung aufweist (ebenfalls Fallstudien) sind die Ergebnisse, die bereits bei den *JCs* gefunden wurden, plausibel. Überraschend ist allerdings zunächst das Ergebnis, dass promovierte *Cs* durchgängig schlechtere Beurteilungen aufweisen (signifikante starke Effekte bei *Quality of Work* und *Behaviour under Pressure*). Eine mögliche Erklärung kann darin liegen,

dass promovierte *Cs* direkt als *C* einsteigen und somit insgesamt mit den Neueinsteigerproblemen zu kämpfen haben (s.o.), die *Cs* mit *JC*-Erfahrung nicht haben (indirekter Effekt).

Tabelle 46: Übersicht Mittelwertunterschiede im Zeitverlauf (Glass'sche Deltas) (N=76)

	Messzeitpunkte		
Kriterienvariablen	**1 vs. 2**	**1 vs. 3**	**2 vs. 3**
Specialist-Know-how	.160	-.175	-.381*
Problem Solution	-.031	-.299*	-.295*
Quality of Work	-.120	-.340*	-.220
Behaviour under Pressure	.148	-.149	-.304
Client Acceptance	-.011	-.230	-.218
Client Communication	.003	-.332*	-.349*
Know-how-Management	.055	-.077	-.174
Teamwork	-.104	-.110	-.007
Durchschnittlicher Gesamtwert	.021	-.315*	-.412*

* Signifikanter Effekt auf 0.05 Niveau (Paarvergleiche der ANOVA)

Abschließend wurden erneut die Mittelwerte der Projektbeurteilungen im Zeitverlauf dargestellt und über eine einfaktorielle Varianzanalyse mit Messwiederholung auf Unterschiede getestet, um mögliche Lerneffekte zu visualisieren. *Tabelle 46* bildet die Effektstärken (Glass'sche Deltas) der Mittelwertunterschiede ab. Die Ergebnisse zeigen, dass die Befunde der *JC-Untersuchung* nur teilweise repliziert werden konnten. So verbessert sich, vermutlich bedingt durch einen Lerneffekt, der *Durchschnittswert der Projektbeurteilung* signifikant mit zunehmender Unternehmenszugehörigkeit. Das Gleiche gilt für die Dimensionen *Problem Solution*, *Quality of Work* und *Client Communication*, Ergebnisse, die auch bei der *JC-Kohorte* vorzufinden waren. Auch für die scheinbar nicht erlernbaren Dimensionen *Client Acceptance* und *Behaviour under Pressure* (die möglicherweise an unveränderbare Persönlichkeitseigenschaften gekoppelt sind) wurden, wie bei der *JC-Kohorte*, keine signifikanten Unterschiede im Zeitverlauf festgestellt. Neu ist hingegen bei den *Cs*, dass sich die Dimensionen *Know-how-Management*, *Teamwork* und *Specialist-Know-how* über die Zeit nicht signifikant ändern. Dies war bei den *JCs* noch der Fall. Bei *Teamwork* ist anzunehmen, dass es sich um eine relativ unveränderbare Dimension handelt. Die signifikanten Effekte bei den *JCs* könnten daran liegen, dass bei den Unterdimensionen dennoch einige Variablen enthalten sind, die mit Lernfortschritt gerade am Anfang der beruflichen Laufbahn (also nach dem Einstieg) in Verbindung stehen und somit bei den *JCs* von größerer Bedeu-

tung sind. Das gleiche kann für *Specialist-Know-how* und *Know-how-Management* angenommen werden, weil sich diese während der *JC-Phase* in einem stärkeren Maße entwickeln, da von einer niedrigeren Ausgangsbasis ausgegangen werden muss.

3.3.3.4 Korrelationsanalyse

Die korrelativen Befunde zeigen auch bei der *C-Kohorte* sowohl für die Dimensionen des Auswahlverfahrens als auch für zahlreiche biographische Informationen nennenswerte prognostische Validitäten (*Fragestellungen (8)* und *(9)*). Bei den Zusammenhängen zwischen ausgewählten *CV-Variablen* und den *Dimensionen der Projektbeurteilung* sind die Ergebnisse allerdings ernüchternd (*Tabelle 47*).

Tabelle 47: Korrelationen Projektbeurteilungsdimensionen mit ausgewählten CV-Variablen (Cs)

Korrelationen (Spearman-Rho)								
	Eintritts-alter	Note-_FSP	Note_D	Note_M	Note_E	Abitur-note	Job_t	Pra_t
Gesamtwert Projektbeurteilung	-.148	.080	-.092	.179	-.094	.072	-.265**	.108
Specialist-Know-how	-.124	.056	-.083	.045	-.115	.063	-.102	.120
Problem Solution	-.264**	-.095	-.190*	-.089	-.123	-.087	-.320**	.107
Quality of Work	-.207*	-.015	-.135	-.011	-.194	-.044	-.296**	.184*
Behaviour under Pressure	-.075	-.055	.034	.143	-.017	.108	-.213*	.079
Client Acceptance	-.014	.211*	.039	.264*	.058	.121	-.097	.227**
Client Communication	-.083	.141	.035	.138	-.090	.147	-.200*	.051
Know-how-Management	-.232**	-.032	-.146	.028	-.102	-.078	-.231**	-.093
Teamwork	-.067	.220*	-.028	.277*	-.061	.160*	-.139	.053
N	114[a]	96[a]	91[a]	64[a]	58	114[a]	113[a]	112[a]

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).
** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).
[a] Für die Dimension uCA_MEAN gilt Stichprobengröße=N-1.
Note_FSP: Note (1-6) im Fächerschwerpunkt; Note_D: Note (1-6) im Fach Deutsch; Note_M: Note (1-6) im Fach Mathematik; Note_E: Note (1-6) im Fach Englisch; Job_t: Zeit in Monaten an Berufserfahrung bei Einstieg in das Unternehmen; Pra_t: Zeit in Monaten an Praktikumserfahrung bei Einstieg in das Unternehmen

So gibt es bei den *Noten* keine oder teilweise in die falsche Richtung gehende Zusammenhänge, die inhaltlich nicht interpretierbar sind. Es ist davon auszugehen, dass die wenigen signifikanten Korrelationen auf Zufallseffekte zurückzuführen sind. Dies steht im Widerspruch zu den gefundenen Korrelationen bei den *JCs.* Es relativiert sich aber bei der Überlegung, dass die Tätigkeit als *JC* als eine Art Ausbildung zum Berater verstanden wer-

den kann. Demgegenüber ist der Zusammenhang zwischen Noten und dem Berufserfolg (der eigentlich auf Grund des Aufgabenspektrums erst ab der Stufe C messbar ist) als nicht vorhanden einzuschätzen, was auch den empirischen Erkenntnissen entspricht (siehe 2.5.2.4).

Das *Eintrittsalter* korreliert negativ signifikant mit *Problem Solution* (r=-.264)*, Quality of Work* (r=-.207) und *Know-how-Management* (r=-.232). Es liegt allerdings die Vermutung nahe, dass es sich um einen indirekten Effekt handelt, da das *Eintrittsalter* mit der *Einstiegsart* (direkt oder über die *JC-Stufe*) hoch signifikant korreliert (r=.458). Hier kann die Ursache also in der größeren Erfahrung derjenigen *Cs* liegen, die bereits als *JCs* für die Firma gearbeitet haben (siehe auch 3.3.2.3). Damit ist eine Konfundierung der Effekte anzunehmen, die nicht im Widerspruch zu den gefundenen Erkenntnissen bei der *JC-Kohorte* steht. Das gleiche Phänomen findet sich bei der *Berufserfahrung* von *Cs*. Diese hat negative signifikante Zusammenhänge mit dem *Gesamtwert* (r=-.265) und den Dimensionen *Problem Solution* (r=-.320), *Quality of Work* (r=-.296), *Know-how-Management* (r=-.231) und *Client Communication* (r=-.200). Da eigentlich davon auszugehen ist, dass sich *Berufserfahrung* (*Job_t*) positiv auf die spätere Leistung auswirkt, ist die negative Korrelation mit zahlreichen Projektbeurteilungsdimensionen zunächst überraschend. Allerdings ist auch hier die Korrelation mit der *Einstiegsart* hoch signifikant (r=.623). Es ist deshalb anzunehmen, dass die schlechtere Leistung ebenfalls an den geschilderten Neueinsteigerproblemen liegt. *Praktikumserfahrung* (*Pra_t*) korreliert hingegen signifikant schwach mit *Client Acceptance* (r= .227) und *Quality of Work* (r=.184). Damit sind diesbezügliche Ergebnisse der *JC-Kohorte* weitgehend repliziert worden.

Unter der Annahme gleicher Symmetrieebenen wurde des Weiteren der Zusammenhang zwischen dem *Gesamtwert aus Recruiting* und der *Abiturnote* mit dem *Gesamtwert aus Projektbeurteilung* betrachtet (*Abbildung 50*). Nach Eliminierung des Ausreißerwertes gibt es einen hoch signifikanten Zusammenhang zwischen den *Gesamtwerten aus Recruiting* und *Projektbeurteilung* i.H.v. r=.341. Diese mittlere Effektgröße erhöht sich durch Rückrechnung des Selektionsfehlers auf die Grundgesamtheit auf r=.369[14]. Dieser Wert ist deutlich höher als bei der *JC-Kohorte*. Dies kann zum Einen an der Verringerung des Beurteilungsfehlers durch die Fokussierung auf mindestens drei *Projektbeurteilungen* pro Mitarbeiter liegen. Zum Anderen ist auch anzunehmen, dass die Beurteilungen insgesamt als *C* genauer sind, da das Aufgabenspektrum breiter und damit die Situationen, in denen valide beurteilt werden kann, häufiger sind (höheres R_{KR}). Die Validität liegt damit sehr nahe bei den meta-

analytisch berichteten Validitäten für *unstrukturierte* Interviews (r=.380) und Assessment Center (r=.370). Bei durchschnittlich 3.18 Beurteilungen pro *C* und einer angenommenen Interrater-Reliabilität von r=.52 für Vorgesetztenurteile (Viswesvaran, et al., 1996) ergibt sich nach *Gleichung (8)* eine Reliabilität von r=.78 und damit attenuationskorrigiert eine höhere Validität von r=.418, die sogar die metaanalytisch berichtete, durchschnittliche Validität von *ACs* übertrifft.

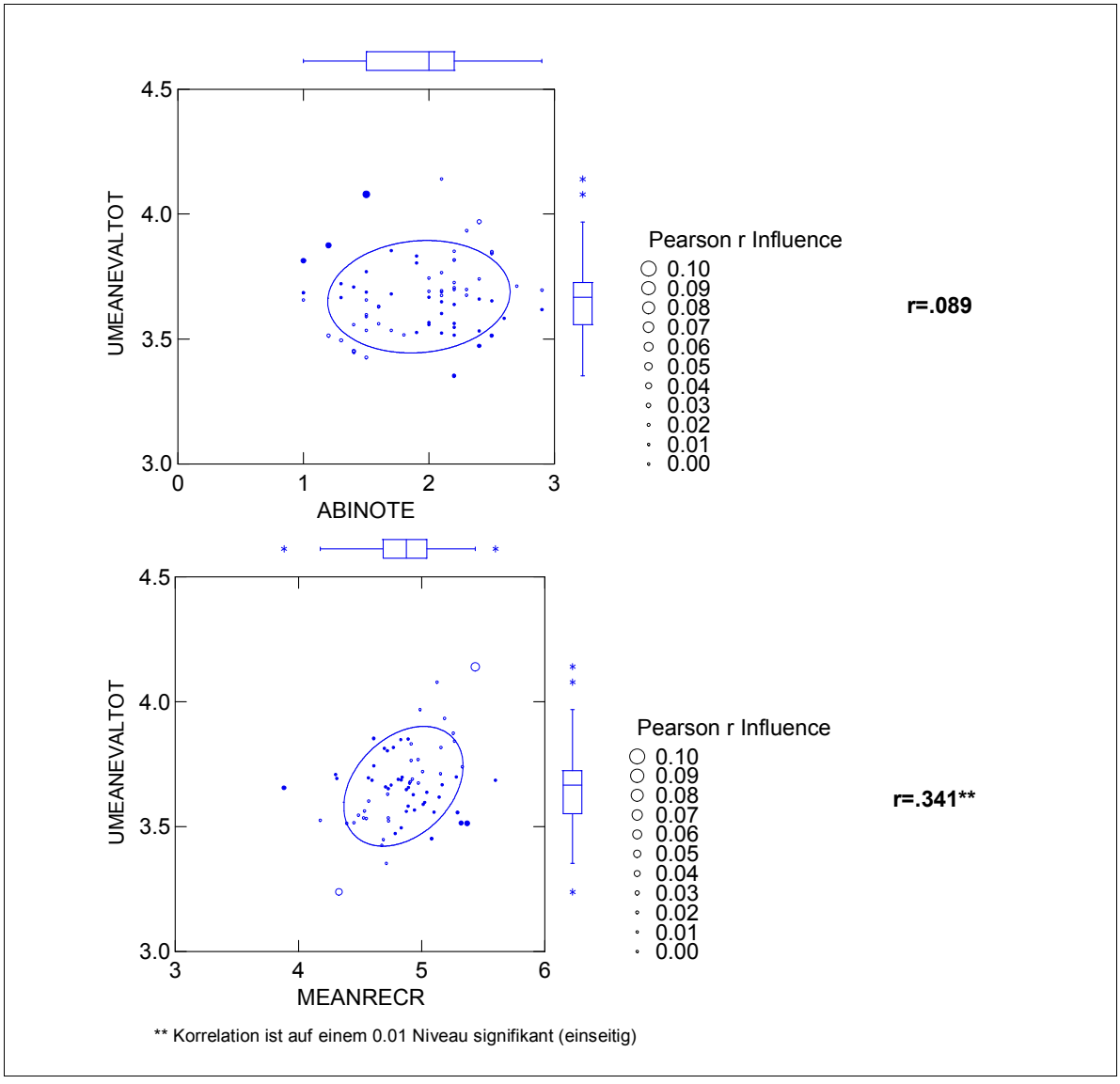

Abbildung 50: Korrelationen Abiturnote und Recruiting-Gesamtwert mit Gesamtwert aus Projektbeurteilungen (C) (N=70 bzw. N=71)

Die *Abiturnote* (*Abbildung 50*) hat keine signifikante Korrelation mit dem *Gesamtwert aus der Projektbeurteilung* oder irgendeiner Unterdimension (*Anhang IX-I*). Damit hat

[14] Die Berechnung des „restriction of range"-Effektes wurde mit der von Paul Barrett entwickelten Software *Correlation Attenuation Corrections v.2.1* vorgenommen und beruht auf einer konservativen SDx-Ermittlung

das *Abitur* als Aggregat vieler Eigenschaften (s.o.) offenbar keinen direkten Zusammenhang mit der Leistung im Beruf als *C*. Die gefundenen Zusammenhänge bei der *JC-Kohorte* stehen im Einklang mit empirischen Befunden, wenn diese Phase als Ausbildung verstanden wird. Der fehlende Zusammenhang mit den *Cs* passt hier auch ins Bild, da der *Abiturnote* insgesamt für den Berufserfolg keine nennenswerte prädiktive Validität zugesprochen wird. Beim Betrachten der Effekte über unterschiedliche Generalitätsebenen (unfaire Validitätstests, *Anhang IX-1*) zeigt sich, dass der *Gesamtwert aus Recruiting* mit den Dimensionen *Quality of Work* (r=.353), *Know-how-Management* (r=.227) und *Teamwork* (r=.261) signifikant korreliert. Dieses Ergebnis wurde auch bei der *JC-Kohorte* festgestellt und interpretiert. Neu ist allerdings die Signifikanz von *Problem Solution* (r=.222), *Client Acceptance* (r=.260) und *Client Communication* (r=.355). Hier wurde bei der *JC-Analyse* bereits vermutet, dass diese Dimensionen auf der Kriterienseite bei den *JCs* schlecht beurteilt werden können. Das Signifikantwerden dieser Dimensionen deutet darauf hin, dass sie bei den *Cs* leichter zu erfassen sind. Nicht mehr signifikant im Vergleich zu den *JCs* sind *Specialist-Know-how* und *Behaviour under Pressure*. Dies kann an der erheblichen Varianzreduktion bei der Stichprobe gegenüber den *JCs* liegen, die darin begründet liegt, dass viele Mitarbeiter, die schlechte Werte auf diesen Dimensionen hatten, bereits ausgeschieden sind. Die Stichprobe mit mindestens drei Projektbeurteilungen ohne dem Ausreißer beinhaltet lediglich noch drei *Forced Exits*.

Da der *Gesamtwert aus der Projektbeurteilung* auch marginal signifikant mit *Charisma*, *Methodenkenntnisse* und *VWL-Kenntnisse* korreliert, ist bei nahezu allen Variablen von Zusammenhängen auszugehen, auf deren Einzelinterpretation bei den Unterdimensionen eingegangen wird. Nicht signifikant, möglicherweise auf Grund des sehr geringen Stichprobenumfanges, sind die Variablen *Selbstorganisation*, *Leadership*, *Führungsfähigkeit*, *Fit ISB* und *Exotenwissen*. Mangelnde Relevanz für die herangezogenen Kriterien wird zudem für *Englischkenntnisse* vermutet. Bei *Konfliktfähigkeit* ist erneut, wie bei der JC-Kohorte, von einer *Suppressorkonstellation* auszugehen. *Fachkenntnisse* korrelieren vermutlich auf Grund der Varianzeinschränkung nicht mit *Specialist-Know-how*, da diese Dimension das inhaltlich relevante Kriterium darstellt (s.o.).

Auf der Variablenebene des *Recruitings* (faire Validitätstest, *Anhang IX-1*) gibt es ebenfalls signifikante Korrelationen. *Specialist-Know-how* korreliert aus den oben genannten Gründen lediglich negativ mit *Konfliktfähigkeit* (r=-.278). *Problem Solution* hängt neben *Konfliktfähigkeit* (r=-.310) mit *logisch-analytischem Denken* (r=.303) und *intellektueller Neugier* (r=.203) signifikant zusammen. Gerade *logisch-analytisches Denken* ist die wich-

tigste Voraussetzung zum Lösen von Problemen und besitzt vermutlich Überschneidungen mit Intelligenzkomponenten (fluide Intelligenz). Die *Intellektuelle Neugier* hilft vermutlich im Sinne des *TIE-Konstruktes* bei der Suche nach Lösungen (Rolfhus & Ackerman, 1999, siehe auch 2.5.1.2). *Quality of Work* hängt mit einer Vielzahl von Variablen zusammen, was den vermutlich hohen Allgemeinheitsgrad und die höhere Generalitätsebene dieser Dimension widerspiegelt. Interessant ist der stärkste Zusammenhang mit *Charisma* (r=.343), dem ja bei der Analyse der *JC*s ebenfalls eine höhere Symmetrieebene unterstellt wurde.

Bei *Behaviour under Pressure* wurde bereits die Varianzreduktion erläutert. Dennoch gibt es signifikante Korrelationen mit *Kommunikationsfähigkeit* (r=.333) und *Frustrationstoleranz* (r=.280). Gerade in stressigen Situationen im Beratungsalltag (oftmals verursacht durch Zeitdruck) ist eine hohe *Frustrationstoleranz*, die vor Aufgabe schützt, vonnöten. Die Rolle *kommunikativer Fähigkeiten* als eher allgemeines Konstrukt wurde bereits bei der Kohorte der *JC* erörtert. Kommunikative Berater scheinen beim Kunden zudem besser anzukommen (r=.262 mit *Client Acceptance*). Da der Informationsaustausch zwischen Kunde und Berater hauptsächlich auf verbalem Wege stattfindet und Beratung viel mit sprachlichem Ausdrucksvermögen zu tun hat, ist dieses Ergebnis wenig überraschend. Darüber hinaus hilft natürlich *unternehmerisches Denken* (r=.218), das größtenteils auf Managementseite beim Kunden vorhanden sein dürfte. Womöglich treffen hier ähnliche Denkweisen aufeinander. *Logisch-analytische Fähigkeiten* und *intellektuelle Neugier* vermitteln im Zusammenhang mit den anderen beiden Eigenschaften gegenüber dem Klienten die notwendige Kompetenz, um akzeptiert zu werden.

Die Beurteilungsdimension *Client Communication* korreliert mit einer Vielzahl von Variablen. Sie beinhaltet neben erfolgreicher Ausdrucksweise auch sehr allgemeine Komponenten wie das Durchführen eines Workshops oder eine erfolgreiche Interviewführung. Aus diesem Grund ist von einer eher höheren Generalitätsebene dieser *Projektbeurteilungsdimension* auszugehen. Auf eine inhaltliche Interpretation jedes einzelnen korrelativen Zusammenhangs wird deshalb verzichtet. *Know-how-Management* korreliert signifikant mit *Konfliktfähigkeit* (r=-.206) und *BWL-Kenntnissen* (r=.250). Da zum guten Wissensmanagement das Vorhandensein von Wissen Voraussetzung ist, ist dieser Zusammenhang plausibel. Die höchste Korrelation mit *Teamwork* hat *Teamfähigkeit* (r=.318). Darüber hinaus scheint plausibel, dass es bei *pragmatischen* Teammitgliedern und solchen mit einer hohen *Frustrationstoleranz* zu weniger Reibungspunkten kommt. Eine Unterdimension von *Teamwork* beinhaltet den Problemlösungsbeitrag im Team. Deshalb ist auch der Zusammenhang mit *intellektueller Neugier* (r=.203) und *logisch-analytischem Denken* (r=.225) erklärbar. Die

Zusammenhänge mit VWL (r=.209) und Allgemeinbildung (r=.304) beruhen wahrscheinlich auf einer höheren Symmetrieebene.

3.3.3.5 Regressionsanalyse

Die bivariaten regressionsanalytischen Befunde variieren auf Grund der etwas kleineren Stichprobe (nur Personen, die auf allen Variablen Werte haben) leicht im Vergleich zu den bivariaten Korrelationen aus dem vorherigen Kapitel. Zunächst wurden Regressionen der einzelnen *Dimensionen der Projektbeurteilung* aus dem *Gesamtwert des Recruiting* und der *Abiturnote* separat vorgenommen. Darüber hinaus wurden inhaltlich-sachlogische Prädiktoren abgeleitet, von denen ein Zusammenhang mit dem jeweiligen Kriterium erwartet wird. Dabei wurden diese explorativ maximierend in eine schrittweise Regression mit den Kriterien eingeführt. Im Anschluss wurde jede *Projektbeurteilungsdimension* explorativ über eine schrittweise Regression durch alle Recruitingvariablen vorhergesagt, um maximale Varianz aufzuklären. *Tabelle 48* zeigt die Ergebnisse im Überblick. Im *Anhang IX* finden sich die Angaben zu den einzelnen Regressionsrechnungen.

Tabelle 48: Gesamtübersicht der Regressionsergebnisse der C-Kohorte (N=67)

Prädiktoren	**Abinote**	**Gesamtwert Recruiting**	**Recruiting-variablen (inhaltlich) (schrittweise)**	**Recruiting-variablen (alle) (schrittweise)**
Kriterien	R^2_{korr}	R^2_{korr}	R^2_{korr}	R^2_{korr}
Durchschnittlicher Gesamtwert	-.015	.107**	.334**	.334**
Specialist-Know-how	-.006	-.010	-,-	.230**
Problem Solution	-.007	.030	.147**	.293**
Quality of Work	-.011	.143**	.135**	.203**
Behaviour under Pressure	.002	-.006	.163**	.206**
Client Acceptance	-.015	.066*	.079*	.079*
Client Communication	-.015	.128**	.108**	.108**
Know-how-Management	-.015	.039	.074*	.074*
Teamwork	-.013	.060*	.238**	.238**

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} :Schrumpfungskorrigierter Determinationskoeffizient

Die Regressionsanalysen ergaben erwartungsgemäß keine signifikanten Zusammenhänge der *Abiturnote* mit dem *Gesamtswert* oder den *Unterdimensionen der Projektbeurteilung*. Die *Abiturnote* hat somit keine prädiktive Validität für die Leistung als *C*. Auch die

Korrelationen und β-Gewichte sind in ihrer Höhe unbedeutend und weit von der Signifikanz entfernt. In dem Korrelationsabschnitt wurde bereits auf die Interpretation dieses Ergebnisses vor dem Hintergrund empirischer Befunde eingegangen. Der Recruiting-Gesamtwert klärt über 10% Varianz des *Gesamtwertes der Projektbeurteilung* auf (R^2=.107). Die Varianzaufklärung ist damit fast doppelt so hoch wie bei der *JC-Kohorte*. Gründe hierfür können an der erwähnten Verringerung des Beurteilungsfehlers durch die Fokussierung auf mindestens drei *Projektbeurteilungen* pro Mitarbeiter und am breiteren Aufgabenspektrum der Cs liegen. Die Varianzaufklärung nimmt außerdem, wie gezeigt, durch Herausrechnung des Selektionsfehlers zu und ist somit als gut zu bewerten.

Darüber hinaus klärt der *Gesamtwert aus Recruiting* bei *Quality of Work* (R^2=.143), *Client Communication* (R^2=.128), *Client Acceptance* (R^2=.066) und *Teamwork* (R^2=.060) Varianz auf. Hier wurde bei der *JC-Analyse* bereits vermutet, dass die Klientendimensionen auf Kriterienseite bei den *JCs* schlecht beurteilt werden können. Auf Grund des oftmals genannten hohen Allgemeinheitsgrades von *Quality of Work* ist nicht verwunderlich, dass hier die höchste Varianz durch den *Gesamtwert aus Recruiting* aufgeklärt wird, gefolgt von der ebenso eher allgemeinen Dimension *Client Communication*. Es ist denkbar, dass beide Dimensionen auf einer Symmetrieebene mit dem *Gesamtwert aus Recruiting* stehen. Besonders bei *Client Communication*, da hier im Vergleich zu allen anderen Prädiktoren am meisten Varianz durch den *Recruiting-Gesamtwert* aufgeklärt wird.

Bei der inhaltlich-sachlogisch abgeleiteten Vorhersage der Dimensionen durch die Recruitingvariablen (*Anhang IX-3 und IX-4*) zeigen sich die höchsten Varianzaufklärungen für den *Gesamtwert* (R^2=.334), *Behaviour under Pressure* (R^2=.163) und *Quality of Work* (R^2=.135). Signifikant beim *Gesamtwert aus Projektbeurteilung* werden dabei die *logisch-analytische Fähigkeit*, die *Kommunikationsfähigkeit* (zur Erklärung siehe 3.3.2.5) und die *Konfliktfähigkeit* als Suppressor. Letztere beide sind auch die aufgenommenen Prädiktoren bei *Behaviour under Pressure*. Erneut sei angemerkt, dass die erhebliche Varianzreduktion von *Behaviour under Pressure* bei der Stichprobe der *Cs* gegenüber den JCs korrelative und prädiktive Zusammenhänge verdecken kann.

Von allen inhaltlich als relevant erachteten Variablen ist bei *Quality of Work* lediglich *Pragmatismus* in die Regression aufgenommen worden, obwohl *Quality of Work* signifikant mit einer Vielzahl von Variablen korreliert. Offensichtlich liegt hier eine sehr hohe *Multikolinearität* der Prädiktoren vor. *Pragmatismus* ist eine Eigenschaft, die in der Beratungsbranche sehr hilfreich ist wenn es darum geht, praktisch umsetzbare Lösungen zu finden. Darüber hinaus hilft *Pragmatismus* auch im Beraterralltag, um das immense Arbeitsvolumen

zu bewältigen. *Problem Solution* wird am besten durch *logisch-analytisches Denken* vorhergesagt. Dabei wird rund 15% der Varianz aufgeklärt (R^2=.147). Dass intellektuell neugierige Menschen beim Klienten ein höhere Akzeptanz haben (R^2=.079) kann damit zusammenhängen, dass der Klient verstärkt das Gefühl bekommt, der Berater habe Interesse an seinen Problemen und an einer Lösungsfindung. Erklärbar ist auch, dass *unternehmerisches Denken*, das eigentlich auch größtenteils auf Managementseite vorherrscht, bei der Kommunikation (Client Communication) hilft, da so gleiche Denkweisen aufeinander treffen (s.o.). Dennoch ist wahrscheinlich, dass die alleinige Aufnahme dieser Prädiktoren bei der Regression mit rund je 10% Varianzaufklärung auf die hohe *Multikolinearität* der Prädiktoren zurückzuführen ist.

Dass zu guten Wissensmanagement das Vorhandensein von *Wissen* Voraussetzung ist, wurde bereits bei den korrelativen Befunden erörtert. Damit ist auch die Aufnahme von *BWL-Know-how* in die Regression von *Know-how-Management* erklärt (R^2=.074). Bei der Dimension *Teamwork* wurden *Teamfähigkeit*, *Konfliktfähigkeit* und *logisch-analytisches Denken* in die Regression aufgenommen. Bei *Konfliktfähigkeit* liegt offenbar wieder ein *Suppressoreffekt* vor. Eine Unterdimension von *Teamwork* beinhaltet den Problemlösungsbeitrag im Team, womit sich auch der Zusammenhang mit *logisch-analytischem Denken* erklärt. Eine inhaltlich-sachlogisch nicht begründbare Varianzaufklärung gibt es bei der Dimension *Specialist-Know-how*. Dies kann an der angesprochenen erheblichen Varianzreduktion durch *Drop-Out-Effekte* liegen.

Bei dem Versuch der maximalen Varianzaufklärung durch schrittweise Regression mit allen Variablen (*Anhang IX-5*) konnte lediglich bei den Dimensionen *Specialist Know-how* (R^2=.230), *Problem Solution* (R^2=.293), *Quality of Work* (R^2=.203) und *Behaviour under Pressure* (R^2=.206) mehr Varianz durch das Gesamtmodell aufgeklärt werden, als im inhaltlich abgeleiteten Fall. Bei *Problem Solution* wird *Konfliktfähigkeit* als Suppressorvariable aufgenommen und *Charisma* mit einem sehr kleinen inkrementellen Varianzanteil von ca. 3%. *Charisma* wurde bei der Analyse der *JCs* bereits eine höhere Generalitätsebene unterstellt. *Specialist-Know-how* ist ebenfalls durch eine Suppressorenkonstellation mit *Konfliktfähigkeit* und *Methodenkenntnissen* gekennzeichnet. Signifikante Prädiktoren werden neben diesen beiden *logisch-analytisches Denken* mit sehr hohen β-Gewichten und *Kommunikationsfähigkeit*. Die Bedeutung der *Kommunikationsfähigkeit* vor dem Hintergrund des Wissensaufbaus und der -entwicklung in dem Unternehmen wurde bereits bei den *JCs* vorgefunden. *Logisch-analytisches Denken* und *Kommunikationsfähigkeit* sind auch bei *Quality of Work* von Bedeutung, wobei Letztere alleine rund 15% Varianz aufklärt und *logisch-*

analytisches Denken weitere 5% inkrementelle Varianzaufklärung beisteuert. Bei *Behaviour under Pressure* wird zu den bereits im inhaltlich-sachlogischen Teil aufgenommenen Prädiktoren *Allgemeinwissen* in die Regression mit aufgenommen. Hier handelt es sich ebenfalls um eine *Suppressorvariable*.

Zusammenfassend lässt sich feststellen, dass nicht unerhebliche Varianzanteile über alle Dimensionen mit unterschiedlichen Prädiktoren des Recruitings aufgeklärt werden konnten. Besonders die inhaltlich abgeleiteten Prädiktoren liefern zufrieden stellende Ergebnisse. Dabei werden aber nur sehr wenige Variablen aufgenommen und *Supressoreffekte* verwischen die eigentlichen Zusammenhänge. Neben der Stichprobengröße ist, wie bei der *JC-Kohorte*, vor allem die hohe *Multikolinearität* sowohl auf der Prädiktoren- als auch auf der Kriterienseite problematisch. Durch sie sind die Personen auf den einzelnen Dimensionen wenig unterscheidbar und die Varianz reduziert. Dies kann an der verwendeten Operationalisierung der Konstrukte liegen, die im vorliegenden Fall sehr anfällig für Beurteilungsfehler ist. Zudem ist auch hier von einer Verletzung der Symmetrieprinzipien nach Brunswik auszugehen, die zu unfairen Validitätstests führt. In einem nächsten Schritt wird versucht, ähnlich wie bei der Analyse der *JC-Kohorte*, über Faktorenanalyse diesem Problem entgegenzutreten.

3.3.3.6 Faktoren- und Pfadanalyse

Für eine detaillierte Interpretation der Faktorenanalyse der *Prädiktorendaten* sei auf 3.3.1.5 verwiesen. *Tabelle 49* zeigt die Korrelationen der unrotierten und rotierten *Recruitingfaktoren* mit den *Projektbeurteilungsdimensionen*. Insgesamt konnten weitere Anhaltspunkte zur Konstruktvalidität der Kriterien und zur Symmetrie gewonnen werden (*Fragestellungen (6)* und *(11)*).

Bei der Kohorte *Einstieg als JC* korreliert der unrotierte Levelfaktor signifikant mit *Quality of Work* (r=.358), *Behaviour under Pressure* (r=.314), *Client Acceptance* (r=.264), *Client Communication* (r=.394) und dem *Gesamtwert aus Projektbeurteilung* (r=.353). Dies verdeutlicht auch bei den *Cs* seinen Level-Charakter. Darüber hinaus deutet es darauf hin, dass die bei der *JC-Analyse* angenommenen höheren Generalitätsebenen von *Quality of Work* und *Behaviour under Pressure* bei der Betrachtung der C-Kohorte um die klientenzentrierten Beurteilungsdimensionen ergänzt werden müssen. Da die *Kommunikation* insgesamt genauso wie die Akzeptanz durch den Klienten unzählige Leistungsbestandteile beinhaltet, ist auch bei diesen Dimensionen ein höheres Generalitätsniveau plausibel.

Bei der rotierten Faktorlösung ist der Allgemeinfaktor derjenige, der die häufigsten signifikanten Zusammenhänge aufweisen. Der rotierte erste Recruiting-*Allgemeinfaktor* der *JC-Einsteiger* korreliert mit nahezu allen Dimensionen, hat aber die höchsten signifikanten Korrelationen bei den oben genannten Dimensionen und *Teamwork*. Der Faktor beinhaltet vermutlich allgemeine Intelligenzkomponenten, da die auf ihm am höchsten ladenden Variablen *Kreativität*, *BWL-*, *VWL-Kenntnisse* und *Allgemeinwissen* sind. Keine Zusammenhänge gibt es mit *Specialist-Know-how* und *Problem Solution.* Hier ist anzunehmen, dass diese Leistungen von spezifischeren Fähigkeiten wie *fluider* und *kristalliner* Intelligenz beeinflusst werden.

Tabelle 49: Korrelation Faktorwerte der Recruitingfaktoren der C-Kohorte mit Projektbeurteilungsdimensionen (Spearman-Rho) (Kohorte: „Einstieg als JC“; N=44)

	Extrahierte rotierte Faktoren			**Extrahierte unrotierte Faktoren[a]**	
Kriterienvariablen	Allgemeinfaktor	Sozialkompetenz	Problemlösungsfähigkeit	Levelfaktor	Profilfaktor
Durchschn. Gesamtwert	.388**	.175	.000	.353**	-.089
Specialist-Know-how	.240	.067	-.061	.189	-.051
Problem Solution	.122	-.010	-.016	.073	-.045
Quality of Work	.287*	.291*	.061	.358**	.105
Behaviour under Pressure	.285*	.224	-.206	.314*	.090
Client Acceptance	.296*	.153	-.061	.264*	-.026
Client Communication	.387**	.171	.050	.394*	-.128
Know-how-Management	.270*	-.039	.081	.185	-.204
Teamwork	.344*	.031	-.155	.219	-.162

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant (einseitig)
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant (einseitig)
a Von den 3 extrahierten Komponenten sind die ersten 2 dargestellt.

Tabelle 50 zeigt die Faktorisierung der Projektbeurteilungsvariablen. Dabei konnten die Ergebnisse aus der *JC-Kohorte* größtenteils repliziert werden, da es ebenso Allgemeinfaktoren, die das *Gesamtlevel* widerspiegeln und *Profilfaktoren* gibt. Die Profilstruktur stellt sich allerdings etwas differenzierter dar. Bei den *Cs* scheint es in der unrotierten Variante ein Kontinuum zwischen den Polen *Behaviour under Pressure* und *Teamwork* auf der einen und *Quality of Work* und *Know-how-Management* auf der anderen Seite zu geben. Demgegenüber sind bei der *JC-Kohorte* die klientenzentrierten Dimensionen als Gegenpol zu *Team-*

work und *Behaviour under Pressure* vorzufinden (*Abbildung 51*). Die Profilvergleiche zeigen aber sehr ähnlich verlaufende Kurven. Allerdings muss hier berücksichtigt werden, dass es sich größtenteils um die identische Stichprobe handelt, da viele ehemalige *JCs* in der *C*-Stichprobe enthalten sind. Der zweite rotierte Faktor beinhaltete bei der *JC-Kohorte* stärker *Problemlösungsverhalten*, *Sozialverhalten* und *Stressverhalten*. Dieser Befund tritt auch bei der *C*-Stichprobe auf und kann somit ebenso repliziert werden.

Tabelle 50: Rotierte und unrotierte Komponentenmatrizen der Projektbeurteilungsvariablen (N=75)

	Extrahierte unrotierte Faktoren[a]		**Extrahierte rotierte Faktoren[a,b]**	
Kriteriendimensionen	1	2	1	2
Specialist-Know-how	.741		.462	.589
Problem Solution	.848		.550	.651
Quality of Work	.714	.466	.838	
Behaviour under Pressure	.611	-.518		.796
Client Acceptance	.564		.493	.301
Client Communication	.681		.618	.338
Know-how-Management	.590	.493	.767	
Teamwork	.700	-.489		.838
VA[c]	47.1%	60.1%	30.8%	60.1%

Faktor unrotiert	1	2	3	4	5
Eigenwert	3.771	1.035	.820	.680	.582

Dargestellt sind nur Faktorladungen mit Betrag >.300.
a Extraktionsmethode: Hauptkomponentenanalyse.
b Rotationsmethode: Varimax mit Kaiser-Normalisierung. Die Rotation ist in 3 Iterationen konvergiert.
c Varianzaufklärung, kumulativ dargestellt

Der *unrotierte Levelfaktor* (1. Faktor aus *Tabelle 50*) korreliert in nahezu identischer Konstellation mit den Recruitingvariablen wie der *Gesamtwert aus Projektbeurteilung* (*Anhang IX-6*), was seinen Level-Charakter unterstreicht. Es werden aber offenbar die *Profile*, im Gegensatz zu der Analyse der *JC-Kohorte* durch das Recruiting nicht besser vorhergesagt als das allgemeine Leistungsniveau. Dies zeigt sich in der nicht signifikanten niedrigen Korrelation mit dem *Gesamtwert aus dem Recruiting*. Bei der *rotierten* Komponentenmatrix gibt es keine eindeutige Interpretation der Faktoren. Stattdessen laden einzelne Variablen auf mehreren Faktoren hoch, ohne dass hier eine eindeutige Beziehung zu erkennen ist. Dabei korreliert, wie bei der unrotierten Lösung, der *Allgemeinfaktor* mit den *Recruitingvariablen*

in gleicher Konstellation wie der *Gesamtwert aus Projektbeurteilung* während der zweite Faktor mit *Kommunikationsverhalten*, *Teamfähigkeit* und *logisch-analytischem Denken* zusammenhängt. Dies sind exakt jene Variablen, die den auf dem Faktor hoch ladenden Dimensionen *Problem Solution*, *Behaviour under Pressure* und *Teamwork* inhaltlich zuzuordnen sind. Bei der Betrachtung der Korrelationen der extrahierten Faktoren untereinander zeigt sich, dass der unrotierte *Levelfaktor* der ehemaligen *JCs* in vergleichbarer Höhe mit dem allgemeinen unrotierten *Recruitingfaktor* korreliert wie die *Gesamtwerte aus Projektbeurteilung* und *Recruiting*. Die Faktorisierung hat somit zu keinem Validitätsgewinn geführt. Dies bestätigt auch die anschließend durchgeführte schrittweise Regression mit den Faktorwerten (*Anhang IX-8*). Hier gab es erwartungsgemäß ebenfalls keine Validitätsgewinne bzw. inkrementellen Varianzaufklärungen im Vergleich zu den Regressionen mit den Variablen (3.3.3.5).

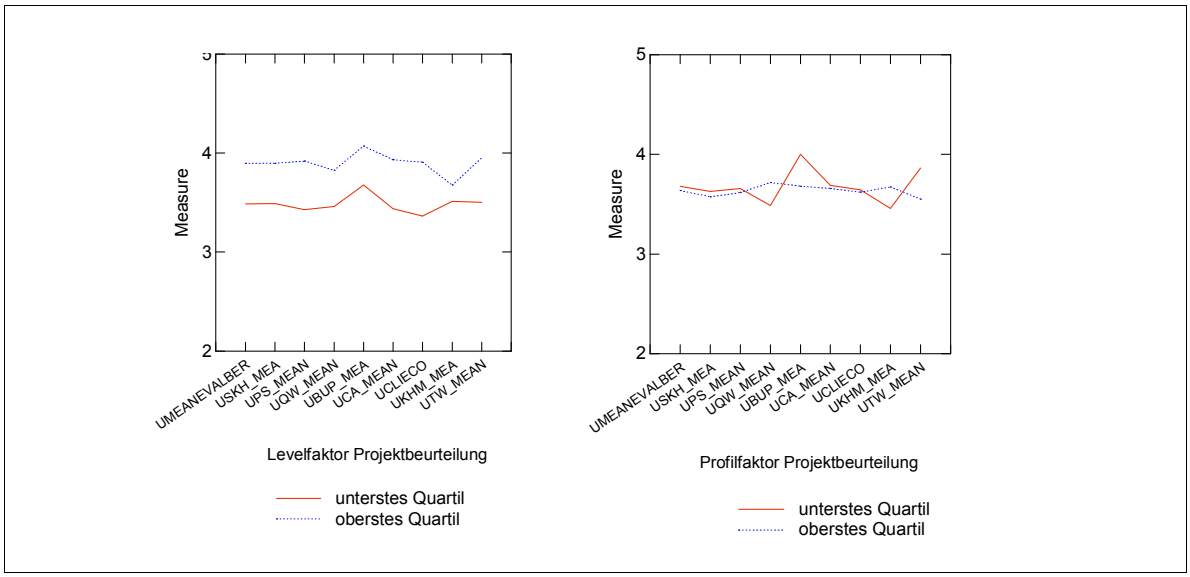

Abbildung 51: Profilvergleich Projektbeurteilung Randquartile Level- und Profilfaktor (unrotierte Faktorlösung)

Darüber hinaus wurde eine zweite Faktorenanalyse aus den Erkenntnissen der korrelativen und regressionsanalytischen Untersuchung heraus gerechnet in der die vermuteten Generalitätsebenen der Kriterienvariablen berücksichtigt wurden (siehe *Abbildung 52*, linker Teil). Dabei wurden über die theoretische Annahme, dass einzelne latente Dimensionen hinter bestimmten Variablenclustern liegen, die Variablen der Kriterienseite anhand von Generalitätsannahmen neu geordnet. Dieses Vorgehen hatte bereits bei der *JC-Analyse* zu höheren Validitätskoeffizienten geführt. Die klientenzentrierten Dimensionen wurden dort aber auf Grund des Bewertungsproblems aus der Betrachtung ausgeschlossen. Im Fall der Tätigkeit als *C* ist allerdings davon auszugehen, dass dieses Bewertungsproblem nicht vorhanden ist,

da *Cs* sehr häufig Kundenkontakt haben. Der neue Faktor *Beratungserfolg* zeigt im Vergleich zu dem Ursprungsfaktor und dem *Gesamtwert aus Projektbeurteilung* erheblich höhere Korrelationen mit den Recruitingvariablen. Die Korrelation mit dem *Gesamtwert aus Recruiting* erreicht dabei eine Höhe von r=.447 (*Tabelle 51*).

Tabelle 51: Korrelationen von Recruitingvariablen mit verschiedenen Kriterien

Korrelationen (Spearman-Rho)	Pragma	Team	Komm	LogAn	Intell	Unt	BWL	MeanRecr
Beratungserfolg-Faktor	.389**	.349**	.461**	.380**	.394**	.338**	.270*	.447**
Unrotierter Levelfaktor	.284*	.282*	.341**	.320**	.359**	.271*	.233*	.358**
Gesamtwert Projektbeurteilung	.284*	.280*	.351**	.304*	.370**	.291*	.213	.369**
N	55	55	55	55	55	54	54	55

Eine darauf folgende *Pfadanalyse* (*Abbildung 52*) führt zu 32% Varianzaufklärung durch die Variablen *Kommunikationsfähigkeit* und *logisch-analytisches Denken*. Diese können wiederum als Surrogat für die in den Profilanalysen bereits besprochenen Dimensionen *verbal-soziale* und *numerisch-analytische Fähigkeiten* verstanden werden. Ihnen kommt offensichtlich, wie in zahlreichen Studien belegt, eine wesentliche Rolle zu.

3.3.4 Analyse der Kohorte Senior-Consultants

Ausgangspunkt des Vorgehens in diesem Abschnitt ist erneut die deskriptive Analyse der Prädiktoren- und Kriterienvariablen. Im Anschluss werden, basierend auf ausgewählten *CV-Variablen*, t-Tests für unabhängige Stichproben auf Mittelwertunterschiede und deren Effektstärken auf allen *Projektbeurteilungsdimensionen* gerechnet. Ziel dabei ist es, signifikante Profilunterschiede zwischen den untersuchten Gruppen auch bei der *SC-Kohorte* festzustellen und somit weitere Anhaltspunkte für eine Verbesserung der Vorauswahl von Bewerbern zu gewinnen. Zudem wird abermals eine Varianzanalyse mit Messwiederholung durchgeführt um festzustellen, inwieweit sich die Bewerber mit zunehmender Funktionsstufenzugehörigkeit auf den Projektbeurteilungen verbessern. In einem weiteren Schritt wird über Korrelationsanalysen die Validität des Auswahlverfahrens und der *CV-Variablen* für den Berufserfolg ermittelt. Die anschließenden Regressionsanalysen sollen aufdecken, in welcher Form die Prädiktoren aus dem Recruiting optimal kombiniert werden können. Ferner sollen im Rahmen von Faktorenanalysen, Anhaltspunkte für die Konstruktvalidität des gemessenen Kriteriums gewonnen werden. Die Vielzahl der unter 3.3.2.1 erwähnten Restriktionen gilt ebenfalls bei der Interpretation der Ergebnisse der *SC-Kohorte*.

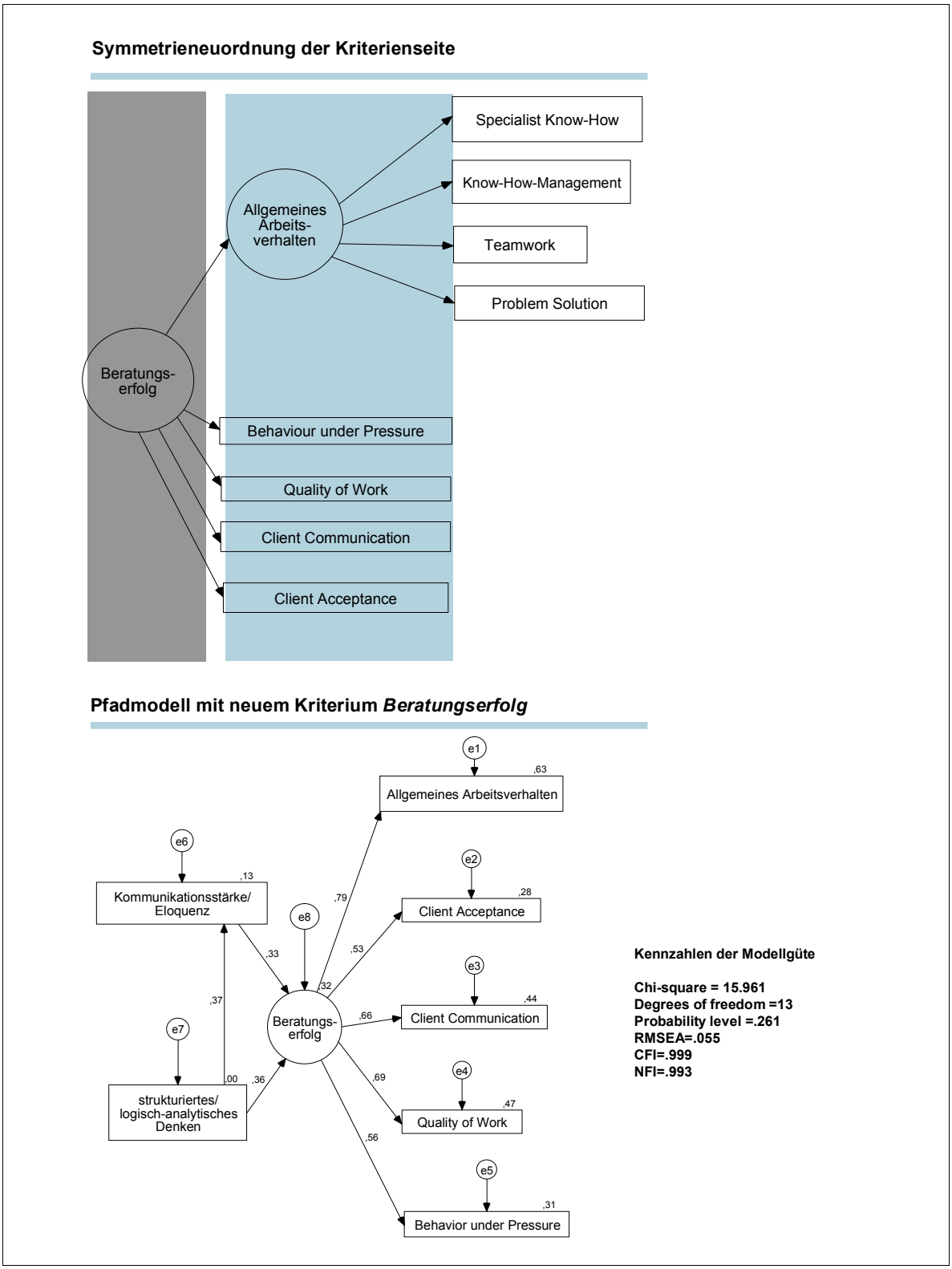

Abbildung 52: Symmetrieneuordnung und Pfadmodell

3.3.4.1 Deskriptive Analyse

Die Ergebnisse der deskriptiven Analyse unterscheiden sich nicht wesentlich von den Befunden in den anderen zwei Kohorten, weshalb sie im Folgenden nur nochmals grob umrissen werden.

Bei der *SC-Kohorte* zeigen sich bei den Recruitingvariablen ebenfalls durchweg hohe Mittelwerte und geringe Standardabweichungen. Zudem finden sich auch hier bei einzelnen Variablen kleine Fallzahlen, da nicht alle Dimensionen bei allen Personen beurteilt wurden und sich die Beurteilungsdimensionen im Zeitverlauf geändert haben. Dies betrifft vor allem die Dimensionen *Leadership*, *Zielorientierung*, *Fit ISB-Team* und *Exotenwissen*. Die Gründe für die in dem deskriptiven Teil der S*C-Kohorte* festgestellte geringe Varianz und die Verteilungseigenschaften der *Prädiktorvariablen* wurden bereits unter 3.3.2.2 beschrieben und gelten auch hier. Bei den Recruitingvariablen gibt es zudem eine Vielzahl an Ausreißerwerten, die auch hier eine Eliminierung auf Grund ihrer hohen Zahl nicht ermöglicht, da diese je nach Variable immer unterschiedliche Personen betreffen. Die Verteilungen der *Recruitingvariablen* sind höchst unterschiedlich. Während einige Variablen signifikant von der Normalverteilung abweichen (z.B. *Pragmatismus*, *Konfliktfähigkeit* und *methodische Kenntnisse*) trifft dies auf andere (z.B. *Charisma* oder den *Gesamtmittelwert*) nicht zu. Die visuelle Inspektion der Histogramme deutet erneut darauf hin, dass die Abweichungen von der Normalverteilung hauptsächlich in einer schmalgipfligen rechtslastigen Verteilung um den Wert 5.0 liegen.

Tabelle 52: Deskriptive Kennwerte der Kriterien (Projektbeurteilungen) bei SCs

Kriterienvariablen	N	MW	Med.	Min.	Max.	SD	NV (j/n)
Specialist-Know-how	89	3.63	3.58	2.80	4.80	.28	j
Problem Solution	89	3.61	3.60	2.60	4.30	.26	j
Quality of Work	89	3.58	3.58	2.50	4.00	.22	j
Behaviour under Pressure	89	3.74	3.75	3.00	4.70	.35	n
Client Acceptance	87	3.70	3.65	3.10	4.50	.26	j
Client Communication	89	3.59	3.58	3.00	4.30	.29	j
Know-how-Management	89	3.51	3.50	3.00	4.10	.26	j
Teamwork	89	3.73	3.74	3.00	4.40	.25	j
Durchschnittlicher Gesamtwert	89	3.63	3.65	2.93	4.03	.17	j

N= Stichprobenumfang; MW=Mittelwert; Med.=Median; Min.=Minimum; Max.=Maximum; SD=Standardabweichung; NV=Normalverteilungsannahme

Bei den *CV-Daten* gibt es auf vielen Variablen höchst unterschiedliche Werte. Die *Berufserfahrung* (*Job_t*) schwankt z.B. stark einzelfallbedingt. Dies betrifft ebenso die Zeit an Praktika (*Pra_t*). Auf diesen Variablen finden sich entsprechend hohe Standardabweichungen. Andererseits differenzieren andere Variablen weniger zwischen den Personen (z.B. *Abiturnote*) und bei manchen Variablen sind nur geringe Fallzahlen vorhanden (z.B. *Mathematiknote*, *Englischnote*). Zudem ist lediglich bei der *Abiturnote*, dem *Eintrittsalter*, der *Ausbildungsnote* und der *Praktikumszeit* von der Normalverteilungsannahme auszugehen. Alle anderen Prädiktoren weichen signifikant von dieser ab, z. T. bedingt durch wenige Kategorien (z.B. *Anzahl der Universitätsabschlüsse*). Die Projektbeurteilungsdaten weichen mit Ausnahme von *Behaviour under Pressure* nicht signifikant von der Normalverteilung ab.

Bei den Projektbeurteilungen halten sich die Ausreißer wiederum in Grenzen. Einzelne Ausreißer finden sich auf allen Dimensionen mit Ausnahme von *Behaviour under Pressure*. Der *Gesamtmittelwert* aus *Projektbeurteilung* ist mit 3.63 gleich hoch wie bei der *C-Kohorte* und ebenfalls weit über dem theoretisch zu erwartenden Wert von 3.0 (*Tabelle 52*). Dies führt zu der Vermutung, dass die Handhabung des Beurteilungsinstrumentes auch auf dieser Stufe wenig streng ist. Zudem zeigt eine geringe Standardabweichung aller Dimensionen die geringe Differenzierung bei den evaluierten Mitarbeitern, die auf die gleichen klassischen Urteilsfehler und auf die Aggregationssystematik, wie auf der Prädiktorenseite beschrieben, zurückzuführen ist. Auf den Unterdimensionen sind, genauso wie bei der *JC-* und der *C-Kohorte*, *Behaviour under Pressure* und *Teamwork* die durchschnittlich am besten beurteilten Dimensionen. Von nahezu allen Mitarbeitern sind auf allen Dimensionen Urteile vorhanden.

3.3.4.2 Mittelwert- und Profilvergleiche

Tabelle 53 zeigt die Ergebnisse explorativer t-Tests bei der Analyse der S*C-Kohorte* mit ausgewählten *CV-Variablen*. Die dargestellten Ergebnisse aus den *Tabellen 53* und *54* wurden zudem zur Interpretationserleichterung als Profilvergleiche visualisiert. Diese werden aus Komplexitätsgründen nicht dargestellt. Insgesamt konnten auf der *SC-Ebene* die in den anderen beiden Kohorten gefundenen Ergebnisse nur vereinzelt repliziert und somit wenige Anhaltspunkte zur Beantwortung der Frage nach charakteristischen Profilen (*Fragestellung (5)*) gewonnen werden. Ein Grund dafür kann darin liegen, dass Mitarbeiter auf der SC-Stufe eine *Selektionsstichprobe* darstellen, da die weniger guten Mitarbeiter zu jenem Zeitpunkt bereits ausgeschieden sind sofern sie nicht erst in dieser Stufe dazukamen (Quereinsteiger). Diese Vermutung wird durch den Vergleich der *Forced Exits* bei der *JC-Kohorte*

(12%) mit den Forced Exits jener *SCs*, die als *JC* eingestiegen sind (3%), verdeutlicht. Hier zeigt sich, dass bei Mitarbeitern der *forcierte* Ausstieg verstärkt in zeitlicher Nähe zu ihrem Einstieg stattfindet. Zudem ist der *forcierte* Ausstieg bei den Mitarbeitern, die als *SC* quer eingestiegen sind, mit 26% sehr hoch.

Tabelle 53: Explorative t-Tests ausgewählter CV-Kategorien für SCs (Effektstärken)

	Gruppen				
Kriterienvariablen	Einstieg als SC (ja vs. nein)	Angestellten-status (angestellt vs. ausgeschieden)	Forced Exit (ja vs. nein)	Universitätsart (öffentlich vs. privat)	Familien-stand (ledig vs. verheir.)
Specialist-Know-how	.240	.112	-.744*	.197	-.310
Problem Solution	-.188	.182	-.834	.019	.016
Quality of Work	-.229	.227	-.891	.014	-.028
Behaviour under Pressure	.236	-.052	.457	.053	-.332
Client Acceptance	-.105	.142	-.487	.046	-.195
Client Communication	-.292	.218	-.382	-.234	-.123
Know-how-Management	-.168	.186	-.425	-.282	-.097
Teamwork	.021	.151	-.679*	.017	.049
Durchschnittlicher Gesamtwert	-.096	.239	-1.002	-.041	-.234

Basis der Effektgrößen d ist der Mittelwertunterschied = 1.Gruppe – 2. Gruppe; zur Anzahl der Fälle siehe Tabelle 20
* Der t-Test ist auf dem 0.05 Niveau signifikant (zweiseitig)
** Der t-Test ist auf dem 0.01 Niveau signifikant (zweiseitig)

Es gibt keine signifikanten Unterschiede zwischen Mitarbeitern, die direkt als *SC* eingestiegen sind im Vergleich zu allen anderen *SCs*. Betrachtet man allerdings nur jene Mitarbeiter, die nicht *forciert* ausgeschieden sind, so zeigt ein zweiseitiger t-Test, dass Quereinsteiger ein höheres *Spezialistenwissen* haben ($N_{SC\text{-}Einstieg}$=20; $N_{keinSC\text{-}Einstieg}$=54; df=72, MW_{Diff}=.147; p=.042). Da *SC-Quereinsteiger* eine mehrjährige Berufserfahrung aufweisen und diese gewöhnlich in einem Bereich oder einer Branche liegt, ist dieses Ergebnis plausibel (hohes Berufswissen). Im Gegensatz zu den Ergebnissen der *JC*- und *C-Kohorten* haben ausgeschiedene Mitarbeiter keine signifikant schlechteren Beurteilungen als Mitarbeiter, die noch im *Angestelltenverhältnis* sind. Hier zeigt sich der Effekt der natürlichen Fluktuation, da ab dem *SC-Level* eine Funktionsstufe erreicht wird bei der Mitarbeiter unabhängig von ihrer Leistung größtenteils freiwillig ausscheiden. Dies ist bei den unteren Kohorten nicht der Fall. Dort waren noch leistungsschwache Personen enthalten, die im Laufe der Zeit ausgeschieden sind. Insgesamt zeigen *forciert* ausgeschiedene Mitarbeiter signifikant schlech-

tere Beurteilungen bei *Specialist-Know-how* und *Teamwork* (starker Effekt). Andere, teilweise starke Effektstärken, sind nicht signifikant, da von ungleichen Varianzen ausgegangen werden muss (*Levene-Test*). Mangelnde *Fachkenntnisse* und fehlende *Teamfähigkeit* scheinen hier die Hauptgründe für ein Ausscheiden auf dem SC-Level zu sein.

Keine signifikanten Unterschiede gibt es bei den Leistungen der Mitarbeiter in Abhängigkeit ihres *Familienstandes*. Hier waren bei der *C-Kohorte* noch signifikant schlechtere Leistungen der verheirateten Mitarbeiter aufgedeckt worden. Hierfür kann es mehrere Gründe geben: *Drop-Out-Effekt* (s.o.), zu kleine Fallzahlen (nur sieben verheiratete Mitarbeiter insgesamt) oder eine nicht neuartig auftretende Situation. So gibt es nur drei verheiratete Quereinsteiger. Die restlichen Mitarbeiter arbeiten schon länger als verheiratete Berater. Zudem kann es durch alle, die nach der Einstellung geheiratet haben, zu einer nicht gemessenen Zunahme von verheirateten Mitarbeitern in der Stichprobe gekommen sein. Dies könnte dazu führen, dass hier keine Differenzierung mehr stattfindet, da sich der gemessene *Familienstand* in der Untersuchungsstichprobe lediglich auf den Zeitpunkt des Eintritts bezieht.

Tabelle 54: Gruppenvergleich (t-Test) biographischer Kategorien für SCs (Effektstärken)

	Gruppen			
Kriterienvariablen	BWL-Studium (ja vs. nein)	MBA (ja vs. nein)	Dr.-Titel (ja vs. nein)	Externe Berufserfahrung (j/n)
Specialist-Know-how	-.275	-.675*	-.182	-.061
Problem Solution	-.055	-.314	.085	-.545*
Quality of Work	-.068	-.294	-.261	-.463*
Behaviour under Pressure	-.093	-.592	-.116	-.094
Client Acceptance	-.004	-.341	.093	-.163
Client Communication	.373	-.315	.290	-.541*
Know-how-Management	-.245	-.110	-.158	-.392
Teamwork	-.127	.178	-.642*	-.161
Durchschnittlicher Gesamtwert	-.106	-.517	.077	-.434

Basis der Effektgrößen d ist der Mittelwertunterschied = 1.Gruppe – 2. Gruppe
* Der t-Test ist auf dem 0.05 Niveau signifikant (zweiseitig)
** Der t-Test ist auf dem 0.01 Niveau signifikant (zweiseitig)
Zur Anzahl der Fälle siehe Tabelle 21

Keine signifikanten Unterschiede als *SC* gibt es ebenfalls in Abhängigkeit der *Universitätsart* auf der studiert wurde. Da es sich bei der *SC-Kohorte* nicht mehr um Berufseinsteiger handelt ist davon auszugehen, dass der bei der *C-Kohorte* aufgedeckte Effekt

mittlerer Stärke im Verlauf der Berufserfahrung an Relevanz verloren hat oder durch andere Effekte überlagert wurde.

Bei den anschließenden Gruppenvergleichen weiterer *biographischer* Kategorien (*Tabelle 54*) gibt es im Gegensatz zu den anderen Kohorten keine signifikanten Unterschiede in Abhängigkeit des absolvierten *BWL-Studiums* (ja/nein). Möglicherweise wurden hier durch die mehrjährige Berufserfahrung vorhandene Schwächen ausgeglichen. Knapp signifikant schlechter sind *MBA-Absolventen* im Bereich *Specialist-Know-how*. Dieses zunächst überraschende Ergebnis relativiert sich, wenn man berücksichtigt, dass die *MBA-Absolventen* (insgesamt nur 11 Mitarbeiter) zwei forciert ausgeschiedene Mitarbeiter beinhalten (18%), die einen größeren Anteil ausmachen als bei den nicht-*MBA*-Absolventen (13%). Nicht überraschend ist die Erkenntnis, dass promovierte *SCs* schlechtere Beurteilungen aufweisen. Einen signifikanten Unterschied gibt es auf der Dimension *Teamwork* (starker Effekt). Eine mögliche Erklärung wurde bereits damit geliefert, dass promovierte *SCs* größtenteils als *SC* oder *C* quer einsteigen und somit insgesamt mit Neueinsteigerproblemen zu kämpfen haben, die *SCs* mit *JC*- oder *C-Erfahrung* nicht haben. Die großen Anteile von *SC*-Quereinsteigern (>40%) und forcierten Abgängen (>28%) unter den promovierten Mitarbeitern stützen diese Annahme. Schließlich sind Personen mit *externer Berufserfahrung* durchweg schlechter, signifikant bei *Problem Solution*, *Quality of Work* und *Client Communication*. Die Gründe liegen vermutlich ebenfalls in der genannten Quereinsteigerproblematik.

Tabelle 55: Mittelwertunterschiede im Zeitverlauf (Glass'sche Deltas) (N=41)

	Messzeitpunkte		
Kriterienvariablen	**1 vs. 2**	**1 vs. 3**	**2 vs. 3**
Specialist-Know-how	-.084	-.006	.078
Problem Solution	-.102*	-.061	.041
Quality of Work	.000	.002	.002
Behaviour under Pressure	-.175	-.124	.051
Client Acceptance	-.100	-.058	.042
Client Communication	-.073	-.061	.012
Know-how-Management	-.094	-.039	.055
Teamwork	-.085	-.061	.024
Durchschnittlicher Gesamtwert	-.088**	-.052	.036

* Signifikanter Effekt auf 0.05 Niveau (Paarvergleiche der ANOVA)
** Signifikanter Effekt auf 0.01 Niveau (Paarvergleiche der ANOVA)

Abschließend wurden, wie bei den anderen Kohorten, erneut die Mittelwerte im Zeitverlauf dargestellt und über eine einfaktorielle Varianzanalyse mit Messwiederholung auf Unterschiede getestet. *Tabelle 55* zeigt die Effektstärken (Glass'sche Deltas) der Mittelwertunterschiede. Dabei konnten die Ergebnisse der anderen Kohorten nicht repliziert werden. So verbessert sich keine der *Projektbeurteilungsdimensionen* signifikant im Zeitverlauf. Neben der Möglichkeit der Nicht-Aufdeckung wahrer Effekte auf Grund anderer methodischer Restriktionen kann die Ursache wiederum in einer Varianzeinschränkung durch *Drop-Out Effekte* liegen. Zudem sind mit zunehmendem Abstand der Prädiktor- und Kriteriumsmessung überlagernde Effekte wahrscheinlich. Der kohortenübergreifende Vergleich der Dimensionen erhärtet die Vermutung, da die Standardabweichung sich massiv verringert. Dies bedeutet, dass womöglich nicht genug Varianz auf der Kriterienseite vorhanden ist, um tatsächliche Zusammenhänge aufzudecken.

3.3.4.3 Korrelationsanalyse

Die Ergebnisse der Korrelationsanalyse bei den *SCs* konnten sowohl für die Dimensionen des Auswahlverfahrens als auch für zahlreiche biographische Informationen nur vereinzelt Anhaltspunkte für prognostische Validitäten (*Fragestellungen (8)* und *(9)*) liefern.

Zunächst wurden für ausgewählte *CV-Variablen* Rangkorrelationen mit den *Projektbeurteilungsdaten* unter Einbezug der gesamten Stichprobe berechnet. *Tabelle 56* zeigt die Ergebnisse. Diese fallen auf Grund der schwachen Effektgrößen ernüchternd aus. Das *Eintrittsalter* korreliert negativ signifikant mit zahlreichen Dimensionen (schwache Effekte) und dem *Gesamtwert* (r=-.206). Es wirkt sich negativ auf die *Problemlösungskompetenz* aus. Bei diesem Zusammenhang ist zu vermuten, dass es sich, wie bei den *Cs*, um einen indirekten Effekt handelt, da auch das *Eintrittsalter* mit der *Einstiegsart* (direkt oder über die *JC*- bzw. *C-Stufe*) korreliert (r=.725). Die Ursache kann hier in der größeren Erfahrung derjenigen *SCs* liegen, die bereits als *JCs* oder *Cs* für die Firma gearbeitet haben (siehe auch 3.3.3.4). Damit steht dieses Ergebnis nicht im Widerspruch zu den gefundenen Erkenntnissen bei der *JC*- und *C-Kohorte*. Das gleiche Phänomen findet sich bei der Berufserfahrung von *SCs*. Auch hier ist die Korrelation mit der Einstiegsart i.H.v. r=.729 sehr signifikant. Darüber hinaus scheinen auch bei der Leistung der *SCs Schulnoten* nicht von wesentlicher Bedeutung zu sein. Signifikante Zusammenhänge gibt es nur bei der *Mathematiknote* mit *Problem Solution* (r=-.275) und *Quality of Work* (r=-.372) sowie bei der Englischnote mit *Quality of Work* (r=-.332). Gute Mathematikleistungen beruhen auf guten analytischen Fähigkeiten, die sich wiederum auf die Problemlösungsfähigkeit günstig auswirken dürften. Auch die *Berufserfahrung* (*Job_t*) korreliert negativ mit *Specialist-Know-how* (r=-.239) und *Client Acceptance*

(r=-.208) während Erfahrung in Form von Praktika schwach positiv mit *Client Acceptance* (r=.218) und *Client Communication* (r=.188) einhergeht. Auf eine Interpretation der Ergebnisse wird an dieser Stelle verzichtet und auf die *C-Kohorte* verwiesen, bei der ähnliche Ergebnisse vorzufinden waren.

Tabelle 56: Korrelationen Projektbeurteilungsdimensionen mit ausgewählten CV-Variablen (SCs)

Korrelationen (Spearman-Rho)								
	Eintritts-alter	No-te_FSP	Note_D	Note_M	Note_E	Abiturnote	Job_t	Pra_t
Gesamtwert Projektbeurteilung	-.206*	.054	-.005	-.188	-.175	.020	-.145	.019
Specialist-Know-how	.042	-.187	-.113	-.226	.031	-.165	-.239*	-.084
Problem Solution	-.284**	-.100	.021	-.275*	.075	-.024	-.117	.063
Quality of Work	-.158	.031	.020	-.372*	-.332*	-.046	.095	.022
Behaviour under Pressure	.086	.085	.073	.046	-.107	.032	-.083	-.159
Client Acceptance	-.074	.090	-.014	-.042	.005	.108	-.208*	.218*
Client Communication	-.243*	.066	-.026	-.073	-.120	.036	-.141	.188*
Know-how-Management	-.239*	-.102	.097	.023	-.224	.014	-.003	-.042
Teamwork	-.059	.161	.050	-.022	-.245	.054	-.145	-.025
N	86[a]	64[a]	65[a]	44[b]	35[a]	84[a]	85[a]	83[a]

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig)
[a] Für die Dimension uCA_MEAN gilt Stichprobengröße=N-2.
[b] Für die Dimension uCA_MEAN gilt Stichprobengröße=N-1.
Note_FSP: Note (1-6) im Fächerschwerpunkt; Note_D: Note (1-6) im Fach Deutsch; Note_M: Note (1-6) im Fach Mathematik; Note_E: Note (1-6) im Fach Englisch; Job_t: Zeit in Monaten an Berufserfahrung bei Einstieg in das Unternehmen; Pra_t: Zeit in Monaten an Praktikumserfahrung bei Einstieg in das Unternehmen

Unter der Annahme gleicher Symmetrieebenen wurde des Weiteren der Zusammenhang zwischen den *Gesamtwerten aus Recruiting* und *aus Projektbeurteilung* betrachtet. Dabei wurde nach Eintrittjahr und Einstiegskohorte mit und ohne Ausreißerwerten differenziert *(Tabelle 57)*. Es zeigt sich, dass es nur bei den Mitarbeitern, die als *JC* in das Unternehmen eingestiegen sind, signifikante Zusammenhänge zwischen dem *Recruiting*- und dem *Projektbeurteilungsgesamtwert* gibt. Allerdings ist hier die durchweg sehr kleine Stichprobe zu berücksichtigen. Bei den anderen Gruppen finden sich keine nennenswerten Effekte. Die Gründe hierfür liegen vermutlich in der starken Varianzreduktion durch einen *Drop-Out-Effekt*, der sich auch in der Verringerung der Standardabweichung des *Gesamtwertes aus Projektbeurteilung* auf nur noch SD=.17 zeigt. Zudem ist von einer Überlagerung der Effek-

te durch andere Moderatoren auszugehen, da zwischen der Messung des Prädiktors und des Kriteriums bei 68% der Mitarbeiter ein Zeitraum von mehr als einem Jahr liegt.

Tabelle 57: Korrelationen Gesamtwerte aus Projektbeurteilung und Recruiting

	Pearson Korrelationen (N)			
Kohorte	Alle Eintrittsjahre	Eintritt nach 2000	Alle Eintrittsjahre ohne Ausreißer	Eintritt nach 2000 ohne Ausreißer
Einstieg als JC	.174 (35)	.214 (21)	.368* (33)	.534** (19)
Einstieg als C	.147 (23)	.059 (17)	-.029 (22)	-.059 (17)
Einstieg als SC	.156 (27)	.083 (23)	-.099 (24)	-.081 (21)
Alle	.175 (85)	.102 (61)	.081 (79)	.096 (57)

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig)
** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig)

Darüber hinaus wurden Korrelationen der *Abiturnote* mit den Dimensionen der Projektbeurteilung nach Ausreißern errechnet. Dies führte zu keinen signifikanten Effekten auf *Gesamtwert*- und nur zu einem schwachen signifikanten Effekt auf *Unterdimensionsebene* bei *Specialist-Know-how* (r=-.189; siehe *Anhang X-1*). Die mangelnde Validität von Noten, die auch in der *C-Kohorte* gefunden und interpretiert wurde, trifft auf die *SC-Kohorte* ebenfalls zu. Im Anschluss wurden Korrelationsmatrizen der Recruiting-Prädiktoren mit den Kriterien (*Projektbeurteilungen*) nach Ausreißern erstellt. Dabei wurde sukzessive von der *Gesamtwertebene* auf die *Unterdimensionen* übergegangen. Die Ergebnisse (*Anhang X-1*) führen zu dem Schluss, dass der *Recruiting-Gesamtwert* lediglich mit *Problem Solution* (r=.228) signifikant schwach korreliert. Entsprechend der Gesamtwertergebnisse finden sich keine Zusammenhänge zwischen den *Recruitingvariablen* und dem *Gesamtwert aus Projektbeurteilung*. Offensichtlich wird auf dem *Gesamtwert* und allen Unterdimensionen zu wenig bei den Beurteilungen differenziert um tatsächliche Leistungsunterschiede zu identifizieren.

Bei der Analyse der unteren Ebenen zeigen sich vereinzelt inhaltlich sinnvolle Zusammenhänge. Während es für *Specialist-Know-how* keine signifikanten *Recruitingprädiktoren* gibt, zeigt sich bei *Problem Solution*, dass *Kreativität* (r=.265) und *BWL-Kenntnisse* (r=.235) signifikant sind (schwache Effekte, selbsterklärend). Darüber hinaus sind Personen mit höheren Fachkenntnissen tendenziell besser bei der generellen Dimension *Quality of Work* (r=.197). *Client Acceptance* korreliert erwartungsgemäß signifikant mit der *Kommunikationsfähigkeit* (r=.210) und *Charisma* (r=.223). *Client Communication* ebenfalls mit *Charisma* (r=.195). Dass kommunikationsstarke und charismatische (als Surrogat für eine Vielzahl von Eigenschaften) Mitarbeiter besser mit dem Klienten umgehen, wurde bereits

bei der Analyse der *JCs* und *Cs* ausführlich erläutert. Keine interpretierbaren Zusammenhänge, da hauptsächlich negative Korrelationen, zeigen sich auf den Dimensionen *Behaviour under Pressure*, *Know-how-Management* und *Teamwork*. Schließlich korreliert mit dem *Gesamtwert aus Projektbeurteilungen* keine Recruitingvariable signifikant.

3.3.4.4 Regressionsanalyse

Die bivariaten regressionsanalytischen Befunde variieren auch bei den *SCs* auf Grund der etwas kleineren Stichprobe (nur Personen, die auf allen Variablen Werte haben) leicht im Vergleich zu den bivariaten Korrelationen aus dem vorherigen Kapitel. *Tabelle 58* gibt eine Übersicht der Regressionsergebnisse (Details siehe *Anhang X-2*).

Tabelle 58: Gesamtübersicht der Regressionsergebnisse Kohorte SC (N=71)

Prädiktoren	**Abiturnote**	**Gesamtwert Recruiting**	**Recruitingvariablen (schrittweise)[a]**
Kriterienvariablen	R^2_{korr}	R^2_{korr}	R^2_{korr}
Durchschnittlicher Gesamtwert	-.008	-.013	-.-
Specialist-Know-how	.043*	-.014	-.-
Problem Solution	-.011	.026	.102**
Quality of Work	-.004	-.007	-.-
Behaviour under Pressure	-.014	.002	-.-
Client Acceptance	-.011	.025	.043*
Client Communication	-.014	-.006	-.-
Know-how-Management	-.014	-.012	-.-
Teamwork	-.014	.005	.162**

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} :Schrumpfungskorrigierter Determinationskoeffizient
[a] Es sind lediglich Modelle dargestellt, die auf Gesamtmodell- und Regressionsgewichtebene signifikant sind.

Die Regressionsanalysen ergaben erwartungsgemäß nahezu keine signifikanten Zusammenhänge der *Abiturnote* mit dem *Gesamtwert* oder den *Unterdimensionen der Projektbeurteilung*, lediglich bei *Specialist-Know-how* konnte eine geringe Varianzaufklärung von rd. 4% erreicht werden. Das heißt, dass die *Abiturnote* kaum prädiktive Validität für die Leistung als *SC* hat. Auch die Korrelationen und β-Gewichte sind in ihrer Höhe unbedeutend und weit von der Signifikanz entfernt. In den Korrelationsabschnitten der anderen Kohorten wurde bereits auf die Interpretation dieses Ergebnisses vor dem Hintergrund empirischer Befunde eingegangen. Ähnlich sind die Ergebnisse bei dem *Recruiting-Gesamtwert*. Hier

konnte keine prädiktive Validität für die *Projektbeurteilungsdimensionen* oder den *Gesamtwert aus Projektbeurteilung* festgestellt werden. Lediglich die explorative schrittweise Regression zeigte vereinzelt Varianzaufklärungen. Analog zu den korrelativen Ergebnissen konnten rd. 10% der Varianz von *Problem Solution* durch *BWL-Kenntnisse* und rd. 4% der Varianz von *Client Acceptance* durch *Kommunikationsfähigkeit* aufgeklärt werden. Darüber hinaus klären *Pragmatismus* und *intellektuelle Neugier* rd. 16% der Varianz von *Teamwork* auf. Da aber *Pragmatismus* hier ein negatives Beta-Gewicht besitzt, bedeutet dies, dass Mitarbeiter, die als pragmatisch eingeschätzt werden tendenziell weniger teamfähig sind, ein Zusammenhang für den es keine ausreichende inhaltliche Erklärung gibt. Es ist deshalb von einer Zufallskorrelation oder einem methodischen Artefakt auszugehen.

Zusammenfassend lässt sich feststellen, dass kaum Varianzanteile bei den Dimensionen mit unterschiedlichen Prädiktoren des Recruitings aufgeklärt wurden. Neben der Stichprobengröße ist auch bei den SCs vor allem die hohe *Multikolinearität* sowohl auf der Prädiktoren- als auch auf der Kriterienseite problematisch. In diesem Zusammenhang sei nochmal auf die erhebliche Varianzreduktion der *Projektbeurteilungsdimensionen* bei der Stichprobe der *SCs* gegenüber den anderen Kohorten hingewiesen, die darin begründet sein kann, dass viele Mitarbeiter, die schlechte Werte auf diesen Dimensionen hatten, bereits ausgeschieden sind (*Drop-Out-Effekte*). Dies verdeckt korrelative und prädiktive Zusammenhänge. Zudem scheinen auf dem *SC-Level* die Beurteilungen an Strenge abzunehmen, so dass keine genügend große Differenzierung der beurteilten Mitarbeiter vorliegt. Schließlich ist wiederum auf Grund der einfachen Durchschnittsbildung über alle Dimensionen von einer Verletzung der *Symmetrieprinzipien nach Brunswik* auszugehen, die zu unfairen Validitätstests führt.

3.3.4.5 Faktorenanalyse

Auf die detaillierte Interpretation der Faktorenanalyse der Prädiktorendaten wird an dieser Stelle verzichtet und auf die Analyse der Prädiktorendaten verwiesen. Die Korrelationsanalysen der Recruitingfaktoren mit den Projektbeurteilungsdimensionen zeigen, dass es sowohl bei der *rotierten* als auch der *unrotierten* Faktorlösung der *Recruitingvariablen* wenige signifikante Korrelationen zu den Projektbeurteilungsdimensionen gibt (*Anhang X-3)*. Zu berücksichtigen sind hier neben den zahlreichen erwähnten methodischen Restriktionen allerdings die sehr geringe Stichprobengröße (N=17). Aus diesem Grund wird im Folgenden auf eine inhaltliche Interpretation verzichtet und lediglich eine Beschreibung der Ergebnisse vorgenommen. Ausschließlich bei den unrotierten und rotierten Levelfaktoren der *JC-*

Einstiegs-Kohorte gibt es signifikante korrelative Zusammenhänge zu dem *Projektbeurteilungsgesamtwert* (r=.551 und r=.517). Dieses Ergebnis spiegelt die bei der Korrelationsanalyse festgestellten Zusammenhänge mit dem Recruiting-Gesamtwert wider. Mit dem unrotierten Allgemeinfaktor der *JC-Einstieg-Kohorte* korreliert zudem stark *Client Acceptance* (r=.554). Diese Ergebnisse sind aus Sicht des Autors inhaltlich nicht interpretierbar.

Tabelle 59: Komponentenmatrizen der Projektbeurteilungsvariablen (N=82)

	Extrahierte unrotierte Faktoren[a]			**Extrahierte rotierte Faktoren[a,b]**		
Kriteriendimensionen	1	2	3	1	2	3
Specialist-Know-how	.538		-.607	.798		
Problem Solution	.710		-.470	.838		
Quality of Work	.811			.513	.570	.310
Behaviour under Pressure	.411		.549		.430	.521
Client Acceptance	.439	-.629	.385			.856
Client Communication	.580	-.461		.438		.598
Know-how-Management	.333	.662			.709	
Teamwork	.571	.470	.380		.814	
VA[c]	32.3%	49.7%	64.7%	22.8%	44.2%	64.7%

Faktor unrotiert	1	2	3	4	5
Eigenwert	2.586	1.387	1.201	.902	.629

Dargestellt sind nur Faktorladungen mit einem Betrag >.300.
a Extraktionsmethode: Hauptkomponentenanalyse.
b Rotationsmethode: Varimax mit Kaiser-Normalisierung. Die Rotation ist in 3 Iterationen konvergiert.
c Varianzaufklärung, kumulativ dargestellt

Tabelle 59 zeigt die Ergebnisse der Faktorenanalyse der Projektbeurteilungsdimensionen. Dabei konnten die Ergebnisse aus der *JC-* und *C-Kohorte* teilweise repliziert werden. Auf der Projektbeurteilungsseite gibt es ebenso einen *Levelfaktor* (bei der unrotierten Lösung), der das *Gesamtlevel* widerspiegelt und *Profilfaktoren*. Die Profilstruktur stellt sich aber etwas differenzierter dar als bei den anderen Kohorten und ist schwer zu interpretieren. Bei den *SCs* scheint in der unrotierten Variante der zweite Faktor schwerpunktmäßig das Klientenverhalten abzubilden (*Client Communication* und *Client Acceptance*). Der dritte Faktor hat hingegen Intelligenzkomponenten als Schwerpunkte (*Specialist-Know-how* als kristalline Intelligenz und *Problem Solution* als fluide Intelligenz). Die rotierte Lösung zeigt bei dem ersten Faktor ebenfalls Intelligenz-, beim zweiten und dritten Faktor Verhaltens-

komponenten. Die dritten Faktoren ergeben sowohl bei der rotierten als auch der unrotierten Variante hingegen wenig interpretierbare Muster.

Bei nur wenigen Recruitingvariablen zeigen sich signifikante Korrelationen mit den *Projektbeurteilungsfaktoren* (*Anhang X-4*). Offenbar hängen *BWL-Kenntnisse* schwach mit dem *unrotierten Levelfaktor* zusammen (r=.232). Der zweite und der dritte Faktor korrelieren wiederum signifikant negativ mit *Pragmatismus* (r=-.409) bzw. *intellektueller Neugier* (r=-.235) und dem *Gesamtwert aus Recruiting* (r=-.201 und r=-.223), Zusammenhänge, die inhaltlich nicht ausreichend interpretierbar sind. Die rotierte Variante korreliert bei dem ersten Faktor mit *methodischen Kenntnissen* (r=.223). Da diese zum Problemlösungsverhalten notwendig sind und *Problem Solution* hoch auf dem Faktor lädt, ist der Zusammenhang erklärbar. Während der zweite Faktor wiederum negativ mit *Pragmatismus* signifikant korreliert (r=-.510), gibt es beim dritten Faktor keine signifikanten Zusammenhänge.

Die korrelativen Vergleiche zwischen den Faktoren fallen ebenso ernüchternd aus. Sie ergeben, dass der unrotierte *Levelfaktor* der ehemaligen *JCs* in vergleichbarer Höhe mit dem *unrotierten Allgemeinfaktor* aus der *Projektbeurteilung* korreliert (r=.525), wie die *Gesamtwerte aus Recruiting und Projektbeurteilung* der JC-Einsteiger (r=.534) (siehe *Tabelle 57* und *Anhang X-5*). Darüber hinaus korreliert der *unrotierte JC-Einstiegsfaktor* mit dem dritten *rotierten Projektbeurteilungsfaktor*, der Stress- und Klientenverhalten widerspiegelt (siehe Faktorladungen *Tabelle 19*). Inhaltlich bedeutet dies, dass Mitarbeiter, die im *JC-Recruiting* insgesamt hoch beurteilt wurden auch in Bezug auf Stressverhalten und im Verhalten gegenüber dem Klienten als *SC* tendenziell besser sind. Darüber hinaus weisen die restlichen Faktoren keine Zusammenhänge auf.

Zusammenfassend lässt sich feststellen, dass die Faktorenanalyse keine Validitätsgewinne gegenüber den einfachen bivariaten Korrelationen bei der *SC-Kohorte* erreichen konnte.

3.4 Erweiterte Fragestellungen

Nachfolgend sollen kurz einige Überlegungen zur Auswahl und Operationalisierung von Prädiktoren und Kriterien dargestellt werden. Dabei wird auf einzelne, im theoretischen Teil der Arbeit erarbeitete Aspekte eingegangen. Im Anschluss wird auf die im vorliegenden Fall angewandte Urteilsbildung und die Folgen für die Verteilung eignungsdiagnostischer Entscheidungen fokussiert.

3.4.1 Zur Auswahl und Operationalisierung von Prädiktor und Kriterium

Unter Kapitel 2.4 wurde erarbeitet, welche Komponenten Beratungserfolg enthält, und darauf aufbauend das *Fünf-Säulen-Modell des Beratungserfolges* abgeleitet (*Abbildung 28*). Vergleicht man nun diese Komponenten mit den in der Unternehmensberatung verwendeten Bögen zur Beurteilung des Beratungserfolges (*Anhang VI*), kommt man zu folgender Erkenntnis: Die *Eigenschaftskriterien* des Beraters scheinen sich inhaltlich auch in den Bögen der *ISB* wieder zu finden. Dies betrifft *analytische Fähigkeiten* genauso wie *Sozialkompetenz*, *Wissensaspekte* und *kommunikative Fähigkeiten*. Für die *Eigenschaftskriterien* des Klienten gilt dies nur bedingt. Zwar ist in den Originalbögen Raum für Schilderung spezieller Probleme mit dem Klienten, diese dienen aber eher der Rechtfertigung für die Vergabe bestimmter Noten. Eine systematische Erfassung der erarbeiteten Klientenaspekte, die auch die Ermittlung einer numerischen Größe beinhaltet, bleibt jedoch aus. Dies ist bedauerlich, da die Besonderheiten des Kunden, wie gezeigt, den Projekterfolg, der unter anderem auch Ergebnis der Interaktion zwischen Klienten und Beratern ist, erheblich beeinflussen können. Ähnliches gilt für die *Kontextvariablen*. Auch hier liefert die angewandte Systematik keine Einschätzung der Rahmenbedingungen. Sowohl *Status* als auch *Zielvereinbarung*, *Ressourcen* oder *Projektorganisation* finden bei der Beurteilung der erbrachten Leistung bestenfalls im Rahmen des klinischen Urteils des Projektmanagers Berücksichtigung. Das klinische urteil ist aber, wie unter 2.2.5.1 gesehen, nur bedingt valide. Die *Ergebniskriterien* werden teilweise inhaltlich abgedeckt. Dies betrifft vor allem die *Zufriedenheit des Klienten*, die sich mit *Client Acceptance* inhaltlich überschneidet. Auch die *UB-Ziele* finden sich ansatzweise unter *Know-how-Management* wieder, das die Bewertung internen Wissenszuwachses beinhaltet. Die *Dokumentation* wird unter *Quality of Work* thematisiert. Alle weiteren Variablen werden nicht berücksichtigt. Der größte Teil der *Prozesskomponenten* bleibt ebenfalls verdeckt. Lediglich *Problem Solution* und *Behaviour under Pressure* greifen einzelne Punkte auf. Ferner gelten für alle verwendeten Kriterien die Vor- und Nachteile, die in *Tabelle 2* dargestellt wurden.

Die mangelnde Inhaltsvalidität auf der Kriterienseite kann schwerwiegende Folgen für die Mitarbeitermotivation haben. So wird die Leistung bzw. der Beratungserfolg an vielen Stellen durch Gegebenheiten beeinflusst, auf die der Berater selbst keinen Einfluss hat. In der Urteilsfindung werden sie allerdings nicht explizit berücksichtigt, so dass hier durchaus die Fairness des Beurteilungssystems in Frage gestellt werden kann. Für die vorliegende Studie bedeutet dies, dass die nicht betrachteten Variablen als Moderatoren auf das Urteil wirken können und somit die wahren Effekte (*NTR-Box aus Abbildungen 2* und *14*) überla-

gern. Zudem ist von einer mangelnden *Konstruktvalidität* des Kriteriums (R_{KR}) auszugehen. Die Analysen zu der *Symmetrie* haben außerdem gezeigt, dass Kriterien unterschiedlicher Generalitätsebenen auf gleicher Ebene angesetzt wurden. Eine Neustrukturierung aus inhaltlichen Überlegungen, wie in ansatzweise in Kapitel 2.4.3 vorgenommen, sowie faktoranalytischen Erkenntnissen ist deshalb zu empfehlen.

Neben der inhaltliche Betrachtung des Kriteriums ist die Operationalisierung entscheidend. Die verwendeten Beurteilungsbögen gewichten alle verwendeten Oberdimensionen und innerhalb der Oberdimensionen die Einzelvariablen gleich. Dass aber alle Aspekte gleich bedeutsam sind, ist mehr als fraglich. Sinnvoller scheint aus Sicht des Autors eine auf Expertenmeinungen basierende Gewichtung der einzelnen Variablen. Die Anwendung von *MAUT* ist hier denkbar und dürfte zudem zu einer hohen Akzeptanz der Beteiligten führen. Auch die mit der Person des Beurteilers einhergehenden Effekte sind nicht unproblematisch. Im vorliegenden Fall wird die Skalierung ausschließlich von den Projektmanagern vorgenommen. Eine Beurteilung durch mehrere Personen wäre aus *MTMM*-Gesichtspunkten sicherlich wünschenswert, allerdings praktisch kaum umsetzbar, da auf einem Projekt selten genug Personen aus den eigenen Reihen sind, die die Arbeit des Beraters beurteilen können. Problematisch ist aus Sicht des Autors eher die Tatsache, dass einzelne Dimensionen von dem Projektmanager nur bedingt oder gar nicht eingeschätzt werden können. So kann die Klientenakzeptanz ohne Zweifel am besten durch den Klienten selbst beurteilt werden, was die Frage aufkommen lässt, warum man diese Daten nicht erhebt. Wenn die Umstände es zulassen, ist auch sicherlich die Durchführung einer 360°-Beurteilung (siehe Scherm, 2004) sinnvoll.

Die im untersuchten Auswahlverfahrens verwendeten Prädiktoren wurden im Rahmen eines Arbeitskreises von Mitarbeitern des Unternehmens festgelegt. Dabei handelt es sich ausschließlich um subjektive Einschätzungen auf Basis impliziter Persönlichkeitstheorien und nicht um ein auf wissenschaftlichen Erkenntnissen basierendes Vorgehen. Dies zeigt sich auch deutlich in den verwendeten Prädiktordimensionen, die nicht wissenschaftlich fundierten Konstrukten entsprechen und auch, wie die Analysen gezeigt haben, vermutlich unterschiedliche Generalitätsebenen besitzen. Inhaltlich betrachtet scheint u.a. die Erfassung folgender Konstrukte, die sich als prädiktiv valide herausgestellt haben, im Rahmen des Auswahlverfahrens viel versprechend zu sein: *Intelligenz* und die Big 5-Faktoren *Gewissenhaftigkeit*, *Extraversion*, *Emotionale Stabilität*, *soziale Kompetenz* und *Verträglichkeit* sowie *Leistungsmotivation*. Darüber hinaus ist die Erfassung von *Wissenskomponenten* und *kommunikativen Fähigkeiten* empfehlenswert. Doch nicht alle der genannten

Konstrukte eignen sich dazu, über *Interviews* operationalisiert zu werden. Aus Sicht des Autors sollte im Rahmen von *Fachinterviews* das *Wissen* abgefragt und die *Intelligenz* im Rahmen von *Intelligenztests* erfasst werden. Die Beobachtungen im Rahmen der Gruppenarbeit eignen sich hingegen dazu, die *soziale Kompetenz* und *Verträglichkeit* des Bewerbers zu beurteilen. Über Persönlichkeitstests könnten ferner *Emotionale Stabilität*, *Leistungsmotivation* und *Gewissenhaftigkeit* erfasst werden. Im Rahmen der *Präsentation* kommen, neben dem *Wissen*, die *kommunikativen Fähigkeiten* am besten zum Ausdruck. Schließlich könnten in dem *Persönlichkeitsinterview* die *Persönlichkeitseinschätzungen* aus der *Gruppenarbeit* und dem *Persönlichkeitstest* gezielt, auch unter Anwendung von Stressinterviewtechniken, verifiziert werden. Unter 3.6 werden die sich ergebenden praktischen Implikationen nochmals aufgegriffen und detailliert dargestellt.

3.4.2 Urteilsbildung und Verteilung diagnostischer Entscheidungen

Im Rahmen des Auswahlverfahrens werden die Entscheidung über Einstellung eines Kandidaten am Ende des Auswahlverfahrens in der Gruppe der Beurteiler diskutiert. Am Ende dieses Diskussionsprozesses entsteht ein Konsensurteil, das, wie berichtet, durch die Diskussion nicht an Güte zunimmt (Schuler, 2000, S.127). Es ist allerdings im Rechtfertigungskontext für das spätere Feedback beim Bewerber sinnvoll. Die Beibehaltung dieses Vorgehens scheint deshalb empfehlenswert. Im Kapitel 2.2.5.1 wurde allerdings ausführlich die Überlegenheit statistischer gegenüber klinischer Urteilsbildung herausgearbeitet. Im vorliegenden Fall bedeutet dies, dass eine Trennung von Bewertung und Urteil vorgenommen und das abschließende Einstellungsurteil rein statistisch errechnet werden sollte. Die Gewichtung der einzelnen Dimensionen könnte dabei auf über multiple Regression gewonnenen Erkenntnissen beruhen. Für diesen Zweck müssen allerdings kreuzvalidierte Regressionsgewichte herangezogen werden, die auch das bisherige kompensatorische Verfahren, das eine Gleichgewichtung aller Recruitingdimensionen vornimmt, ersetzen sollten.

Ein weiterer wesentlicher Aspekt bei der Entscheidungsfindung sind die Folgen im Sinne der Verteilung diagnostischer Entscheidungen. Es soll im Folgenden bei der Kohorte der *Cs* davon ausgegangen werden, dass alle Berater, die einen Evaluationsmittelwert kleiner als 3.5 haben, zu den FP gehören. Setzt man als tetrachorischen Korrelationskoeffizienten den Validitätskoeffizienten von r=.369 in *Gleichung (12)* ein, so ergeben sich bei einer Selektionsrate von SQ=40%, eine Basisrate von 58% (durch Vorauswahl) und die Verteilung und Gütekriterien analog *Abbildung 51*.

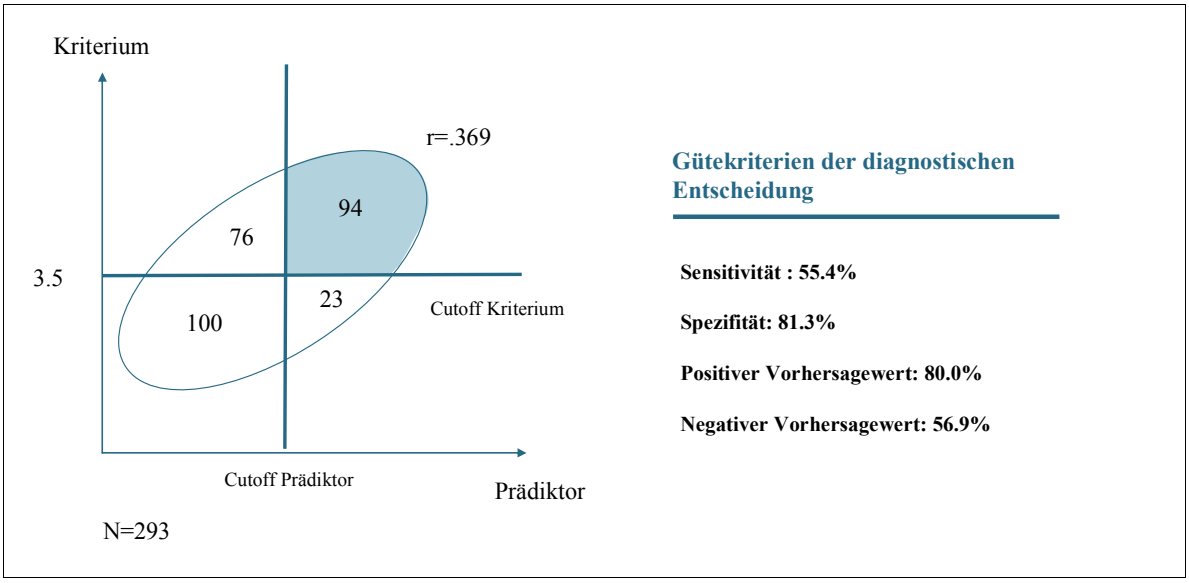

Abbildung 53: Verteilung der diagnostischen Entscheidungen bei der C-Kohorte

Dabei zeigt sich deutlich, dass die Schwäche des Auswahlverfahrens hauptsächlich in dem negativen Prädiktionswert liegt. So ist die Wahrscheinlichkeit, dass eine negative Prognose richtig ist, mit 56.9% ziemlich gering, da nur 55.4% der positiven Zustände auch als solche erkannt werden (Spezifität). Im Kapitel 2.3 wurde bereits auf die besondere Problematik der FN bei Unternehmensberatungen eingegangen. Ziel sollte deshalb auch sein, diese durch ein Vorgehen höherer Validität zu verringern, um negative Nutzeneffekte, die mit derzeitigen Methoden in ihrer Höhe nicht berechenbar, aber dennoch wahrscheinlich sind, zu reduzieren. Welche realistischen Nutzeneffekte durch mögliche Änderungen und eine einhergehende Validitätserhöhung zu erreichen sind, soll nachfolgend aufgezeigt werden.

3.5 Nutzenanalyse

Die Nutzenkalkulation des vorliegenden Auswahlverfahrens erfolgt analog dem Boudreau-Modell aus *Gleichung 19* und berücksichtigt somit Investitionsgesichtspunkte. Die für die Operationalisierung der Modellparameter notwendige Datenbasis wurde von der Personalabteilung der Unternehmensberatung zur Verfügung gestellt oder (z.B. beim Kalkulationszinsfuss) durch den Autor auf Basis plausibler Annahmen geschätzt. Darüber hinaus wurden in der Nutzenkalkulation die Beförderungen der Mitarbeiter und somit ein kohortenspezifischer Nutzeneffekt im Zeitverlauf modelliert. Zusätzlich wurde über Variation der Selektionsrate *SQ*, des Validitätsparameters r_{xy} und der Standardabweichung *SDy* eine Sensitivitätsanalyse durchgeführt. Inhalt der Nutzenberechnung sind lediglich die Kohorten *JC*, *C* und *SC,* die auch Bestandteil der empirischen Untersuchung waren sowie ausschließlich das Assessment Center, d.h. die Vorauswahl wird als gegeben angesehen. Außerdem wurde le-

diglich der quantifizierbare Nutzen ermittelt und der Nutzen 2. oder 3. Art, (*Abbildung 25*) nicht erfasst.

3.5.1 Operationalisierung der Parameter

Tabelle 60 gibt eine Übersicht der Operationalisierung der Parameter. Die Schätzgrößen für *SDy* ergeben für *JCs*, Cs und *SCs* jeweils durchschnittlich Werte in Höhe von 76 071, 89 248 und 111 078 EUR. Für das bisher verwendete Auswahlverfahren, das z.T. einige kleinere Änderungen im Laufe der Jahre erfahren hat, wurde eine Anwendungszeitraum von insgesamt 5 Jahren angenommen. Grundlage der Berechnungen ist zudem eine geschätzte mittlere Verweildauer der Mitarbeiter im Unternehmen von 3.0 Jahren. Der Einfachheit halber wurde analog zu Boudreau (1983a) angenommen, dass die durchschnittliche Zahl der jährlich eingestellten Mitarbeiter am Anfang des Jahres eingestellt wird und am Ende der mittleren Verweildauer (3 Jahre) ausscheidet. Daraus ergibt sich ein Betrachtungszeitraum von insgesamt 7 Jahren. Die Verweildauer wurde zudem als unabhängig von der Einstiegsstufe angenommen. Ferner wurde eine Beförderung in die nächste Funktionsstufe im Jahresrhythmus vorausgesetzt. Die Anzahl der *ACs* basiert auf historischen Daten der Personalabteilung und wurde für die Perioden 1 und 2 geschätzt. Gleiches gilt für die Selektionsquote *SQ* und folglich für die Anzahl akzeptierter Bewerber. Für die Periode 5 wurde ein Schätzwert der Personalabteilung angesetzt. Die bewerberabhängigen Kosten wurden für die Perioden 1 und 2 aus dem Durchschnitt der historischen Werte der Perioden 3 bis 5 hochgerechnet.

Für die beobachterabhängigen variablen Kosten wurden jeweils sechs Beobachter- bzw. Interviewer-Manntage pro durchgeführtem Auswahlverfahren angenommen und pro Person 1000 EUR, inklusive Reisekosten, angesetzt. Die fixen Kosten aus dem Personalbereich für die ersten 2 Perioden wurden aus dem historischen Durchschnitt der Perioden 3 bis 5 berechnet. Sie beinhalten Personalkosten der mit dem Auswahlverfahren betrauten Mitarbeiter sowie anteilsmäßige Gemeinkosten (Overhead). Die fixen Kosten wurden in die Berechnung mit einbezogen, da davon ausgegangen wurde, dass bei einer ausschließlichen Auswahl nach Bewerbungsunterlagen und keinem weiteren Verfahren, diese nicht mehr anfallen, da entsprechend Kapazitäten freigesetzt würden. Für die Entwicklung des Verfahrens wurden drei Monate zu drei Manntagen á 250 EUR angesetzt und die angefallenen Kosten dem ersten Jahr zugerechnet. *Tabelle 61* fasst das Dargestellte zusammen.

Tabelle 60: Operationalisierung der Parameter zur Nutzenberechnung

Parameter	Vorgehen
Q	Ermittlung aus historischen empirischen Werten aus *1999* bis *2001* sowie als Schätzgröße der Personalabteilung. Sie variiert jahresabhängig und wurde für alle Kohorten als gleich groß angenommen
λ	Der Ordinatenwert der Standardnormalverteilung an der Stelle Q wurde tabellarisch ermittelt.
$\overline{Z}_X^*$	Ergibt sich für jedes Jahr rechnerisch aus *Gleichung (15)*. Für die Perioden 6 und 7 wurde zur Nutzenberechnung als Basis die Durchschnittsselektionsquote der vorherigen Perioden verwendet (konservative Schätzung)
i	Der Kalkulationszinsfuß für die Diskontierung künftig anfallender Gewinne orientiert sich an Marktgegebenheiten und wurde auf 8% festgelegt
s	Der Steuersatz setzt sich aus Körperschafts- (Spitzengrenzsteuersatz) und Gewerbesteuerhebesatz zusammen und wurde für die Vergangenheitswerte mit 66% angesetzt. Für die prognostische Betrachtung wurde ein Gewerbersteuerhebesatz von 490%, ein Körperschaftssteuersatz von 25% und ein Solidaritätszuschlag von 5.5% angenommen (entspricht 40.8% effektiver Steuersatz)
(1-v)	Der Einfluss der variablen Kosten der Mehrleistung auf den Gewinn wurde als nicht vorhanden angesehen, da durch Nebenkostenbudgets die variablen Kostenanteile durch den Klienten separat abgedeckt werden
SD_{Prod}	Die Standardabweichung der Produktivität wurde empirisch für die einzelnen Kohorten für die Jahre 2001 und 2002 festgestellt und für die anderen Jahre als Durchschnitt angesetzt
Tagessatz	Die Tagessätze wurden für jede Kohorte einzeln entsprechend üblichen Marktsätzen festgelegt: JC: 1800 EUR, C: 2250 EUR und SC: 3000 EUR
SDy	SDy wurde aus den Tagessätzen und den Produktivitäten errechnet über die Formel: $SDy=SD_{Prod}$*220 Arbeitstage*Tagessatz

Tabelle 61: Kennwerte der einzelnen Behandlungsperioden

			Bewerber				**Kosten ['000 EUR]**			
							Variabel (Cv)		Fix (Cf)	
Periode k	Zeitraum	Anzahl ACs	Gesamt	SQ	Nat	Nst	Bewerber	Beobachter	HR	Entwicklungskosten
1	1-12/1998	52[1]	221[1]	0.45[1]	100[1]	0	1 198[1]	315[1]	343[1]	45
2	1-12/1999	81[1]	340[1]	0.48	163[1]	0	1 842[1]	484[1]	343[1]	0
3	1-12/2000	63	181	0.56	102	100	983	378[1]	304	0
4	1-12/2001	51	263	0.32	83	163	841	306[1]	437	0
5	1-12/2002	14	95	0.20[1]	19	102	315	84[1]	288	0
6	1-12/2003	0	0	0	0	83	0	0	0	0
7	1-12/2004	0	0	0	0	19	0	0	0	0

[1] Schätzwerte; SQ= Selektionsquote; Nat= Akzeptierte Bewerber; Nst= Ausgeschiedene Bewerber, Cv= variable Kosten; Cf= fixe Kosten

3.5.2 Nutzenberechnung

In einem ersten Schritt wurde der Nutzen des vorliegenden Verfahrens im Vergleich zur Zufallsauswahl (nach Vorauswahl) berechnet. Dabei wurde, separat nach Kohorten, auf der Annahme aufgebaut, dass Mitarbeiter in der Kohorte der *JC* und *C* jeweils ein Jahr und die restliche Zeit in der Kohorte der *SC* verbleiben (*Tabelle 62*).

Tabelle 62: Mitarbeiterfluss bei der Nutzenberechnung

		Zugänge				**Anfangsbestand**			
Periode k	Zeitraum	JC	C	SC	Gesamt	JC	C	SC	Gesamt
1	1-12/1998	32	37	31	100	32	37	31	100
2	1-12/1999	53	60	50	163	53	92	118	263
3	1-12/2000	33	38	31	102	33	90	242	365
4	1-12/2001	27	31	26	83	27	64	258	348
5	1-12/2002	6	7	6	19	6	34	164	204
6	1-12/2003	0	0	0	0	0	6	96	102
7	1-12/2004	0	0	0	0	0	0	0	19

Zudem basiert die Berechnung auf der Prämisse, dass die Validität des Auswahlverfahrens bei der *SC-Kohorte* der Validität bei der *C-Kohorte* entspricht (r=.369). Die Analysen aus 3.3 hatten Anlass zu der Vermutung gegeben, dass auf Grund großer Varianzeinschränkung der *SC*-Projektbeurteilungen in der vorhandenen Stichprobe und weiterer methodischer Artefakte sowie Überlagerungseffekten wahre Zusammenhänge nicht aufgedeckt werden konnten. Die Annahme gleich hoher Validität ist auf Grund des ähnlichen Aufgabenspektrums der *SCs* und *Cs* vertretbar. Zudem konnten für jene *SCs*, die als *JC* eingetreten sind und in der Realität einen Großteil der *SCs* ausmachen dürften, beachtliche und weitaus höhere Korrelationen festgestellt werden, auch wenn diese auf einer sehr kleinen Stichprobe beruhen (siehe dazu *Tabelle 57*). Die Annahme gleicher Validitätshöhen für beide Kohorten scheint deshalb einen eher konservativen Charakter zu haben. In *Tabelle 63* sind die Kosten des Auswahlverfahrens nach Perioden dargestellt. In dem angegebenen Zeitraum des Einsatzes des Auswahlverfahrens fallen demnach, unter Berücksichtigung steuerlicher und investitionsrechnerischer Aspekte, Kosten in Höhe von rd. 2.7 Mio. EUR an.

Tabelle 63: Kostenübersicht nach Perioden (in '000 EUR)

Periode k	Nk	C_v	C_f	(1-s)	$1/(1+i)^{k-1}$	Σ Kosten
1	100	1 512	388	0.333	1.000	634
2	263	2 327	343	0.333	0.974	867
3	365	1 361	304	0.333	0.949	527
4	348	1 147	437	0.333	0.924	488
5	204	399	288	0.333	0.900	206
6	102	0	0	0.333	0.877	0
7	19	0	0	0.333	0.854	0
Σ						**2 721**

Nk= Verbliebene eingestellte Mitarbeiter; die restlichen Bezeichnungen gelten analog zu Gleichung 18 und 19

Tabelle 64: Nutzenübersicht und Ergebnisbeitrag nach Perioden (in '000 EUR)

Periode k	JC	C	SC	Σ Nutzen	Σ Kosten	Ergebnisbeitrag
1	210	346	359	914	634	281
2	317	801	1 270	2 387	867	1 520
3	161	636	2 120	2 917	527	2 391
4	216	711	3 592	4 519	488	4 031
5	56	460	2 767	3 283	206	3 077
6	0	56	1 081	1 137	0	1 137
7	0	0	209	209	0	209
Σ	**960**	**3 009**	**11 397**	**15 367**	**2 721**	**12 646**

Demgegenüber stehen die Nutzenbeträge nach Kohorte aus *Tabelle 64* in Höhe von rd. 15.4 Mio. EUR (Details siehe *Anhang XI-1*). Dabei zeigt sich, dass mit Abstand der größte Nutzen von der Kohorte der *SCs* ausgeht. Dies liegt sicherlich teilweise an dem erheblich höheren Tagessatz, den SCs fakturieren. Zudem verbringen die eingestellten Mitarbeiter einen Großteil ihrer Zeit, die sie für das Unternehmen arbeiten, auf der SC Stufe. Insgesamt ergibt sich ein Ergebnisbeitrag nach Steuern in Höhe von 12.6 Mio. EUR.

3.5.3 Sensitivitäts- und Break-Even-Analyse ausgewählter Parameter

Ausgangspunkt der Sensitivitäts- und Break-Even-Analysen ist eine erneute Berechnung des Nutzens des aktuellen Auswahlverfahrens, allerdings nicht aus historischer, son-

dern prognostischer Perspektive. Dabei wurde die gleiche Laufzeit des Programms für 5 Jahre sowie eine Selektionsrate von SQ=20 % bei einem Bedarf von insgesamt 50 Beratern pro Jahr (30 *JCs*, 15 *Cs* und 5 *SCs*) angenommen.

Die variablen Kosten wurden entsprechend hochgerechnet und die fixen Kosten der Personalabteilung inkl. Gemeinkosten auf 400 000 EUR angesetzt. Als SDy wurden die Durchschnittswerte der historischen Betrachtung angesetzt. Die Entwicklungskosten des Verfahrens wurden nicht mehr in der Rechnung berücksichtigt, da das Verfahren nicht neu entwickelt wird. Grundlage der Berechnungen ist zudem ein den aktuellen Gegebenheiten angepasster Körperschaftssteuersatz von 25%, Gewerbesteuerhebesatz von 490%, sowie ein Solidaritätszuschlag von 5.5%. Alle weiteren Parameter wurden konstant gehalten. Vor dem Hintergrund dieser Ausgangssituation errechnen sich ein Ergebnisbeitrag von rd. 12.3 Mio. EUR nach Steuern (*Tabelle 65*).

Bei der Sensitivitätsanalyse wurden die Selektionsrate und die Validität, die für alle Kohorten als gleich angenommen wurde, jeweils bei Konstanthaltung aller anderen Parameter variiert. *Abbildung 54* visualisiert das Ergebnis (siehe auch *Anhang XI-3*).

Tabelle 65: Prognostische Nutzenübersicht nach Perioden [in '000 EUR]

	Kosten	**Nutzen**				**Ergebnisbeitrag**
Periode k	gesamt	JC	C	SC	Σ	gesamt
1	1 248	544	391	162	1 097	-152
2	1 192	519	1 120	774	2 413	1 221
3	1 138	496	1 069	2 218	3 783	2 645
4	1 087	473	1 021	2 118	3 612	2 525
5	1 038	452	975	2 022	3 449	2 411
6	0	0	621	1 802	2 422	2 422
7	0	0	0	1 229	1 229	1 229
Σ	**5 702**	**2 483**	**5 196**	**10 325**	**18 005**	**12 302**

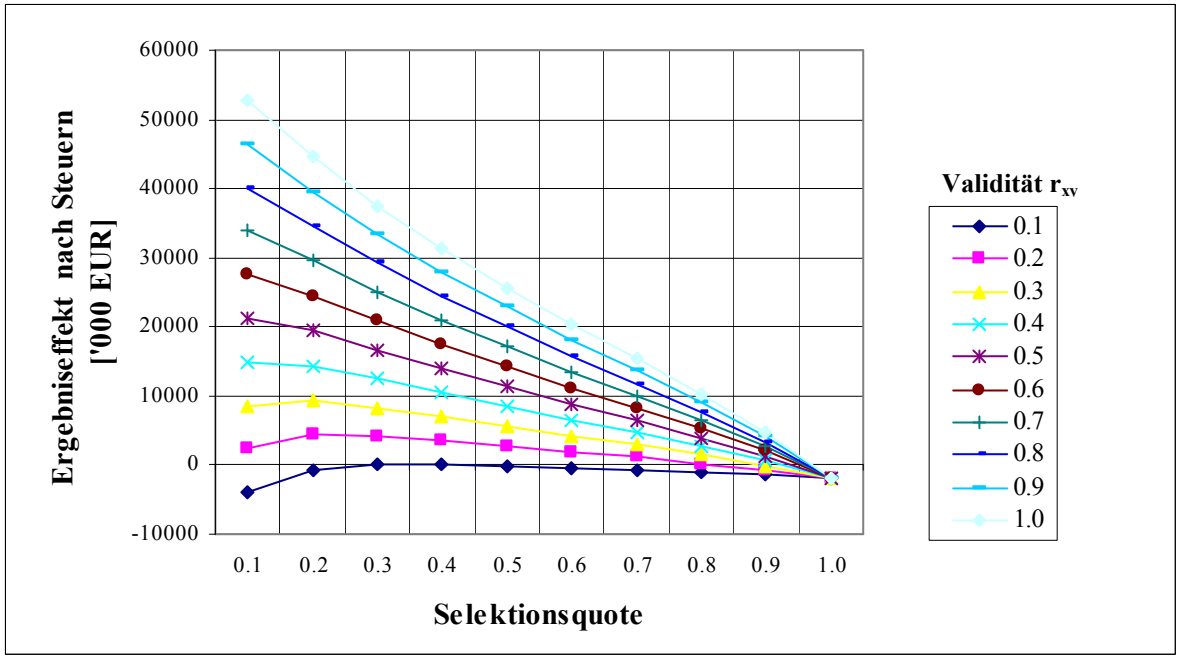

Abbildung 54: Sensitivitätsanalyse mit Validität und Selektionsquote

Dabei zeigt sich, dass mit zunehmender Selektionsquote, die Ergebniseffekte zunehmender Validität geringer werden und ab einer Validität von r=.20 bei einer Selektionsquote von maximal SQ=80% bereits positive Effekte durch das Personalauswahlverfahren zu erreichen sind. Betrachtet man das Auswahlverfahren unter den eingangs geschilderten Bedingungen, so liegt der Break-Even des Ergebniseffektes bei einer Selektionsrate von SQ=.958 bzw. SQ=.042. Bei einer konstanten Selektionsrate von SQ=.200 wird der Break-Even bei einer kohortenübergreifenden Validität von r=.113 erreicht (siehe *Anhang XI-8*).

In einem weiteren Schritt wurde der Parameter der Produktivität *SDy* unter Konstanthaltung aller anderen Parameter variiert. *SDy* wurde dabei für alle Funktionsstufen als gleich angenommen. *Abbildung 55* visualisiert die Ergebnisse für verschiedene Selektionsquoten (siehe auch *Anhang XII-4*). Erwartungsgemäß steigt der Ergebnisbeitrag mit zunehmendem *SDy* linear an, wobei die Steigung der Geraden mit abnehmender Selektionsquote stärker wird. Bei einer Selektionsquote von SQ=20% liegt der Break-Even-Point bei SDy=31 020.

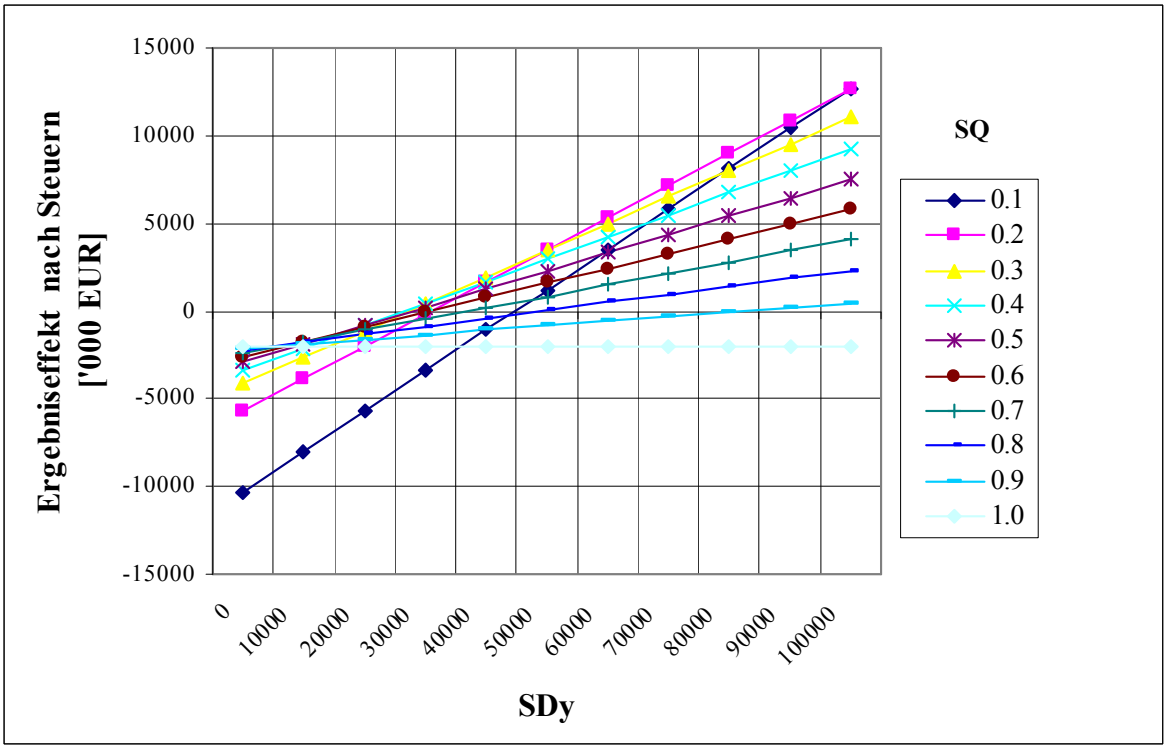

Abbildung 55: Sensitivitätsanalyse mit SDy und Selektionsquote SQ

3.5.4 Validitätserhöhung und Nutzenvorteil

Im vorherigen Kapitel wurde der prognostische Nutzen des aktuell benutzten Auswahlverfahrens berechnet und bei einer Anwendung in den nächsten fünf Jahren vor den gegebenen Prämissen auf 12.3 Mio. EUR taxiert.

Durch Änderungen im Vorgehen und bei der Auswahl der Instrumente ist eine Validitätssteigerung zu erwarten, deren ökonomische Auswirkung aufgezeigt werden soll. Als Ausgangsbasis dienen die metaanalytisch ermittelten prognostischen Validitäten aus *Tabelle 14*. Dabei dürften die zusätzlichen Kosten durch die Einführung des neuen Verfahrens variieren. Die Einführung von Tests dürfte, im Zeitalter der Computertechnologie, im Vergleich zu den restlichen Kosten und Nutzen, eher marginale zusätzliche Kosten verursachen. Dies gilt natürlich nur unter der Prämisse, dass man einen aus der Vielfalt vorhandener Tests käuflich erwirbt. Die Entwicklung von *ACs, Arbeitsproben* und *strukturierten Interviews* ist hingegen aufwendiger. Für den vorliegenden Fall wurde davon ausgegangen, dass die variablen Reisekosten des Bewerbers in gleicher Höhe bestehen bleiben. Zudem wurde angenommen, dass bei der Durchführung des *ACs*, *Arbeitsproben* und *strukturierten Interviews* ebenfalls Berater des Unternehmens in gleichem Verhältnis erforderlich sind wie bisher, auch wenn die von ihnen übernommene Aufgaben variieren dürften.

Tabelle 66: Angenommene differierende variablen Kosten

Instrument	Annahme	EUR
Zufallsauswahl	keine Kosten	0
Intelligenztest	keine Kosten (wird von HR wie bisher durchgeführt)	0
Wissenstest	keine Kosten (wird von HR wie bisher durchgeführt)	0
Arbeitsprobe	6 Berater pro Recruitingtag á 1000 EUR (inkl. Reisekosten)	6 000 (pro RT)
Assessment Center	6 Berater pro Recruitingtag á 1000 EUR (inkl. Reisekosten)	6 000 (pro RT)
Interview strukturiert	6 Berater pro Recruitingtag á 1000 EUR (inkl. Reisekosten)	6 000 (pro RT)
Integritätstest	keine Kosten (wird von HR wie bisher durchgeführt)	0

RT= Recruitingtag; HR= Human Resources

Tabelle 67: Angenommene Entwicklungskosten

Instrument	Annahme	EUR
Zufallsauswahl	keine Kosten	0
Intelligenztest	Kosten durch einmaligen Kauf einer Testbatterie	10 000
Wissenstest	Entwicklung eines Wissenstest durch 5 Berater á 10 Tage (50 Manntage á 1000 EUR)	50 000
Arbeitsprobe	Entwicklung einer Arbeitsprobe durch 5 Berater á 10 Tage (50 Manntage á 1000 EUR)	50 000
Assessment Center	Entwicklung eines Assessment Centers durch 5 Berater á 30 Tage (150 Manntage á 1000 EUR)	150 000
Interview strukturiert	Entwicklung eines Leitfadens durch 5 Berater á 10 Tage (50 Manntage á 1000 EUR)	50 000
Integritätstest	Kosten durch einmaligen Kauf einer Testbatterie	10 000

Für die Anwendung der Tests geht die Berechnung davon aus, dass die Mitarbeiter der Personalabteilung diese durchführen und die Kosten wie bisher in gleicher Höhe bestehen bleiben. Für die Zufallsauswahl wurden keine variablen Kosten angesetzt und angenommen, dass die mit der Personalauswahl einhergehenden Fixkosten von 400 000 EUR p.a. ebenfalls entfallen. Die Annahmen variabler Kosten sind in *Tabelle 66* dargestellt. Die Entwicklungskosten für die einzelnen Verfahren wurden analog *Tabelle 67* geschätzt. Alle weiteren Parameter wurden aus der prognostischen Validitätsrechnung übernommen.

Abbildung 56 visualisiert die monetären Auswirkungen in Abhängigkeit des neu eingeführten Instrumentes als Ersatz für das bisherige Vorgehen. Dabei zeigen sich erhebliche Effekte von über 10.0 Mio. EUR bei Einführung von *Intelligenztests* mit *Arbeitsproben*, *strukturierten Interviews* oder *Integritätstests* (siehe auch *Anhang XI-5*). Durch die im Rah-

men dieser Arbeit empfohlene Einführung von Intelligenztests in Verbindung mit der Anwendung strukturierter Interviews (siehe 3.6) scheinen realistische Ergebniseffekte von zusätzlich 13.7 Mio. EUR möglich zu sein.

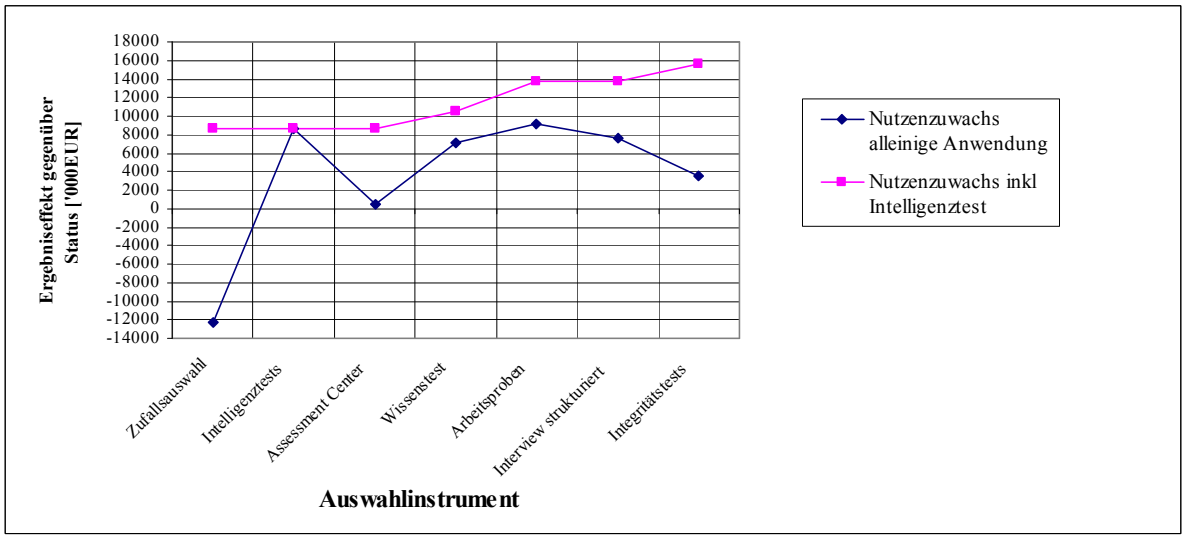

Abbildung 56: Inkrementeller Ergebniseffekt bei Änderung des Auswahlinstrumentes

3.5.5 Fazit

Die Nutzenanalyse zeigt, dass das untersuchte Auswahlverfahren einen positiven Ergebnisbeitrag erbringt. Die Höhe dieses Beitrages dürfte für die vergangenen Jahre, bei einer Anwendung des Auswahlverfahrens von 5 Jahren, eine Höhe von um die 12.6 Mio. EUR erreichen. Der zukünftige Ergebnisbeitrag bei Verwendung des gleichen Verfahrens unter den neu definierten Bedingungen liegt gegenüber der Zufallsauswahl etwas geringer bei 12.3 Mio. EUR. Dieses Ergebnis steht im Einklang mit zahlreichen empirischen Arbeiten, die durchweg sehr hohe Ergebniseffekte für unterschiedliche Auswahlverfahren gefunden haben (siehe *Anhang II*) und überrascht deshalb nicht. Allerdings ist die absolute Höhe vor dem Hintergrund einiger Prämissen zu interpretieren. So wurden im Rahmen der Operationalisierung einzelne Parameter, wenn auch konservativ, dennoch geschätzt, was mit einem Unsicherheitsfaktor einhergeht. Dies betrifft im Wesentlichen die Parameter *SQ*, *SDy* und den Validitätskoeffizienten r_{xy}.

Der Validitätskoeffizient r_{xy} wurde kohortenspezifisch aus den Analysen in dem Kapitel 3.3 übernommen, die ja, wie berichtet, zahlreichen methodischen Restriktionen unterliegen. Darüber hinaus ist davon ausgegangen worden, dass das verwendete Kriterium auch tatsächlich dem entspricht, was unter reellem Beratungserfolg zu verstehen ist. Unter 3.4.1 wurde aufgezeigt, dass Zweifel daran angebracht sind. Da das Kriterium allerdings die be-

triebliche Wirklichkeit darstellt, weil es auch als Grundlage für Beförderungen herangezogen wird, ist dieses Vorgehen vertretbar. Für die Kohorte der *SCs* wurde zudem der Validitätskoeffizient der *C-Kohorte* angewandt, da sich die vorhandene Datenbasis auf Grund der starken Varianzeinschränkung zur Berechnung des Koeffizienten wenig eignet. Ein Validitätskoeffizient anderer Höhe ist hier durchaus denkbar. Allerdings hat die Sensitivitätsanalyse gezeigt, dass bei *SQs*, die unter 80% liegen, bereits positive Ergebniseffekte zu verzeichnen sind, wenn die Validität des Verfahrens (kohortenübergreifend) mindestens r=.20 ist. Da historisch betrachtet die *SQs* immer niedriger waren und die empirischen Daten Grund zur Annahme liefern, dass die Validität höher ist, dürfte ein positiver Ergebnisbeitrag sehr wahrscheinlich sein. Darüber hinaus zeigte die Sensitivitätsanalyse auch, dass bereits kleine Validitätssprünge zu erheblichen Nutzendifferenzen führen können. Dies verdeutlicht, dass Investitionen in ein valides Auswahlverfahren, einen erheblichen Ergebnisbeitrag in dem Anwendungszeitraum liefern können. Bei einer *SQ* von 20% und einer Validitätserhöhung von r=.30 auf r=.40 beträgt sie im vorliegenden Fall rd. 5.0 Mio. EUR. Die Anwendung metaanalytisch fundierter und somit realistischer Werte führt zu der Erkenntnis, dass vermutlich zusätzliche Ergebnisverbesserungen im zweistelligen Millionenbereich bei einer Anwendung alternativer Instrumente möglich sind. So kann die Einführung von Intelligenztests und strukturierten Interviews anstelle des bisherigen Vorgehens unter den geschilderten Rahmenbedingungen zu einem zusätzlichen Ergebniseffekt von 13.7 Mio. EUR führen. Dieses umsetzbare Nahziel sollte für jeden Personalverantwortlichen Ansporn genug sein, an einer Verbesserung des aktuellen Vorgehens, unter Einbezug wissenschaftlicher Erkenntnissen und Befunde, zu arbeiten und die notwendigen Investitionen zu tätigen um die „brachliegenden“ Potenziale zu heben.

Die in den Berechnungen verwendeten *SQs* wurden teilweise aus historischen Daten übernommen und teilweise geschätzt. Ihre unterschiedliche Höhe spiegelt den unterschiedlichen Bedarf an Mitarbeitern wider. Offensichtlich sind über die Stellgröße *SQ* in der Vergangenheit, je nach Bedarf, mehr oder weniger Mitarbeiter eingestellt worden. Die katastrophalen finanziellen Auswirkungen dieses Vorgehens, das zwangsläufig mit einer Einstellung weniger qualifizierter Mitarbeiter einhergeht, wurde durch die Sensitivitätsanalyse deutlich. Sie zeigt, dass beispielsweise eine Erhöhung von *SQ=.30* auf *SQ=.50*, bei einer gegebenen Validität von r=.30, mit einem negativen Effekt von rd. 2.7 Mio. EUR einhergeht. Dies dürfte als Warnung dienen, nicht willkürlich den Personalbedarf über die Selektionsquote zu regulieren.

Die Höhe von *SDy* wurde über die Produktivitätskennzahl der Mitarbeiter gerechnet und ist höher als das durchschnittliche Jahreseinkommen in den Kohorten. Damit liegt es

auch außerhalb der von Hunter und Schmidt (1982) empfohlenen proportionalen Regel (40%-70%). Die Besonderheit der Beratungstätigkeit, mit ihren, relativ zum Einkommen des Beraters hohen Tagessätzen, führt allerdings entsprechend zu wesentlich höheren ökonomischen Effekten durch geringe Produktivitätssprünge. Eine Orientierung an dem Gehalt des Beraters ist deshalb wenig sinnvoll. Somit ist die Größe von *SDy* durchaus plausibel. Allerdings bleibt die Kritik, dass das Modell auf der Annahme basiert, dass die Produktivität ausschließlich durch die Leistung bestimmt wird und damit die Validität des Verfahrens ebenso auf die Produktivität zutrifft (vgl. Law, 1995, siehe dazu auch 2.3.5). Die Sensitivitätsanalyse verdeutlicht den linearen Zusammenhang der Veränderung von SDy zum Ergebniseffekt und zeigt, dass eine Fehleinschätzung der Größe zu eklatanten Ergebnisverschiebungen führen kann. Zudem wird deutlich, dass die Kosten des Auswahlverfahrens bei zahlreichen Selektionsquoten nur bei SDy-Werten über 50 000 EUR vertretbar sind. Der Grund liegt in der sehr kurzen durchschnittlichen Verweildauer der Mitarbeiter von drei Jahren, die den Nutzen, bei gleich bleibenden Kosten, im Vergleich zu längeren Verweildauern stark reduziert. Eine weitere Prämisse der Berechnung ist die fehlende Berücksichtigung der Ablehnungen durch die Bewerber, die ein Angebot bekommen haben. In der Praxis wird aber in einem solchen Fall nicht einem vorher abgelehnten Bewerber, also dem, der nach dem Ablehnenden am besten abgeschnitten hat, die Stelle angeboten. Stattdessen wird im Rahmen weiterer Auswahlverfahren ein vergleichbar guter Bewerber angenommen, so dass dieser Effekt für die Berechnungen vernachlässigt werden kann.

Zusammenfassend lässt sich feststellen, dass die eindrucksvollen Ergebniseffekte des Auswahlverfahrens und die angewandten Sensitivitätsanalysen die ökonomische Bedeutung eines validen Vorgehens mit einer angemessenen Selektionsquote verdeutlichen. Sie liefern somit den Personalverantwortlichen Argumente, um Investitionen zur Verbesserung des Auswahlverfahrens zu rechtfertigen.

3.6 Implikationen für die Praxis

Die evaluativen Ausführungen der vorherigen Kapitel lassen sich auf eine einfache Aussage reduzieren: Das Auswahlverfahren ist nützlich, es gibt aber noch erhebliches Verbesserungspotenzial. Die Untersuchungsergebnisse bieten zahlreiche Anhaltspunkte für Verbesserungsmöglichkeiten des Auswahlverfahrens auf verschiedenen Ebenen. Betrachtet man *Abbildung 57*, so ergeben sich aus den Analysen Ansätze in vier Bereichen: *Vorauswahl der Bewerber*, *Prädiktorenauswahl*, *Operationalisierung der Prädiktoren* und *Entscheidungsfin-*

dung. Die nachfolgenden Empfehlungen sollten allerdings auf Grund ihres explorativen Erkenntniscarakters anhand einer weiteren Stichprobe validiert werden.

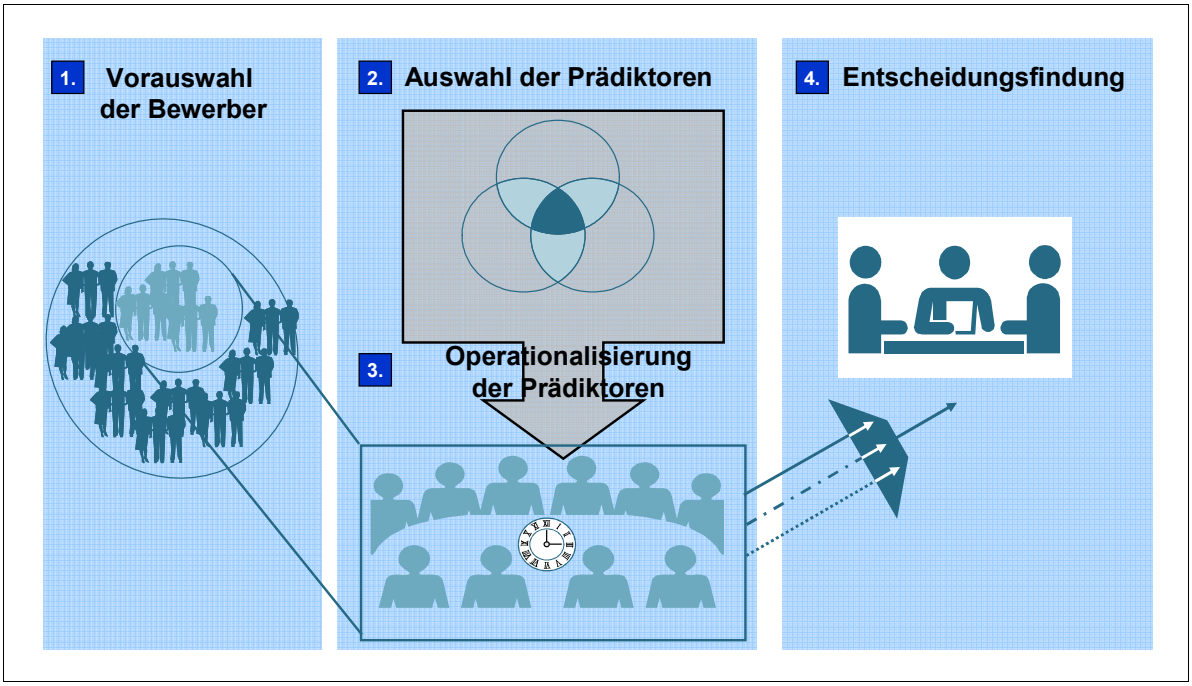

Abbildung 57: Praktische Ansatzpunkte zur Verbesserung des Auswahlverfahrens

Die Untersuchung der einzelnen Kohorten hat mögliche Hebel bei der *Vorauswahl* auf Basis *biographischer* Informationen aufgedeckt. Annahme der nachfolgenden Empfehlungen ist die Konstruktvalidität des Kriteriums, d.h. dass das bei der Evaluation herangezogene Kriterium tatsächlich Berufserfolg in Beratungen misst. So führen die t-Tests und Korrelationsanalysen zu dem Schluss, dass bei den *JC-Anwärtern*, die Vorauswahl *älterer* Mitarbeiter sowie solcher mit einem sehr guten *Abitur*, einem *MBA-Titel*, einem Studium auf einer *Privatuniversität* und bzw. oder langer *Praxiserfahrung* die Validität des Vorgehens erhöhen kann. Zudem führen die Analysen der *C-* und *SC-Kohorten* zu der Empfehlung, tendenziell weniger Quereinsteiger einzustellen, und wenn doch welche ausgewählt werden, der *Abiturnote* dabei keine Bedeutung beizumessen. Vielmehr sollte auch hier vorzugsweise auf *MBA-Absolventen* oder Absolventen mit einem *betriebswirtschaftlichen Studium*, bestenfalls auf einer *Privatuniversität* und mit *Familienstand ledig* fokussiert werden. Zudem könnten die Komponenten des Spezialistenwissens und des wirtschaftlichen Interesses im SC-Recruiting stärker gewichtet werden, da hier vermutlich Ursachen für *forciertes* Ausscheiden vorliegen. Dabei ist allerdings zu beachten, dass die aufgezählten Aspekte eher der Sensibilisierung für bestimmte vorteilhafte Konstellationen und nicht als Absolutkriterium dienen sollten.

Unter 3.4.1 wurde bereits die *Auswahl der Prädiktoren* im vorliegenden Fall kritisch beleuchtet. Das Auswahlverfahren greift hier auf grob definierte Eigenschaften zurück, die als hilfreich erachtet wurden, ohne einen Abgleich mit den Kriterien vorzunehmen. Auf wissenschaftliche Erkenntnisse hat man sich dabei ebenso wenig gestützt. Dies kann dazu führen, dass Prädiktoren herangezogen werden, die gar keinen Zusammenhang zur späteren Leistung haben, oder im schlimmsten Fall sogar negativ mit dem Erfolg zusammenhängen, obwohl man von einem positiven Zusammenhang ausgeht. Die verwendeten Prädiktordimensionen, die nicht wissenschaftlich fundierten Konstrukten entsprechen, ließen sich außerdem faktoranalytisch nicht replizieren. Deshalb ist fraglich, ob hinter diesen Variablen überhaupt einzelne latente Dimensionen zu finden sind. Scheinbar werden außerdem unterschiedliche Generalitätsebenen miteinander vermengt, wie sich aus den korrelativen Befunden ergibt (siehe z.B. Anhang XI-1). Die vorgenommene subjektive Auswahl auf Basis impliziter Persönlichkeitstheorien ist folglich einem auf wissenschaftlichen Erkenntnissen basierendem Vorgehen unterlegen. Abhilfe kann die Hinzunahme von Ergebnissen aus Studien und Metaanalysen (siehe 2.5), die auf die Tätigkeit des Beraters übertragen werden können, bieten. Dabei hat der Beraterberuf hohe Affinität zu Verkauf- und Managementberufen. Für jedes Kriterium des Berufserfolges sollten deshalb, wissenschaftlich gestützt, sinnvolle Prädiktoren definiert werden, die eine hinreichende Vorhersagevalidität besitzen.

Grundsätzlich scheint es bei den *JCs* zwei, sich auch faktoranalytisch andeutende Dimensionen zu geben, die im Recruiting erfasst wurden: *Soziale Kompetenz* und *Wissenskomponenten*. Darüber hinaus hat die Faktorenanalyse der Prädiktoren den Verdacht erhärtet, dass hinter den extrahierten unrotierten Allgemeinfaktoren Intelligenzkomponenten stehen. Hinzukommt bei den *JCs* die hohe prädiktive Validität der *Abiturnote*, die sicherlich gemeinsame Varianz mit der *Intelligenz* besitzt und darauf hindeutet, dass Intelligenz für das gemessene Kriterium von Bedeutung ist.

Zusätzlich gibt es Anhaltspunkte dafür, dass bestimmte Persönlichkeitseigenschaften (*Charisma*, *Kommunikationsfähigkeit*) bei den *JCs* Zusammenhänge mit der beruflichen Leistung aufweisen. Es ist zudem bei der C-Kohorte zu vermuten, dass bestimmten Variablen eine besondere Bedeutung zukommt: *Pragmatismus*, *Teamfähigkeit*, *Kommunikationsfähigkeit*, *logisch-analytisches Denken*, *intellektuelle Neugier*, *unternehmerisches Denken* und *BWL-Kenntnisse*. Bei den Quereinsteigern scheinen zusätzlich *Fachkenntnisse*, *Kreativität*, *Methodische Kenntnisse* und *Charisma* bedeutsam zu sein. Diese Variablen korrelieren signifikant mit dem *Gesamtwert aus Projektbeurteilung* oder den Unterdimensionen, werden teilweise in die Regressionsrechnungen aufgenommen und kehren auch immer wieder bei

der korrelativen Betrachtung der Unterdimensionen. Darüber hinaus lassen sie sich in vier Obergruppen einteilen: *Sozialkompetenz*, *logisch-analytisches Denken*, *Stressresistenz* (als wahrscheinlich wenig veränderbare persönliche Eigenschaft) und *Wissen*. Auch ist bei den genannten Variablen auf Grund ihrer semantischen Ähnlichkeit davon auszugehen, dass sie persönlichkeits- und intelligenzaffine Bestandteile beinhalten (siehe *Tabelle 17*). Dies alles steht im Einklang mit zahlreichen metaanalytischen Befunden und Einzelstudien, die breiten Persönlichkeitsmerkmalen und *Intelligenz* prädiktive Validität zusprechen. Wissenschaftlich betrachtet scheint, wie unter 3.4.1 erläutert, die Erfassung folgender Konstrukte, die sich als prädiktiv valide herausgestellt haben, empfehlenswert: *Intelligenz* und die Big 5-Faktoren *Gewissenhaftigkeit*, *Extraversion*, *Emotionale Stabilität*, *soziale Kompetenz* und *Verträglichkeit* sowie *Leistungsmotivation*. Zusätzlich ist die Erfassung von *Wissenskomponenten* und *kommunikativen Fähigkeiten* sinnvoll.

Betrachtet man die Operationalisierung des vorliegenden Verfahrens, so sollte, wie erwähnt, aus Sicht des Autors im Rahmen von *Fachinterviews* das *Wissen* erfasst werden. Die *Beobachtungen* im Rahmen der Gruppenarbeit (Gruppendiskussion, -fallbearbeitung) eignen sich hingegen dazu, die *soziale Kompetenz* und *Verträglichkeit* des Bewerbers zu beurteilen. Die *Präsentation* eignet sich zur Erfassung von *Wissen* und *kommunikativen Fähigkeiten*. Doch nicht alle Konstrukte eignen sich dazu, über *Interviews* oder *Beobachtung operationalisiert* zu werden. Empfehlenswert ist deshalb die Einführung von weiteren, weitaus objektiveren *Methoden*, wie *Intelligenz*- und *Persönlichkeitstests*. Idealerweise sollten zu dem Zeitpunkt der *Interviews* und *Beobachtung*, die Ergebnisse aus den Tests bereits vorliegen. So könnten beispielsweise in dem *Persönlichkeitsinterview* die *Persönlichkeitseinschätzungen* aus der *Gruppenarbeit* und dem *Persönlichkeitstest* gezielt, auch unter Anwendung von Stressinterviewtechniken, verifiziert werden (*MTMM-Ansatz*). Gerade bei *Persönlichkeitstests* kann von der angesprochene Gefahr der geringen Akzeptanz eines solchen Instrumentariums bei der Personalauswahl bei Unternehmensberatungen nur bedingt ausgegangen werden. Die Bewerbergruppe dürfte sich nämlich größtenteils der Besonderheit des Beraterberufes und dem damit verbundenen harten Auswahlverfahren bewusst sein. Über Persönlichkeitstests könnten folglich *Emotionale Stabilität*, *Leistungsmotivation*, und *Gewissenhaftigkeit* erfasst werden. Allerdings ist es nicht immer sinnvoll, alle Variablen mit mehreren Instrumenten zu messen. Einzelne Instrumente eignen sich nicht für die Messung und könnten durch die behaftete Fehlervarianzquelle bei der Aggregation die Validität verringern. Insgesamt sollten also vier Instrumente bei einer Neugestaltung des Auswahlverfahrens zum Einsatz kommen: *Interviews*, *Beobachtung* (Gruppendiskussion, -fallbearbeitung,

Präsentation), *Intelligenztests* und *Persönlichkeitstests*. Eine getrennte Durchführung von Auswahlverfahren je nach Kohorte ist dabei nicht zielführend. Da im Beratungsalltag auch unterschiedliche Funktionsstufen zusammenarbeiten, entspricht eine gemeinsame Aufgabendurchführung unter Beteiligung seniorer Quereinsteiger und Berufsanfänger sicherlich eher der Realität. Sie führt somit zu einer besseren Inhaltsvalidität der Auswahlverfahrens.

Tabelle 68: Komponenten strukturierter Interviews (siehe Campion, et al., 1997)

	Komponente
1	Fragenentwicklung auf Basis einer AA
2	Fragengleichheit bei allen Kandidaten
3	Begrenzung von Nachfragen, Nachhelfen und Ausarbeitung von Fragen nur auf das Nötigste
4	Verwendung von anforderungsorientierten Fragen
5	Angemessene Interviewlänge (mindestens 60 Min.)
6	Kontrolle oder Elimination flankierender (Vorab-) Informationen über den Bewerber
7	Zulassen von Bewerberfragen erst nach Abschluss des Interviews
8	Bewertung jeder einzelnen Antwort oder Einsatz multipler Skalen
9	Verwendung detaillierter verankerter Beurteilungsskalen
10	Detaillierte Protokollierung des Interviews
11	Einsatz mehrerer Interviewer
12	Einsatz des gleichen Interviewers für alle Bewerber
13	Keine Diskussionen über die Kandidaten oder Antworten zwischen den Interviews
14	Durchführung ausführlicher Interviewtrainings
15	Statistische statt klinische Vorhersage

Betrachtet man nun die angewandte *Interviewtechnik*, so ist diese im vorliegenden Fall als suboptimal zu bezeichnen. Unter 2.5.3 wurden die problematischen Aspekte *offener Interviews* thematisiert. Diese dürften auf die semi-strukturierte Variante ebenfalls in großen Teilen zutreffen. Die Überlegenheit *standardisierter* bzw. *strukturierter Interviews* sollte deshalb auch bei der Unternehmensberatung ausgenutzt werden (*Tabelle 56*). Bei den *Interviews* scheint es also sinnvoll, strukturierte Leitfäden zu verwenden, die die Interviewer zwingen, sich stärker an einen bestimmte „Fahrplan“ zu halten, der die zu beurteilenden Dimensionen „abarbeitet“. Zahlreiche Studien haben hier bedeutende Validitätszuwächse im Vergleich zu der offenen Interviewform festgestellt (siehe 2.5.3.2). *Tabelle 68* fasst Empfehlungen zum Inhalt und der Durchführung strukturierter Interviews zusammen (Campion, et

al., 1997), die allerdings nicht alle praktisch umsetzbar sind (z.B. Punkte 2, 6 oder 12). Denkbar ist auch der Einsatz eines *MMI* (siehe 2.5.3.2), das allerdings in seiner Entwicklung sehr aufwendig ist.

Eine ebenso hohe Standardisierung sollte bei den anderen Übungen (Gruppenbeobachtung, Präsentation) eingehalten werden. Das heißt, dass durch identische Gestaltung der Rahmenbedingungen bei allen Bewerbern weitere Fehlerquellen bei der Einschätzung der Bewerber ausgeschaltet werden (Verringerung $\sigma^2{}_{ZwS}{}^{PR}$). Im Fall der *Beobachtung* ist hier bereits die Vorgehensweise ansatzweise richtig. Eine mögliche Verbesserung kann dadurch erzielt werden, dass die Beobachter sich die zu beurteilenden Dimensionen aufteilen und jeder für einen gewissen Zeitraum nur auf zwei bis drei Aspekte achtet. Nach einem *Rotationsprinzip* beurteilt dann jeder Beobachter alle Dimensionen bei dem Bewerber im Laufe des gesamten Beobachtungszeitraums. Somit wird eine höhere Genauigkeit der Einschätzungen auf den Skalen erzielt, da die Konzentration auf bestimmte Merkmale und Verhaltensweisen größer ist (siehe auch Kolk, Born & Van der Flier, 2002).

Eine hohe Konstruktreliabilität (R_{PR}) der gemessenen Werte setzt zudem voraus, dass die im Recruiting, insbesondere bei *Interview* und *Beobachtung* gemessenen Variablen, exakt definiert werden. Nur so kann ein einheitliches Verständnis bei allen Beteiligten erzeugt werden. Hilfreich kann hier ein Definitionsblatt mit genauer Beschreibung jeder Variablen und dazugehörigen Beispielen sein. Dabei sollte auf sehr allgemeine Konstrukte, wie beispielsweise das des *Charismas*, verzichtet und diese genauer detailliert und durch mehrere andere Variablen ersetzt werden. Zusätzlich sollte die Anzahl zu beurteilender Dimensionen reduziert werden. Insgesamt scheint es auf der Prädiktorenseite zu viele Variablen zu geben, die zu dem eigentlichen Beratungserfolg in keinem Zusammenhang stehen. Neben den methodischen Restriktionen und der mangelnden Reliabilität der Urteile kann die Ursache darin liegen, dass ein Beurteiler im Rahmen seiner klinischen Urteilsbildung eine kleine Anzahl an Variablen heranzieht, die er persönlich für sinnvoll hält. Die *Reliabilitätsuntersuchung* und die *Faktorenanalyse* der Recruitingvariablen haben außerdem Grund zu der Annahme geliefert, dass durch unterschiedliches Verständnis bei den Beurteilern, insbesondere bezüglich der allgemeinen Variablen, auch unterschiedliche Dinge gemessen werden. Alternativ werden vermutlich zur Urteilsbildung auch generellere Hilfsvariablen (z.B. Intelligenz oder globale Leistung) hinzugezogen. Zu empfehlen ist auch, dass jeder Beurteiler jede Dimension beurteilt und auch die nötige Gelegenheit dazu bekommt, die für die Beurteilung notwendigen Verhaltensweisen zu beobachten. Bei dem Feststellen der Ausprägungen der Bewerber

auf den *Recruitingvariablen* sollte deshalb stärker darauf geachtet werden, dass die verwendeten Instrumente sich auch für die Messung eignen (s.o.).

Zusätzlich sollten die Beurteiler angehalten werden, die komplette Spannweite der Beurteilungsskala auszunutzen, um eine deutlichere Differenzierung zwischen den zu beurteilenden Personen zu erreichen. Die Skalierungsmethode sollte deshalb stärker verhaltensverankert sein (siehe Guion, 1998, S.542ff.). Das Rating sollte dabei, zumindest bei den Beobachtungen, während den Übungen stattfinden (siehe Silverman, Dalessio, Woods & Johnson, 1986). Schließlich ist zu empfehlen, die zu beurteilenden Dimensionen den Bewerbern vorher bekannt zu geben (Kleinmann, Kuptsch & Koller, 1996). Darüber hinaus ist es sinnvoll, zur weiteren Reduktion der Fehlerquellen beim Urteil (Verringerung $\sigma^2_{Fehler}{}^{PR}$), nur geschulte Spezialistenteams zum Recruiting zuzulassen (siehe Kanning & Holling, 2002). Denkbar ist hier der Einsatz erfahrener Mitarbeiter (höhere Funktionsstufen) aus allen CCs des Unternehmens, die regelmäßig geschult werden. Berater dürften sich wegen ihrer durch Berufserfahrung angereicherten kognitiven Struktur und ihren analytischen Fähigkeiten gut dazu eignen (Guion, 1998, S.621f.). Außerdem könnten dadurch mögliche *CC*- und *funktionsstufenabhängige* Strengeeffekte (*Abbildung 40*) eliminiert werden. Eine Unterstützung durch Psychologen und erfahrene Evaluationsforscher, die mögliche Fehlentwicklungen rechtzeitig erkennen und Impulse zum Gegensteuern liefern, sollte angestrebt werden.

Der letzte zu erwähnende Ansatzpunkt betrifft die *Entscheidungsfindung* über die Einstellung des Bewerbers. Zunächst ist eine konsequente *Cutoff-Setzung* teilweise auch auf der Einzelvariablenebene empfehlenswert. Dies verhindert, dass auf Gesamtwertebene Schwächen des Bewerbers in einem Bereich (z.B. *BWL-Know-how*) durch Stärken in einem anderen (z.B. *Charisma*) ausgeglichen werden können. Darüber hinaus sollten gerade für die Quereinsteiger einheitlich definierte Erwartungslevels formuliert werden. Dies ist bisher nicht vorgenommen worden, so dass jeder Recruiter für sich selbst einen *Cutoff* definiert, wie hoch beispielsweise das *Spezialistenwissen* eines SC sein muss. Dies sollte vor allem für die *sozialen* Komponenten (insb. *Frustrationstoleranz*, *Konfliktfähigkeit*) und die *Wissenskomponenten* gelten, da hier vermutlich Ursachen für forciertes Ausscheiden vorliegen. Die Faktorenanalyse hat zudem für *Einzelvariablen des Recruitings* keine Bestätigung für dahinter liegende Konstrukte geliefert, die die Unabhängigkeit der Variablen bestätigt und somit eine kompensatorische Aggregation rechtfertigen würde. Eine Alternative kann allerdings auch die stärkere Berücksichtigung der *Profile* bei der Personalentwicklung der Mitarbeiter sein. So sollten Mitarbeiter gezielter bezüglich ihrer im Recruiting identifizierten Schwächen gefördert werden. Dies spart Ressourcen und führt zu einer effektiveren Weiterbildung. Bei

den sozialen Komponenten dürfte dies mit größeren Schwierigkeiten verbunden sein, da diese nur begrenzt erlernbar sind. Anders bei wissensrelevanten Komponenten, bei denen durch gezielte, profilabhängige Weiterbildung, Schwächen reduziert werden können (z.B. BWL-Schulungen bei Fachfremden).

Wichtig ist zudem die strikte Trennung zwischen Bewertung und Entscheidung, da eine Entscheidung des einzelnen Urteilers bei der Urteilsbildung die Objektivität der Einschätzung verwässern kann. Das abschließende Entscheidungsurteil sollte dabei statistisch errechnet werden (siehe dazu 3.4.1)

Neben den genannten Optimierungsmöglichkeiten des Auswahlverfahrens in den vier Bereichen aus *Abbildung 57* gibt es noch weiter gehende Empfehlungen für die Praxis. Dies betrifft zunächst das verwendete Projektbeurteilungsinstrument. Unter 3.4.1 wurden Defizite bei der aktuellen Ermittlung des Beratungserfolges aufgedeckt. Normalerweise werden dazu AA als Ausgangsbasis (2.2.1) angefertigt. Dies ist allerdings im vorliegenden Fall nicht geschehen. Die Evaluationsdimensionen sind im Rahmen eines Workshops von einer Gruppe von Mitarbeitern festgelegt worden. Hierbei ist augenscheinlich, wie der Vergleich zu dem erarbeiteten *Fünf-Säulen-Modell* zeigte, von einer teilweise Übereinstimmung zu den tatsächlichen Erfolgskriterien auszugehen. Allerdings wurden, wie aus den Analysen z.T. ersichtlich, Symmetrieprinzipien missachtet und somit Daten unterschiedlicher Symmetrieebenen aggregiert. Dies gilt z.B. für die sehr generelle *Quality of Work*- und die sehr spezifische *Problem Solution-Dimension*. Durch die Missachtung der Symmetrieebenen kam es zu Validitätseinbußen in der Evaluationsstudie. Zwar betraf die Symmetrieanpassung lediglich die Kriterienseite, die zunächst nur als Variable zur Erfolgsmessung herangezogen wurde. Auf die tatsächliche Validität des Recruitingverfahrens hat sie deshalb lediglich insofern einen Einfluss, als sie durch ihre unterschiedliche Gestaltung bzw. Aggregation die Korrelation (Validität) beeinflusst. Allerdings wirkt sie natürlich indirekt auf das Auswahlverfahren, da auf ihrer Basis auch personalpolitische Entscheidungen getroffen werden (*Forced Exits*, Beförderungen). Die Kriterien sind somit Grundlage für das Urteil, ob eine Person erfolgreich ist oder nicht. Dies verdeutlicht die Wichtigkeit der Einhaltung der Symmetrieprinzipien der zu messenden Prädiktoren und Kriterien. Erste Anhaltspunkte über die Generalitätsebenen der vorliegenden Variablen konnten anhand der korrelativen Befunde gewonnen werden. Die Hinzunahme wissenschaftlicher Erkenntnisse bezüglich symmetrischer Anordnung möglicher Konstrukte im semantischen Raum der genannten Variablen und der dazugehörigen Prädiktoren scheint aber unumgänglich.

Um diese allerdings mit der notwendigen Güte vornehmen zu können, ist natürlich auch eine präzise und valide Auswahl und Operationalisierung der Kriterien vonnöten, die vorhergesagt werden sollen. Die Faktorenanalyse hat erhebliche Zweifel an der Konstruktvalidität des aktuellen Beurteilungsinstrumentes aufkommen lassen. Die zweifelhafte Unabhängigkeit der Variablen rechtfertigt außerdem nicht eine Aggregation bzw. Durchschnittsbildung. Problematisch ist zudem, dass einzelne Dimensionen gar nicht beurteilt werden können, da sie der Beurteilende nicht in ausreichendem Maße perzipiert oder der Beurteilte gar keine Gelegenheit zur Leistungserbringung hat. Dies wurde bei den klientenzentrierten Evaluationsdimensionen bereits erörtert. Die verwendete Operationalisierung des Kriteriums durch die Urteile des Projektmanagers könnte beispielsweise durch *Klientenbefragungen* ergänzt werden, wenn Klientenkontakt zustande kommt, oder durch eine 360°-Beurteilung ersetzt wird (siehe 3.4.1). Flankierend sollte auch hier durch eine detaillierte Beispielgebung pro Dimension, durch die bereits angesprochene Verhaltensverankerung der Ratings die Varianz der Urteile erhöht werden. Vorteilhaft sind auch Schulungen, die bisher auch auf der Projektbeurteilungsseite ausgeblieben sind, und mit denen der Versuch der Einführung eines einheitlichen Strengemaßstabes unternommen werden kann.

Der letzte in diesem Abschnitt zu nennende Bereich betrifft die organisatorischen Implikationen für das Unternehmen. In Abbildung 5 wurde visualisiert, dass ein systematischer, umfassender Evaluationsprozess eines Personalauswahlverfahrens über die in dieser Arbeit betrachteten Aspekte hinausgeht. Ein solcher Prozess sollte nicht einen einmaligen Charakter haben, sondern organisatorisch verankert werden und kontinuierlich, analog Abbildung 6 in Form eines Regelkreises, ablaufen. Sinnvollerweise sollte dies durch erfahrene Evaluationsforscher und Psychologen innerhalb des Personalbereichs im Unternehmen übernommen werden. Diese sollten die einzelnen Stufen des eignungsdiagnostischen Prozesses regelmäßig mit der betrieblichen Wirklichkeit abgleichen. Dazu gehört, dass die Ergebnisse aus einer einmal durchgeführten AA regelmäßig mit den sich wandelnden Anforderungen, die der Markt an die Berater stellt, verglichen werden. Vorhandene Diskrepanzen müssen entsprechend, mit den Folgen für die darauf aufbauenden Stufen, berücksichtigt werden. Zudem ist empfehlenswert, die vorgenommene Auswahl und Operationalisierung der Prädiktoren und Kriterien vor dem Hintergrund laufend neu berichteter wissenschaftlicher Erkenntnisse zu hinterfragen und anzupassen. Zusätzlich könnten eigene statistische Untersuchungen mit den gewonnen Daten zu weiteren Erkenntnissen führen. Dies gilt natürlich ebenso für die Validitätsanalyse. Eine regelmäßige Validierung und Erweiterung des Datensatzes, der durch immer neue Bewerber an Aussagekraft zunehmen dürfte, hilft bei der Op-

timierung des Vorgehens. Schließlich sollte der errechnete Nutzen des Auswahlverfahrens und seiner Anpassungen auch mit den finanzwirtschaftlichen Daten des Unternehmens abgeglichen werden. Der Einsatz einer *Balanced Scorecard* (Kaplan & Norton, 1992) ist hier denkbar und wäre auch im Rechtfertigungskontext für die durch die Veränderungen anfallenden Investitionen im Personalbereich sehr hilfreich.

4 Zusammenfassung und Ausblick

Die Eignungsdiagnostik bei der Personalauswahl blickt auf eine lange, mehrere tausend Jahre andauernde Geschichte zurück und ist heute, gerade in der Beratungsbranche, ein wesentlicher Einflussfaktor auf den ökonomischen Erfolg der Unternehmen. Obwohl bereits Anfang des 20. Jahrhunderts die ökonomische Tragweite personalpolitischer Entscheidungen thematisiert wurde (Münsterberg, 1912), nutzen dennoch viele Unternehmen nicht die zahlreichen, aus dem wissenschaftlich-psychologischen Bereich generierten Erkenntnisse. Viele dieser Erkenntnisse basieren auf Evaluationsstudien, die sich nach unterschiedlichen Typologien klassifizieren lassen. Ein Systematisierungsversuch ist das *Fünf-Datenboxenmodell* (Wittmann, 1985), das sich gut zur Darstellung methodischer Fragestellungen eignet. Mithilfe der Evaluationstypologie von Chen (1990) kann anhand des typischen Verlaufes eines Personalauswahlverfahrens gezeigt werden, dass dabei die reine Fokussierung der Evaluation auf die Eignungsdiagnostik zu kurz greift.

Der Begriff der Diagnostik wird vor zahlreichen Kontexten verwendet und hat viele Anwendungsgebiete. Dabei werden teilweise komplexe Messinstrumente eingesetzt, deren messtheoretische Grundlage meistens die klassische Testtheorie (KTT) ist. Die KTT ermöglicht die Berechnung von Testgütekriterien, zu denen die Validität, Reliabilität und Objektivität als die wichtigsten gehören.

Die Schritte, die bei der Berufseignungsdiagnostik im Rahmen der Personalauswahl vorgenommen werden, können anhand des *eignungsdiagnostischen Prozesses* nach Maukisch (1978) dargestellt werden. Im Rahmen einer *Arbeits-* und *Anforderungsanalyse* (AA) können dabei zunächst *Eigenschafts-*, *Verhaltens-*, *Qualifikations-* und *Ergebnisanforderungen* ermittelt werden. Grundsätzlich stehen dazu *erfahrungsgeleitet-intuitive*, *personenbezogen-empirische* und *arbeitsplatzanalytisch-empirische* Methoden zur Verfügung. Alternativ kann das *O*Net* zur Beschreibung einer Arbeitsstelle anhand einer Vielzahl von Dimensionen herangezogen werden. Bei der *Generierung* und *Operationalisierung* der Kriterien können analog *Ergebnis-*, *Verhaltens-* und *Eigenschaftskriterien* unterschieden werden. Dabei gibt es, sowohl für objektive als auch für subjektive Kriterien zahlreiche potenzielle Fehlerquellen, die die Gütemaßstäbe beeinträchtigen können. In Abhängigkeit der Kenntnis über ihr Wirkungsgefüge mit den Kriterien besteht der nächste Schritt in der Auswahl geeigneter Prädiktoren. Für die *Operationalisierung der Prädiktoren* stehen dabei die Prinzipien der *rationalen* (deduktiven), der *externalen* (kriteriumsbezogenen), der *induktiven* (internalen) *Methode* und der *Prototypenansatz* zur Verfügung. Die Prädiktoren werden anschließend im Rahmen der *Validitätsanalyse* hinsichtlich ihrer Vorhersagevalidität untersucht. Dabei kön-

nen das *Fünf-Datenboxenmodell*, die *Reliabilitätstheorie* sowie das *Brunswik´sche Linsenmodell* (Wittmann, 1987; 1990) zur systematischen Analyse herangezogen werden. Die Kombination der ausgewählten Prädiktoren mit dem Ziel, die Vorhersageleistung zu optimieren, kann *klinisch* und *statistisch* erfolgen. Zahlreiche Untersuchungen von Wiggins (1973), Sawyer (1966) und Meehl (1954) deuten dabei auf eine Überlegenheit statistischer Verfahren hin. Diese können in *kompensatorische* und *nichtkompensatorische* Verfahren unterteilt werden. Kausalanalytische Betrachtungen lassen sich mit Hilfe des *LISREL-Ansatzes* und der *Pfadanalyse* durchführen. Die prognostischen Verfahren liefern Informationen, auf deren Basis Entscheidungen getroffen werden, die nach Cronbach und Gleser (1965) in vier Entscheidungsarten klassifiziert werden können. Unterschieden werden zudem *kompensatorische* und *konjunktive* sowie *ein-* und *mehrstufige Strategien*.

Mithilfe von Cronbach und Glesers *Vierfelderschema* (1965) kann die eingeschränkte Interpretierbarkeit des Validitätskoeffizienten eines Prädiktors aufgezeigt werden. Noack und Petermann (1992) empfehlen deshalb spezifischere Gütekriterien, während Taylor und Russel (1939) die Beurteilung der Effektivität einer Entscheidungsstrategie mit Hilfe der Validität, der Basisrate und der Selektionsrate aufzeigen. Die Fokussierung auf solche Statistiken vernachlässigt jedoch die Folgen einer Entscheidung. Abhilfe kann hier die Anwendung von Nutzenmodellen schaffen.

Auswahlverfahren sollten dabei nicht anhand ihres absoluten, sondern ihres inkrementellen Nutzens, relativ zu bereits vorhandenen Informationen und der besten a-priori-Strategie bewertet werden. Historische Ansätze der Nutzenanalyse gehen auf Taylor und Russel (1939), Sands (1973), Cronbach und Gleser (1965) und Brogden (1949) zurück. Diese bilden die Realität allerdings in nur unzureichendem Maße ab. Das bisher brauchbarste Nutzenmodell stammt von Boudreau (1983b) und integriert in die Nutzenbetrachtung das Kapitalwertmodell der dynamischen Investitionsrechnung. Problematisch bei allen Vorgehensweisen ist vor allem die Operationalisierung der Standardabweichung der Produktivität SD_y, der Achillesferse der Entscheidungstheorie. Dabei können drei Ansätze zur Ermittlung des „pay-off“ durch Produktivitätsgewinn unterschieden werden: die *Kostenreduktion*, der *Verkaufswert-* und der in den meisten Fällen adäquate *Gewinnbeitragsansatz*. SD_y wird i.d.R. durch Schätzverfahren bestimmt. Diese lassen sich in die *Kostenrechnungsmethode*, *proportionale Regeln*, das *CREPID-Verfahren* und die am häufigsten angewandten *globalen Schätzverfahren* einteilen. Alle genannten Modelle quantifizieren allerdings den ökonomischen Nutzen lediglich anhand von Produktivitätsgewinnen. Es sind jedoch auch zahlreiche weitere ökonomische Vorteile durch eine bessere Personalauswahl denkbar. Neben den sta-

tistischen Modellen gibt es noch eine Reihe weiterer Alternativen, die zur Entscheidungsfindung herangezogen werden können. Dazu gehören z.B. *Benchmarking*, *MAUT* und *Balanced Scorecard*.

Die hohe Komplexität und mangelnde Akzeptanz der Nutzenmodelle führte bisher zu wenigen veröffentlichten empirischen Arbeiten, keine davon im Unternehmensberatungskontext. Dieser geht mit einigen Besonderheiten einher, die hauptsächlich die Konsequenzen personeller Fehlentscheidungen und die vereinfachte Ermittlung der SD_y, die direkt gemessen werden kann, betreffen. Allerdings müssen die Ergebnisse immer vor dem Hintergrund einiger beratungsspezifischer Prämissen beurteilt werden.

Die klassischen Kriterien des Berufserfolges, wie Gehalt oder Beförderungslevel sind bei Unternehmensberatungen nur begrenzt einsetzbar. Eine Alternative stellt die Betrachtung projektbezogener Beratungstätigkeit zur Ableitung von Kriterien, die diese als erfolgreich einstufen lässt, dar. Beratungstätigkeiten sind Dienstleistungen, deren Erfolg nicht exakt messbar oder zuordenbar ist. Aus diesem Grund wird in der Praxis eine Erfolgsmessung selten vorgenommen. Das Heranziehen des O*Net und die Analyse von zahlreichen Studien zur Evaluation von Beratungsleistungen helfen bei einer Ordnung der Kriterienvielfalt. Die verwendeten *Verhaltenskriterien* können dabei dem *Beratungsprozess*, die *Eigenschaftskriterien* dem *Berater* oder dem *Klienten* und die *Ergebniskriterien* dem *Ergebnis* zugeordnet werden. Zudem ist eine erfolgreiche Beratung auch vom *Kontext* abhängig, der als fünfte Kriteriumsperspektive das *Fünf-Säulen-Modell* erfolgreicher Beratung komplettiert. Dieses kann bei der Beurteilung helfen, ob bei einem verwendeten Leistungsbeurteilungssystem, alle relevanten inhaltlichen Aspekte betrachtet und auf geeignete Weise operationalisiert wurden.

Die bisherige Forschung hat zu der Erkenntnis geführt, dass die geeigneten Modelle der Intelligenz, hierarchisch mit einem Generalfaktor (*g*) auf oberster Ebene, aufgebaut sind. Im Bereich der Persönlichkeitsforschung hat sich das *Fünf-Faktoren-Modell* der Persönlichkeit als Leitmodell durchgesetzt, obwohl plausible und empirisch fundierte Modelle zum Zusammenhang mit Berufserfolg bisher ausstehen. Bei der Interessensforschung ist die auf Holland (1996) zurückgehende *RIASEC-Taxonomie* im Personalbereich am weitesten verbreitet. Die im Zusammenhang mit Berufserfolg bedeutendsten Motive sind das *Leistungs-* und das *Machtmotiv*. Schließlich wird beim Zusammenhang von Wissen und Berufserfolg häufig *prozedurales* und *deklaratives* sowie *Berufswissen* konzeptionell unterschieden.

Das Konstrukt *g* ist auf Grund einer umfassenden intensiven Forschung im Vergleich zu anderen Prädiktoren am besten untersucht und theoretisch fundiert. Unterschiede auf den spezifischen kognitiven Fähigkeitsprofilen können unterschiedliche Präferenzen der Infor-

mationsverarbeitung (Datenart, Prozesse) bedingen und somit ebenso wertvolle Informationen im Auswahlkontext liefern. Varianzaufklärungen gemeinsam mit *g* von bis zu R^2=.50 an der Berufsleistung sind dabei möglich. Bei den Persönlichkeitskonstrukten wurden hingegen bisher eher moderate prognostische Validitäten berichtet. In Abhängigkeit des Kriteriums und der symmetrischen Konstellation können dabei sowohl spezifische als auch generelle Konstrukte prognostisch wertvoll sein. Auch von Zusammenhängen zwischen *Machtmotivation*, *Leistungsmotivation* sowie *Selbstkontrolle* und Berufs- bzw. Führungserfolg ist auszugehen. Diese bedürfen aber weiterer empirischer Validierung.

In der Praxis hat sich die *Analyse der Bewerbungsunterlagen* als Vorauswahlinstrument etabliert. Sehr häufig findet ebenso das *unstrukturierte Interview* Anwendung, obwohl die strukturierte Variante sich psychometrisch als weitaus bessere Alternative erwiesen hat. *Assessment Center*, die als mehrere Instrumente integrierende Vorgehen zu verstehen sind, haben eine gute prognostische Validität und auf Grund des *Validitätsparadoxons* geringe Konstruktvalidität. *Intelligenztests* gelten aus Validitätsaspekten allerdings als beste Methode zur Personalauswahl für Arbeitsplätze, die Informationsverarbeitung erfordern. Andere Verfahren sollten, sofern sie inkrementelle Varianz aufklären, als Ergänzung zu *Intelligenztests* verstanden werden.

Die empirische Untersuchung dieser Arbeit hat gezeigt, dass die Leistungsbeurteilungen in dem angewandten AC eine geringe Objektivität aufweisen. Die Ursache liegt dabei vermutlich an den zahlreichen, mit dem Urteil einhergehenden potenziellen Fehlerquellen. Es konnten somit, unter Berücksichtigung methodischer Restriktionen, Befunde zur geringen Reliabilität offener bzw. semi-strukturierter Interviews und Beobachterratings (siehe Schuler, 2002) repliziert werden. Zudem wurde gezeigt, dass das angewandte kompensatorische Verfahren die Bildung unterschiedlicher Profile bei den eingestellten Bewerbern begünstigt. Diese stehen in Verbindung mit bio- und demographischen Informationen. So konnten bei Frauen tendenziell stärkere Ausprägungen auf kommunikativen, kreativen und sozialen Bewertungsdimensionen festgestellt werden. Dieser Befund ist ebenso in anderen empirischen Studien berichtet worden, wie das Vorliegen gekippter Profile (tilted profiles; siehe dazu Lubinski, 2004; Achter, et al., 1999; Shea, et al., 2001; Webb, et al., 2002). Im vorliegenden Fall zeigten sich die gekippten Profile bei den Recruitingvariablen *Kommunikationsfähigkeit* und *logisch-analytisches Denken*, die als Surrogat der genannten Dimensionen verstanden werden können. Auch beim Ausbildungsstand wurden Profilunterschiede festgestellt. Während promovierte Bewerber und solche mit einem *MBA-Abschluss* im *AC* durchweg besser beurteilt wurden, traf dies auf Absolventen eines *BWL-Studiums* allerdings nur bei wissens-

relevanten Dimensionen zu. Inwiefern die besseren Urteile auf die Bildung von Stereotypen zurückzuführen sind, konnte nicht abschließend geklärt werden.

Ferner wurde erwartungsgemäß faktoranalytisch gezeigt, dass in dem *AC* der Untersuchung ein Hauptkonstrukt, das den größten Teil der Varianz aufklärt, gemessen wird. Dabei handelt es sich vermutlich um ein Globalurteil der Eignung oder um allgemeine Intelligenz. Dieser Befund steht ebenfalls im Einklang mit anderen Untersuchungen zur *Konstruktvalidität* von ACs, die geringe *diskriminante* und hohe *konvergente Validitäten* der Einzeldimensionen berichten (*Validitätsparadoxon;* Maukisch, 1989; Sackett & Dreher, 1982; Schuler, 2000). Dies ist sicherlich teilweise auf die sich an globalen Kriterien orientierende Urteilsbildung der Beurteiler und der fehlenden Trennung von Beobachtung und Bewertung zurückzuführen. Zudem werden anforderungs- und verfahrensbedingte Mängel, sowie Mängel bei der Erfassung relevanter Verhaltensweisen als Grund für mangelnde Konstruktvalidität vermutet (Höft & Funke, 2001; Kleinmann, 1997). Der Schluss auf mangelnde Konstruktvalidität ist allerdings dabei nicht unproblematisch, da wie Schuler (2000) betont, gerade die multimodale Messung der Eignungsdiagnostik die Repräsentativität der AC-Tätigkeiten erhöhen soll. Die Urteile sind deshalb immer eine Kombination zwischen Verhalten, Eigenschaft und Messmethode. Neben dem geschilderten Hauptkonstrukt zeigte die Extraktion des unrotierten zweiten Faktors der AC-Variablen, dass dieser erneut ein verbal-numerisches Kontinuum abbildet, das wiederum an die gefundenen *tilted profiles* erinnert.

Die Extraktion von einem *Generalfaktor*, der einen Großteil der Varianz aufklärt, wurde in der vorliegenden Untersuchung auch auf den Kriteriendaten, also den Leistungsbeurteilungen durch Vorgesetzte, gefunden. Dieses Ergebnis ist auch in anderen Untersuchungen bestätigt worden (Viswesvaran, et al., 2005). Es wird neben den klassischen Urteilsfehlern (insb. *Halo-Effekt*) darauf zurückgeführt, dass symmetrisch sehr generell angesiedelte Faktoren (wie *g* oder die Persönlichkeitseigenschaft *Gewissenhaftigkeit*) den Berufserfolg auf nahezu allen Dimensionen beeinflussen und somit zu hohen Interkorrelationen führen (Feldmann, 1981; Motowidlo, et al., 1997).

Bezüglich der Determinanten des Berufserfolgs bestätigen die Ergebnisse teilweise die Vermutung, dass Bewerber mit einem entsprechenden beratungsspezifischen Vorwissen oder beruflicher Erfahrung, insgesamt bessere Leistungen in den Projekten erbringen. Korrelationsanalysen und t-Tests hatten beispielsweise für MBA-Absolventen, die ein sehr praktisch orientiertes Studium durchlaufen, tendenziell bessere Projektbeurteilungen ergeben. Im Gegensatz hierzu wurden promovierte Mitarbeiter durchweg schlechter beurteilt. Somit

scheint das dort angesammelte Wissen nur wenig berufsrelevante (*job knowledge*, siehe Hunter & Hunter, 1984) oder praktische Komponenten (*tacit knowledge*, siehe Sternberg, et al., 2000) zu beinhalten. Der sich teilweise in der Untersuchung zeigende Zusammenhang von *Praxiserfahrung* und *Berufsleistung* unterstreicht die Vermutung und steht im Einklang mit Befunden von Hunter (1986) und Schmidt, et al. (1986).

Neben der beruflichen Erfahrung zeigten sich auch bei Noten, insbesondere der *Abiturnote*, prognostische Validitäten für den Berufserfolg. Dies galt aber nur für die Leistung als *JC*, also für Hochschulabsolventen ohne Berufserfahrung. Die Korrelationen lagen allerdings in einem Bereich, der eigentlich eher bei der Vorhersage von Ausbildungsleistungen und nicht von Berufserfolg angenommen wird (Baron-Boldt, et al., 1989). Da die Tätigkeit als *JC* aber mit einer Art Ausbildungsphase vergleichbar ist und andere intellektuelle Anforderungen als eine Beratungstätigkeit auf einer höheren Funktionsstufe hat (siehe dazu Sieglen, 1998), kann dieses Ergebnis als Bestätigung für die genannte Annahme gelten.

Darüber hinaus ergab die Untersuchung Hinweise darauf, dass Intelligenz und breite Persönlichkeitsfaktoren valide Prädiktoren für die Leistung als Berater sein dürften. Zwar sind weder die Intelligenz noch Persönlichkeitseigenschaften im vorliegenden Fall über einen Test erfasst worden. Es ist jedoch bei zahlreichen Variablen, auf Grund ihrer semantischen Ähnlichkeit davon auszugehen, dass sie persönlichkeits- und intelligenzaffine Bestandteile beinhalten. Das faktoranalytisch ermittelte Hauptkonstrukt beim AC, hinter dem Intelligenz vermutet wird (s.o), korreliert außerdem signifikant mit der späteren Berufsleistung. Ferner zeigte sich die Variable *Charisma*, die per Definition eine sehr breite Persönlichkeitseigenschaft ist, als prädiktiv valide für den Berufserfolg. Dies alles steht im Einklang mit zahlreichen metaanalytischen Befunden und Einzelstudien, die breiten Persönlichkeitsmerkmalen und *Intelligenz* prädiktive Validität zusprechen.

Die Ergebnisse zur Validität des angewandten Instrumentes entsprechen ebenfalls den wissenschaftlichen Erkenntnissen. Die metaanalytischen Befunde von Schmidt und Hunter (1998) führen zu dem Schluss, dass die durchschnittliche prognostische Validität von *ACs* bei r=.37 liegt. Die im vorliegenden Fall gefundene Validität schwankt je untersuchter Kohorte und erreicht bei den *Cs* eine um Selektionsfehler korrigierte Höhe von r=.369, die sehr nahe an dem metaanalytischen Wert von Schmidt und Hunter liegt.

Zusätzlich wurde in der Untersuchung gezeigt, dass die Einhaltung von *Symmetrieprinzipien* auf Prädiktor- und Kriterienseite für die Validitätsanalyse entscheidend ist. Nach Wittmann (1985; 1990) sollten Kriterium und Prädiktor hinsichtlich ihrer Breite und inhaltlichem Bezug übereinstimmen, da es ansonsten zu erheblichen Validitätseinbußen kommen

kann. Durch eine Neuordnung der Kriterienseite, bei der aufbauend auf korrelativen Befunden die Variablen unterschiedlichen Generalitätsniveaus zugeordnet und faktorisiert wurden, konnte in der Untersuchung die Validität erhöht werden.

Ferner wurde in der Untersuchung die Auswahl und Operationalisierung von Prädiktor und Kriterium kritisch hinterfragt. Der Vergleich des *Fünf-Säulen-Modells* erfolgreicher Beratung mit den in der Unternehmensberatung verwendeten Bögen zur Beurteilung des Beratungserfolges ergab nur teilweise Überschneidungen. Die sich dadurch ergebende mangelnde Inhaltsvalidität auf der Kriterienseite kann schwerwiegende Folgen auf die Mitarbeitermotivation haben, da der gemessene Beratungserfolg an vielen Stellen durch Gegebenheiten beeinflusst wird, auf die der Berater selbst keinen Einfluss hat. Die Analysen zu der *Symmetrie* haben außerdem gezeigt, dass Kriterien unterschiedlicher Generalitätsebenen auf gleicher Ebene angesetzt wurden. Eine Neustrukturierung aus inhaltlichen Überlegungen sowie faktoranalytischen Erkenntnissen wurde deshalb empfohlen. Bei der Operationalisierung der Kriterien fiel auf, dass die verwendeten Beurteilungsbögen alle verwendeten Oberdimensionen und innerhalb der Oberdimensionen die Einzelvariablen gleich gewichten. Dass aber alle Aspekte gleich bedeutsam sind, ist unwahrscheinlich. Eine auf Expertenmeinungen basierende Gewichtung der einzelnen Variablen, etwa durch Anwendung von *MAUT*, kann hier helfen und dürfte zudem zu einer hohen Akzeptanz der Beteiligten führen. Außerdem wäre eine Beurteilung durch mehrere Personen aus *MTMM*-Gesichtspunkten wünschenswert, u.U. im Rahmen einer 360°-Beurteilung (siehe Scherm, 2004).

Die verwendeten Prädiktoren wurden durch subjektive Einschätzungen auf Basis impliziter Persönlichkeitstheorien und nicht anhand wissenschaftlicher Erkenntnissen generiert. Aus wissenschaftlicher Perspektive ist allerdings u.a. die Erfassung folgender Konstrukte, viel versprechend: *Intelligenz* und die Big 5-Faktoren *Gewissenhaftigkeit*, *Extraversion*, *Emotionale Stabilität*, *soziale Kompetenz* und *Verträglichkeit* sowie *Leistungsmotivation*. Darüber hinaus ist die Erfassung von *Wissenskomponenten* und *kommunikativen Fähigkeiten* empfehlenswert. Dabei sollten im Rahmen von *Fachinterviews* das *Wissen,* die *Intelligenz* durch *Intelligenztests* und über Persönlichkeitstests *Emotionale Stabilität*, *Leistungsmotivation* und *Gewissenhaftigkeit* erfasst werden. Die Beobachtungen im Rahmen der Gruppenarbeit eignen sich, die *soziale Kompetenz* und *Verträglichkeit* des Bewerbers zu beurteilen. Im Rahmen der *Präsentation* kommen, neben dem *Wissen*, die *kommunikativen Fähigkeiten* am besten zum Ausdruck. Schließlich könnten in einem *Persönlichkeitsinterview* die *Persönlichkeitseinschätzungen* aus der *Gruppenarbeit* und dem *Persönlichkeitstest* verifiziert werden.

Bei der Entscheidungsfindung über die Einstellung eines Bewerbers scheint die Beibehaltung eines Diskussionsprozesses unter den Beurteilern für die spätere Rechtfertigung der Entscheidung empfehlenswert (Schuler, 2000, S.127). Die finale Entscheidung sollte allerdings eine Trennung von Bewertung und Urteil beinhalten und das abschließende Einstellungsurteil rein statistisch errechnen. Die Gewichtung der einzelnen Dimensionen müsste dann, an Stelle des bisherigen kompensatorischen Vorgehens, auf über multiple Regression gewonnenen Erkenntnissen beruhen. Es zeigt sich außerdem, dass das bisherige Auswahlverfahrens Schwächen beim negativen Prädiktionswert aufweist. Ziel sollte deshalb auch sein, diese durch ein Vorgehen höherer Validität zu verringern, um negative Nutzeneffekte zu reduzieren.

Die Berechnung des Nutzens des Auswahlverfahrens hat für die vergangenen fünf Jahre einen Ergebnisbeitrag von 12.6 Mio. EUR erbracht. Der zukünftige Ergebnisbeitrag bei Verwendung des gleichen Verfahrens unter neu definierten Bedingungen liegt etwas geringer bei 12.3 Mio. EUR gegenüber der Zufallsauswahl. Dieses Ergebnis steht im Einklang mit zahlreichen empirischen Arbeiten, die durchweg sehr hohe Ergebniseffekte für unterschiedliche Auswahlverfahren gefunden haben. Allerdings ist die absolute Höhe vor dem Hintergrund einiger Prämissen zu interpretieren. So wurden im Rahmen der Operationalisierung einzelne Parameter, wenn auch konservativ, dennoch geschätzt, was mit einem Unsicherheitsfaktor einhergeht. Dies betrifft im Wesentlichen die Parameter *SQ*, *SDy* und den Validitätskoeffizienten r_{xy}. Die Berechnung monetärer Auswirkungen in Abhängigkeit eines neu eingeführten Instrumentes als Ersatz für das bisherige Vorgehen ergibt erhebliche zusätzliche Ergebniseffekte von über 10.0 Mio. EUR. Dies gilt für die Einführung von *Intelligenztests* mit *Arbeitsproben*, *strukturierten Interviews* oder *Integritätstests*. Dieses Ergebnis und durchgeführte Sensitivitätsanalysen verdeutlichen die ökonomische Bedeutung eines validen Vorgehens mit einer angemessenen Selektionsquote und eröffnen den Personalverantwortlichen die Möglichkeit, Investitionen zur Verbesserung des Auswahlverfahrens zu rechtfertigen.

Aus den dargestellten Ergebnissen konnten zahlreiche, noch zu validierende praktische Empfehlungen für das untersuchte Unternehmen abgeleitet werden. Diese betreffen im Wesentlichen vier Bereiche: *Vorauswahl der Bewerber*, *Prädiktorenauswahl*, *Operationalisierung der Prädiktoren* und *Entscheidungsfindung.* So wurde z.B. bei der *Vorauswahl* der *JCs* die Bevorzugung *älterer* Mitarbeiter sowie solcher mit einem sehr guten *Abitur*, einem *MBA-Titel*, einem Studium auf einer *Privatuniversität* und bzw. oder langer *Praxiserfahrung* empfohlen. Zudem ist es sinnvoll, für jedes Kriterium des Berufserfolges, wissenschaftlich

gestützt, valide Prädiktoren zu definieren und diese mit geeigneten Instrumenten zu operationalisieren (siehe dazu weiter oben). Zu den empfohlenen Instrumenten gehören dabei *strukturierte Interviews*, *Beobachtung* (*Gruppendiskussion*, *Fallbearbeitung* und *Präsentation*), *Persönlichkeitstests* und *Intelligenztests*.

Bezüglich der Ausgestaltung der Durchführung sollten die Aufgaben der Beobachtung gemeinsam von senioren Quereinsteigern und Berufsanfängern durchgeführt werden. Alle Übungen sollten dabei mit einer hohen Standardisierung einhergehen. Das heißt, dass durch identische Gestaltung der Rahmenbedingungen bei allen Bewerbern, weitere ungewünschte Varianz bei der Einschätzung der Bewerber eliminiert werden sollte. Darüber hinaus ist die Einführung eines *Rotationsprinzips* unter den Beurteilern hilfreich. Die im Recruiting gemessenen Variablen sollten zusätzlich exakt definiert werden, um ein einheitliches Verständnis bei allen Beteiligten zu erzeugen. Auch die Anzahl zu beurteilender Dimensionen sollte reduziert werden. Zusätzlich sollten die Beurteiler angehalten werden, die komplette Spannweite der Beurteilungsskala auszunutzen, um eine deutlichere Differenzierung zwischen den zu beurteilenden Personen zu erreichen. Die Skalierungsmethode müsste deshalb stärker verhaltensverankert sein und das Rating während der Übungen stattfinden. Schließlich ist zu empfehlen, die zu beurteilenden Dimensionen den Bewerbern vorher bekannt zu geben und nur geschulte Spezialistenteams zum Recruiting zuzulassen. Eine Unterstützung durch Psychologen und erfahrene Evaluationsforscher wäre dabei hilfreich.

Bei der *Entscheidungsfindung* über die Einstellung des Bewerbers ist, zusätzlich zu dem bereits Erwähnten, eine konsequente *Cutoff-Setzung* teilweise auch auf der Einzelvariablenebene empfehlenswert. Darüber hinaus sollten, gerade für die Quereinsteiger, einheitlich definierte Erwartungslevels formuliert werden. Eine Alternative kann auch die stärkere Berücksichtigung der *Profile* bei der Personalentwicklung der Mitarbeiter sein, um Mitarbeiter gezielter bezüglich ihrer Schwächen zu fördern. Wichtig ist zudem die strikte Trennung zwischen Bewertung und einer statistisch berechneten Entscheidung (s.o.). Zusätzlich sollte die Evaluation des Auswahlverfahrens, unterstützt durch Psychologen und Evaluationsforscher als rollierender Prozess organisatorisch verankert werden.

Wirft man einen Blick in die Zukunft, so ergibt sich sowohl aus praktischer, als auch aus wissenschaftlicher Sicht weiterer Aufklärungsbedarf.

Bezogen auf das untersuchte Unternehmen wäre eine Umsetzung der geschilderten Empfehlungen wünschenswert. Diese sollte allerdings, wie bereits mehrfach erwähnt, mit einer fortlaufenden Validierung der hier gefundenen und weiterer sich ergebender Erkenntnisse einhergehen. Die untersuchte Unternehmensberatung hat hier bereits erste Maßnahmen

umgesetzt. Neben der inhaltlichen Überarbeitung der Beurteilungsbögen wurde ein Intelligenztest als weiteres Auswahlinstrument eingeführt. Darüber hinaus wurden geschulte Spezialistenteams für das Recruiting gebildet, die sich aus senioren Beratern zusammensetzen. Geplant ist zudem, weitere Änderungen sukzessive zu implementieren und fortlaufend zu validieren. Es bleibt zu hoffen, dass die notwendigen Veränderungen, die, wie in dieser Arbeit gezeigt, zu erheblichen ökonomischen Effekten führen dürften, auch organisationsintern Zuspruch finden. Die Ergebnisse dieser Arbeit könnten dabei einen Beitrag dazu leisten, interne Blockaden zu lösen. Dies ist die Voraussetzung, um auch die finanziellen Ressourcen zu erhalten, die, im Vergleich zu dem potenziellen Nutzen, eher marginaler Höhe sein dürften, aber für die Implementierung unabdingbar sind.

Aus wissenschaftlicher Perspektive herrscht weiterer Forschungsbedarf in mehreren Bereichen. So gibt es zwar zahlreiche Untersuchungen, die unterschiedlichste Kriterien des Beratungserfolges verwenden. Diese sind allerdings oftmals aus methodischen Gesichtspunkten mangelhaft oder verlassen nicht die deskriptive Ebene. Wünschenswert wäre hier die Durchführung einer empirischen Studie, die alle Facetten des Beratungserfolges erfasst und Hinweise auf die Struktur und die kausalen Zusammenhänge innerhalb des Kriteriums Beratungserfolg aufdeckt. Mit dem *Fünf-Säulen-Modell* erfolgreicher Beratung ist ein erster Schritt in diese Richtung vorgenommen worden. Allerdings beruht die Kategorisierung, da nicht Hauptanliegen dieser Arbeit, auf der subjektiven Einschätzung des Autors. Sie ist somit, ohne empirische Fundierung beispielsweise durch eine Befragung von Führungskräften in zahlreichen Beratungsunternehmen und eine Analyse durchgeführter Projekte, wenig geeignet zur Verwendung als Erfolgskriterium.

Weiteren Forschungsbedarf gibt es bei Erklärungsmodellen zum Berufserfolg im Allgemeinen bzw. dem Beratungserfolg im Speziellen. Während im Bereich der Intelligenzforschung bereits einige theoretischen Modelle empirisch fundiert wurden, gibt es z.B. bei den Persönlichkeitsmerkmalen größtenteils lediglich Validitätsstudien zum Berufserfolg. Theoretische Erklärungsmodelle sind hier bisher weitgehend ausgeblieben. Integrative Ansätze, wie die PPIK-Theorie von Ackerman (1996) gehen in die erwünschte Richtung und sollten hinsichtlich ihres kausalen Zusammenhangs mit dem Berufserfolg näher untersucht und empirisch validiert werden. Dabei sollten Unternehmensberatungen stärker Untersuchungsgegenstand von Studien werden. Dem Autor ist bisher keine Studie bekannt, in der eine Unternehmensberatung im Betrachtungsfokus stand, obwohl gerade in Beratungen zweifelsohne das Humankapital die wichtigste Ressource darstellt. Empirische Arbeiten wie die

vorliegende, die weitere Erkenntnisse zu brauchbaren Prädiktoren und Personalauswahlinstrumenten bei Unternehmensberatungen generieren, sind folglich erstrebenswert.

Darüber hinaus sollte auch die Entwicklung neuartiger Auswahlinstrumente, die die Besonderheiten des Beraterberufes berücksichtigen, vorangetrieben werden. Erheblichen Forschungsbedarf gibt es vor allem in dem Bereich der Auswahl von Mitarbeitern, die in Teams arbeiten, wie es in Unternehmensberatungen der Fall ist. Hier dürften sich die Auswahlinstrumente von den traditionellen Methoden unterscheiden (siehe Hough & Oswald, 2000) und somit stärker den Bedürfnissen der Beratungsunternehmen entsprechen. Impulse könnten zudem von der in den letzten zwei Jahrzehnten aufgekommenen *computergestützten Eignungsdiagnostik* kommen (vgl. Schuler & Funke, 1991a, S.67ff.). Sie geht mit einer Vielzahl an Vorteilen einher, die im Wesentlichen den hohen Standardisierungsgrad, die Kontrolle vieler potenzieller Fehlerquellen, die Möglichkeit adaptiven Testens, sowie die einfache und ökonomische Durchführung und Auswertung betreffen (Schuler & Funke, 1995; Wottawa, 1993). Zudem ist mit ihrer Hilfe die Applikation von komplexen dynamischen Problemlösungsszenarien (Strauß & Kleinmann, 1995) möglich, die sich ja auch im Beratungsalltag zeigen. Denkbar ist in diesem Zusammenhang auch die Entwicklung von netzbasierten Auswahlverfahren (*E-Recruiting*), die insbesondere in den USA bereits Anwendung finden. Eine für den Beratungsbereich ebenfalls interessante Weiterentwicklung sind außerdem sog. „Lernfähigkeits-“ oder „Lernpotenzial“-*ACs* (Sarges, 1995b). Sie stehen in Zusammenhang mit der Erforschung der Veränderbarkeit von Eignungsmerkmalen durch Training (Guthke, 1991). Vor dem Hintergrund einer sich immer stärker und schneller wandelnden Umwelt und damit Veränderung der Anforderungen an die Berater, scheinen diese Verfahren viel versprechend zu sein, auch wenn sie sich bisher nur begrenzt empirisch bewährt haben (Schuler, 2000, S.105).

Des Weiteren gibt es Forschungslücken bei den Nutzenmodellen. Aus Sicht des Autors ist die fortlaufende Entwicklung von Modellen, die immer mehr zusätzliche Parameter in die Berechnung integrieren, ein Weg in die falsche Richtung. Die geringe Anzahl bisheriger Studien zur Nutzenberechnung (in Unternehmensberatungen mit Ausnahme der vorliegenden offenbar keine) liegt nicht zuletzt an der fehlenden Akzeptanz der Modelle in der Praxis. Die Gründe dafür hängen vermutlich mit der hohen Komplexität der angewandten Formeln, die auf Verständnisschwierigkeiten bei den Personalverantwortlichen und dem Management stoßen, zusammen. Zwar ist, wie dargestellt, die Ermittlung monetärer Größen im Rechtfertigungskontext unverzichtbar. Genauso wichtig ist allerdings aus Sicht des Autors, dass die zugrunde liegende Berechnungssystematik von der Praxisseite verstanden wird.

Ein entscheidender Aspekt ist dabei die Validierung der in vielen Studien prognostizierten monetären Ergebniseffekte (siehe auch Vance & Colella, 1990). Dieser kann, wie im vorliegenden Fall gezeigt, erhebliche monetäre Größen annehmen, die für den Unternehmenserfolg entscheidend sein können. Leider ermöglichen die bisherigen, im Finanzbereich der Unternehmen angewandten betriebswirtschaftlichen Ansätze keine zufrieden stellende ökonomische Erfolgsmessung von Personalmaßnahmen oder Humankapital (Weatherly, 2003). Eine Vernetzung der Nutzenanalyse mit strategischem Personalmanagement und der Finanzbuchhaltung wäre der Idealfall (Boudreau & Ramstad, 2002), blieb aber bisher in der Praxis aus. Insofern ist auch eine Verifizierung der mit den komplexen Formeln berechneten Nutzeneffekte derzeit nicht zweifelsfrei möglich. Die Antwort auf diese Problematik kann nicht das Heranziehen von Modellen eines höheren Komplexitätsgrades sein. Die Zweifel werden noch verstärkt durch die Tatsache, dass die bisherigen Nutzenmodelle lediglich den Produktivitätsgewinn der eingestellten Mitarbeiter betrachten. Die teilweise gerade im Beratungsbereich eklatanten weiteren Folgen, die beispielsweise mit zu Unrecht abgelehnten Bewerbern einhergehen, werden jedoch völlig vernachlässigt. Es ist deshalb wünschenswert, dass neue Modelle diese Aspekte sowie den Nutzen zweiter und dritter Art in ihre Betrachtung integrieren.

In der vorliegenden Arbeit ist ein bestehendes Auswahlverfahren mit Hilfe eines umfassenden statistischen Instrumentariums analysiert und praktische Empfehlungen für Verbesserungen des Vorgehens abgeleitet worden. Diese Verbesserungen sollen dazu beitragen, dass die Qualität des Auswahlverfahrens und damit der ausgewählten Bewerber erhöht wird, einhergehend mit den genannten monetären Nutzeneffekten. Eine derartige detaillierte Analyse der Auswahlverfahren eines jeden Unternehmens ist zwar wünschenswert, dürfte sich allerdings als kaum praktikabel erweisen. Zum Einen müssten die notwendigen Daten für eine statistische Analyse vorhanden und somit in der Vergangenheit in ausreichender Qualität erfasst worden sein. Zum Anderen scheitert auf Grund der geringen akzeptierten Bewerberanzahl gerade bei kleineren Unternehmen eine Validierung an der geringen Stichprobengröße. Gefragt sind also aus empirischen Befunden abgeleitete allgemeingültige Ansätze und Richtlinien die jedem Unternehmen bei der Entwicklung des eigenen Auswahlverfahrens als Orientierung dienen. Einen wichtigen Schritt in diese Richtung ist durch die DIN 33430 vorgenommen worden (Westhoff, et al., 2004), die Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen festlegt. Mit ihrer Hilfe können die internen Personalprozesse sowie die Entscheidungssicherheit bei der Personalauswahl verbessert und die Rechtssicherheit bei der Eignungsbeurteilung erhöht werden.. Zudem bedeutet aus Bewerbersicht die Anwendung eines DIN-konformen Auswahlprozes-

ses hohe Fairness und gerechte Beurteilungen. Die DIN 33430 könnte somit, bei einer umfassenden Einführung am Markt, die Schließung der Lücke zwischen Wissenschaft und Praxis vorantreiben und eine qualitätsgeprüfte Eignungsbeurteilung sicherstellen. Es bleibt abzuwarten, in wie weit eine solche Norm Akzeptanz und Anwendung bei den Personalverantwortlichen in den Unternehmen findet und auf Basis neu aufkommender Erkenntnisse konsequent weiterentwickelt wird.

Alles in allem ist zu erwarten, dass das Thema Evaluation und Nutzen von Personalauswahlverfahren weiterhin ein spannendes Forschungsfeld bleiben und viele weitere neue Erkenntnisse liefern wird, denen sich die Praxis auf Dauer nicht entziehen kann.

Anhang

Anhang I: Beschreibungsdimensionen des O*Net 280
Anhang II: Nutzenanalyse zur Personalauswahl 289
Anhang III: Ausgewählte Kriterien erfolgreicher Beratung 291
Anhang IV: Kriterienzuordnung 293
Anhang V: Detaillierung der Recruitingdimensionen und Aggregation 297
Anhang VI: Inhalte der Projektbeurteilungsbögen 302
Anhang VII: Variablenbezeichnung und Skalierung der Beurteilungsdimensionen 304
Anhang VIII: Detaillierung der Ergebnisse zur JC-Kohorte 305
Anhang IX: Detaillierung der Ergebnisse zur C-Kohorte 311
Anhang X: Detaillierung der Ergebnisse zur SC-Kohorte 318
Anhang XI: Detaillierung der Nutzenanalyse 323

Anhang I: Beschreibungsdimensionen des O*Net

General description: Conduct organizational studies and evaluations, design systems and procedures, conduct work simplifications and measurement studies, and prepare operations and procedures manuals to assist management in operating more efficiently and effectively. *Includes* program analysts and *management consultants*.

1. Worker Characteristics

A. Abilities[a]

Ability	Description	Importance
Oral Expression	The ability to communicate information and ideas in speaking so others will understand.	75
Speech Clarity	The ability to speak clearly so others can understand you.	75
Written Expression	The ability to communicate information and ideas in writing so others will understand.	73
Problem Sensitivity	The ability to tell when something is wrong or is likely to go wrong. It does not involve solving the problem, only recognizing there is a problem.	65
Oral Comprehension	The ability to listen to and understand information and ideas presented through spoken words and sentences.	63
Written Comprehension	The ability to read and understand information and ideas presented in writing.	63
Fluency of Ideas	The ability to come up with a number of ideas about a topic (the number of ideas is important, not their quality, correctness, or creativity).	58
Information Ordering	The ability to arrange things or actions in a certain order or pattern according to a specific rule or set of rules (e.g., patterns of numbers. Letters, words, pictures, mathematical operations).	53
Deductive Reasoning	The ability to apply general rules to specific problems to produce answers that make sense.	50
Inductive Reasoning	The ability to combine pieces of information to form general rules or conclusions (includes finding a relationship among seemingly unrelated events).	50
Mathematical Reasoning	The ability to choose the right mathematical methods or formulas to solve a problem.	50
Originality	The ability to come up with unusual or clever ideas about a given topic or situation, or to develop creative ways to solve a problem.	48
Visualization	The ability to imagine how something will look after it is moved around or when its parts are moved or rearranged.	48
Near Vision	The ability to see details at close range (within a few feet of the observer).	48
Category Flexibility	The ability to generate or use different sets of rules for combining or grouping things in different ways.	45

Number Facility	The ability to add, subtract, multiply, or divide quickly and correctly.	45
Speed of Closure	The ability to quickly make sense of, combine, and organize information into meaningful patterns.	40
Speech Recognition	The ability to identify and understand the speech of another person.	35
Selective Attention	The ability to concentrate on a task over a period of time without being distracted.	33
Time Sharing	The ability to shift back and forth between two or more activities or sources of information (such as speech, sounds, touch, or other sources).	28
Auditory Attention	The ability to focus on a single source of sound in the presence of other distracting sounds.	28
Memorization	The ability to remember information such as words, numbers, pictures, and procedures.	25
Perceptual Speed	The ability to quickly and accurately compare similarities and differences among sets of letters, numbers, objects, pictures, or patterns. The things to be compared may be presented at the same time or one after the other. This ability also includes comparing a presented object with a remembered object.	20

[a]Anmerkung**:** Es sind nur Fähigkeiten mit einem Gewicht von mindestens 20 dargestellt.

B. Interests - Occupational Interests

Interest	**Description**	**Importance**
Investigative	Investigative occupations frequently involve working with ideas, and require an extensive amount of thinking. These occupations can involve searching for facts and figuring out problems mentally.	83
Enterprising	Enterprising occupations frequently involve starting up and carrying out projects. These occupations can involve leading people and making many decisions. Sometimes they require risk taking and often deal with business.	83
Conventional	Conventional occupations frequently involve following set procedures and routines. These occupations can include working with data and details more than with ideas. Usually there is a clear line of authority to follow.	83
Social	Social occupations frequently involve working with, communicating with, and teaching people. These occupations often involve helping or providing service to others.	44
Realistic	Realistic occupations frequently involve work activities that include practical, hands-on problems and solutions. They often deal with plants, animals, and real-world materials like wood, tools, and machinery. Many of the occupations require working outside, and do not involve a lot of paperwork or working closely with others.	33
Artistic	Artistic occupations frequently involve working with forms, designs and patterns. They often require self-expression and the work can be done without following a clear set of rules.	17

B. Interests - Occupational Values

Interest	Description	Importance
Achievement	Occupations that satisfy this work value are results oriented and allow employees to use their strongest abilities, giving them a feeling of accomplishment. Corresponding needs are Ability Utilization and Achievement.	78
Independence	Occupations that satisfy this work value allow employs to work on their own and make decisions. Corresponding needs are Creativity, Responsibility and Autonomy.	75
Recognition	Occupations that satisfy this work value offer advancement, potential for leadership, and are often considered prestigious. Corresponding needs are Advancement, Authority, Recognition and Social Status.	64
Working Conditions	Occupations that satisfy this work value offer job security and good working conditions. Corresponding needs are Activity, Compensation, Independence, Security, Variety and Working Conditions.	63
Relationships	Occupations that satisfy this work value allow employees to provide service to others and work with co-workers in a friendly non-competitive environment. Corresponding needs are Co-workers, Moral Values and Social Service.	57
Support	Occupations that satisfy this work value offer supportive management that stands behind employees. Corresponding needs are Company Policies, Supervision: Human Relations and Supervision: Technical.	52

2. Worker Requirements

A. Basic Skills[a]

Basic Skill	Description	Importance
Writing	Communicating effectively in writing as appropriate for the needs of the audience.	89
Reading Comprehension	Understanding written sentences and paragraphs in work related documents.	86
Speaking	Talking to others to convey information effectively.	84
Systems Evaluation	Identifying measures or indicators of system performance and the actions needed to improve or correct performance, relative to the goals of the system.	83
Active Listening	Giving full attention to what other people are saying, taking time to understand the points being made, asking questions as appropriate, and not interrupting at inappropriate times.	82
Critical Thinking	Using logic and reasoning to identify the strengths and weaknesses of alternative solutions, conclusions or approaches to problems.	82
Operations Analysis	Analyzing needs and product requirements to create a design.	80
Judgment and Decision Making	Considering the relative costs and benefits of potential actions to choose the most appropriate one.	80

Complex Problem Solving	Identifying complex problems and reviewing related information to develop and evaluate options and implement solutions.	77
Systems Analysis	Determining how a system should work and how changes in conditions, operations. and the environment will affect outcomes.	77
Active Learning	Understanding the implications of new information for both current and future problem-solving and decision-making.	75
Management of Personnel Resources	Motivating, developing, and directing people as they work, identifying the best people for the job.	73
Monitoring	Monitoring/Assessing performance of yourself, other individuals, or organizations to make improvements or take corrective action.	70
Coordination	Adjusting actions in relation to others' actions.	68
Management of Material Resources	Obtaining and seeing to the appropriate use of equipment, facilities, and materials needed to do certain work.	68
Learning Strategies	Selecting and using training/instructional methods and procedures appropriate for the situation when learning or teaching new things.	66
Time Management	Managing one's own time and the time of others.	64
Instructing	Teaching others how to do something.	61
Equipment Selection	Determining the kind of tools and equipment needed to do a job.	59
Mathematics	Using mathematics to solve problems.	52
Social Perceptiveness	Being aware of others' reactions and understanding why they react as they do.	50
Persuasion	Persuading others to change their minds or behavior.	45
Quality Control Analysis	Conducting tests and inspections of products, services, or processes to evaluate quality or performance.	41
Negotiation	Bringing others together and trying to reconcile differences.	34
Service Orientation	Actively looking for ways to help people.	30
Technology Design	Generating or adapting equipment and technology to serve user needs.	30
Management of Financial Resources	Determining how money will be spent to get the work done. and accounting for these expenditures.	30

[a]Anmerkung**:** Es sind nur Qualifikationen mit einem Gewicht von mindestens 20 dargestellt.

C. Knowledge[a]

Knowledge	Description	Importance
Administration and Management	Knowledge of business and management principles involved in strategic planning, resource allocation, human resources modeling, leadership technique, production methods, and coordination of people and resources.	91
English Language	Knowledge of the structure and content of the English language including the meaning and spelling of words, rules of composition. and grammar.	64
Education and Training	Knowledge of principles and methods for curriculum and training design. teaching and instruction for individuals and groups, and the measurement of training effects.	61
Mathematics	Knowledge of arithmetic, algebra, geometry, calculus, statistics, and their applications.	59
Personnel and Human Resources	Knowledge of principles and procedures for personnel recruitment, selection, training, compensation and benefits, labor relations and negotiation. and personnel information systems.	52
Clerical	Knowledge of administrative and clerical procedures and systems such as word processing, managing files and records, stenography and transcription, designing forms, and other office procedures and terminology.	48
Economics and Accounting	Knowledge of economic and accounting principles and practices, the financial markets, banking and the analysis and reporting of financial data.	45
Computers and Electronics	Knowledge of circuit boards. processors, chips, electronic equipment, and computer hardware and software, including applications and programming.	39
Psychology	Knowledge of human behavior and performance. individual differences in ability. personality. and interests. learning and motivation. psychological research methods. and the assessment and treatment of behavioral and affective disorders.	34
Production and Processing	Knowledge of raw materials, production processes, quality control, costs, and other techniques for maximizing the effective manufacture and distribution of goods.	30
Design	Knowledge of design techniques, tools, and principles involved in production of precision technical plans. blueprints, drawings, and models.	23
Communications and Media	Knowledge of media production, communication. and dissemination techniques and methods. This includes alternative ways to inform and entertain via written, oral, and visual media.	23
Law and Government	Knowledge of laws, legal codes, court procedures, precedents, government regulations, executive orders, agency rules, and the democratic political process.	20
Telecommunications	Knowledge of transmission, broadcasting, switching, control, and operation of telecommunications systems.	20

[a]Anmerkung**:** Es sind nur Wissensdimensionen mit einem Gewicht von mindestens 20 dargestellt. Eine komplette Darstellung dieser Tabelle findet sich in der dieser Arbeit beigelegten CD-Rom

4. Occupational Requirements

A. Generalized Work Activities[a]

Work Activity	Description	Importance
Provide Consultation and Advice to Others	recommend improvements to work methods or procedures.	89
Provide Consultation and Advice to Others	recommend purchase or repair of furnishings or equipment.	89
Provide Consultation and Advice to Others	recommend solutions of administrative problems.	89
Provide Consultation and Advice to Others	work with management or employees to improve organizational system.	89
Getting Information	obtain information from individuals.	86
Getting Information	study time, motion, or work methods of workers.	86
Communicating with Supervisors, Peers, or Subordinates	Providing information to supervisors, co-workers, and subordinates by telephone, in written form, e-mail, or in person.	86
Analyzing Data or Information	analyze operational or management reports or records.	82
Analyzing Data or Information	analyze organizational operating practices or procedures.	82
Analyzing Data or Information	analyze workflow.	82
Analyzing Data or Information	conduct organizational studies.	82
Analyzing Data or Information	evaluate office operations.	82
Making Decisions and Solving Problems	select software for clerical activities.	77
Organizing, Planning, and Prioritizing Work	plan study of work problems or procedures.	73
Organizing, Planning, and Prioritizing Work	prepare workflow chart.	73
Monitor Processes, Materials, or Surroundings	identify problems or improvements.	68
Identifying Objects, Actions, and Events	understand technical operating, service or repair manuals.	68
Developing Objectives and Strategies	develop plans for programs or projects.	66
Developing Objectives and Strategies	develop policies, procedures, methods, or standards.	66
Processing Information	Compiling, coding, categorizing, calculating, tabulating, auditing, or verifying information or data.	64

Documenting/Recording Information	Entering, transcribing, recording, storing, or maintaining information in written or electronic/magnetic form.	64
Establishing and Maintaining Interpersonal Relationships	Developing constructive and cooperative working relationships with others, and maintaining them over time.	64
Evaluating Information to Determine Compliance with Standards	Using relevant information and individual judgment to determine whether events or processes comply with laws, regulations, or standards.	61
Thinking Creatively	design office layout.	61
Thinking Creatively	develop new office forms.	61
Interpreting the Meaning of Information for Others	prepare instruction manuals.	61
Performing Administrative Activities	develop records management system.	61
Performing Administrative Activities	maintain records, reports, or files.	61
Performing Administrative Activities	prepare reports.	61
Performing Administrative Activities	prepare reports for management.	61
Performing Administrative Activities	write administrative procedures services manual.	61
Estimating the Quantifiable Characteristics of Products, Events, or Information	Estimating sizes, distances, and quantities or determining time, costs, resources, or materials needed to perform a work activity.	59
Judging the Qualities of Things, Services, or People	Assessing the value, importance, or quality of things or people.	59
Updating and Using Relevant Knowledge	use interviewing procedures.	57
Updating and Using Relevant Knowledge	use inventory control procedures.	57
Updating and Using Relevant Knowledge	use oral or written communication techniques.	57
Coordinating the Work and Activities of Others	direct implementation of new procedures, policies, or programs.	57
Training and Teaching Others	conduct training for personnel.	57
Training and Teaching Others	write employee orientation or training materials.	57
Scheduling Work and Activities	Scheduling events, programs, and activities, as well as the work of others.	55
Communicating with Persons Outside Organization	communicate technical information.	52
Communicating with Persons Outside Organization	compile information through interviews.	52
Communicating with Persons Outside Organization	confer with management or users	52

Monitoring and Controlling Resources	Monitoring and controlling resources and overseeing the spending of money.	50

[a]Anmerkung**:** Es sind nur Aktivitäten mit einem Gewicht von mindestens 50 dargestellt. Die Beschreibung bezieht sich immer auf die höchste differenzierende Dimension innerhalb der Aktivität.

C. Work Context[a]

Work Context	**Description**	**Importance**
Indoors, Environmentally Controlled	How often does this job require working indoors in environmentally controlled conditions?	100
Coordinate or Lead Others	How important is it to coordinate or lead others in accomplishing work activities in this job?	75
Contact With Others	How much does this job require the worker to be in contact with others (face-to-face, by telephone, or otherwise) in order to perform it?	73
Responsibility for Outcomes and Results	How responsible is the worker for work outcomes and results of other workers?	68
Spend Time Sitting	How much does this job require sitting?	65
Importance of Being Exact or Accurate	How important is being very exact or highly accurate in performing this job?	60
Consequence of Error	How serious would the result usually be if the worker made a mistake that was not readily correctable?	55
Spend Time Standing	How much does this job require standing?	40
Frequency of Conflict Situations	How often are there conflict situations the employee has to face in this job?	35
Responsible for Others' Health and Safety	How much responsibility is there for the health and safety of others in this job?	31
Deal With External Customers	How important is it to work with external customers or the public in this job?	30
Deal With Unpleasant or Angry People	How frequently does the worker have to deal with unpleasant, angry, or discourteous individuals as part of the job requirements?	30
Spend Time Walking and Running	How much does this job require walking and running?	30
Spend Time Using Your Hands to Handle, Control, or Feel Objects, Tools, or Controls	How much does this job require using your hands to handle, control, or feel objects, tools or controls?	25

[a]Anmerkung**:** Es sind nur Kontextdimensionen mit einem Gewicht von mindestens 20 dargestellt.

5. Occupation Specific Information

A. Tasks

Task
Review forms and reports, and confer with management and users about format, distribution, and purpose, and to identify problems and improvements.
Develop and implement records management program for filing, protection, and retrieval of records, and assure compliance with program.
Interview personnel and conduct on-site observation to ascertain unit functions, work performed, and methods, equipment, and personnel used.
Prepare manuals and train workers in use of new forms, reports, procedures or equipment, according to organizational policy.
Design, evaluate, recommend, and approve changes of forms and reports.
Recommend purchase of storage equipment, and design area layout to locate equipment in space available.
Plan study of work problems and procedures, such as organizational change, communications, information flow, integrated production methods, inventory control, or cost analysis.
Gather and organize information on problems or procedures.
Analyze data gathered and develop solutions or alternative methods of proceeding.
Document findings of study and prepare recommendations for implementation of new systems, procedures, or organizational changes.
Confer with personnel concerned to ensure successful functioning of newly implemented systems or procedures.

Anhang II: Nutzenanalyse zur Personalauswahl (erweitert in Anlehnung an Holling (1998, S.12ff.))

Autoren	Inhalt	SD_y-Schätzung	Nutzenmodell[a]
Roche (1965)	Nutzen psychologischer Verfahren bei der Auswahl von Bohrmaschinenbedienern	Kostenrechnungs-methode	BCG-Modellvariante
v. Naerssen (1965)	Nutzen psychologischer Verfahren bei der Auswahl von Fahrschülern für die holländische Arme	Kostenrechnungs-methode	BCG-Modellvariante
Cascio und Silbey (1979)	Nutzen eines AC bei der Auswahl von Vertriebsmanagern in der Nahrungsmittelindustrie	Globales Schätz-verfahren	BCG-Modellvariante
Schmidt, Hunter. McKenzie und Muldrow (1979)	Nutzen eines Tests zur Auswahl von Programmierern	Globales Schätz-verfahren	BCG-Modellvariante
Arnold. Rauschenberger, Soubel und Guion (1982)	Nutzen eines Tests zur Messung physischer Kraft bei der Auswahl von Stahlarbeitern	Kostenrechnungs-methode	BCG-Modellvariante
Dunette. et. al. (1982)	Nutzen eines Testbatterie für die Auswahl von Bedienern von Atomkraftwerken	Globales Schätz-verfahren	BCG-Modellvariante
Ledvinka, Simonet. Neiner und Kruse (1983)	Nutzen eines Tests bei der Auswahl von Mitarbeiter einer Lebensversicherung	Proportionale Regel	BCG-Modellvariante
Schmidt, Mack und Hunter (1984)	Nutzen eines Tests bei der Auswahl von Nationalparkaufsehern	Globales Schätz-verfahren und proportionale Regel	BCG-Modellvariante
Wroten (1984)	Nutzen mehrerer Tests bei der Auswahl von Bedienern von Pumpen	Proportionale Regel	BCG-Modellvariante
Weekley, O'Connor, Frank und Peters (1985)	Nutzen einer Testbatterie bei der Auswahl von Leitern von 24h-Läden	Globales Schätz-verfahren und CREPID	BCG-Modellvariante
Cascio und Ramos (1986)	Nutzen eines AC bei der Auswahl von Büroleitern einer Telefongesellschaft	CREPID	BCG-Modellvariante
Schmidt, Hunter, Outerbridge und Trattner (1986)	Nutzen eines Test kognitiver Fähigkeiten bei der Auswahl von Behördenmitarbeitern	Proportionale Regel	BCG-Modellvariante
Burke und Frederik (1986)	Nutzen eines AC bei der Auswahl von mittlerem Management im Vertrieb	Kostenrechnungs-methode und globales Schätz-verfahren	Boudreau-Modellvariante
Cronshaw, Alexander, Wiesner und Barrick (1987)	Nutzen eines kognitiven Fähigkeitstest bei der Auswahl von Verwaltungsangestellten des kanadischen Militärs	Proportionale Regel	Boudreau-Modellvariante
Rich und Boudreau (1987)	Nutzen eines Eignungstests bei der Auswahl von Programmierern	Globales Schätz-verfahren	Boudreau-Modellvariante

Barthel (1988)	Nutzen biographischer Fragebögen bei der Auswahl von Versicherungsaußendienstmitarbeitern	Empirische Ermittlung basierend auf Verkaufszahlen	Boudreau-Modellvariante
Gerpott (1990)	Nutzen eines AC bei der Auswahl von Managern in der chemischen Industrie	Proportionale Regel	Boudreau-Modellvariante
Hakstian, Wooley. Woolsey und Kryger (1991)	Nutzen einer Testbatterie bei der Auswahl von Managern	Globales Schätzverfahren	BCG-Modellvariante
Tziner, et. al. (1994)	Nutzen eines AC bei der Auswahl von Managern eines israelischen Unternehmens	k. A. (Quelle: Buchhaltung)	Boudreau-Modellvariante
Holling und Reiners (1994)	Nutzen eines AC bei der Auswahl von Versicherungsaußendienstmitarbeitern	Empirische Ermittlung	BCG-Modellvariante
Schuler, Moser, Funke und Donat (1995)	Nutzen verschiedener Verfahren bei der Auswahl von Personal aus Forschung und Entwicklung	Globales Schätzverfahren und CREPID	Boudreau-Modellvariante
Jacobs, Conte, Day, Silva und Harris (1996)	Nutzen eines Verfahrens für die Auswahl von Busfahrern im öffentlichen Dienst	Keine	Herleitung des Nutzens über Extremgruppenvergleiche
Hoffmann und Thornton (1997)	Nutzen eines AC bei der Auswahl von Mitarbeitern in einem öffentlichen Versorgungsbetrieb	Proportionale Regel	BCG-Modellvariante
Gast und McMillan (1999)	Nutzen eines neuen Auswahlverfahrens für Führungskräfte	Proportionale Regel	BCG-Modellvariante
Lindstam (2005)	Nutzen eines Personalauswahlprogramm für den Außendienst eines mittelständischen Unternehmens der Bauindustrie	Kostenrechnungsmethode (erwirtschafteter Deckungsbeitrag)	BCG-Modellvariante und Boudreau-Modellvariante

[a]Anmerkung**:** Bei der Spalte *Nutzenmodell* wurde das Boudreau-Modell, dass im Grunde auch eine Variante des BCG-Modells darstellt, separat ausgewiesen, um den Investitionsgedanken, der in der jeweiligen Berechnung steckt, hervorzuheben.

Anhang III: Ausgewählte Kriterien erfolgreicher Beratung

Autoren	**Ausgewählte Kriterien**[a]	**Kriterienart**
Klein (1978)	Umsatz des Beraters mit dem Klienten	Ergebniskriterium
	Zufriedenheit des Kunden	Ergebniskriterium
	Quantitative Zielvariablen	Ergebniskriterium
Kienbaum (1979)	Allgemeiner Vergleich von Konzept und Realisation	Ergebniskriterium
	Soll-Ist-Vergleich vorgeschlagener Maßnahmen	Ergebniskriterium
	Kosten-Nutzen-Vergleich	Ergebniskriterium
	Zufriedenheit des Klienten	Ergebniskriterium
	Bereitschaft des Klienten. als Referenz zu dienen	Ergebniskriterium
	Folgeprojekte	Ergebniskriterium
Klaile (1984)[b]	Information über die Hintergründe und Zusammenhänge der Ziele und Tätigkeiten	Verhaltenskriterium
	Detaillierte Zielvereinbarung bzw. –vorgabe durch Klienten	Kontext
	Detaillierte Untersuchung der zu lösenden Aufgabe	Verhaltenskriterium
	Adäquate Berücksichtigung der Fähigkeiten der Mitarbeiter	Verhaltenskriterium
	Lösung von Rollenkonflikten durch Beseitigung der Ursachen	Verhaltenskriterium
	Bildung von Konsens	Verhaltenskriterium
	Ausrichtung der Aufgabenzuweisung an den Pk-faktoren	Verhaltenskriterium
	Berücksichtigung der Einstellungen und Werte aller Beteiligten	Verhaltenskriterium
	Erfassung relevanter Erwartungen aller Beteiligten	Verhaltenskriterium
	Kreativität bei der Problemlösung	Verhaltenskriterium
	Fähigkeit zur Beeinflussung anderer	Eigenschaftskriterium
	Bereitschaft. Konsequenzen von Entscheidungen zu akzeptieren	Eigenschaftskriterium
Fridrich (1985)	Wirtschaftlichkeit	Ergebniskriterium
	Effizienzsteigerung	Ergebniskriterium
	Realisierungsgrad der Vorschläge	Ergebniskriterium
Gabele und Hirsch (1986)	Benutzerfreundlichkeit der (End-)Dokumentation	Ergebniskriterium
	Präzision der (End)-Dokumentation	Ergebniskriterium
Mugler und Lampe (1987)	Angemessene zeitliche Ressourcenallokation im Projekt	Kontext
	Angemessene personelle Ressourcenallokation im Projekt	Kontext
	Berücksichtigung der kognitiven Fähigkeiten des Unternehmers bei der Ergebnisdarstellung	Ergebniskriterium
	Divergenzen über Einschätzung der Ausgangslage	Kontext
Meffert (1990)	Kundenzufriedenheit	Ergebniskriterium
	Richtigkeit der Empfehlung	Ergebniskriterium
	Grad der Umsetzung	Ergebniskriterium
	Anschlussaufträge	Ergebniskriterium
	Know-how-Zuwachs beim Klienten	Ergebniskriterium
	Know-how-Zuwachs bei der Beratung	Ergebniskriterium
	Imagezuwachs bei der Beratung	Ergebniskriterium
	Akzeptanz der Beratungsergebnisse	Ergebniskriterium
	Erfahrungsgewinnung beim Berater	Ergebniskriterium
	Qualitative Zielerreichung	Ergebniskriterium
	Kosten-Nutzen Relation des Projektes	Ergebniskriterium

Hill (1990)	Angemessener Detaillierungsgrad der an den Klienten übergebenen Informationen	Ergebniskriterium
	Überprüfbarkeit der Informationserhebung und logische Nachvollziehbarkeit der Auswertung	Verhaltenskriterium
	Klare Definition der benötigten Information	Verhaltenskriterium
	Methodenkenntnis des Beraters	Eigenschaftskriterium
	Know-how des Beraters	Eigenschaftskriterium
	Abstimmung der Ziele zwischen Berater und Klienten	Kontext
	Kooperationsbereitschaft und Engagement des Klienten	Eigenschaftskriterium
	Zugang zu Informationsquellen	Kontext
Hoffmann (1991)	Finanzielle Ressourcen des Klienten	Kontext
	Bereitschaft des Klienten zur aktiven Mitarbeit	Eigenschaftskriterium
	Lern- und Wandelfähigkeit des Klienten	Eigenschaftskriterium
	Angemessenes übergreifendes Projektcontrolling	Kontext
	Übereinstimmung Beraterverhalten und Klientenerwartung	Ergebniskriterium
	Zügigkeit des Prozesses	Verhaltenskriterium
	Zufriedenheit des Klienten	Ergebniskriterium
	Zufriedenheit der Mitarbeiter	Ergebniskriterium
	Umsetzungsgrad der Empfehlungen	Ergebniskriterium
Effenberger (1998)	Zufriedenheit des Kunden	Ergebniskriterium
Gable (1996)	Akzeptanz der Empfehlungen	Eigenschaftskriterium
	Angemessenheit und Güte der Empfehlungen des Beraters	Verhaltenskriterium
	Lerneffekt des Kunden	Ergebniskriterium
	Zufriedenheit des Kunden	Ergebniskriterium
	Technische Beraterkompetenz	Eigenschaftskriterium
	Verständnis der Anforderungen	Eigenschaftskriterium
	Vollständigkeit der Dokumentation	Ergebniskriterium
	Vertraulichkeit	Verhaltenskriterium
Köppen (1999)	Wandlungsbereitschaft des Klienten	Eigenschaftskriterium
	Klientenengagement	Eigenschaftskriterium
	Fachkompetenz	Eigenschaftskriterium
	Konfliktgehalt der Aufgabe	Kontext
	Soziale Kompetenz	Eigenschaftskriterium
	Seniorität des Auftretens des Beraters	Verhaltenskriterium
Phillips (2000)	Zufriedenheit des Klienten	Ergebniskriterium
	Lerneffekt	Ergebniskriterium
	Implementierungserfolg	Ergebniskriterium
	Geschäftsrelevante Veränderungen (Business Impact)	Ergebniskriterium
	Kosten-Nutzen-Relation	Ergebniskriterium
Höck und Keuper (2001)	Termineinhaltung	Verhaltenskriterium
	Kommunikationsfähigkeit	Eigenschaftskriterium

[a] Es sind jeweils nur die aus Sicht des Autors für die vorliegende Arbeit relevanten Kriterien genannt. Diese sind zudem teilweise aus Verständlichkeitsgründen umformuliert worden.

[b] Obwohl Klaile die Kriterien auf die Führungsfähigkeit im Unternehmen bezieht, können einige davon nach Meinung des Autors sehr gut auf die Projektarbeit des Beraters übertragen werden, die ja ebenfalls die Führung des Projektes oder von Teilmodulen beinhaltet

Anhang IV: Kriterienzuordnung

1. Eigenschaftskriterien Berater

Kategorie	Dimensionen	Zugeordnete Kriterien
Fachwissen	Branchenwissen Funktionales Wissen	Fachkompetenz (Köppen, 1999), technische Beraterkompetenz (Gable, 1996)
Methoden-Wissen	Kenntnisse der Analyseinstrumente Technische Kenntnisse (EDV, Software)	Methodenkenntnisse des Beraters (Hill, 1990), Know-how des Beraters (Hill, 1990)
Soziale Kompetenz	Kooperationsbereitschaft Solidarität Toleranz Konfliktfähigkeit Empathie	Soziale Kompetenz (Köppen, 1999), Fähigkeit zur Beeinflussung anderer (Klaile, 1984)
Analytische Fähigkeiten	Mündliche Auffassungsgabe Schriftliche Auffassungsgabe Problemsensibilität Strukturiertheit Rechnerische Fähigkeiten Schlussfolgerndes Denken	Verständnis der Anforderungen (Gable, 1996), Written Comprehension (O*Net), Oral Comprehension (O*Net), Problem Sensitivity (O*Net), Deductive Reasoning (O*Net), Inductive Reasoning (O*Net), Mathematical Reasoning (O*Net),Number Facility (O*Net) Speed for Closure (O*Net), Information Ordering (O*Net)
Kommunikative Fähigkeiten	Mündliche Ausdrucksfähigkeit Klarheit der Aussprache Schriftliche Ausdrucksfähigkeit Visualisierungsfähigkeit	Kommunikationsfähigkeit (Höck & Keuper, 2001), Oral Expression (O*Net), Speech Clarity (O*Net), Written Expression (O*Net)

2. Eigenschaftskriterien Klient[a]

Kategorie	Dimensionen	Zugeordnete Kriterien
Engagement	Leistungswille Identifikation mit Projektzielen	Klientenengagement (Köppen, 1999), Kooperationsbereitschaft und Engagement des Klienten (Hill, 1990)
Kooperations-bereitschaft	Veränderungswille Akzeptanz getroffener Projekt-entscheidungen Wille zur Zusammenarbeit	Kooperationsbereitschaft und Engagement des Klienten (Hill, 1990), Akzeptanz der Empfehlungen (Gable, 1996), Wandlungsbereitschaft des Klienten (Köppen, 1999), Bereitschaft zur aktiven Mitarbeit (Hoffmann, 1991), Bereitschaft, Konsequenzen von Entscheidungen zu akzeptieren (Klaile, 1984)
Wandlungsfähigkeit	Lernfähigkeit Flexibilität	Lern- und Wandlungsfähigkeit des Klienten (Hoffmann, 1991)

[a] Es werden lediglich solche Eigenschaften genannt, die nach Ansicht des Autors einen Einfluss auf die Arbeit und somit die Ergebnisse des Beraters haben und in der Evaluation der Beraterleistung besondere Berücksichtigung finden sollten.

3. Kontext

Kategorie	Dimensionen	Zugeordnete Kriterien
Status	Einheitliches Verständnis über Ausgangslage Klare Kommunikation der Ausgangslage Konfliktgehalt der Aufgabe	Divergenzen über Einschätzung der Ausgangslage (Mugler & Lampe, 1987), Konfliktgehalt der Aufgabe (Köppen, 1999)
Zielvereinbarung	Klare Zielabsprache Klare Zieldefinition Klare Zielkommunikation	Detaillierte Zielvereinbarung und –vorgabe durch Klienten (Klaile, 1984), Abstimmung der Ziele zwischen Berater und Klienten (Hill, 1990)
Ressourcen	Angemessene zeitliche Ressourcen Angemessene persönliche Ressourcen Angemessene finanzielle Ressourcen Zugang zu Informationsquellen	Angemessene zeitliche Ressourcenallokation im Projekt (Mugler & Lampe, 1987), Angemessene persönliche Ressourcenallokation im Projekt (Mugler & Lampe, 1987), Finanzielle Ressourcen des Klienten (Hoffmann, 1991)
Projektorganisation	Klare Projektstruktur Eindeutige Projektverantwortlichkeiten Professionelles Projektcontrolling	Angemessenes übergreifendes Projektcontrolling (Hoffmann, 1991)

4. Ergebnis

Kategorie	Dimensionen	Zugeordnete Kriterien
UB-Ziele	Umsatz des Beraters mit dem Klienten Anzahl/Umfang der Folgeprojekte Interner Wissenszuwachs Imagezuwachs	Umsatz des Beraters mit dem Klienten (Klein, 1978), Folgeprojekte (Kienbaum, 1979), Anschlussaufträge (Meffert, 1990), Know-how-Zuwachs bei der Beratung (Meffert, 1990), Imagezuwachs bei der Beratung (Meffert, 1990)
Zielerreichung	Kosten-Nutzen-Relation des Projektes Implementierungsgrad der Maßnahmen Quantitative Zielvariablen (projektspezifisch) Qualitative Zielvariablen (projektspezifisch) Akzeptanz der Ergebnisse Know-how-Zuwachs beim Klienten	Quantitative Zielvariablen (Klein, 1978), Kosten-Nutzen-Vergleich (Kienbaum, 1979, Philipps, 2000, Fridrich, 1985, Meffert, 1990), Soll-Ist-Vergleich vorgeschlagener Maßnahmen (Kienbaum, 1979), Effizienzsteigerung (Fridrich, 1985), Realisierungsgrad der Vorschläge (Kienbaum, 1979, Fridrich, 1985), Grad der Umsetzung (Meffert, 1990, Hoffmann, 1991), Know-how-Zuwachs beim Klienten (Meffert, 1990, Philipps, 2000, Gable 1996), Akzeptanz der Beratungsergebnisse (Meffert, 1990), Qualitative Zielerreichung (Meffert, 1990), Übereinstimmung Beraterverhalten und Klientenerwartung (Hoffmann, 1991), Implementierungserfolg (Philipps, 2000), Geschäftsrelevante Veränderung (Philipps, 2000)
Zufriedenheit	Zufriedenheit des Auftraggebers Zufriedenheit der Klientenmitarbeiter Bereitschaft des Klienten. als Referenzprojekt zu dienen	Zufriedenheit des Kunden (Klein, 1978, Kienbaum, 1979, Hoffmann, 1991, Effenberger, 1998, Meffert, 1990, Phillips, 2000, Gable, 1996), Bereitschaft des Klienten, als Referenzprojekt zu dienen (Kienbaum, 1979), Zufriedenheit der Mitarbeiter (Hoffmann, 1991)

Dokumentation	Qualität der Charts Struktur der Präsentation Verständlichkeit für Auftraggeber Angemessener Detaillierungsgrad Präzision/Fehlerfreiheit der Präsentation Vollständigkeit Einhaltung der Standards	Benutzerfreundlichkeit der (End-) Dokumentation (Gabele & Hirsch, 1986), Präzision der Enddokumentation (Gabele & Hirsch, 1986), Angemessener Detaillierungsgrad der an den Klienten übergebenen Information (Hill, 1990), Berücksichtigung der kognitiven Fähigkeiten des Unternehmers bei der Ergebnisdarstellung (Mugler & Lampe, 1987), Vollständigkeit der Präsentation (Gable, 1996)
Qualität	Richtigkeit der Empfehlung	Richtigkeit der Empfehlung (Meffert, 1990)

5. Prozess

Prozessstufe	**Dimensionen**	**Zugeordnete Kriterien**
Problemdefinition	Umfassende Erfassung der Ausgangslage Identifikation der Erwartungen der Beteiligten Identifikation des Problems Exakte Beschreibung des Problems	Erfassung relevanter Erwartung der Beteiligten (Klaile, 1984), Active Listening (O*Net)
Erste Maßnahmen	Identifikation von „Quick-Hits“ Schnelle Umsetzung von „Quick-Hits“ Effektivität der „Quick-Hits“	k. A.
Datensammlung	Klare Definition benötigter Information Identifikation relevanter Quellen Sammlung und Aufbereitung aller Daten Überprüfbarkeit und Nachvollziehbarkeit des Vorgehens	Detaillierte Untersuchung der zu lösenden Aufgabe (Klaile, 1984), Klare Definition benötigter Information (Hill, 1990)
Datenanalyse	Logische Strukturierung der Information Auswahl/Entwicklung geeigneter Analyseraster und -instrumente Fehlerfreiheit/Präzision der Analysen Überprüfbarkeit und Nachvollziehbarkeit des Vorgehens	Detaillierte Untersuchung der zu lösenden Aufgabe (Klaile, 1984), Complex Problem Solving (O*Net), Equipment Selection (O*Net), Mathematics (O*Net)
Ursachenanalyse	Identifikation der Ursache-Wirkungs-Zusammenhänge Herausarbeitung der „Key-Messages“	Lösung von Rollenkonflikten durch Beseitigung der Ursachen (Klaile, 1984), Operational Analysis (O*Net), System Analysis (O*Net)
Lösungssuche	Flüssigkeit bei der Ideengenerierung Originalität bei der Ideengenerierung Angemessenheit der Ideen	Angemessenheit und Güte der Empfehlungen des Beraters (Gable, 1996), Kreativität bei der Problemlösung (Klaile, 1984), Fluency of Ideas (O*Net), Originality (O*Net),
Bewertung	Identifikation relevanter Bewertungskriterien Systematische Beurteilung der Lösungen Ableitung geeigneter Maßnahmen	Systems Evaluation (O*Net), Judgement and Decision Making (O*Net), Critical Thinking (O*Net)
Identifikation möglicher Hindernisse	Identifikation potenzieller Hindernisse Erarbeitung möglicher Auswege	Social Perceptiveness (O*Net)

Implementierung	Umfassende Einweisung und Information der Beteiligten Berücksichtigung der Fähigkeiten der Mitarbeiter bei der Aufgabenverteilung Professionelle Konfliktlösung in kritischen Situationen Positive Beeinflussung der Beteiligten Koordination der Umsetzung Überprüfbarkeit und Nachvollziehbarkeit des Vorgehens	Information über Hintergrunde und Zusammenhänge der Ziele und Tätigkeiten (Klaile, 1984), adäquate Berücksichtigung der Fähigkeiten der Mitarbeiter (Klaile, 1984), Ausrichtung der Aufgabenzuweisung an den Pk-Faktoren (Klaile, 1984), Bildung von Konsens (Klaile, 1984), Coordination (O*Net), Management of Personnel Resources (O*Net), Instructing (O*Net), Persuasion (O*Net)
Kontrolle und Verbesserung	Nachhalten der Implementierungserfolge Abweichungsanalysen Gegensteuern bei Fehlentwicklungen	Monitoring (O*Net), Quality Control Analysis (O*Net)
Stufenübergreifende Professionalität	Vertraulichkeit im Verhalten Seniorität des Auftretens Pünktlichkeit/Termineinhaltung Zuverlässigkeit Zügigkeit der Bearbeitung Ausreichender Informationsaustausch mit anderen Teammitgliedern Verhandlungsgeschick Kostenbewusstes Verhalten Durchsetzungsstärke Stressverhalten Akzeptanz durch den Klienten	Vertraulichkeit (Gable, 1996), Seniorität des Auftretens des Beraters (Köppen, 1999), Zügigkeit des Prozesses (Hoffmann, 1991), Time Management (O*Net)

Anhang V: Detaillierung der Recruitingdimensionen und Aggregation

1. Beurteilungsbögen

Bewerberbeurteilung – Persönlichkeitsinterview

Bewerber: _____ Datum: _____

Interviewer: _____ Ort: _____

		nicht entwickelt 1	2	3	4	*hoch entwickelt* 5
Pragmatismus/ Ergebnisorientierung	• Steht der Prozess oder das Ergebnis im Vordergrund • Denkt der Bewerber in Details oder global • Entscheidungsverhalten	☐	☐	☐	☐	☐
Selbstorganisation Engagement, Biss	• Engagement auch außerhalb des Studiums • Biss auch bei schwierigen Aufgaben • Verantwortungsbewusstsein	☐	☐	☐	☐	☐
Frustrationstoleranz	• Steht auch schwierige, frustrierende Situationen durch • Kann Durststrecken überwinden/Leistungsbereitschaft • Bleibt im Stress relativ ruhig/Stressstabilität	☐	☐	☐	☐	☐
Fit zu ISB	• Passt ins ISB Team	☐	☐	☐	☐	☐
Team-/Integrations-fähigkeit	• Kann sehr gut im Team arbeiten • Handelt integrativ im Team • Übernimmt zur richtigen Zeit die Führung des Teams	☐	☐	☐	☐	☐
Kreativität/Flexibilität	• Geistig beweglich, stellt sich schnell auf neue Themen und Situationen ein • Entwickelt unterschiedliche Problemslösungsansätze	☐	☐	☐	☐	☐
Durchsetzungskraft/ Führungsfähigkeit	• Selbstbewusstsein, natürliche Autorität • Außerberufliche Führungserfahrungen • Zielorientierung als Grundlage der Durchsetzungsfähigkeit	☐	☐	☐	☐	☐
Konflikt- und Kritikfähigkeit	• Umgang und Lösung von Konflikten • Wie geht der Bewerber mit Kritik an seiner Person um • Emotionale Stabilität	☐	☐	☐	☐	☐
Kommunikationsstärke/ Eloquenz	• Aussagenklarheit, Flüssigkeit im Sprechen • Schlüssigkeit der Argumentation, Überzeugungskraft • Formulierung, Redegewandtheit	☐	☐	☐	☐	☐
Ausstrahlung/Charisma	• Auftreten/Eindruck des Bewerbers • Persönliche Reife	☐	☐	☐	☐	☐

Entscheidung für Einstellung: ☐ **ja** ☐ **nein**

Bemerkungen:

Erläuterungen:

1 Die Leistung entsprach nicht den Erwartungen und Anforderungen
2 Die Erwartungen und Anforderungen wurden teilweise nicht ganz erfüllt
3 Die Leistung entsprach stets den Erwartungen und Anforderungen
4 Die Leistung entsprach stets den Erwartungen und Anforderungen und hat sie häufig übertroffen
5 Die Leistung übertraf stets die E. und A.

Bewerberbeurteilung – Fachinterview

Bewerber: Datum:

Interviewer: Ort:

		nicht entwickelt				*hoch entwickelt*
		1	**2**	**3**	**4**	**5**
Pragmatismus/ Ergebnisorientierung	• Steht der Prozess oder das Ergebnis im Vordergrund • Denkt der Bewerber in Details oder global • Entscheidungsverhalten	☐	☐	☐	☐	☐
Frustrationstoleranz	• Steht auch schwierige, frustrierende Situationen durch • Kann Durststrecken überwinden/Leistungsbereitschaft • Bleibt im Streß relativ ruhig/Streßstabilität	☐	☐	☐	☐	☐
Kreativität/Flexibilität	• Geistig beweglich, stellt sich schnell auf neue Themen, Situationen ein • Entwickelt unterschiedliche Problemlösungsansätze	☐	☐	☐	☐	☐
Kommunikationsstärke/ Eloquenz	• Aussagenklarheit, Flüssigkeit im Sprechen • Schlüssigkeit der Argumentation, Überzeugungskraft • Formulierung, Redegewandheit	☐	☐	☐	☐	☐
Strukturiertes/logisch-analytisches Denken	• Analytische Fähigkeiten • Auffassungsgabe • Strukturierung eines komplexen Problems	☐	☐	☐	☐	☐
Unternehmerisches Denken	• Kann sich in unternehmerischen Fragestellungen bewegen • Erfasst die Gesamtsituation • Unternehmerische Problemlösungsansätze	☐	☐	☐	☐	☐
Team-/Integrations-fähigkeit	• Kann sehr gut im Team arbeiten • Handelt integrativ im Team • Übernimmt zur richtigen Zeit die Führung	☐	☐	☐	☐	☐
Fachkenntnisse Exoten	• Kennt sich in seinem Fachgebiet aus	☐	☐	☐	☐	☐
Fachkenntnisse BWL	• Kosten- und Leistungsrechnung • Marketing, Organisation etc.	☐	☐	☐	☐	☐
Fachkenntnisse VWL	• Kenntnis der größten Player in der Wirtschaft u. deren Bedeutung • Grundverständnis wirtschaftlicher Zusammenhänge	☐	☐	☐	☐	☐
Allgemeine Kenntnisse	• Wirtschaftsteil der FAZ • Aktuelle wirtschaftliche Themen	☐	☐	☐	☐	☐
Methoden-Kenntnisse	• Kenntnis u. Anwendung analytischer/konzeptioneller Methoden	☐	☐	☐	☐	☐
Fit zu ISB	• Passt ins ISB Team	☐	☐	☐	☐	☐
Englischkenntnisse		☐	☐	☐	☐	☐

Entscheidung für Einstellung: ☐ **ja** ☐ **nein**

Erläuterungen:

1 Die Leistung entsprach nicht den Erwartungen und Anforderungen
2 Die Erwartungen und Anforderungen wurden teilweise nicht ganz erfüllt
3 Die Leistung entsprach stets den Erwartungen und Anforderungen
4 Die Leistung entsprach stets den Erwartungen und Anforderungen und hat sie häufig übertroffen
5 Die Leistung übertraf stets die E. und A.

Bewerberbeurteilung – Fallvorbereitung

Bewerber: _____ Datum: _____

Interviewer: _____ Ort: _____

		nicht entwickelt 1	2	3	4	*hoch entwickelt* 5
Pragmatismus/ Ergebnisorientierung	• Klar erkennbare Zielausrichtung • Orientierung an den wesentlichen Zielinhalten • Durchsetzungsvermögen	☐	☐	☐	☐	☐
Selbstorganisation Engagement, Biss	• Vertritt seinen Standpunkt mit Engagement • Bleibt auch in schwierigen Situationen seiner Linie treu • Reflektiert verantwortungsbewusst die ganze Bandbreite seines Standpunktes	☐	☐	☐	☐	☐
Leadership	• Übernimmt im richtigen Moment auch mal die Führungsrolle in der Diskussion • Integriert schwache Teammitglieder/bietet Hilfe an	☐	☐	☐	☐	☐
Zielorientierung	• Achtet auf die Zeiteinteilung aller Mitglieder • Arbeitet pragmatisch/ ergebnisorientiert • Bringt die Gruppe nach vorne	☐	☐	☐	☐	☐
Kooperation	• Geht Kompromisse ein • Denkt für die anderen mit • Achtet auf gerechte Verteilung	☐	☐	☐	☐	☐
Durchsetzungskraft/ Führungsfähigkeit	• Selbstbewusstsein, natürliche Autorität • Bekommt Anerkennung der Teammitglieder	☐	☐	☐	☐	☐
Konflikt- und Kritikfähigkeit	• Moderiert bei Konflikten • Denkt teamorientiert	☐	☐	☐	☐	☐
Kommunikationsstärke/ Eloquenz	• Aussagenklarheit, Flüssigkeit im Sprechen • Schlüssigkeit der Argumentation, Überzeugungskraft • Formulierung, Redegewandtheit	☐	☐	☐	☐	☐
Fachkenntnisse	• Kenntnisse zur Problemlösung • Inhaltliche Qualität des Inputs	☐	☐	☐	☐	☐
Strukturiertes/logisch-analytisches Denken	• Strukturiert komplexe Sachverhalte und trägt so zum Verständnis bei • Nimmt Argumentation auf und nutzt sie für seine Argumentation • Analytische Fähigkeiten	☐	☐	☐	☐	☐
Intellektuelle Neugierde	• Ist an Position anderer interessiert und verarbeitet diese • Flexibel und kreativ in der Denkweise	☐	☐	☐	☐	☐

Entscheidung für Einstellung: ☐ **ja** ☐ **nein**

Bemerkungen:

Erläuterungen:

1 Die Leistung entsprach nicht den Erwartungen und Anforderungen
2 Die Erwartungen und Anforderungen wurden teilweise nicht ganz erfüllt
3 Die Leistung entsprach stets den Erwartungen und Anforderungen
4 Die Leistung entsprach stets den Erwartungen und Anforderungen und hat sie häufig übertroffen
5 Die Leistung übertraf stets die E. und A.

Bewerberbeurteilung – Fallpräsentation

Bewerber: _____ Datum: _____

Interviewer: _____ Ort: _____

		nicht entwickelt				*hoch entwickelt*
		1	**2**	**3**	**4**	**5**
Kritikfähigkeit	• Umgang mit Kritik an Falllösung • Emotionale Stabilität	☐	☐	☐	☐	☐
Ausstrahlung, Charisma	• Auftreten/persönliche Reife • Welchen Eindruck würde der Bewerber in der Präsentation auf Klienten machen	☐	☐	☐	☐	☐
Kommunikationsverhalten	• Eloquenz • Überzeugungskraft • Vortragsvorbereitung/-darstellung • Formulierung/Klarheit im Ausdruck	☐	☐	☐	☐	☐
Strukturierungsfähigkeit	• Problemerfassung/-strukturierung und -lösung • Analytische/logische Fähigkeiten	☐	☐	☐	☐	☐
Intellektuelle Neugier	• Interesse an fachlichem Neuland • Geistige Flexibilität • Ideenfluß, innovative Fähigkeiten	☐	☐	☐	☐	☐
Methodenkenntnisse	• Inhaltliche Qualität des Falles • Kenntnisse zur Problemlösung	☐	☐	☐	☐	☐
Fachkenntnisse	• Inhaltliche Qualität des Falles • Kenntnisse zur Problemlösung	☐	☐	☐	☐	☐

Entscheidung für Einstellung: ☐ **ja** ☐ **nein**

Bemerkungen:

Erläuterungen:

1 Die Leistung entsprach nicht den Erwartungen und Anforderungen
2 Die Erwartungen und Anforderungen wurden teilweise nicht ganz erfüllt
3 Die Leistung entsprach stets den Erwartungen und Anforderungen
4 Die Leistung entsprach stets den Erwartungen und Anforderungen und hat sie häufig übertroffen
5 Die Leistung übertraf stets die E. und A.

2. Aggregationssystematik über die verwendeten Auswahlinstrumente

Aggregierte Dimension	**Fachinterview**	**Persönlichkeits-interview**	**Fallvorbereitung**	**Fallpräsentation**
Team-/Integrations-fähigkeit	Team-/Integrations-fähigkeit	Team-/Integrations-fähigkeit	Kooperation	-
Konflikt- und Kritikfähigkeit	-	Konflikt- und Kritikfähigkeit	Konflikt- und Kritikfähigkeit	Kritikfähigkeit
Kommunikations-stärke/Eloquenz	Kommunikations-stärke/Eloquenz	Kommunikations-stärke/Eloquenz	Kommunikations-stärke/Eloquenz	Kommunikations-verhalten
Ausstrahlung/ Charisma	-	Ausstrahlung/ Charisma	-	Ausstrahlung/ Charisma
VWL-Kenntnisse	Fachkenntnisse VWL	-	-	-
Allgemeinwissen	Allgemeine Kenntnisse	-	-	-
Pragmatismus/ Ergebnisorientierung	Pragmatismus/ Ergebnisorientierung	Pragmatismus/ Ergebnisorientierung	Pragmatismus/ Ergebnisorientierung	-
Unternehmerisches Denken	Unternehmerisches Denken	-	-	-
Strukturiertes/ logisch-analytisches Denken	Strukturiertes/ logisch-analytisches Denken	-	Strukturiertes/ logisch-analytisches Denken	Strukturierungs-fähigkeit
Methoden-Kenntnisse	Methoden-Kenntnisse	-	-	Methoden-Kenntnisse
Kreativität/ Flexibilität	Kreativität/ Flexibilität	Kreativität/ Flexibilität	-	-
Fachkenntnisse	-	-	Fachkenntnisse	Fachkenntnisse
BWL-Kenntnisse	Fachkenntnisse BWL	-	-	-
Intellektuelle Neugierde	-	-	Intellektuelle Neugierde	Intellektuelle Neugier
Frustrationstoleranz	Frustrationstoleranz	Frustrationstoleranz	-	-
Englischkenntnisse	-	-	-	-
Durchsetzungskraft/ Führungsfähigkeit	-	Durchsetzungskraft/ Führungsfähigkeit	Durchsetzungskraft/ Führungsfähigkeit	-
Allgemeinbildung	-	-	-	-
Selbstorganisation, Engagement, Biss	-	Selbstorganisation, Engagement, Biss	Selbstorganisation, Engagement, Biss	-
Leadership	-	-	Leadership	-
Zielorientierung	-	-	Zielorientierung	-
Fit ISB-Team	Fit zu ISB	Fit zu ISB	-	-
Exotenwissen	Fachkenntnisse Exoten	-	-	-

Anhang VI: Inhalte der Projektbeurteilungsbögen

1. Junior Consultant und Consultant

Dimension	Einzelvariablen
Specialist Know-how	Is developing the necessary functional Know-how including tools and methods Is developing the necessary industry Know-how Is developing the necessary methodological Know-how Is developing the necessary EDP Know-how
Problem Solution	Searches in a targeted, creative and efficient way for the relevant data Is able to differentiate between essential and non-essential Develops key messages based on the analyses Makes a valuable contribution to the development of measures
Quality of Work	Produces high quality charts (and storylines) Ensures project language is of a high quality Clearly structures own work and data Efficiently and reliably completes the tasks assigned Adheres to schedule, incl. non-project work Works in a cost budget conscious way Documents own work in a structured. comprehensive and efficient way in line with company standards
Behaviour under Pressure	Remains calm and rational under pressure Shows consistent quality of work even when the volume of work is high Remains motivated. supportive and committed in difficult situations
Client Acceptance	Behaves in a suitable professional manner when in contact with the client Enjoys a good working relationship with the client Is accepted as a contact by the client top management Is accepted as a contact by the client's working team Develops an understanding of the client's problem and possible scope of solutions Respects the clients' staff
Client Communi-cation	Is self-assured when dealing with clients Holds successful workshops with prepared agenda Holds successful interviews with prepared agenda Efficiently evaluates the interview Expresses him/herself in an appropriate clear, comprehensible and precise way (written and spoken language)
Know-how-Management	Uses available knowledge within the company Records work results promptly and in a orderly manner and files them in a comprehensible and justifiable fashion Collects and records CC-related Know-how beyond the simple requirements of the project Forwards all necessary information to the team
Teamwork	Is cooperative when working with the team Is cooperative when working with Service Staff Plays an active role in finding new ideas Recognizes the areas in which he/she has to improve and works on improving his/her own abilities Accepts improvement proposals/criticism of team members Is respected by the team both on a personal and professional level

2. Senior Consultant und Project Manager

Dimension	Einzelvariablen
Specialist Know-how	Has the necessary functional Know-how including tools and methods Has the necessary industry Know-how Has the necessary methodological Know-how Has the necessary EDP Know-how
Problem Solution	Structures the problem in a logical and convincing way Compiles the necessary analyses Is able to deal with comprehensive and complex tasks Analyzes in a precise and systematic way Develops and structures the key messages/ recommendations in a logical and convincing way Structures the entire presentation in a logical and convincing way
Quality of Work	Adheres to the milestones when preparing documentation (schedule, storyline/storyboard, master, archiving) Provides high quality charts (and storylines) Ensures written language is of a high standard (expressions, grammar and spelling) Clearly structures own work and data Efficiently and reliably completes the tasks assigned Adheres to schedule, incl. non-project work Works in a cost budget conscious way Documents own work in a structured, comprehensive and efficient way
Behaviour under Pressure	Remains calm and rational under pressure Shows consistent quality of work even when the volume of work is high Remains motivated. supportive and committed in difficult situations
Client Acceptance	Enjoys acceptance by the client both on a personal and professional level (incl. top management) Solves conflicts with the client in a professional and constructive way Takes into account client problems as well as the clients perception of the possible scope of solution Respects the clients staff
Client Communication	Expresses him/herself in an appropriate clear, comprehensible and precise way (written and spoken language) Actively seeks to build up long term customer relationship and achieves client loyalty Communicates the key messages/ recommendations in a convincing and practical way
Know-how-Management	Uses available implicit knowledge Initiates and supervises prompt documentation; creates the freedom to do this Transfers and records practical experience gained during the project Promotes the learning effect of consultants Ensures an efficient flow of information within the team
Teamwork	Ensures there is a cooperative atmosphere within the team Motivates the team members and helps the to identify with the task at hand Always shows a high level of commitment Actively occupies him/herself with the improvement of his/her weaknesses Is respected by the team both on a personal and professional level
Project-management[a]	Clearly communicates the objectives and allocates tasks systematically within the team Develops and supervises work plan and coordinates schedules Plans staff-assignment in a target-oriented, efficient and realistic way Plans budget assignment in a targeted and efficient way Communicates saving potentials within the budget to the team (fees and incidentals) Recognizes conflicts and crises in time and supports the team in overcoming them Encourages consultants to develop creative ideas Gives fair and constructive feedback - not only at the end of the project Helps other staff to recognize their own weaknesses and improve on them Provides fair, differentiated and justified evaluation

[a] Die Dimension Projektmanagement wurde wegen zu kleiner Fallzahl (N=26) aus der Betrachtung ausgeschlossen

Anhang VII: Variablenbezeichnung und Skalierung der Beurteilungsdimensionen

Variablenname[a]	Dimension	Skalierung
uSKH_MEAN	Specialist Know-how	1 bis 5
uPS_MEAN	Problem Solution	1 bis 5
uQW_MEAN	Quality of Work	1 bis 5
uBUP_MEAN	Behaviour under Pressure	1 bis 5
uCA_MEAN	Client Acceptance	1 bis 5
uCLIECO_MEAN	Client Communication	1 bis 5
uKHM_MEAN	Know-how-Management	1 bis 5
uTW_MEAN	Teamwork	1 bis 5
uMeanEvaltot	Ungewichteter Durchschnittsscore über alle Variablen	1 bis 5

[a] Das Kürzel *u* vor dem Variablennamen kennzeichnet, dass es sich um die ungewichtete Variante handelt
Die ab dem SC-Level zusätzlich eingeführte Dimension „Projektmanagement“ wurde wegen zu kleiner Fallzahlen (N=26) aus der Betrachtung ausgeschlossen

Anhang VIII: Detaillierung der Ergebnisse zur JC-Kohorte

1. Korrelationen Projektbeurteilungsdimensionen mit Recruitingvariablen

Korrelationen (Spearman-Rho)

	N	uSKH_ MEAN	uPS_ MEAN.	uQW_ MEAN	uBUP_ MEAN	uCA_ MEAN	uCLIECO _MEAN	uKHM_ MEAN	uTW_ MEAN	uMean Evaltot
Pragma	59	.082	.167	.175	.355**	-.010	-.100	.158	.167	.143
SelOrg	35	.115	-.036	.161	.022	.189	.114	.402**	.318*	.162
Leader	25	.030	.221	.102	.002	.183	.048	.110	.345*	.097
Zielor	25	-.157	-.138	-.181	-.132	-.132	-.419*	-.331	-.049	-.306
Team	59	.112	.050	.264*	.122	-.037	-.133	.245*	.377*	.191
Fuehr	35	.050	.167	.128	.229	.153	.002	.114	.370*	.128
Konfl	59	-.105	.057	.078	-.007	-.067	-.127	-.099	.080	-.005
Komm	59	.148	.094	.286*	.295*	.131	-.003	.260*	.278*	.248*
Fach	58	.252*	.170	.079	.183	-.088	.051	.053	.070	.101
LogAn	59	.157	.137	.187	.200	-.127	-.068	.101	.161	.134
Intell	59	.290*	.151	.302*	.096	.094	.088	.205	.349*	.237*
Charis	59	.183	.233*	.354**	.298*	.164	.057	.279*	.329**	.335**
Frust	59	-.051	-.073	.119	.136	-.112	.050	-.059	.190	.011
Kreat	59	.097	-.090	.175	.140	-.119	-.005	.154	.196	.116
Meth	59	.161	.111	.244*	.290*	-.068	.081	.129	.201	.204
Fit_ISB	28	-.140	-.058	-.091	.079	-.239	-.163	-.070	-.036	-.184
Unt	58	.079	.026	.157	.367**	.017	.144	.176	.126	.197
Exot	19	.342	.132	.040	.030	.308	.173	.021	.145	.246
BWL	58	.243*	.129	.248*	.305**	-.012	.004	.043	.237*	.213
VWL	57	.313*	-.029	.292*	.206	.243*	.145	.338**	.244*	.305*
Allg	55	.178	-.065	.150	.176	-.066	.054	.201	.211	.122
Engl	42	.239	.199	.111	.276*	.108	.200	.078	.168	.215
AllBil	28	.245	.099	.349*	.341*	-.060	.115	.173	.150	.274
MeanRecr[a]	59	.235*	.093	.300*	.410**	.008	.035	.244*	.299*	.276*
Abinote[a]	59	-.417**	-.309**	-.331**	-.297*	-.266*	-.191	-.204	-.347**	-.381**

[a] Korrelationskoeffizient nach Pearson
* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).
** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).

2. Einfache Regressionen mit Abiturnote und Gesamtwert aus Recruiting (N=52)

Prädiktoren	**Abinote**			**Gesamtwert Recruiting**		
Kriterien	R^2	R^2_{korr}	β-Gewicht	R^2	R^2_{korr}	β-Gewicht
Durchschnittlicher Gesamtwert	.162	.145**	-.402**	.077	.058**	.277**
Specialist-Know-how	.156	.139**	-.395**	.053	.034	.230
Problem Solution	.118	.101*	-.344*	.013	-.007	.112
Quality of Work	.117	.100*	-.342*	.096	.078*	.310*
Behaviour under Pressure	.092	.074*	-.303*	.188	.172**	.434**
Client Acceptance	.085	.067*	-.292*	.000	-.020	.004
Client Communication	.057	.038	-.239	.001	-.019	.036
Know-how-Management	.044	.025	-.209	.056	.037	.236
Teamwork	.089	.071*	-.298*	.071	.053	.267

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} :Schrumpfungskorrigierter Determinationskoeffizient

3. Hierarchische Regressionen mit Abiturnote und Gesamtwert aus Recruiting (N=52)

Prädiktoren	**Abinote**	**Abinote und Gesamtwert Recruiting[a]**				
Kriterien	R^2_{korr}	R^2	R^2_{korr}	R^2_{inkr}	$\beta_{Abinote}$	$\beta_{RecrutingGW}$
Durchschnittlicher Gesamtwert	.145**	.162	.145**	-.-	-.402**	-.-
Specialist-Know-how	.139**	.156	.139**	-.-	-.395**	-.-
Problem Solution	.101*	.118	.101*	-.-	-.344*	-.-
Quality of Work	.100*	.201	.168**	.068*	-.324*	.289*
Behaviour under Pressure	.074*	.265	.235**	.161**	-.276	.417**
Client Acceptance	.067*	.085	.067*	-.-	-.292*	-.-
Client Communication	.038	-.-	-.-	-.-	-.-	-.-
Know-how-Management	.025	-.-	-.-	-.-	-.-	-.-
Teamwork	.071*	.089	.071*	-.-	-.298*	-.-

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} : Schrumpfungskorrigierter Determinationskoeffizient
R^2_{inkr} : Inkrementelle Varianzaufklärung
[a] Es sind lediglich die Modelle dargestellt, die auf Gesamtmodell- und Koeffizientenebene signifikant sind

4. Explorative schrittweise Regression mit Recruitingvariablen (N=52)

Prädiktoren	**Recrutingvariablen**[a]					**β-Gewichte**[a]	
Kriterien	Modell	R^2	R^2_{korr}	R^2_{inkr}	aufgenommene Variablen	Modell 1	Modell 2
Durchschnittlicher Gesamtwert	1	.096	.078*	-.-	Charis	.309*	-.-
Specialist-Know-how	1	.077	.059*	-.-	Intell	.278*	-.-
Problem Solution	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Quality of Work	1	.132	.115**	-.-	Charis	.364**	
Behaviour under Pressure	1	.181	.165**	-.-	Unt	.425**	.348**
	2	.279	.249**	.084*	+Komm	-.-	.322*
Client Acceptance	1	.181	.165**	-.-	Unt	.425**	.348**
Client Communication	1	.279	.249**	.084*	+Komm	-.-	.322*
Know-how-Management	1	.079	.060*	-,-	VWL	.281*	-.-
Teamwork	1	-.-	-.-	-.-	-.-	-.-	-.-

***** Die Koeffizienten sind auf dem 0.05 Niveau signifikant
****** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} : Schrumpfungskorrigierter Determinationskoeffizient
R^2_{inkr} : Inkrementelle Varianzaufklärung
[a] Es sind lediglich die Modelle dargestellt, die auf Gesamtmodell- und Koeffizientenebene signifikant sind

5. Explorative Schrittweise Regression mit Recruitingvariablen und Abiturnote (N=52)

Prädiktoren	**Recruitingvariablen und Abiturnote[a]**					**β-Gewichte[a]**			
Kriterien	Modell	R^2	R^2_{korr}	R^2_{inkr}	aufgenommene Variablen	Modell 1	Modell 2	Modell 3	Modell 4
Durchschn. Gesamtwert	1	.162	.145**	-.-	Abinote	-.402**	-.428**	-.448**	-.448**
	2	.278	.248**	.103**	+Charis	-.-	.341**	.491**	.494**
	3	.346	.305**	.057**	+Konfl	-.-	-.-	-.302**	-.368**
	4	.401	.350**	.045*	+BWL	-.-	-.-	-.-	.242*
Specialist-Know-how	1	.156	.139**	-.-	Abinote	-.395**	-.412**	-.-	-.-
	2	.233	.201**	.062*	+Fach	-.-	.278*	-.-	-.-
Problem Solution	1	.188	.101*	-.-	Abinote	-.344*	-.364**	-.-	-.-
	2	.190	.157*	.056*	+Charis	-.-	.268*	-.-	-.-
Quality of Work	1	.117	.100	-.-	Abinote	-.342*	-.371**	-.-	-.-
	2	.269	.239	.139**	+Charis	-.-	.391**	-.-	-.-
Behaviour under Pressure	1	.092	.074*	-.-	Abinote	-.303*	-.324**	-.319**	-.376**
	2	.285	.256**	.182**	+Unt	-.-	.441**	.364**	.367
	3	.380	.341**	.085**	+Komm	-.-	-.-	.316**	.490**
	4	.437	.389**	.048*	+Intell	-.-	-.-	-.-	-.301*
Client Acceptance	1	.085	.067*	-.-	Abinote	-.292*	-.286*	-.335*	-.-
	2	.161	.126*	.059*	+VWL	-.-	-.275*	.382**	-.-
	3	.235	.187**	.061*	+LogAn	-.-	-.-	-.297*	-.-
Client Communication	-.-	-.-	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Know-how-Management	-.-	-.-	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Teamwork	1	.089	.071*	-.-	Abinote	-.298*	-.321*	-.-	-.-
	2	.181	.148**		+Charis	-.-	.304**	-.-	-.-

***** Die Koeffizienten sind auf dem 0.05 Niveau signifikant
****** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} : Schrumpfungskorrigierter Determinationskoeffizient
R^2_{inkr} : Inkrementelle Varianzaufklärung
[a] Es sind lediglich die Modelle dargestellt, die auf Gesamtmodell- und Koeffizientenebene signifikant sind

6. Explorative Schrittweise Regression mit Setvariablen und Abiturnote (N=52)[a]

Prädiktoren	**Recruitingvariablen und Abiturnote**				
Kriterien	Modell	R^2	R^2_{korr}	R^2_{inkr}	aufgenommene Variablen
Durchschn. Gesamtwert	1	.162	.145**	-.-	Abinote
	2	.365	.296**	.151	+Soziale Fähigkeiten
	3	.422	.330**	.034	+Spezialwissen
Specialist-Know-how	1	.156	.139**	-.-	Abinote
	2	.265	.219**	.080	+Spezialwissen
	3	.306	.213**	-.006	+Prozessuales Wissen
	4	.494	.371**	.158	+Soziale Fähigkeiten
Problem Solution	1	.188	.101*	-.-	Abinote
	2	.192	.123*	.022	+Prozessuales Wissen
Quality of Work	1	.117	.100*	-.-	Abinote
	2	.169	.098*	-.002	+Soziale Fähigkeiten
	3	.319	.192*	.094	+Spezialwissen +Fokussierung
Behaviour under Pressure	1	.092	.074*	-.-	Abinote
	2	.331	.258**	.184	+Soziale Fähigkeiten
	3	.426	.334**	.076	+Spezialwissen
	4	.472	.359**	.025	+Fokussierung
Client Acceptance	1	.085	.067*	-.-	Abinote
	2	.200	.150**	.083	+Allgemeinwissen
Client Communication	-.-	-.-	-.-	-.-	-.-
Know-how-Management	-.-	-.-	-.-	-.-	-.-
Teamwork	1	.089	.071*	-.-	Abinote
	2	.220	.135*	.064	+Soziale Fähigkeiten

***** Die Koeffizienten sind auf dem 0.05 Niveau signifikant
****** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} : Schrumpfungskorrigierter Determinationskoeffizient
R^2_{inkr} : Inkrementelle Varianzaufklärung
[a] Es sind lediglich die Werte signifikanter Blöcke dargestellt.

7. Explorative Schrittweise Regression mit extrahierten Recruitingfaktoren (N=54)

Prädiktoren	**Rotierte Recruitingfaktoren**		**Unrotierte Recruitingfaktoren**	
Kriterien	R^2_{korr}	aufgenommene Faktoren	R^2_{korr}	aufgenommene Faktoren
Levelfaktor Projektbeurteilung	-.-	-.-	.059*	Levelfaktor Recruiting
Profilfaktor Projektbeurteilung	-.-	-.-	.079*	Levelfaktor Recruiting
F1 rotiert Projektbeurteilung	-.-	-.-	-.-	-.-
F2 rotiert Projektbeurteilung	.077*	Sozialkompetenz	.154**	Levelfaktor Recruiting
Durchschnittlicher Gesamtwert	-.-	-.-	.061*	Levelfaktor Recruiting
Specialist-Know-how	-.-	-.-	-.-	-.-
Problem Solution	-.-	-.-	-.-	-.-
Quality of Work	.063*	Sozialkompetenz	.094*	Levelfaktor Recruiting
Behaviour under Pressure	.079*	Allgemeinfaktor	.156*	Levelfaktor Recruiting
Client Acceptance	-.-	-.-	-.-	-.-
Client Communication	-.-	-.-	-.-	-.-
Know-how-Management	-.-	-.-	-.-	-.-
Teamwork	.079*	Sozialkompetenz	.053*	Profilfaktor Recruiting

***** Die Koeffizienten sind auf dem 0.05 Niveau signifikant
****** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} : Schrumpfungskorrigierter Determinationskoeffizient

Anhang IX: Detaillierung der Ergebnisse zur C-Kohorte

1. Korrelative Befunde Projektbeurteilungsdimensionen mit Recruitingvariablen

Korrelationen (Spearman-Rho)

	N	uSKH_ MEAN	uPS_ MEAN.	uQW_ MEAN	uBUP_ MEAN	uCA_ MEAN	uCLIECO _MEAN	uKHM_ MEAN	uTW_ MEAN	uMean Evaltot
Pragma	70	.103	.213*	.320**	.062	.166	.263*	.097	.214*	.243*
SelOrg	33	.075	.091	.031	.068	.010	.189	.168	-.108	.083
Leader	13	-.235	-.063	.134	-.215	.384	-.141	-.340	-.216	-.065
Zielor	13	-.156	.146	.352	-.077	.175	-.214	-.407	-.161	-.009
Team	70	.073	.136	.217*	.180	.025	.216*	.110	.318**	.206*
Fuehr	33	-.284	-.073	.077	-.204	.179	.139	.019	-.287	-.052
Konfl	70	-.278**	-.310**	.009	-.147	.087	-.052	-.206*	-.205*	-.156
Komm	70	.060	.048	.342**	.333**	.262*	.338**	.054	.180	.304**
Fach	70	-.084	.095	-.034	-.091	.152	-.042	.052	.006	.024
LogAn	70	.136	.303**	.278*	.082	.221*	.238*	.099	.225*	.284**
Intell	70	.008	.203*	.293**	.055	.260*	.296**	.146	.203*	.266*
Charis	70	.043	.046	.343**	.146	.103	.214*	-.008	.029	.172
Frust	70	.154	.129	.254*	.280**	.123	.208*	.173	.205*	.229
Kreat	70	.111	.102	.104	.072	.108	.238*	.126	.113	.142
Meth	70	-.011	.118	.061	.041	.191	.228*	.106	.155	.166
Fit_ISB	14	.342	.099	-.104	.193	.044	.010	-.430	-.119	.069
Unt	69	.082	.071	.276*	.050	.218*	.348**	.130	.081	.224*
Exot	10	-.442	-.094	.354	-.224	.186	-.006	-.081	-.313	-.025
BWL	69	.068	.097	.326**	.066	-.014	.281**	.250*	.049	.209*
VWL	67	-.058	-.015	.147	.091	.144	.238*	.140	.209*	.188
Allg	68	-.035	-.036	.213*	.050	.111	.130	.083	.186	.123
Engl	48	-.117	.070	.051	-.187	.093	.108	.115	-.003	.049
AllBil	52	.187	.180	.113	.068	.121	.300*	.051	.304*	.245*
MeanRecr[a]	70	.070	.222*	.353**	.097	.260*	.355**	.227*	.261*	.341**
Abinote[a]	71	-.054	-.067	.030	.147	.082	.118	.095	.142	.089

[a] Korrelationskoeffizient nach Pearson
* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).
** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).

2. Gegenüberstellung Regressionen mit Abiturnote und Gesamtwert aus Recruiting (N=67)

Prädiktoren	Abinote			Gesamtwert Recruiting		
Kriterien	R^2	R^2_{korr}	β-Gewicht	R^2	R^2_{korr}	β-Gewicht
Durchschnittlicher Gesamtwert	.000	-.015	-.005	.121	.107**	.347**
Specialist-Know-how	.009	-.006	-.095	.005	-.010	.070
Problem Solution	.008	-.007	-.090	-.045	.030	.211
Quality of Work	.004	-.011	-.063	.156	.143**	.395**
Behaviour under Pressure	.017	.002	.131	.010	-.006	.098
Client Acceptance	.000	-.015	-.001	.080	.066*	.283*
Client Communication	.001	-.015	.028	.141	.128**	.376**
Know-how-Management	.000	-.015	-.009	.053	.039	.231
Teamwork	.002	-.013	.050	.074	.060*	.272*

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} :Schrumpfungskorrigierter Determinationskoeffizient

3. Schrittweise Regression mit inhaltlich-sachlogisch abgeleiteten Recruitingvariablen (N=67)[a]

Kriterien	Inhaltlich zugeordnete Variablen	R^2	R^2_{korr}	Aufgenommene Variablen
Durchschnittlicher Gesamtwert	Alle	.364	.334**	LogAn, Konfl, Komm
Specialist-Know-how	Fach, Intell, Meth, BWL,VWL	-.-	-.-	
Problem Solution	Pragma, LogAn, Fach, Intell, Frust, Kreat, Meth, Unt, BWL	.160	.147**	LogAn
Quality of Work	Pragma, Fach, LogAn, Kreat, Meth, Unt, BWL, VWL	.148	.135**	Pragma
Behaviour under Pressure	Pragma, Team, Konfl, Komm, Frust, Kreat, Unt	.189	.163**	Komm, Konfl
Client Acceptance	Alle	.093	.079**	Intell
Client Communication	Alle	.121	.108**	Unt
Know-how-Management	Fach, Meth, BWL, Pragma, Team	.088	.074*	BWL
Teamwork	Team, Konfl, Komm, Charis, LogAn, Kreat	.272	.238**	Team, Konfl, LogAn

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} :Schrumpfungskorrigierter Determinationskoeffizient
[a] Es sind lediglich die Modelle dargestellt, die auf Gesamtmodell- und Koeffizientenebene signifikant sind

4. Schrittweise Regression mit inhaltlich-sachlogisch abgeleiteten Recruitingvariablen(N=67)[a]

Kriterien	**Recruitingvariablen**					**ß-Gewichte**		
	Modell	R^2	R^2_{korr}	R^2_{inkr}	Aufgenommene Variablen	Modell 1	Modell 2	Modell 3
Durchschnittlicher Gesamtwert	1	.150	.137**	-.-	LogAn	.388**	.482**	.374**
	2	.236	.212**	.075	+ Konfl.		-.307**	-.427**
	3	.364	.334**	.122	+ Komm			.403**
Specialist-Know-how	-.-	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Problem Solution	1	.160	.147**	-.-	LogAn	.400**	-.-	-.-
Quality of Work	1	.148	.135**	-.-	Pragma	.385**	-.-	-.-
Behaviour under Pressure	1	.102	.089**	-.-	Komm	.320**	.440**	-.-
	2	.189	.163**	.074	+ Konfl.		-.318*	-.-
Client Acceptance	1	.093	.079**	-.-	Intell	.306*	-.-	-.-
Client Communication	1	.121	.108**	-.-	Unt	.349*	-.-	-.-
Know-how-Management	1	.088	.074*	-.-	BWL	.296*	-.-	-.-
Teamwork	1	.125	.111**		Team	.353**	.398**	.319**
	2	.188	.162**	.051	+ Konfl		-.255*	.338**
	3	.272	.238**	.076	+ LogAn			.316**

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} :Schrumpfungskorrigierter Determinationskoeffizient
[a] Es sind lediglich die Modelle dargestellt, die auf Gesamtmodell- und Koeffizientenebene signifikant sind

5. Schrittweise Regression mit allen Recruitingvariablen (N=67)[a]

Kriterien	Recruitingvariablen					ß-Gewichte			
	Modell	R^2	R^2_{korr}	R^2_{inkr}	Aufgenommene Variablen	Modell 1	Modell 2	Modell 3	Modell 4
Durchschnittlicher Gesamtwert	1	.150	.137**	-.-	LogAn	.388**	.482**	.374**	-.-
	2	.236	.212**	.075	+ Konfl		-.307**	-.427**	-.-
	3	.364	.334**	.122	+ Komm			.403**	-.-
Specialist-Know-how	1	.064	.049**	-.-	LogAn	.252**	.353**	.535**	.510**
	2	.162	.135**	.086	+ Konfl.		-.329**	-.320**	-.402**
	3	.217	.179**	.044	+ Meth			-.299**	-.383**
	4	.277	.230**	.051	+ Komm				.286*
Problem Solution	1	.160	.147**	-.-	LogAn	.400**	-.-	-.-	
	2	.282	.260**	.113	+ Konfl				
	3	.325	.293**	.033	+ Charis				
Quality of Work	1	.162	.149**	-.-	Komm	.402**	.304*	-.-	-.-
	2	.227	.203**	.054	+ LogAn		.274*	-.-	-.-
Behaviour under Pressure	1	.102	.089**	-.-	Komm	.312**	.440**	.595**	-.-
	2	.189	.163**	.075	+ Konfl		-.318*	-.352**	-.-
	3	.242	.206**	.043	+ Allg			-.273*	-.-
Client Acceptance	1	.093	.079**	-.-	Intell	.306*	-.-	-.-	-.-
Client Communication	1	.121	.108**	-.-	Unt	.349*	-.-	-.-	-.-
Know-how-Management	1	.088	.074**	-.-	BWL	.296*	-.-	-.-	-.-
Teamwork	1	.125	.111**	-.-	Team	.353**	.398**	.319**	-.-
	2	.188	.162**	.051	+ Konfl		-.255*	-.338**	-.-
	3	.272	.238**	.075	+ LogAn			.316**	-.-

***** Die Koeffizienten sind auf dem 0.05 Niveau signifikant
****** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} :Schrumpfungskorrigierter Determinationskoeffizient
[a] Es sind lediglich die Modelle dargestellt, die auf Gesamtmodell- und Koeffizientenebene signifikant sind

6. Korrelationen extrahierter Projektbeurteilungsfaktoren mit Recruitingvariablen

Korrelationen (Spearman-Rho)

		Unrotierte Faktoren		Rotierte Faktoren	
	N	Levelfaktor	Profilfaktor	Allgemeinfaktor 1	Allgemeinfaktor 2
Pragma	55	.284*	.100	.316*	.128
Team	55	.282*	-.107	.199	.252*
Konfl	55	-.148	.023	-.113	-.191
Komm	55	.341**	-.105	.239*	.254*
Fach	55	.065	.061	.151	-.068
LogAn	55	.320**	-.057	.237*	.242*
Intell	55	.359**	.181	.423**	.053
Charis	55	.193	.029	.235*	.098
Frust	55	.177	-.030	.151	.208
Kreat	55	.184	.067	.198	.091
Meth	55	.190	-.060	.133	.125
Unt	55	.271*	.148	.348*	.027
BWL	55	.233*	.196	.343**	.052
Allg	55	.110	.048	.111	.025
VWL	55	.166	-.019	.147	.105
MeanRecr.		.358**	.030	.350**	.186

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).
** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).

7. Korrelationen extrahierter Projektbeurteilungsfaktoren mit Recruitingfaktoren

Korrelationen (Spearman-Rho)

				Projektbeurteilungsfaktoren			
				Unrotierte Faktoren		Rotierte Faktoren	
Recruitingfaktoren			N	Levelfaktor	Profilfaktor	Allgemein-faktor 1	Allgemein-faktor 2
Unrotierte Faktorlösung	Einstieg als JC	Allgemeinfaktor	44	.346*	-.045	.247	.203
		Profilfaktor	44	-.088	-.069	-.071	-.002
Rotierte Faktorlösung	Einstieg als JC	Allgemeinfaktor	44	.382**	-.046	.242	.253*
		Sozialkompetenz	44	.172	-.079	.109	.113
		Problemlösungsfähigkeit	44	.000	.191	.163	-.225

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).
** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).

8. Schrittweise Regression extrahierter Faktoren C-Kohorte (Einstieg als JC) (N=44)

Prädiktoren	**Rotierte Recruitingfaktoren**		**Unrotierte Recruitingfaktoren**	
Kriterien	R^2_{korr}	aufgenommene Faktoren	R^2_{korr}	aufgenommene Faktoren
Levelfaktor Projektbeurteilung	.101*	Allgemeinfaktor	.129	Allgemeinfaktor
Profilfaktor Projektbeurteilung	-.-	-,-	-,-	-,-
Allgemeinfaktor 1 rotiert	-,-	-,-	.105*	Allgemeinfaktor
Allgemeinfaktor 2 rotiert	-,-	-,-	-,-	-,-
Durchschnittlicher Gesamtwert	.105*	Allgemeinfaktor	.134**	Allgemeinfaktor
Specialist-Know-how	-,-	-,-	-,-	-,-
Problem Solution	-,-	-,-	-,-	-,-
Quality of Work	.074*	Allgemeinfaktor	.175**	Allgemeinfaktor
Behaviour under Pressure	-,-	-,-	-,-	-,-
Client Acceptance	-,-	-,-	.123**	Allgemeinfaktor
Client Communication	-,-	-,-	-,-	-,-
Know-how-Management	-,-	-,-	-,-	-,-
Teamwork	.089*	Allgemeinfaktor	-,-	-,-

***** Die Koeffizienten sind auf dem 0.05 Niveau signifikant
****** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} : Schrumpfungskorrigierter Determinationskoeffizient

Anhang X: Detaillierung der Ergebnisse zur SC-Kohorte

1. Korrelative Befunde Projektbeurteilungsdimensionen mit Recruitingvariablen

Korrelationen (Spearman-Rho)

	N	uSKH_MEAN	uPS_MEAN.	uQW_MEAN	uBUP_MEAN	uCA_MEAN	uCLIECO_MEAN	uKHM_MEAN	uTW_MEAN	uMean Evaltot
Pragma	78	.020	.185	-.059	.012	.172	.055	-.227*	-.424**	-.057
Team	78	.006	.067	.006	.177	.139	.059	.180	.048	.177
Konfl	79	-.062	.132	.013	.027	.078	.119	-.175	-.010	.027
Komm	79	-.039	.141	.104	-.160	.210*	.155	-.046	-.068	.035
Fach	78	.015	.121	.197**	-.229*	-.043	-.113	.038	.066	-.044
LogAn	79	.089	.107	.126	-.010	.034	-.047	-.123	-.095	.019
Intell	79	-.042	.155	.071	-.056	.067	-.012	.018	.060	.064
Charis	79	-.043	.142	-.015	.016	.223*	.195*	.048	-.114	.079
Frust	78	-.071	-.060	-.152	-.083	.073	-.044	-.011	-.204*	-.093
Kreat	78	.133	.265**	.015	.030	.120	.070	-.059	-.056	.100
Meth	79	.114	.176	.178	-.211*	-.055	.075	-.041	-.038	.040
Unt	77	-.009	.157	.020	-.214*	.047	.120	.060	-.162	.005
BWL	77	.065	.235*	.127	-.044	.001	.014	-.037	-.094	.066
VWL	73	-.052	.098	.063	-.108	.010	.135	.089	-.132	.031
Allg	76	-.150	-.100	-.092	.003	-.081	.016	.112	-.144	-.101
Exot	7	.158	.239	.000	.492	.525	-.249	-.316	.239	.316
Engl	55	.056	.170	-.090	-.271*	-.044	.020	-.111	-.072	-.059
AllBil	63	-.073	-.097	-.099	-.127	.001	-.027	-.074	-.157	-.111
Fit_ISB	7	.316	-.080	-.656	.239	-.420	.161	.316	.319	.474
Fuehr	33	-.031	.028	.077	-.076	.091	-.184	-.218	-.447**	-.188
SelOrg	33	-.044	.156	.005	-.229	.096	-.004	.021	-.247	.002
Leader	5	.211	-.081	-.295	-.211	.351	.189	-.369	.211	-.211
Zielor	5	-.447	.574	.750	.447	.344	-.344	-.671	-.447	-.224
MeanRecr[a]	78	.088	.228*	.065	-.109[b]	.173[c]	.102	-.053	-.133	0.081[d]
Abinote[a]	78	-.189*	-.024	-.078	-.037	.117[c]	.011	-.008	-.012	-0.027

Alle dargestellten Werte sind Zusammenhänge nach Eliminierung von Ausreißern
[a] Korrelationskoeffizient nach Pearson; [b] Spearman-Rho, da *Behaviour under Pressure* nicht normalverteilt; [c] N=77; [d] N=79
* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).
** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).

2. Schrittweise Regression mit Recruitingvariablen [a] (N=71)

	Recruitingvariablen					ß-Gewichte	
Kriterien	Modell	R^2	R^2_{korr}	R^2_{inkr}	Aufgenommene Variablen	Modell 1	Modell 2
Durchschnittlicher Gesamtwert	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Specialist-Know-how	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Problem Solution	1	.115	.102**	-.-	BWL	.339**	-.-
Quality of Work	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Behaviour under Pressure	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Client Acceptance	1	.057	.043*	-.-	Komm	.238*	-.-
Client Communi-cation	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Know-how-Management	-.-	-.-	-.-	-.-	-.-	-.-	-.-
Teamwork	1 2	.115 .186	.102** .162**	-.- .060	Pragma Intel	-.339**	-.469** .269*

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant
R^2_{korr} :Schrumpfungskorrigierter Determinationskoeffizient
[a] Es sind lediglich die Modelle dargestellt, die auf Gesamtmodell- und Koeffizientenebene signifikant sind

3. Korrelationsmatrix Recruitingfaktoren SC-Kohorte mit Projektbeurteilungsdimensionen (Spearman-Rho)

	Extrahierte rotierte Faktoren			Extrahierte unrotierte Faktoren[a]	
	Einstieg als JC (N=17)			Einstieg als JC (N=17)	
Kriterien	Allgemeinfaktor	Sozialkompetenz	Problemlösefähigkeit	Levelfaktor	Profilfaktor
Durchschn. Gesamtwert	551*	.130	.037	.517*	-.120
Specialist-Know-how	.314	.036	-.191	.102	-.050
Problem Solution	.286	-.058	.354	.352	-.163
Quality of Work	.348	.128	.042	.386	.023
Behaviour under Pressure	.367	.048	-.294	.174	-.114
Client Acceptance	.147	.369	.268	.554*	.280
Client Communication	.261	.402	-.066	.351	.267
Know-how-Management	.339	-.025	-.118	.020	-.189
Teamwork	.036	-.284	-.141	-.136	-.318

* Die Koeffizienten sind auf dem 0.05 Niveau signifikant (einseitig)
** Die Koeffizienten sind auf dem 0.01 Niveau signifikant (einseitig)
a Von den 3 extrahierten Komponenten sind die ersten 2 dargestellt.

4. Korrelationen extrahierter Projektbeurteilungsfaktoren mit Recruitingvariablen

Korrelationen (Spearman-Rho)							
		Unrotierte Faktoren			Rotierte Faktoren		
	N	1	2	3	1	2	3
Pragma	55	-.040	-.409**	-.136	.088	-.510**	.185
Team	55	.036	-.080	.038	-.001	-.077	.019
Konfl	56	.068	-.187	.010	.034	-.128	.146
Komm	56	.053	-.191	-.158	.124	-.219	.097
Fach	55	.033	.094	-.192	.168	.073	-.033
LogAn	56	.152	-.189	-.077	.110	-.112	.153
Intell	56	.017	.012	-.235*	.132	-.085	-.073
Charis	56	.037	-.208	.058	.031	-.143	.179
Frust	55	-.158	-.081	.056	-.154	-.192	-.008
Kreat	55	.091	-.205	-.180	.186	-.163	.119
Meth	56	.163	-.025	-.209	.223*	-.094	-.031
Unt	54	.051	-.052	-.110	.169	-.118	-.007
BWL	54	.232*	-.040	-.101	.215	.000	.048
VWL	52	.150	-.065	-.074	.157	-.083	.038
Allg	54	-.122	.045	.095	-.106	.051	.011
MeanRecr	78	.050	-.201*	-.223*	.184	-234*	.005*

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).
** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).

5. Korrelationsmatrix der Recruiting- und Projektbeurteilungsfaktoren

		Extrahierte rotierte Faktoren			Extrahierte unrotierte	
		Einstieg als JC (N=17)			Einstieg als JC (N=17)	
	Projekt-beurtei-lungsfak-toren	Allgemein-faktor	Sozial-kompetenz	Problemlö-sefähigkeit	Levelfaktor	Profilfaktor
Unrotierte Faktoren	1	.461*	.145	.108	.525*	-.066
	2	.088	-.265	-.252	-.292	-.292
	3	-.108	.044	-.228	-.132	.061
Rotierte Faktoren	1	.417*	.081	.184	.395	-.093
	2	.194	-.120	-.260	-.088	-.216
	3	.191	.282	.142	.451*	.174

Anhang XI: Detaillierung der Nutzenanalyse

1. Basisdaten der historischen Berechnung

	SDy			λ			Zx*		
Periode k	JC	C	SC	JC	C	SC	JC	C	SC
1[a]	76	89	111	.396	.396	.396	.875	.875	.875
2[a]	76	89	111	.398	.398	.398	.831	.831	.831
3[a]	76	89	111	.390	.390	.390	.694	.694	.694
4	78	89	111	.359	.359	.359	1.137	1.137	1.137
5[a]	74	90	112	.280	.280	.280	1.402	1.402	1.402
6[a]	76	89	111	.000	.000	.000	.000	.000	.000
7[a]	76	89	111	.000	.000	.000	.000	.000	.000

[a]Durchschnittswerte
Bezeichnungen gelten analog Gleichung (18)

2. Historische Nutzenberechnung nach Kohorten

A. Junior Consultants

Periode k	N_k	$1/(1+i)^k$	rxy	Zx*	SDy [EUR]	(1+v)	(1-s)	Nutzen U [EUR]
1	32	0.974	0.301	0.875	76 072	1.000	0.333	210 288
2	53	0.949	0.301	0.831	76 072	1.000	0.333	316 953
3	33	0.924	0.301	0.694	76 072	1.000	0.333	161 422
4	27	0.900	0.301	1.137	78 487	1.000	0.333	216 055
5	6	0.877	0.301	1.402	73 656	1.000	0.333	55 749
6	0	0.854	0.301	0.967	76 072	1.000	0.333	0
7	0	0.832	0.301	0.967	76 072	1.000	0.333	0

Bezeichnungen gelten analog Gleichung (18)

B. Consultants

Periode k	N_k	$1/(1+i)^k$	rxy	Zx*	SDy [EUR]	(1+v)	(1-s)	Nutzen U [EUR]
1	37	0.974	0.369	0.875	89 249	1.000	0.333	345 657
2	92	0.949	0.369	0.831	89 249	1.000	0.333	800 656
3	90	0.924	0.369	0.694	89 249	1.000	0.333	636 346
4	64	0.900	0.369	1.137	88 803	1.000	0.333	710 764
5	34	0.877	0.369	1.402	89 991	1.000	0.333	460 190
6	6	0.854	0.369	0.967	89 249	1.000	0.333	55 639
7	0	0.832	0.369	0.967	89 249	1.000	0.333	0

Bezeichnungen gelten analog Gleichung (18)

C. Senior Consultants

Periode k	N_k	$1/(1+i)^k$	rxy	Zx*	SDy [EUR]	(1+v)	(1-s)	Nutzen U [EUR]
1	31	0.974	0.369	0.875	111 078	1.000	0.333	358 501
2	118	0.949	0.369	0.831	111 078	1.000	0.333	1 269 646
3	242	0.924	0.369	0.694	111 078	1.000	0.333	2 119 532
4	258	0.900	0.369	1.137	110 814	1.000	0.333	3 592 158
5	164	0.877	0.369	1.402	111 606	1.000	0.333	2 767 192
6	96	0.854	0.369	0.967	111 078	1.000	0.333	1 081 405
7	19	0.832	0.369	0.967	111 078	1.000	0.333	208 769

3. Sensitivitätsanalyse Ergebnisbeitrag nach Selektionsquote und Validität ['000 EUR]

		Selektionsquote									
		0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
r_{xy}	**0.1**	-4 013	-671	9	79	-65	-313	-623	-989	-1 415	-2 005
	0.2	2 298	4 360	4 179	3 550	2 799	1 994	1 155	259	-723	-2 005
	0.3	8 609	9 392	8 349	7 020	5 663	4 302	2 933	1 506	-30	-2 005
	0.4	14 920	14 423	12 520	10 491	8 527	6 610	4 711	2 754	662	-2 005
	0.5	21 232	19 454	16 690	13 961	11 391	8 918	6 489	4 001	1 355	-2 005
	0.6	27 543	24 486	20 860	17 432	14 255	11 225	8 267	5 249	2 047	-2 005
	0.7	33 854	29 517	25 031	20 903	17 119	13 533	10 045	6 497	2 740	-2 005
	0.8	40 165	34 548	29 201	24 373	19 983	15 841	11 823	7 744	3 432	-2 005
	0.9	46 476	39 579	33 371	27 844	22 847	18 148	13 601	8 992	4 124	-2 005
	1.0	52 787	44 611	37 541	31 314	25 711	20 456	15 379	10 239	4 817	-2 005

4. Sensitivitätsanalyse Ergebnisbeitrag nach Selektionsquote und SDy ['000 EUR]

		SDy ['000 EUR]									
		10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
SQ	**0.1**	-8 018	-5 713	-3 407	-1 101	1 205	3 510	5 816	8 122	10 428	12 734
	0.2	-3 864	-2 026	-188	1 650	3 489	5 327	7 165	9 003	10 841	12 680
	0.3	-2 638	-1 115	409	1 933	3 456	4 980	6 504	8 027	9 551	11 074
	0.4	-2 124	-856	412	1 680	2 948	4 216	5 484	6 752	8 020	9 288
	0.5	-1 883	-837	210	1 256	2 303	3 349	4 395	5 442	6 488	7 535
	0.6	-1 778	-935	-92	751	1 594	2 438	3 281	4 124	4 967	5 810
	0.7	-1 751	-1 102	-452	197	847	1 497	2 146	2 796	3 445	4 095
	0.8	-1 780	-1 324	-869	-413	43	499	954	1 410	1 866	2 322
	0.9	-1 855	-1 602	-1 349	-1 096	-843	-590	-337	-84	169	422
	1.0	-2 005	-2 005	-2 005	-2 005	-2 005	-2 005	-2 005	-2 005	-2 005	-2 005

5. Inkrementeller Nutzen nach Einführung eines neuen Instrumentes ['000 EUR]

Instrument	Prognosegüte	Prognosegüte inkl. Intelligenztest	Nutzen alleine ['000 EUR]	Nutzenzuwachs alleine ['000 EUR]	Nutzen inkl. Intelligenztest ['000 EUR]	Nutzenzuwachs inkl. IT ['000 EUR]
Zufallsauswahl	0	0	0	-12 302	20 913	8 611
Allgemeine Intelligenztests	.51	.51	20 913	8 611	20 913	8 611
Assessment Center	.37	.53	12 830	528	20 875	8 572
Test zur Erfassung berufsrelevanten Wissens	.48	.55	19 380	7 078	22 896	10 594
Arbeitsproben	.54	.63	21 443	9 141	25 965	13 663
Interview strukturiert	.51	.63	19 933	7 631	25 965	13 663
Integritätstests	.41	.65	15 882	3 580	27 951	15 649

Literaturverzeichnis

Abramson, T. (1979). Issues and models in vocational education evaluation. In Abramson, T., Kehr Title, C. & Cohen, L. Handbook of vocational educational evaluation. Beverly Hills, CA: Sage.

Achter, J. A., Lubinski, D., Benbow, C. P., et al. (1999). Assessing vocational preferences among gifted adolescents adds incremental validity to abilities: A discriminant analysis of educational outcomes over a 10-year interval. Journal of Educational Psychology, 91, 777-786.

Ackerman, P. L. (1996). A Theory of Adult Intellectual Development: Process, Personality, Interests, and Knowledge. Intelligence, 22, 227-257.

Algera, J. A. & Greuter, A. M. (1998). Job Analysis. In Drenth, P. J. D., Thierry, H. & de Wolff, C. J. (Eds.), Handbook of Work and Organizational Psychology: Personnel Psychology, Vol. 3,, 2nd ed.

Amelang, M. & Bartussek, D. (1981). Differentielle Psychologie und Persönlichkeitsforschung. Stuttgart: Kohlhammer.

Amelang, M. & Zielinski, W. (1997). Psychologische Diagnostik und Intervention (2. korr., aktual. und überarb. Auflage). Berlin; Heidelberg; New York; Barcelona; Budapest; Hong Kong; London; Mailand; Paris; Santa Clara; Singapur; Tokio: Springer.

Amelang, M. & Zielinski, W. (2002). Psychologische Diagnostik und Intervention (3. korr., aktual. und überarb. Aufl.). Berlin: Springer.

Anastasi, A. (1988). Psychological testing (6th ed.). New York: McMillan.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review , 89(4), 369-406.

Andrews, J. D. W. (1967). The achievement motive and advancement in two types of organizations. Journal of Personality and Social Psychology, 6, 163-168.

Arnold, J. D., Rauschenberger, J. M., Soubel, W. G., et al. (1982). Validation and Utility of a Strength Test for Selecting Steelworkers. Journal of Applied Psychology, 67, 588-604.

Arthur, W., Jr., Woehr, D. J. & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical re-examination of the assessment center construct-related validity paradox. Journal of Management, 26, 813-835.

Arvey, R. D. & Campion, J. E. (1982). The employment interview: A summary and review of recent research. Personnel Psychology, 35, 281-322.

Arvey, R. D. & Murphy, K. R. (1998). Performance evaluation in work settings. Annual Review of Psychology, 49, 141-168.

Asendorpf, J. B. (1999). Psychologie der Persönlichkeit (2. Aufl.). Heidelberg: Springer.

Ashton, M. C. (1998). Personality and job performance: the importance of narrow traits. Journal of Organizational Behavior, 19, 189-303.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. Psychological Review, 64, 359-372.

Atkinson, J. W. (1978). Motivational determinants of intellective performance and cumulative achievement. In Atkinson, J. W. & Rynor, J. O. (Eds.), Personality, motivation, and achievement, 221-242. Washington: Hemisphere.

Backhaus, K., Erichson, B., Plinke, W., et al. (1996). Multivariate Analysemethoden - Eine anwendungsorientierte Einführung. Berlin: Springer.

Baehr, M. E. & Williams, G. B. (1967). Underlying dimensions of personal background data and their relation to occupational classification. Journal of Applied Psychology, 51, 481-490.

Baron-Boldt, J., Funke, U. & Schuler, H. (1989). Prognostische Validität von Schulnoten. Eine Metaanalyse der Prognose der Studien- und Ausbildungsleistung. In Jäger, R. S., Horn, R. & Ingenkamp, K. (Hrsg.), Test und Trends 7, 11-39. Weinheim: Beltz.

Barret, G. V. & Depinet, R. L. (1991). A reconsideration of 'Testing of competence rather than for intelligence'. American Psychologist, 46, 1012-1024.

Barrett, G. V., Caldwell, M. S. & Alexander, R. A. (1985). The concept of dynamic criteria: a critical reanalysis. Personnel Psychology, 38, 41-56.

Barrick, M. R. & Alexander, R. A. (1991). Assessing the utility of stochastic employee movements. Decision Sciences, 22, 171-180.

Barrick, M. R. & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1-26.

Barrick, M. R., Mount, M. K. & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives. Journal of Applied Psychology, 78, 715-722.

Barrick, M. R., Patton, G. K. & Haughland, S. N. (2000). Accuracy of interviewer judgements of job applicant personality traits. Personnel Psychology, 53, 925-951.

Barthel, E. (1988). Nutzen eignungsdiagnostischer Verfahren bei der Bewerberauswahl. Frankfurt am Main: Lang.

Becker, B. E. & Huselid, M. A. (1998). High performance work systems and firm performance: A synthesis of research and managerial implications. Research in Personnel and Human Resource Management, 16, 53-101.

Becker, B. E., Huselid, M. A. & Ulrich, D. (2001). The HR Scorecard: Linking People, Strategy and Performance. Boston: Harvard Business School Press.

Berthel, J. (1995). Personal-Management. Stuttgart: Schäffer-Poeschel.

Binning, J. F., LeBreton, J. M. & Adorno, A. J. (1999). Assessing Personality. In Eder, R. W. & Ferris, G. R. (Eds.), The Employment Interview Handbook, 105-123. Thousand Oaks: Sage.

Bliesener, T. (1991). Die Validität biographischer Daten bei der Prognose des Berufserfolges: Erste Ergebnisse einer metaanalytischen Studie zur Validitätsgeneralisierung. In Schuler, H. & Funke, U. (Hrsg.), Eignungsdiagnostik in Forschung und Praxis, Band 10, 63-66. Stuttgart: Verlag für Angewandte Psychologie.

Bliesener, T. (1992). Ist die Validität biographischer Daten ein methodisches Artefakt? Ergebnisse einer meta-analytischen Studie. Zeitschrift für Arbeits- und Organisationspsychologie, 36, 12-21.

Bliesener, T. (1996). Methodological moderators in validating biographical data in personnel selection. Journal of Occupational and Organizational Psychology, 69, 107-120.

Blum, M. L. & Naylor, J. C. (1968). Industrial psychology. Its theoretical and social foundations. New York: Harper and Row.

Bobko, P., Karren, R. & Parkington, J. J. (1983). Estimation of Standard Deviations in Utility Analysis: An Empirical Test. Journal of Applied Psychology, 68, 170-176.

Bobko, P., Shetzer, L. & Russel, C. (1991). Estimating the standard deviation of professor's worth: The effects for frame and presentation order in utility analysis. Journal of Occupational Psychology, 64, 179-188.

Boerlijst, G. & Meijboom, G. (1989). Matching the Individual and the Organization. In Herriot, P. Handbook of Assessment in Organizations. Chichester, New York, Brisbane, Toronto, Singapore: John Wiley & Sons.

Borman, W. C. (1978). Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 63, 135-144.

Borman, W. C., Hanson, M. A. & Hedge, J. W. (1997). Personnel Selection. Annual Review of Psychology, 299-337.

Borman, W. C. & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In Schmitt, N. & Borman, W. C. Personnel selection in organizations, 71-98. San Francisco: Jossey-Bass.

Borman, W. C. & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. Human Performance, 10, 99-109.

Bortz, J. (1993). Statistik für Sozialwissenschaftler (4., vollst. überarb. Aufl.). Berlin; Heidelberg; New York; London; Paris; Tokio; Hong Kong; Barcelona; Budapest: Springer.

Bortz, J. & Döring, N. (1995). Forschungsmethoden und Evaluation für Sozialwissenschaftler (2. vollst. überarb. und aktual. Aufl.). Berlin; Heidelberg; New York; Barcelona; Budapest; Hong Kong; London; Mailand; Paris; Tokio: 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 Dunette, M. D. & Hough, L. M. ((Eds.)), Handbook of Industrial and Organisational Psychology, Vol. 2, 2nd ed., 621-745. Palo Alto: Consulting Psychologists Press.

Boudreau, J. W. (1993). Selection utility analysis. In Smith, M. & Robertson, I. T. (Eds.), Advances in selection and assessment, 227-257. New York: Wiley.

Boudreau, J. W. & Ramstad, P. M. (1999). Human resource metrics: Can measures be strategic? Research in Personnel and Human Resources Management, Supplement, 4, 75-98.

Boudreau, J. W. & Ramstad, P. M. (2002). Strategic I/O Psychology and the Role of Utility Analysis Models. Center for Advanced Human Resource Studies, Working Paper.

Brambring, M. (1983). Spezielle Eignungsdiagnostik. In Groffmann, K. J. M. & L., M. (Hrsg.), Intelligenz- und Leistungsdiagnostik. Enzyklopädie der Psychologie, BII, 2, 414-481. Göttingen: Hogrefe.

Brandstätter, H. (1983). Sozialpsychologie. Stuttgart: Kohlhammer.

Bray, D. W., Campbell, R. J. & Grant, D. L. (1974). Formative years in business: A long-term AT&T study of managerial lives. New York: Wiley.

Brody, E. B. & Brody, N. (1976). Intelligence: Nature, determinants, and consequences. New York: Wiley.

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. The Journal of Educational Psychology, 37, 65-76.

Brogden, H. E. (1949). When Testing Pays Off. Personnel Psychology, 2, 171-185.

Brogden, H. E. & Taylor, E. K. (1950). The dollar criterion - applying the cost accounting concept to criterion construction. Personnel Psychology, 3, 133-154.

Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. Psychological Review, 62, 193-217.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. Psychological Review, 62, 193-217.

Bühner, M. (2004). Einführung in die Test- und Fragebogenkonstruktion. München: Pearson Studium.

Bungard, W. (1987). Zur Problematik von Reaktivitätseffekten bei der Durchführung eines Assessment Centers. In Schuler, H. & Stehle, W. (Hrsg.), Assessment Center als Methode der Personalentwicklung, 99-125. Göttingen: Hogrefe/Verlag für Allgemeine Psychologie.

Burke, M. J. & Frederick, J. T. (1984). Two modified procedures for estimating standard deviations in utility analyses. Journal of Applied Psychology, 71, 334-339.

Burke, M. J. & Frederick, J. T. (1986). A comparison of economic utility estimates for alternative SDy estimate 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 generalization procedures. Journal of Applied Psychology, 71, 349-353.

Burnett, J. R. & Motowidlo, S. J. (1998). Relations between different sources of information in the structured selection interview. Personnel Psychology, 51, 963-983.

Campbell, D. T. (1990). The role of theory in industrial and organizational psychology. In Dunette, M. D. & Hough, L. M. (Eds.), Handbook of I/O Psychology, 2nd ed., 39-74. Palo Alto, CA: Consulting Psychology Press.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Campbell, J. P. (1994). Alternative models of job performance and their implications for selection and classification. In Rumsey, M. G., Walker, C. B. & Harris, J. H. Personnel Selection and Classification, Eds., 33-52. Hillsdale, NJ: Lawrence Erlbaum Associates.

Campbell, J. P., McCloy, R. A., Oppler, S. H., et al. (1993). A theory of performance. In Schmitt, N. & Borman, W. C. Personnel selection in organizations, 35-70. San Francisco: Jossey-Bass.

Campion, M. A. (1994). Job analysis for the future. In Rumsey, M. G., Walker, C. B. & Harris, J. H. Personnel Selection and Classification, 1-12. Hillsdale, NJ: Lawrence Erlbaum Associates.

Campion, M. A., Campion, J. E. & Hudson, J. P. (1994). Structured interviewing: a note on incremental validity and alternative question types. Journal of Applied Psychology, 79, 998-1002.

Campion, M. A., Palmer, D. K. & Campion, J. E. (1997). A review of structure in the selection interview. Personnel Psychology, 50, 655-702.

Campion, M. A., Pursell, E. D. & Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. Personnel Psychology, 41, 25-42.

Canibano, L., Garcia-Ayuso, M. & Sanchez, P. (2000). Accounting for intangibles: A literature review. Journal of Accounting Literature, 19, 102-130.

Carroll, J. B. (1968). Review of *The Nature of Human Intelligence* by J. P. Guilford. American Educational Research Journal, 5, 249-256.

Carroll, J. B. (1993). Human cognitive abilities: A survey of factor analytic studies. Cambridge, MA: Cambridge University Press.

Carson, K. P., Becker, J. S. & Henderson, J. A. (1998). Is utility really futile? A failure to replicate and an extension. Journal of Applied Psychology, 83(1), 84-96.

Cascio, W. F. (1980). Responding to the Demand of Accountability: A Critical Analysis of Three Utility Models. Organizational Behaviour & Human Performance, 25, 32-45.

Cascio, W. F. (1982). Costing human resources: The financial impact of behaviour in organizations. Boston: Kent.

Cascio, W. F. (1996). The role of utility analysis in the strategic management of organizations. Journal of Human Resource Costing and Accounting, 1(2), 85-95.

Cascio, W. F. (2000). The Cost Factor: The Financial Impact of Human Resources in Corporations. Thomson Learnings.

Cascio, W. F. & Morris, J. K. (1990). A critical reanalysis of Hunter, Schmidt, and Coggins' 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.

Cascio, W. F. & Ramos, R. A. (1986). Development and application of a new method for assessing job performance in behavioural/economic terms. Journal of Applied Psychology, 71, 20-28.

Cascio, W. F. & Silbey, V. (1979). Utility of the assessment center as a selection device. Journal of Applied Psychology, 64, 107-118.

Cassens, F. (1966). Cross-cultural dimensions of executive life history antecedents. Greensboro, NC: Creativity Research Institute Richardson Foundation.

Cattell, R. B. (Ed.) (1966). Handbook of multivariate experimental psychology. Chicago: Rand McNally.

Cattell, R. B. (1971). Abilities: Their structure, growth, and action. New York: Houghton Mifflin.

Cattell, R. B. (1972). The 16PF and basic personality structures: a reply to Eysenck! Journal of Behavioural Science, 17, 169-187.

Cesare, S. J., Blankenship, M. H. & Gianetto, P. W. (1994). A dual focus on SDy estimations: A test of linearity assumption and multivariate application. Human Performance, 7(4), 235-255.

Chen, H.-T. (1990). Theory-driven evaluations. Newbury Park, Calif.: Sage.

Clark, T. (1995). Managing Consultants - Consultancy as the Management of Impressions. Buckingham: Open University Press.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned so far. American Psychologist, 45, 1304-1312.

Cohen, J. & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Committee on Techniques for the Enhancement of Human Performance: Occupational Analysis, Commission on Behavioral and Social Sciences and Education & National Research Council (1999). The Changing Nature of Work: Implications for Occupational Analysis. Washington: National Academic Press.

Conrad, W. (1992). Diagnostik als Messung. In Jäger, R. S. & Petermann, F. Psychologische Diagnostik, 245-256. Weinheim: Psychologie Verlags Union.

Conway, J. M. & Huffcut, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. Human Performance, 10, 331-360.

Conway, J. M., Jako, R. A. & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. Journal of Applied Psychology, 80, 565-579.

Conway, J. M., Lombardo, K. & Sanders, K. C. (2001). A meta-analysis of incremental validity and nomological networks for subordinate and peer ratings. Human Performance, 14, 267-303.

Conway, J. M. & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. Journal of Business and Psychology, 13, 485-506.

Cook, M. (1993). Personnel selection and productivity (2. ed.). Chichester: Wiley.

Cook, M. (1995). Personnel selection and productivity (2. ed.). Chichester: Wiley.

Cook, M. (1998). Personnel selection: adding value through people (3rd ed.). Chichester: Wiley.

Cook, T. D. & Campbell, D. T. (1976). The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings. In Dunette, M. D. (Ed.), Handbook of Industrial and Organizational Research, 223-326. Skokie, IL: Rand McNally.

Cortina, J. M., Goldstein, N. B., Payne, S. C., et al. (2000). The incremental validity of interview sources over and above cognitive ability and conscientiousness scores. Personnel Psychology, 53, 325-351.

Cronbach, L. J. (1949). Essentials of psychological testing. New York: Harper & Row.

Cronbach, L. J. (1982). Designing Evaluations of Educational and Social Programs. San Francisco: Jossey-Bass.

Cronbach, L. J. & Gleser, G. S. (1965). Psychological tests and personnel decisions (2nd ed.). Urbana, J.L.: University of Illinois Press.

Cronshaw, S. F. (1997). Lo! The stimulus speaks: The insider's view on Whyte and Latham's "the futility of utility analysis". Personnel Psychology, 50(3), 611-615.

Cronshaw, S. F. & Alexander, R. A. (1985). One answer to the demand for accountability: Selection utility as an investment decision. Organizational Behaviour & Human Decision Processes, 35, 102-118.

Cronshaw, S. F., Alexander, R. A., Wiesner, W. H., et al. (1987). Incorporating risk into selection utility: Two models for sensitivity analysis and risk simulation. Organizational Behaviour & Human Decision Processes, 40, 270-286.

Day, A. L. & Carroll, S. A. (2003). Situational and Patterned Behaviour Description Interviews: A Comparison of Their Validity, Correlates, and Perceived Fairness. Human Performance, 16(1), 25-47.

De Corte, W. (1994). Utility analysis for the one-cohort selection-retention decision with a probationary period. Journal of Applied Psychology, 79(3), 402-411.

De Corte, W. (1997). Utility analysis for the probationary selection decision to obtain a fixed quota of successful selectees. Journal of Organizational Behavior, 18(1), 67-82.

De Corte, W. (1998a). An analytic procedure to estimate the benefits of a particular type of personnel classification decision. Educational and Psychological Measurement, 58(6), 929-943.

De Corte, W. (1998b). Estimating and maximizing the utility of sequential selection decisions with probationary period. British Journal of Mathematical & Statistical Psychology, 51, 101-121.

De Corte, W. (2000a). Estimating the classification efficiency of a test battery. Educational and Psychological Measurement, 60(1), 73-85.

De Corte, W. (2000b). Using order statistics to assess the sampling variability of personnel selection utility estimates. Journal of Applied Statistics, 27(6), 703-713.

de Wolff, C. J. (1989). The Changing Role of Psychologists in Selection. In Herriot, P. Assessment and Selection in Organizations, 81-93. Chichester, New York, Brisbane, Toronto, Singapore: John Wiley & Sons.

Deadrick, D. L. & Madigan, R. M. (1990). Dynamic criteria revisited: a longitudinal study of performance stability and prediction validity. Personnel Psychology, 43, 717-744.

Delery, D. E. & Doty, D. H. (1996). Modes of theorizing in strategic human resource management: Tests of universalistic, contingency, and configurational predictions. Academy of Management Journal, 39, 802-835.

DeSimone, R. L., Alexander, R. A. & Cronshaw, S. F. (1986). Accuracy and reliability of SDy estimates in utility analysis. Journal of Industrial Psychology, 59, 93-102.

Diemand, A. & Schuler, H. (1991). Sozial erwünschtes Verhalten in eignungsdiagnostischen Situationen. In Schuler, H. & Funke, U. (Hrsg.), Eignungsdiagnostik in Forschung und Praxis, Band 10, 242-248. Stuttgart: Verlag für Angewandte Psychologie.

Digman, J. M. (1990). Personality Structure: Emergence of the five factor model. Annual Review of Psychology, 41, 417-440.

Donabedian, A. (1980). The Definition of Quality and Approaches to its Assessment. Ann Arbour (MI): Health Administration Press.

Dreher, G. F., Ash, R. A. & Hancock, P. (1988). The role of traditional research design in underestimating the validity of the employment interview. Personnel Psychology, 41, 315-325.

DuBois, P. H. (1970). A history of psychological testing. Boston: Allyn & Bacon.

Dunette, M. D., Rosse, R. L., Houston, J. S., et al. (1982). Development and validation of a industry-wide electric power plant operator selection system. Washington, D.C: Edison Electric Institute.

Dye, D. A., Reck, M. & McDaniel, M. (1993). The Validity of Job Knowledge Measures. International Journal of Selection and Assessment, 1(3), 153-157.

Eaton, N. K., Wing, H. & Mitchell, K. J. (1985). Alternate Methods of Estimating the Dollar Value of Performance. Personnel Psychology, 38, 27-40.

Eckardt, H. H. & Schuler, H. (1992). Berufseignungsdiagnostik. In Jäger, R. S. & Petermann, F. (Hrsg.), Psychologische Diagnostik, (2. Aufl.), 533-551. Weinheim: Psychologie Verlags Union.

Effenberger, J. (1998). Erfolgsfaktoren der Strategieberatung: Analyse einer Leistung von Unternehmensberatern aus Klientensicht. Stuttgart: M und P Verlag für Wissenschaft und Forschung.

Eilles-Matthiessen, C., elHage, N., Janssen, S., et al. (2002). Schlüsselqualifikationen in Personalauswahl und Personalentwicklung: ein Arbeitsbuch für die Praxis (1. Auflage). Bern; Göttingen: Huber.

Eisenhardt, K. M. (1989). Agency Theory. An Assessment and Review. Academy of Management Review, 14, 57-74.

Elfgen, R. & Klaile, B. (1987). Unternehmensberatung. Angebot, Nachfrage, Zusammenarbeit. Stuttgart: Poeschel Verlag.

Ernst, B. (2002). Die Evaluation von Beratungsleistungen: Prozesse der Wahrnehmung und Bewertung (1. Auflage). Wiesbaden: Deutscher Universitäts-Verlag.

Etzel, S. & Küppers, A. (2002). Innovative Managementdiagnostik. Göttingen; Bern: Hogrefe.

Eysenck, H. J. (1991). Dimensions of personality: 16, 5, or 3? - Criteria for a taxonomic paradigm. Personality and Individual Differences, 12, 773-790.

Facteau, J. D., Facteau, C. L., Jackson, K. A., et al. (2000). Do structured interviews measure KSA? An investigation of the construct validity of interview ratings in two organi-

sations. Paper presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Feldmann, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.

Fiedler-Winter, R. (2004). Mit intelligenten Lösungen in die Zukunft - Interview mit Prof. Dr. Knut Bleicher. Personal, 4, 32.

Fine, S. A. & Wiley, W. W. (1971). An Introduction to Functional Job Analysis. Kalamazoo, MI: The W.E. Upjohn Institute for Employment Research.

Fisseni, H. J. (1990). Lehrbuch der psychologischen Eignungsdiagnostik. Göttingen: Hogrefe.

Fisseni, H. J. & Fennekels, G. P. (1995). Das Assessment Center. Göttingen: Verlag für Angewandte Psychologie.

Fitz-Enz, J. (2000). The ROI of Human Capital: Measuring the Economic Value of Employee Performance. New York: Amacom.

Fitz-Enz, J. & Davison, B. (2002). How to Measure Human Resources Management (3. ed.). Mc Graw Hill.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-358.

Fleishman, E. A. & Quaintance, M. K. (1984). Taxonomies of human performance. Orlando: Academic Press.

Fortmüller, R. (1997). Wissen und Problemlösen. Wien: MANZ.

Franklin, J. L. & Trasher, J. H. (1976). An introduction to program evaluation. New York: Wiley.

Fridrich, A. (1985). Marketing- und Managementberatungen in mittelständischen Unternehmen. Stuttgart: Erich Schmidt Verlag.

Frieling, E. & Hoyos, C. G. (1978). Fragebogen zur Arbeitsanalyse (FAA): Deutsche Bearbeitung des Position Analysis Questionnaire (PAQ). Bern: Huber.

Fritz, W. & Effenberger, J. (1998). Strategische Unternehmensberatung. Verlauf und Erfolg von Projekten der Strategieberatung. Die Betriebswirtschaft, 58, 103-118.

Funder, D. C. (1991). Global traits: A neon-Allportian approach to personality. Psychological Science, 2, 31-39.

Funke, U. (1986). Die Validität verschiedener eingungsdiagnostischer Verfahren bei Lehrstellenbewerbern. Zeitschrift für Arbeits- und Organisationspsychologie, 30, 92-97.

Funke, U. & Barthel, E. (1990). Nutzen und Kosten von Personalentscheidungen. In Sarges, W. (Hrsg.), Management-Diagnostik, 647-658. Göttingen: Hogrefe.

Funke, U., Krauss, J., Schuler, H., et al. (1987). Zur Prognostizierbarkeit wissenschaftlich-technischer Leistungen mittels Personenvariablen: Eine Metaanalyse der Validität diagnostischer Verfahren im Bereich Forschung und Entwicklung. Gruppendynamik, 18 (4), 407-428.

Gabele, E. & Hirsch, I. (1986). Zur Qualität betriebswirtschaftlicher Beratungsleistungen für kleine Unternehmen: Ergebnisse einer Inhaltsanalyse von Beratungsberichten. Die Betriebswirtschaft, 46(4), 486-500.

Gable, G. G. (1996). A multidimensional model of client success when engaging external consultants. Management Science, 42(8), 1175-1198.

Gast, I. & McMillan, L. (1999). Assessing the Return for Investment of a New Executive Selection Program. IPMAAC Conference on Professional Personnel Assessment, Conference Presentation.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. Behavioural and Brain Sciences, 19, 229-284.

Geary, D. C. (1998). Male, female: The evolution of human sex differences. Washington, DC: American Psychological Association.

Gerpott, T. J. (1990). Erfolgswirkung von Personalauswahlverfahren. Zeitschrift für Führung und Organisation, 59, 37-44.

Ghiselli, E. E. (1964). Dr. Ghiselli comments in Dr. Tupes' note. Personnel Psychology, 17, 61-63.

Ghiselli, E. E. (1966). The validity of occupational aptitude tests. New York: Wiley.

Ghiselli, E. E. & Haire, M. (1960). The validation of selection tests in the light of the dynamic character of criteria. Personnel Psychology, 13, 225-231.

Giese, F. (1924). Die Arbeitsprobe in der Psychodiagnostik. Zeitschrift für Angewandte Psychologie, 23, 3-4.

Glass, G. V. & Ellet, F. S. J. (1980). Evaluation research. Annual Review of Psychology, 31, 211-228.

Goff, M. & Ackerman, P. L. (1992). Personality-Intelligence Relations: Assessment of Typical Intellectual Engagement. Journal of Educational Psychology, 84(4), 537-552.

Goldberg, L. R. (1970). Man versus model of man: a rationale plus evidence for a method of improving on clinical inferences. Psychological Bulletin, 73, 422-432.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. American Psychologist, 48, 26-34.

Goodale, J. G. (1989). Effective employment interviewing. In Eder, R. W. & Ferris, G. R. (Eds.), The Employment Interview: Theory, Research, and Practice, 307-323. Newbury Park: Sage.

Gorsuch, R. (1983). Factor analysis. Hillsdale, NJ: Erlbaum.

Gottfredson, L. S. (2002). Where and Why g Matters: Not a Mistery. Human Performance, 15(1/2), 25-46.

Gottfredson, L. S. (2003a). Dissecting practical intelligence theory: Its claims and evidence. Intelligence, 31, 343-397.

Gottfredson, L. S. (2003b). Practical Intelligence. In Fernandez-Ballesteros, R. (Ed.), Encyclopedia of Psychological Assessment, 740-745. London: Sage.

Greer, O. L. & Cascio, W. F. (1987). Is cost accounting the answer? Comparison of two behaviourally 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.

Greuter, A. M. & Algera, J. A. (1989). Criterion Development and Job Analysis. In Herriot, P. (Ed.), Assessment and Selection in Organizations, 143-182. Chichester, New York, Brisbane, Toronto, Singapore: Wiley.

Grönross, C. (1982). Strategic Marketing and Management in the Service Sector. Research Report No. 8 of Swedish School of Economics and Business Administration, Helsingfors,

Gross, P. (2003). Waffenstillstand im war for talents. Personal, 6, 42.

Grossmann, R. J. (2000). Measuring up. HRMagazin, 45(1), 28-35.

Grove, W. M. & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. Psychology, Public Policy, and Law, 2, 293-323.

Guba, E. G. (1969). The failure of educational evaluation. Educational Technology, 9, 29-38.

Guilford, J. P. (1956). The structure of intellect. Psychological Bulletin, 53, 267-293.

Guion, R. M. (1965). Personnel Testing. New York: McGraw-Hill.

Guion, R. M. (1998). Assessment, measurement, and prediction for personnel decisions. Mahwah, NJ: Erlbaum.

Guthke, J. (1991). Das Lerntestkonzept in der Eignungsdiagnostik. In Schuler, H. & Funke, U. ((Hrsg.)), Eignungsdiagnostik in Forschung und Praxis, 33-35. Göttingen: Hogrefe.

Guzzo, R. A., Jette, R. D. & Katzell, R. A. (1985). The effects of psychologically based intervention programs on worker productivity: A meta-analysis. Personnel Psychology, 38, 275-292.

Hakstian, A. R., Wooley, R. M., Woolsey, L. K., et al. (1991). Management selection by multiple-domain assessment: II. Utility to the organisation. Educational and Psychological Measurement, 51, 899-911.

Halpern, D. F. (2000). Sex differences in cognitive abilities (3rd. ed.). Mahwah, NJ: Erlbaum.

Harrel, T. W. & Harrel, M. S. (1945). Army General Classification Test scores for civilian occupations. Educational and Psychological Measurement, 5, 229-239.

Hartmann, H. A., Haubl, R., Neuberger, O., et al. (1984). Diagnostische Probleme psychologischer Begutachtung. In Hartmann, H. A. & Haubl, R. (Hrsg.), Psychologische Begutachtung: Probleme und Praxisfelder, 75-126. München: Urban & Schwarzenberg.

Harvey, R. J. (1991). Job Analysis. In Dunette, M. D. & Hough, L. M. (Eds.), Handbook of industrial and organizational psychology, Vol. 2, 2nd. ed., 71-164. Palo Alto, CA: Consulting Psychology Press.

Hazer, J. T. & Highhouse, S. (1997). Factors influencing managers' reactions to utility analysis: Effects of SDy method, information frame, and focal intervention. Journal of Applied Psychology, 82(1), 104-112.

Heckhausen, H. (1989). Motivation und Handeln (2. völlig überarb. u. erg. Aufl.). Berlin; Heidelberg; New York; London; Paris; Tokio; Hong Kong: Springer.

Helmreich, R. L., L., S. L. & Carsrud, A. L. (1986). The honeymoon effect in job performance: Temporal increases in the predictive power of achievement motivation. Journal of Applied Psychology, 71, 185-188.

Heneman, R. L. (1986). The relationship between supervisory ratings and results-oriented measures of performance: a meta-analysis. Personnel Psychology, 39, 811-826.

Herrmann, T. (1991). Lehrbuch der empirischen Persönlichkeitsforschung (6. Aufl.). Göttingen, Toronto, Zürich: Hogrefe.

Hill, W. (1990). Der Stellenwert der Unternehmensberatung für die Unternehmensführung. Die Betriebswirtschaft, 50(2), 171-180.

Hochholdinger, S. (2002). Evaluation von Managementdiagnostik: Methoden, Konstrukte, Beispiele. Mannheim: Fakultät für Sozialwissenschaften der Universität Mannheim.

Höck, M. & Keuper, F. (2001). Empirische Untersuchung zur Auswahl und Kompetenz von Beratungsgesellschaften. Die Betriebswirtschaft, 61(4), 427-443.

Hoffmann, C. C. & Thornton, G. C. (1997). Examining selection utility where competing predictors differ in adverse impact. Personnel Psychology, 50, 455-470.

Hoffmann, W. H. (1991). Faktoren erfolgreicher Unternehmensberatung. Wiesbaden: Deutscher Universitäts-Verlag.

Höft, S. & Funke, U. (2001). Simulationsorientierte Verfahren der Personalauswahl. In Schuler, H. (Hrsg.), Lehrbuch der Personalpsychologie. Göttingen: Hogrefe.

Hogan, R. T., Hogan, J. & Roberts, B. W. (1996). Personality measurement and employment decisions. American Psychologist, 54 (5), 469-477.

Holland, J. (1959). A Theory of Vocational Choice. Journal of Counseling Psychology, 6(1), 35-45.

Holland, J. (1973). Making Vocational Choices: A Theory of Careers. Englewood Cliffs, NJ: Prentice-Hall.

Holland, J. (1985). Making Vocational Choices: A Theory of Careers (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Holland, J. (1997). Making vocational choices: A theory of vocational personality and work environments (3rd ed.). Odessa: Psychological Assessment Resources.

Holland, J. L. (1996). Exploring careers with a typology: What we have learned and some new directions. American Psychologist, 51, 397-406.

Holling, H. (1998). Utility Analysis Of Personnel Selection - An Overview And Empirical Study Based On Objective Performance Measures. Methods of Psychological Research Online, Vol. 3, 1.

Holling, H., Preckel, F. & Vock, M. (2004). Intelligenzdiagnostik. In Holling, H. & Petermann, F. Kompendien psychologischer Diagnostik, Band 6. Göttingen, Bern, Toronto, Seattle, Oxford, Prag: Hogrefe.

Holling, H. & Reiners, W. (1994). Utility Analysis of Assessment Centers using objective profit data. Web published document.

Hollmann, H. (1991). Validität in der Eignungsdiagnostik - Neue Ansätze einer sachgerechten Bewertung und effizienten Verbesserung. Göttingen; Toronto; Zürich: Hogrefe.

Hollmann, H. & Reitzig, G. (1995). Referenzen und Dokumentenanalyse. In Sarges, W. (Hrsg.), Management-Diagnostik, 463-470. Göttingen: Hogrefe.

Horn, J. L. & Knapp, J. R. (1974). Thirty wrongs do not make a right: Reply to Guilford. Psychological Bulletin, 81, 502-504.

Hossiep, R. (1995). Berufseignungsdiagnostische Entscheidungen: zur Bewährung eignungsdiagnostischer Ansätze. Göttingen: Hogrefe.

Hossiep, R. & Wottawa, H. (1993). Diagnostik. In Schorr, A. (Hrsg.), Handwörterbuch der Angewandten Psychologie, 131-136. Bonn: Deutscher Psychologen Verlag.

Hough, L. M. (1992). The "Big Five" personality variables construct confusion: Description vs. prediction. Human Performance, 5, 139-155.

Hough, L. M., Eaton, N. K., Dunette, M. D., et al. (1990). Criterion related validities of personality constructs and the effect of response distortion on those validities. Journal of Applied Psychology, 75, 581-595.

Hough, L. M. & Oswald, F. L. (2000). Personnel selection: Looking toward the future - remembering the past. Annual Review of Psychology, 51, 81-108.

House, E. R. (1980). Evaluating with validity. Beverly Hills: Sage.

Huffcut, A. I. & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. Journal of Applied Psychology, 79, 184-190.

Huffcut, A. I., Conway, J., Roth, P. L., et al. (2001). Identification and meta-analysis for constructs measured in employment interviews. Journal of Applied Psychology, 86, 897-913.

Huffcut, A. I., Roth, P. L. & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. Journal of Applied Psychology, 81, 459-473.

Huffcut, A. I., Weekley, J. A., Wiesner, W. H., et al. (2001). Comparison of situational and behaviour description interview questions for higher level positions. Personnel Psychology, 54, 619-644.

Hulin, C. L. (1982). Some reflections on general performance dimensions and halo rating error. Journal of Applied Psychology, 67, 165-170.

Hull, C. L. (1928). Aptitude Testing. Yonkers: World Book.

Humphreys, L. G., Lubinski, D. & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. Journal of Applied Psychology, 78, 250-261.

Hunt, E. (1995). Will We Be Smart Enough? New York: Russel Sage Foundation.

Hunter, J. E. (1986). Cognitive Ability, Cognitive Aptitudes, Job Knowledge, and Job Performance. Journal of Vocational Behaviour, 29, 340-362.

Hunter, J. E. & Hirsh, H. R. (1987). Application of meta-analysis. In Cooper, C. L. & Robertson, I. T. (Eds.), International review of industrial and organizational psychology 1987, 321-357. Chichester: Wiley.

Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternate predictors of job performance. Psychological Bulletin, 96, 72-98.

Hunter, J. E. & Schmidt, F. (1996). Intelligence and job performance: Economic and social implications. Psychology, Public Policy, and Law, 2, 447-472.

Hunter, J. E., Schmidt, F. & Jackson, G. B. (1982). Meta-analysis. Cumulating research findings across studies. Beverly Hills: Sage Publications.

Hunter, J. E. & Schmidt, F. L. (1982). Fitting people to jobs: The impact personnel selection on national productivity. In Fleishman, E. A. & Dunnette, M. D. (Eds.), Human performance and productivity, 1, 233-284. Hillsdale: Erlbaum.

Hunter, J. E. & Schmidt, F. L. (1989). Meta-Analysis: Facts and theories. In Smith, M. & Robertson, I. T. (Eds.), Advances in selection and assessment, 203-216. New York: Wiley.

Hunter, J. E. & Schmidt, F. L. (1990). Methods of Meta-Analysis - Correcting Error and Bias in Research Findings. Newbury Park; London; New Delhi: Sage.

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. & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. Journal of Applied Psychology, 75, 28-42.

Huselid, M. A., Jackson, S. E. & Schuler, R. S. (1997). Organizational characteristics as predictors of personnel practices. Personnel Psychology, 42, 727-786.

Hyde, J. S., Fennema, E. & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. Psychological Bulletin, 107, 139-153.

Ichniowski, C., Shaw, K. & Prennushi, G. (1997). The effects of human resource management practices on productivity: A study of steel finishing lines. American Economic Review, 87(3), 291-313.

Jacobs, R. R., Conte, J. M., Day, D. V., et al. (1996). Selecting Bus Drivers: Multiple Predictors, Multiple Perspectives on Validity, and Multiple Estimates of Utility. Human Performance, 9 (3), 199-217.

Jäger, A. O. (1967). Dimensionen der Intelligenz. Göttingen: Hogrefe.

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. Diagnostica, 28, 195-226.

Jäger, A. O. (1986). Validität von Intelligenztests. Diagnostica, 32, 272-289.

Jäger, A. O., Holling, H., Preckel, F., et al. (2004). Berliner Intelligenzstrukturtest für Jugendliche: Begabungs- und Hochbegabungsdiagnostik (BIS-HB). Göttingen: Hogrefe (in Druck).

Jäger, R. (1995). Biographische Daten. In Jäger, R. & Petermann, F. (Hrsg.), Psychologische Diagnostik, 350-361. Weinheim: Psychologie Verlags Union.

Jäger, R. S. (1992). Statusdiagnostik. In Jäger, R. S. & Petermann, F., 200-202. Weinheim: Psychologie Verlags Union.

Jäger, R. S. & Petermann, F. (1992). Einleitung. In Jäger, R. S. & Petermann, F. Psychologische Diagnostik, 11-15. Weinheim: Psychologie Verlags Union.

Jäger, R. S. & Scheurer, H. (1992). Prozeßdiagnostik. In Jäger, R. S. & Petermann, F. Psychologische Diagnostik, 202-208. Weinheim: Psychologie Verlags Union.

Janz, J. T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. Journal of Applied Psychology, 57, 577-580.

Janz, T., Hellervik, L. & Gilmore, D. C. (1986). Behaviour description interviewing: New accurate, cost-effective. Boston, MA: Allyn & Bacon.

Jeserich, W. (1991). Mitarbeiter auswählen und fördern. Assessment Center Verfahren. Handbuch der Weiterbildung für die Praxis in Wirtschaft und Verwaltung. (6. Aufl.). München: Hanser.

Jones, A. (1989). Introduction to Section 3: The Tools of Selection. In Herriot, P. (Ed.), Assessment and Selection in Organizations, 275-279. Chichester, New York, Brisbane, Toronto, Singapore: John Wiley & Sons.

Judge, T. A., Higgins, C. A., Thoresen, C. J., et al. (1999). The Big Five Personality Traits, General Mental Ability, and career success across the life span. Personnel Psychology, 52, 621-652.

Judiesch, M. K., Schmidt, F. & Mount, M. K. (1992). Estimates of the Dollar Value of Employee Output in Utility. Journal of Applied Psychology, 77(3), 234-250.

Kaas, K. P. & Schade, C. (1995). Unternehmensberater im Wettbewerb - Eine empirische Untersuchung aus der Perspektive der Neuen Institutionslehre. Zeitschrift für Betriebswirtschaft, 10, 1076-1089.

Kailer, N. & Falter, C. (2000). Sicherung der Qualität von Beratungsprozessen in kleinen und mittleren Unternehmen - Ergebnisse einer Qualitätsmessung. In Kailer, N. & Walger, G. (Hrsg.), Perspektiven der Unternehmensberatung für kleinere und mittlere Betriebe, 129-165. Wien: Linde Verlag.

Kailer, N. & Scheff, J. (2000). Beratung als Dienstleistung: Die Zusammenarbeit von kleinen und mittleren Unternehmen und Unternehmensberatern. In Kailer, N. & Walger, G. (Hrsg.), Perspektiven der Unternehmensberatung für kleinere und mittlere Betriebe, 41-104. Wien: Linde Verlag.

Kanning, U. P. (2002). Tipps für die Anwendung nicht-standardisierter Methoden. In Kanning, U. P. & Holling, H. (Hrsg.), Handbuch personaldiagnostischer Instrumente. Göttingen, Bern, Toronto, Seattle: Hogrefe.

Kanning, U. P. & Holling, H. (Hrsg.) (2002). Handbuch personaldiagnostischer Instrumente. Göttingen, Bern, Toronto, Seattle: Hogrefe.

Kaplan, R. S. & Norton, D. P. (1992). The balanced scorecard - measures that drive performance. Harvard Business Review, January/February, 71-79.

Kataoka, H. C., Latham, G. P. & Whyte, G. (1997). The Relative Resistance of the Situational, Patterned Behavior, and Conventional Structured Interviews to Anchoring Effects. Human Performance, 10 (1), 47-63.

Kelley, T. L. (1923). Statistical method. New York: Macmillan.

Kienbaum, G. & Meissner, D. (1979). Zur Problematik des Effizienznachweises von Beratungen. Ansätze im Beratungsprozess. Betriebswirtschaftliche Forschung und Praxis, 31(2), 109-116.

Kieser, A. (1998). Unternehmensberater - Händler in Problemen, Praktiken und Sinn. In Glaser, H., Schröder, E. F. & von Werder, A. (Hrsg.), Organisation im Wandel der Märkte. Erich Frese zum 60. Geburtstag, 191-225. Wiesbaden: Gabler.

Kiker, D. S. & Motowidlo, S. J. (1998). Effects of rating strategy on interdimensional variance, reliability, and validity of interview ratings. Journal of Applied Psychology, 83(2), 179-189.

Kimura, D. (1999). Sex and cognition. Cambridge, MA: MIT Press.

Kirkpatrick, J. J. (1960). Background factors that lead to executive success. American Psychologist, 15, 477.

Kirkpatrick, J. J., Ewen, R. B., Barrett, R. S., et al. (1968). Testing and fair employment: fairness and validity of personnel tests for different ethnic groups. New York: New York Univ. Press.

Klaile, B. (1984). Managementberatung in mittelständischen Unternehmen - Bessere Führung durch externe Unternehmensberatung. Berlin: Erich Schmidt Verlag.

Klein, H. (1978). Zur Messung des Beratungserfolges. Zeitschrift für Organisation, 47(2), 105-110.

Kleinmann, M. (1997). Assessment Center, Stand der Forschung – Konsequenzen für die Praxis. Göttingen: Verlag für Angewandte Psychologie.

Kleinmann, M., Exler, C., Kuptsch, C., et al. (1995). Unabhängigkeit und Beobachtbarkeit von Anforderungsdimensionen im Assessment Center als Moderatoren der Konstruktvalidität. Zeitschrift für Arbeits- und Organisationspsychologie, 39, 22-28.

Kleinmann, M. & Koller, O. (1997). Improving the construct validity of assessement centers: appropriate use of confirmatory factor analysis and suitable construction principles. Journal of Social Behaviour and Personality, 12, 65-84.

Kleinmann, M., Kuptsch, C. & Koller, O. (1996). Transparency: a necessary requirement for the construct validity of assessment centres. Applied Psychology: An International Review, 45, 67-84.

Kleinmutz, B. (1963). MMPI decision rules for the identification of college maladjustments: a digital computer approach. Psychological Monographs, 77, 14.

Kolk, N. J., Born, M. P. & Van der Flier, H. (2002). Impact of Common Rater Variance on Construct Validity of Assessment Center Dimension Judgements. Human Performance, 15(4), 325-337.

Kompa, A. (1989). Personalbeschaffung und Personalauswahl. Stuttgart: Enke.

Köppen, O. R. (1999). Erfolgsfaktoren von Unternehmensberatungen - Die Nachfolgeregelung in kleinen und mittleren Unternehmen. Wiesbaden: Deutsche Universitätsverlag.

Krapp, A. (1986). Diagnose und Prognose. In Weidemann, B., Krapp, A., Hofer, M., et al. (Hrsg.), Pädagogische Psychologie. Ein Lehrbuch, 566-630. Weinheim: Psychologie Verlags Union.

Kristof, A. (1996). Person-Organisation Fit: An Integrative Review of its Conceptualizations, Measurement, and Implications. Personal Psychology, 49, 1-49.

Kröber, H. W. (1991). Der Beratungsbegriff in der Fachliteratur. In Hofmann, M., von Rosenstiel, L. & Zapotoczky, K. (Hrsg.), Die sozio-kulturellen Rahmenbedingungen für Unternehmensberater. Stuttgart: Kohlhammer.

Krystofiak, F., Newman, J. M. & Anderson, G. (1979). A quantified approach to measurement of job content: procedures and payoffs. Personnel Psychology, 32, 341-357.

Lanyon, R. I. & Goodstein, L. D. (1997). Personality assessment (3rd ed.). New York: John Wiley.

Latham, G. P., Saari, L. M., Pursell, E. D., et al. (1980). The situational interview. Journal of Applied Psychology, 65, 422-427.

Latham, G. P. & Whyte, G. (1994). The Futility of Utility Analysis. Personnel Psychology, 47, 31-46.

Law, K. S. (1995). Estimating the Dollar Value Contribution of Human Resource Intervention Programs: Some Comments on the Brogden Utility Equation. Australian Journal of Management, 20(2), 197-206.

Lechner, C., Gesing, N., Kreutzer, M., et al. (2005). Berater unter Druck. Harvard Business Manager, August, 6-8.

Ledvinka, J., Simonet, J. K., Neiner, A. G., et al. (1983). The dollar value for JEPS at Life of Georgia. (unpublished technical report). Aus: Boudreau, J. W. (1991). Utility analysis for decisions in human resource management. In Dunette, M. D. & Hough, L. M. (Eds.), Handbook of Industrial and Organisational Psychology, 621-745. Palo Alto: Consulting Psychologists Press.

Lee, R. & Booth, J. M. (1974). A Utility Analysis of a Weighted Application Blank Designed To Predict Turnover for Clerical Employees. Journal of Applied Psychology, 59, 516-518.

Lindstam, S. (2005). Utility Analysis for Decisions in Human Resource Management. In Beauducel, A., Biehl, B., Bosnjak, M., et al. Multivariate Research Strategies. Festschrift in Honor of Werner W. Wittmann, 329. Aachen: Shaker Verlag.

Litwin, G. H. & Stringer, R. A. (1968). Motivation and organizational climate. Boston: Harvard University, Graduate School of Business Administration, Division Research.

Lochner-Hannen, P. (2002). Die prognostische Validität des Interviews in der Führungskräfteauswahl. Aachen: Shaker Verlag.

Loehlin, J. C. (2000). Group differences in intelligence. In Sternberg, R. J. (Ed.), Handbook of Intelligence, 176-196. Cambridge: Cambridge University Press.

Lopez, F. M. (1965). Personnel interviewing. Theory and practice. New York: McGraw-Hill.

Lord, F. M. & Novick, M. R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley.

Lubinski, D. (2004). Introduction to the Special Section on Cognitive Abilities: 100 Years After Spearman's (1904) "'General Intelligence', Objectively Determined and Measured". Journal of Personality and Social Psychology, 86 (1), 96-111.

Lubinski, D., Benbow, C. P., Shea, D. L., et al. (2001). Men and Women At Promise For Scientific Excellence: Similarity Not Dissimilarity. Psychological Science, 12(4), 309-317.

Lubinski, D., Webb, R. M., Morelock, M. J., et al. (2001). Top 1 in 10,000: A 10-Year Follow-Up of the Pofoundly Gifted. Journal of Applied Psychology, 86(4), 718-729.

Mabon, H. (1996). The cost of downsizing in an enterprise with job security. Journal of Human Resource Costing and Accounting, 1, 35-62.

Mabon, H. (1998). Utility aspects of personality and performance. Human Performance, 11(2/3), 289-304.

Macan, T. H. & Highhouse, S. (1994). Communicating the utility of HR activities: A survey of I/O and HR professionals. Journal of Business and Psychology, 8 (4), 425-436.

Machwirth, U., Schuler, H. & Moser, K. (1996). Entscheidungsprozesse bei der Analyse von Bewerbungsunterlagen. Diagnostica, 42, 220-241.

Marchese, M. C. & Muchinsky, P. M. (1993). The validity of the employment interview: A meta-analysis. International Journal of Selection and Assessment, 1, 18-26.

Marcus, B., Funke, U. & Schuler, H. (1997). Integrity Tests als spezielle Gruppe eignungsdiagnostischer Verfahren: Literaturüberblick und Metaanalytische Befunde zur Konstruktvalidität. Zeitschrift für Arbeits- und Organisationspsychologie, 41 (1), 2-17.

Marcus, B. & Schuler, H. (2001). Leistungsbeurteilung. In Schuler, H. Lehrbuch der Personalpsychologie, Hrsg., 397-431. Göttingen, Bern, Toronto, Seattle: Hogrefe.

Maukisch, H. (1978). Einführung in die Eignungsdiagnostik. In Mayer, A. (Ed.), Organisationspsychologie. Stuttgart: Poeschl.

Maukisch, H. (1989). Informationswert und Ökonomie der diagnostischen Prinzipien von Assessment Center-Systemen zur Erfassung von Managementpotenzial. In Lattmann, C. (Hrsg.), Das Assessment-Center-Verfahren der Eignungsbeurteilung, 259-289. Heidelberg: Physica.

Mayfield, E. C. (1964). The selection interview: A re-evaluation of published research. Personnel Psychology, 17, 239-260.

McClelland, D. C. (1961). The achieving society. Princeton, NJ: Van Nostrand.

McClelland, D. C. (1973). Testing for competence rather than for 'intelligence'. American Psychologist, 28, 1-14.

McClelland, D. C. (1985). How motives, skills, and values determine what people do. American Psychologist, 41, 812-825.

McClelland, D. C., Atkinson, J. W., Clark, R. A., et al. (1953). The achievement motive. New York: Appleton-Century-Crofts.

McClelland, D. C. & Boyatzis, R. E. (1982). Leadership motive pattern and long term success in management. Journal of Applied Psychology, 67, 737-743.

McCloy, R. A., Campbell, J. P. & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. Journal of Applied Psychology, 79(4), 493-505.

McCormick, E. J. & Jeanneret, P. R. (1988). Position analysis questionnaire (PAQ). In Gael, S. (Ed.), The job analysis handbook for business, industry, and government. New York: Wiley.

McCrae, R. R. & John, O. P. (1992). An introduction to the five-factor model and its applications. Journal of Personality, 60, 175-215.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., et al. (1994). The Validity of Employment Interviews: A Comprehensive Review and Meta-Analysis. Journal of Applied Psychology, 79 (4), 599-616.

McGraw, K. O. & Wong, S. P. (1996). Forming Inferences About Some Intraclass Correlation Coefficients. Psychological Methods, 1 (1), 30-46.

McHenry, J. J., Hough, L. M., Toquam, J., et al. (1990). Project A validity results: the relationship between predictor and criterion domains. Personnel Psychology, 43, 335-354.

McKinley, W. & Scherer, A. G. (2000). Some unanticipated consequences of organizational restructuring. Academy of Management Review, 25(4), 735-752.

McManus, M. & Kelly, M. (1999). Personality measures and biodata: evidence regarding their incremental predictive value in the life insurance industry. Personnel Psychology, 52, 137-148.

Meehl, P. (1954). Clinical vs. statistical prediction. Minneapolis: University of Minnesota Press.

Meehl, P. (1965). See over sign: The first good example. Journal of Experimental Research in Personality, 1, 27-32.

Meffert, H. (1990). Unternehmensberatung und Unternehmensführung - eine empirische Bestandsaufnahme. Die Betriebswirtschaft, 50(2), 181-197.

Meyer, A. & Mattmüller, R. (1987). Qualität von Dienstleistungen – Entwurf eines praxisorientierten Qualitätsmodells. Marketing ZFP Zeitschrift für Forschung und Praxis, 9(3), 187-195.

Michel, L. & Conrad, W. (1982). Theoretische Grundlagen psychometrischer Tests. In Groffmann, K. J. M. & Michel, L. (Hrsg.), Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie II: Psychologische Diagnostik, Band 1: Grundlagen psychologischer Diagnostik, 11-29. Göttingen: Hogrefe.

Moser, K. & Reuter, N. (2001). Wird durch Aggregation von Interviewerdaten die Validität von Einstellungsinterviews unterschätzt? Zeitschrift für Arbeits- und Organisationspsychologie, 4, 188-201.

Moser, K. & Rhyssen, D. (2001). Referenzen als eignungsdiagnostische Methode. Zeitschrift für Arbeits- und Organisationspsychologie, 45, 40-46.

Moser, K. & Schuler, H. (1989). The Nature of Psychological Measurement. In Herriot, P. Assessment and Selection in Organizations, 281-305. Chichester, New York, Brisbane, Toronto, Singapore: John Wiley & Sons.

Motowidlo, S. J., Borman, W. C. & Schmit, M. J. (1997). A theory of individual differences in task and conceptual performance. Human Performance, 10, 71-83.

Motowidlo, S. J. & Burnett, J. R. (1995). Aural and visual sources of validity in structured employment interviews. Organizational Behaviour & Human Decision Process, 61, 239-249.

Motowidlo, S. J., Carter, G. W., Dunnette, M. D., et al. (1992). Studies of the structured behavioural interview. Journal of Applied Psychology, 77, 571-587.

Mount, M. K. & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in human resources management. In Rowland, K. M. & Ferris, G. R. (Eds.), Research in personnel and human resources management, 13, 153-200. Greenwick: JAI Press.

Mount, M. K. & Barrick, M. R. (1998). Five reasons why the "big five" article has been frequently cited. Personnel Psychology, 51, 849-857.

Mount, M. K., Barrick, M. R. & Stewart, G. L. (1998). Five-Factor Model of Personality and Performance in Jobs Involving Interpersonal Interactions. Human Performance, 11 (2/3), 145-165.

Mount, M. K., Judge, T. A., Scullen, S., et al. (1998). Trait, rater and level effects in 360-degree performance ratings. Personnel Psychology, 51, 557-576.

Mount, M. K., Witt, L. A. & Barrick, M. L. (2000). Incremental validity of empirical keyed Biodata scales over GMA and the five factor personality constructs. Personnel Psychology, 53, 299-323.

Mugler, J. & Lampe, R. (1987). Betriebswirtschaftliche Beratung von Klein- und Mittelbetrieben. Betriebswirtschaftliche Forschung und Praxis, 39(6), 477-493.

Mumford, M. D. & Owens, W. A. (1987). Methodology review: principles, procedures, and findings in the application of background measures. Applied Psychological Measurement, 11, 1-31.

Mumford, M. D., Snell, A. F. & Reiter-Palmon, R. (1994). Personality and background data: life history and self-concepts in an ecological model. In Stokes, G., Mumford, M. & Owens, W. (Eds.), Biodata handbook: theory, research, and use of biographical information in selection performance prediction, 583-625. Palo Alto: CPP Books.

Münsterberg, H. (1912). Psychologie und Wirtschaftsleben. Leipzig: Barth.

Murphy, K. R. (1986). When your top choice turns you down: The effect of rejected offers on the utility of selection tests. Psychological Bulletin, 99, 133-138.

Murphy, K. R. (1996). Individual differences and behaviour in organizations. Much more than *g*. In Murphy, K. R. (Ed.), Individual differences and behaviour in organizations, 3-30. San Francisco: Jossey-Bass.

Murphy, K. R., Cleveland, J. N., Skattebo, A. L., et al. (2004). Raters Who Pursue Different Goals Give Different Ratings. Journal of Applied Psychology, 89(1), 158-164.

Murphy, K. R. & De Shon, R. (2000a). Interrater Correlations do not estimate the reliability of job performance ratings. Personnel Psychology, 53, 873-900.

Murphy, K. R. & De Shon, R. (2000b). Progress in psychometrics: Can I-O-psychology catch up? Personnel Psychology, 33, 913-924.

Murray, C. (1998). Income, inequality and IQ. Washington: Enterprise Institute.

Nachreiner, F. & Müller, G. F. (1995). Führungsmotivation. In Sarges, W. (Hrsg.), Managementdiagnostik, 2. vollst. überarb. und erw. Auflage, 271-278. Göttingen: Hogrefe.

Naerssen, R. F. v. (1965). Application of the Decision Theoretical Approach to the Selection of Drivers. In Cronbach, L. J. & Gleser, G. S. Psychological Tests and Personnel Decisions, 273-290. Urbana: University of Illinois Press.

Naylor, J. C. & Shine, L. C. (1965). A table determining the increase in mean criterion score obtained by using a selection device. Journal of Industrial Psychology, 3, 33-42.

Niewiem, S. & Richter, A. (2004). The changing balance of power in the consulting market. Business Strategy Review, 15 (1), 8-13.

Noack, H. & Petermann, F. (1992). Entscheidungstheorie. In Jäger, R. S. & Petermann, F. (Hrsg.), Psychologische Diagnostik, 2., veränderte Auflage, 295-310. Weinheim: Psychologie Verlags Union.

O*Net-Center (2005). onet_content_model_reference.txt. http://www.onetcenter.org/database.html#download,

Oesterreich, R., Leitner, K. & Resch, M. (2000). Analyse psychischer Anforderungen und Belastungen in der Produktionsarbeit. Das Verfahren RHIA/VERA-Produktion. Göttingen: Hogrefe.

O'Leary, B. S. (1980). College GPA as an indicator of occupational success: An update (PRR 80-23). Washington, DC: United States Office of Personnel Management, Personnel Research and Development Center.

Ones, D. S. & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. Journal of Organizational Behavior, 17, 609-626.

Ones, D. S., Viswesvaran, C. & Dilchert, S. (2004). Cognitive Ability in Selection Decisions. In Wilhelm, O. & Engle, R. (Eds.), Understanding and Measuring Intelligence. London: Sage.

Ones, D. S., Viswesvaran, C. & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. Journal of Applied Psychology, 78, 679-703.

Owens, W. A. (1976). Background data. In Dunette, M. D. (Ed.), Handbook of industrial and organizational psychology, 609-644. Chicago: Rand McNally.

Pawlik, K. (1976). Modell und Praxisdimensionen psychologischer Diagnostik. In Pawlik, K. (Hrsg.), Diagnose der Diagnostik: Beiträge zur Diskussion der psychologischen Diagnostik in der Verhaltensmodifikation, 13-43. Stuttgart: Klett.

Petermann, F. (1992). Psychologische Diagnostik in der Anwendung. In Jäger, R. S. & Petermann, F. Psychologische Diagnostik, 493-495. Weinheim: Psychologie Verlags Union.

Peterson, N. G., Mumford, M. D., Borman, W. C., et al. (2001). Understanding Work Using the Occupational Information Network (O*Net): Implications for Practice and Research. Personnel Psychology, 54, 451-492.

Phillips, J. S. (2000). The Consultant´s Scorecard. New York: McGraw-Hill.

Pulakos, E. D. & Schmitt, N. (1995). Experience-based and situational interview questions: Studies of validity. Personal Psychology, 48, 289-308.

Pulakos, E. D., Schmitt, N., Whitney, D., et al. (1996). Individual differences in interviewer ratings: The impact of standardization, consensus discussion, and sampling error on validity of a structured interview. Personal Psychology, 49, 85-102.

Pursell, E. D., Campion, M. A. & Gaylord, S. R. (1980). Structured interviewing: Avoiding selection problems. Personnel Journal, 59, 907-912.

Raju, N. S., Burke, M. J. & Normand, J. (1990). A new approach for utility analysis. Journal of Applied Psychology, 75, 3-12.

Raju, N. S., Steinhaus, S. D., Edwards, J. E., et al. (1991). A logistic regression model for personnel selection. Applied Psychological Measurement, 15, 139-152.

Rastetter, D. (1999). Bewerbungsunterlagenscreening. Analyse eines hochselektiven, kaum erforschten Auswahlinstrumentes. , , . Zeitschrift für Arbeitswissenschaft, 53, 37-44.

Rauschenberger, J. M. & Schmidt, F. (1987). Measuring the economic impact of human resource programs. Journal of Business and Psychology, 2(1), 50-59.

Raymark., P. H., Schmit, M. J. & Guion, R. M. (1997). Identifying Potentially Useful Personality Constructs For Employee Selection. Personnel Psychology, 50, 723-736.

Ree, M. J. & Earles, J. A. (1991). Predicting training success: Not much more than *g*. Personnel Psychology, 44, 321-332.

Ree, M. J. & Earles, J. A. (1992). Intelligence is the best predictor of job performance. Current Directions in Psychological Science, 1, 86-89.

Ree, M. J. & Earles, J. A. (1994). The ubiquitous predictiveness of *g*. In Rumsey, M. G., Walker, C. B. & Harris, J. H. Personnel Selection and Classification. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ree, M. J., Earles, J. A. & Teachout, M. S. (1994). Predicting Job Performance: Not Much More Than *g*. Journal of Applied Psychology, 79 (4), 518-524.

Reilly, R. R. & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. Personnel Psychology, 35, 1-62.

Rettler, H. (1992). Psychologische Diagnostik. In Jäger, R. S. & Petermann, F., 221-26. Weinheim: Psychologie Verlags Union.

Rich, J. R. & Boudreau, J. W. (1987). The effects of variability and risk in selection utility analysis: an empirical comparison. Personnel Psychology, 40, 55-84.

Risch, S. & Sommer, C. (1996). ...und raus bist du! managermagazin, 26, 5, 220-229.

Robertson, I. T. & Kinder, A. (1993). Personality and job competences: The criterion-related validity of some personality variables. Journal of Occupational and Organizational Psychology, 66, 225-244.

Robertson, I. T. & Smith, M. (2001). Personnel selection. Journal of Occupational and Organizational Psychology, 74, 441-472.

Roche, W. J. J. (1965). A Dollar Criterion in Fixed Treatment Employee Selection. In Cronbach, L. J. & Gleser, G. S. Psychological Tests and Personnel Decisions, 254-268. Urbana, Chicago, London: University of Illinois Press.

Rolfhus, E. L. & Ackerman, P. L. (1999). Assessing Individual Differences in Knowledge: Knowledge, Intelligence, and Related Traits. Journal of Educational Psychology, 91(3), 511-526.

Rossi, P. H. & Freeman, H. E. (1985). Evaluation: A systematic Approach. New York: Sage.

Roth, E. & Bobko, P. (1997). A research agenda for multi-attribute utility analysis in human resource management. Human Resource Management Review, 7(3), 341-368.

Roth, P., Segers, A. & Wright, P. (2000). The acceptance of utility analysis: Designing a model. Paper presented at the academy of Management Annual Meeting, Toronto.

Roth, P. L. (1994). Multi-attribute utility analysis using the Promes approach. Journal of Business and Psychology, 9(1), 69-80.

Roth, P. L., Pritchard, R. D., Stout, J. D., et al. (1994). Estimating the impact of variable costs on SDy in complex situations. Journal of Business and Psychology, 8(4), 437-454.

Russel, C. J. (1985). Individual decision process in an assessment center. Journal of Applied Psychology, 70, 737-746.

Ryan, A. M., McFarland, L., Baron, H., et al. (1999). An international look at selection practices: nation and culture as explanations for variability. Personnel Psychology, 52, 359-391.

Sackett, P. R. & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. Journal of Applied Psychology, 67, 401-110.

Sackett, P. R., Laczo, R. M. & Arvey, R. D. (2002). The effects of range restriction on estimates of criterion interrater reliability: Implications for validation research. Personnel Psychology, 55, 807-825.

Sagie, A. & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories, and assessment center construct validity. Journal of Occupational and Organizational Psychology, 70, 103-108.

Salas, E. & Canon-Bowers, J. A. (2001). The sciences of training: A decade of progress. Annual Review of Psychology, 52, 471-499.

Salgado, J. F. (1997). The Five Factor Model of Personality and job performance in the European community. Journal of Applied Psychology, 82 (1), 30-43.

Salgado, J. F., Anderson, N., Moscoso, S., et al. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. Personnel Psychology, 56, 573-605.

Salgado, J. F., Anderson, N., Moscoso, S., et al. (2003). A meta-analytic study of general mental ability for different occupations in the European community. Journal of Applied Psychology, 88, 1068-1081.

Salgado, J. F. & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. European Journal of Work & Organizational Psychology, 11(3), 299-324.

Sandberg, R. & Werr, A. (2003). The Three Challenges of Corporate Consulting. MIT Sloan Management Review, Spring 2003, 59-66.

Sands, W. A. (1973). A Method of Evaluating Alternative Recruiting-Selection Strategies: The CAPER-Modell. Journal of Applied Psychology, 57, 222-227.

Sangüesa, M. (2003). Modell zur Evaluierung von Beratungsprojekten (Diss.). Berlin: Fakultät V Verkehrs- und Maschinensysteme der Technischen Universität Berlin.

Sarbin, T. R. (1944). The logic of prediction in psychology. Psychological Review, 51, 210-228.

Sarges, W. (1994). Eignungsdiagnostische Überlegungen für den Management Bereich. In Bartussek, D. & Amelang, M. (Hrsg.), Fortschritte der Differentiellen Psychologie und psychologischen Diagnostik, 415-434. Göttingen: Hogrefe.

Sarges, W. (1995a). Interviews. In Sarges, W. (Hrsg.), Management-Diagnostik, (2. Aufl.), 475-489. Göttingen: Hogrefe.

Sarges, W. (1995b). Lernpotential-AC. In Sarges, W. (Hrsg.), Management-Diagnostik, 2. Aufl., 728-739. Göttingen: Hogrefe.

Sarges, W. (1996). Weiterentwicklung der Assessment Center-Methode. Göttingen: Verlag für Angewandte Psychologie.

Sarges, W. & Wottawa, H. (2001). Handbuch wirtschaftspsychologischer Testverfahren. Lengerich: Pabst Science Publishers.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. Psychological Bulletin, 66, 178-200.

Scherm, M. (2004). 360°-Beurteilung. In Schuler, H. Beurteilung und Förderung beruflicher Leistung. Göttingen
Bern
Toronto
Seattle: Hogrefe.

Schettgen, P. (1996). Arbeit, Leistung, Lohn. Analyse- und Bewertungsmethoden aus sozioökonomischer Perspektive. Stuttgart: Enke.

Scheurer, H. (1992). Diagnostik als Testung. In Jäger, R. S. & Petermann, F. Psychologische Diagnostik, 257-263. Weinheim: Psychologie Verlags Union.

Schmidt, F. (2002). The Role of General Cognitive Ability and Job Performance: Why There Cannot Be a Debate. Human Performance, 15(1/2), 187-210.

Schmidt, F. L. (1988). The problem of group differences in ability scores in employment selection. Journal of Vocational Behaviour, 33, 272-292.

Schmidt, F. L. & Hunter, J. E. (1992). Development of a Causal Model of Processes Determining Job Performance. Current Directions in Psychological Science, 1 (3), 89-92.

Schmidt, F. L. & Hunter, J. E. (1993). Tacit Knowledge, Practical Intelligence, General Mental Ability, and Job Knowledge. Current Directions in Psychological Science, 8-9.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology. Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124, 262-274.

Schmidt, F. L., Hunter, J. E., Hirsh, H. R., et al. (1985). Forty questions about validity generalization and meta-analysis. Personnel Psychology, 38, 697-798.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., et al. (1979). Impact of Valid Selection Procedures on Work-Force Productivity. Journal of Applied Psychology, 6, 609-629.

Schmidt, F. L., Hunter, J. E. & Ones, D. S. (1992). Personnel Selection. Annual Review of Psychology, 43, 627-670.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., et al. (1988). Joint relation of experience and ability with job performance: Test of three hypothesis. Journal of Applied Psychology, 73, 46-57.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., et al. (1986). The economic impact of job selection methods on size, productivity, and payroll costs of the federal work force: an empirical based demonstration. Personnel Psychology, 39(1), 1-29.

Schmidt, F. L., Hunter, J. E. & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. Journal of Applied Psychology, 38, 166-185.

Schmidt, F. L., Hunter, J. E. & Urry, V. W. (1976). Statistical power in criterion-related validation studies. Journal of Applied Psychology, 61, 473-485.

Schmidt, F. L., Law, K., Hunter, J. E., et al. (1993). Refinements in validity generalization methods: Implications for the situational specifity hypothesis. Journal of Applied Psychology, 78, 3-12.

Schmidt, F. L., Mack, M. J. & Hunter, J. E. (1984). Selection Utility in the Occupation of U.S. Park Ranger for Three Modes of Test Use. Journal of Applied Psychology, 69, 490-497.

Schmidt, F. L. & Rader, M. (1999). Exploring the boundary conditions for interview validity: Meta-analytic validity findings for a new interview type. Personnel Psychology, 52, 445-464.

Schmidt, F. L., Viswesvaran, C. & Ones, D. S. (2000). Reliability is not validity and validity is not reliability. Personnel Psychology, 53, 901-912.

Schmidt, F. L. & Zimmermann, R. D. (2004). A Counterintuitive Hypothesis About Employment Interview Validity and Supporting Evidence. Journal of Applied Psychology, 89(3), 553-561.

Schmidt-Atzert, L., Deter, B. & Jaeckel, S. (2004). Prädiktoren von Ausbildungserfolg: Allgemeine Intelligenz (g) oder spezifische kognitive Fähigkeiten? Zeitschrift für Personalpsychologie, 3(4), 147-158.

Schmidtke, A. (1992). Diagnostik als Inventarisieren. In Jäger, R. S. & Petermann, F. Psychologische Diagnostik, 263-267. Weinheim: Psychologie Verlags Union.

Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. Personnel Psychology, 29, 79-101.

Schmitt, N. & Chan, D. (1998). Personnel Selection: A Theoretical Approach. Thousand Oaks: Sage.

Schmitt, N., Gooding, R. Z., Noe, R. D., et al. (1984). Metaanalysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Schneider, R. J. & Hough, L. M. (1995). Personality and Industrial/ Organizational Psychology. In Cooper, C. L. & Robertson, I. T. (Eds.), International Review of Industrial and Organizational Psychology 10, 75-129. Chichester: Wiley & Sons.

Schneider, R. J., Hough, L. M. & Dunette, M. D. (1996). Broadside by broad traits: how to sink science in five dimensions or less. Journal of Organizational Behavior, 17 (6), 639-655.

Scholz, G. & Schuler, H. (1993). Das nomologische Netzwerk des Assessment Centers: Eine Metaanalyse. Zeitschrift für Arbeits- und Organisationspsychologie, 37, 73-85.

Schott, F. (1992). Kriteriumsorientierte Diagnostik. In Jäger, R. S. & Petermann, F. Psychologische Diagnostik, 226-232. Weinheim: Psychologie Verlags Union.

Schrädler, J. (1996). Unternehmensberatung aus organisatorischer Sicht. Wiesbaden: Deutscher Universitäts-Verlag.

Schuler, H. (1988). Berufseignungsdiagnostik. Zeitschrift für Differentielle und Diagnostische Psychologie, 9, 201-213.

Schuler, H. (1991a). Auswahl von Mitarbeitern. In Rosenstiel, L. v., Regnet, E. & Domsch, M. (Hrsg.), Führung von Mitarbeitern - Handbuch für erfolgreiches Personalmanagement, 100-125. Stuttgart: Schäffer.

Schuler, H. (1991b). Der Funktionskreis "Leistungsförderung"- eine Skizze. In Schuler, H. (Hrsg.), Beurteilung und Förderung beruflicher Leistung, 171-190. Stuttgart: Verlag für Angewandte Psychologie.

Schuler, H. (1992). Assessment-Center als Auswahl und Entwicklungsinstrument: Einleitung und Überblick. In Schuler, H. & Stehle, W. Assessment-Center als Methode der Personalentwicklung. Beiträge zur Organisationspsychologie, 2. Aufl. Göttingen, Stuttgart: Verlag für Angewandte Psychologie.

Schuler, H. (1992a). Das Multimodale Einstellungsinterview. Diagnostica, 38, 281-300.

Schuler, H. (1993). Lehrbuch der Organisationspsychologie. In Schuler, H. (Hrsg.), Lehrbuch der Organisationspsychologie, (1. Aufl.). Bern: Huber.

Schuler, H. (2000). Psychologische Personalauswahl. Göttingen; Bern; Toronto; Seattle: Hogrefe.

Schuler, H. (2001a). Arbeits- und Anforderungsanalyse. In Schuler, H. (Hrsg.), Lehrbuch der Personalpsychologie, 44-61. Göttingen, Bern, Toronto, Seattle: Hogrefe.

Schuler, H. (2001b). Lehrbuch der Personalpsychologie. In Schuler, H. (Hrsg.), Lehrbuch der Personalpsychologie. Göttingen; Bern: Hogrefe.

Schuler, H. (2002). Das Einstellungsinterview. Göttingen, Bern, Toronto, Seattle: Hogrefe.

Schuler, H. (2004a). Der Prozess der Urteilsbildung und die Qualität der Beurteilungen. In Schuler, H. (Hrsg.), Beurteilung und Förderung beruflicher Leistung, 2., vollst. überarb. und erw. Auflage, 33-60. Göttingen, Bern, Toronto, Seattle: Hogrefe.

Schuler, H. (2004b). Leistungsbeurteilung - Gegenstand, Funktionen, Formen. In Schuler, H. (Hrsg.), Beurteilung und Förderung beruflicher Leistung, 2., vollst. überarb. und erw. Auflage, 1-24. Göttingen, Bern, Toronto, Seattle: Hogrefe.

Schuler, H., Frier, D. & Kaufmann, M. (1991). Validität, Praktikabilität und Akzeptanz eignungsdiagnostischer Verfahren in der Einschätzung der Verwender. In Schuler, H. & Funke, U. (Hrsg.), Eignungsdiagnostik in Forschung und Praxis, Band 10, 26-32. Stuttgart: Verlag für Angewandte Psychologie.

Schuler, H., Frier, D. & Kaufmann, M. (1993). Personalauswahl im europäischen Vergleich. Göttingen: Verlag für Angewandte Psychologie.

Schuler, H. & Funke, U. (1989). Berufseignungsdiagnostik. In Roth, E. (Hrsg.), Organisationspsychologie, Enzyklopädie der Psychologie, III, 3, 281-320. Göttingen: Hogrefe.

Schuler, H. & Funke, U. (1991a). Eignungsdiagnostik in Forschung und Praxis. Psychologische Information für Auswahl, Beratung und Förderung von Mitarbeitern. In Schuler, H. & Stehle, W. (Hrsg.), Beiträge zur Organisationspsychologie, 10. Stuttgart: Verlag für angewandte Psychologie.

Schuler, H. & Funke, U. (1991b). Zu diesem Band. In Schuler, H. & Funke, U. Eignungsdiagnostik in Forschung und Praxis, V-VII. Stuttgart: Verlag für Angewandte Psychologie.

Schuler, H. & Funke, U. (1993). Diagnose beruflicher Eignung und Leistung. In Schuler, H. (Hrsg.), Lehrbuch Organisationspsychologie, 253-283. Bern: Huber.

Schuler, H. & Funke, U. (1995). Diagnose beruflicher Eignung und Leistung. In Schuler, H. (Hrsg.), Lehrbuch Organisationspsychologie, 235-283. Göttingen: Verlag Hans Huber.

Schuler, H., Moser, K., Funke, U., et al. (1995). Personalauswahl in Forschung und Entwicklung - Eignung und Leistung von Wissenschaftlern und Ingenieuren. Göttingen, Bern, Toronto, Seattle: Hogrefe.

Schuler, H., Muck, P. M., Hell, B., et al. (2004). Entwicklung eines multimodalen Systems zur Beurteilung von Individualleistung. In Schuler, H. (Hrsg.), Beurteilung und Förderung beruflicher Leistung, 2., vollst. überarb. und erw. Auflage, 125-132. Göttingen, Bern, Toronto, Seattle: Hogrefe.

Schuler, H. & Prochaska, M. (2000). Entwicklung und Konstruktvalidierung eines berufsbezogenen Leistungsmotivationstests. Diagnostica, 46, 61-72.

Schuler, H. & Schmitt, N. (1987). Multimodale Messung in der Personalpsychologie. Diagnostica, 33, 259-271.

Schuler, H. & Stehle, W. (1983). Neuere Entwicklungen des Assessment-Center-Ansatzes - beurteilt unter dem Aspekt der sozialen Validität. Psychologie und Praxis. Zeitschrift für Arbeits- und Organisationspsychologie, 27, 33-44.

Scriven, M. (1991). Evaluation Thesaurus. Thousand Oaks: Sage.

Sechrest, L. & Figueredo, A. J. (1993). Program Evaluation. Annual Review of Psychology, 44, 645-674.

Seibt, H. & Kleinmann, M. (1991). Personalauswahl von Bewerbern: Derzeitiger Stand und Alternativen. In Schuler, H. & Funke, U. Eignungsdiagnostik in Forschung und Praxis, 174-177. Stuttgart: Verlag für angewandte Psychologie.

Selchert, M. (1997). Organisationsstrukturen und Professionalität. Formen und Funktionen professioneller Inhouse Dienstleistungen. Hamburg: Kovac.

Shea, D. L., Lubinski, D. & Benbow, C. P. (2001). Importance of Assessing Spatial Ability in Intellectually Talented Young Adolescents: A 20-Year Longitudinal Study. Journal of Educational Psychology, 93, 604-614.

Sieglen, J. (1998). Erklärung und Vorhersage außergewöhnlicher beruflicher Leistung. Frankfurt am Main; Berlin; New York; Paris; Wien: Lang.

Siemann, C. (2004). Karriere und Management - Auf der Suche nach Sicherheit. Handelsblatt, 74, 6.

Silverman, W. H., Dalessio, A., Woods, S. B., et al. (1986). Influence of assessment center methods on assessors ratings. Personnel Psychology, 39, 565-578.

Spearman, C. (1904). "General Intelligence" objectively determined and measured. American Journal of Psychology, 15, 201-293.

Spenner, K. I., Otto, L. B. & Call, V. R. A. (1980). Estimates of Third Edition DOT Job Characteristics for 1970 Census Occupation-Industry Categories. Omaha, NE: Boys Town Center.

Stamps, D. (2000). Measuring minds. Training, May, 1-7.

Starbuck, W. H. (1992). Learning by knowledge-intensive firms. Journal of Management Studies, 29, 713-740.

Stehle, W. (1986). Personalauswahl mittels biographischer Fragebögen. In Schuler, H. & Stehle, W. (Hrsg.), Biographische Fragebögen als Methode der Personalauswahl, 17-57. Stuttgart: Verlag für Angewandte Psychologie.

Sternberg, R. J. (1997a). Managerial Intelligence: Why IQ Isn't Enough. In Vecchio, R. P. (Eds.), Leadership: Understanding the Dynamics of Power and Influences in Organizations. Notre Dame, Indiana: University of Notre Dame Press.

Sternberg, R. J. (1997b). The triarchic theory of intelligence. In Flanagen, D. P., Genshaft, J. L. & Harrison, P. L. (Eds.), Contemporary intellectual assessment: Theories, tests, and issues. New York: Wiley.

Sternberg, R. J., Forsthe, G. B., Hedlund, J., et al. (2000). Practical Intelligence in everyday life. New York: Cambridge University Press.

Sternberg, R. J. & Hedlund, J. (2002). Practical Intelligence, g, and Work Psychology. Human Performance, 15(1/2), 143-160.

Steyrer, J. (1991). Unternehmensberatung - Stand der deutschsprachigen Theoriebildung und empirischen Forschung. In Hofmann, M. Theorie und Praxis der Unternehmensberatung: Bestandsaufnahme und Entwicklungsperspektiven, 1-44. Heidelberg:

Stief, M. (2001). Selbstwirksamkeitserwartungen, Ziele und Berufserfolg. Aachen: Shaker Verlag.

Strasser, H. (1993). Unternehmensberatung aus der Sicht des Kunden – Eine resultatorientierte Gestaltung der Beratungsbeziehung und des Beratungsprozesses (Band 177). Zürich: Schulthess Polygraphischer Verlag.

Strauß, B. & Kleinmann, M. (1995). Computersimulierte Szenarien in der Personalarbeit. Göttingen: Hogrefe/Verlag für Angewandte Psychologie.

Stufflebeam, D. L. (2000). The CIPP model for evaluation. In Stufflebeam, D. L., Madaus, G. F. & Kellaghan, T. (Eds.), Evaluation Models, 2nd ed. Boston: Kluwer Academic Publishers.

Stufflebeam, D. L. (2001). Evaluation models. San Francisco: Jossey-Bass.

Suchman, E. A. (1967). Evaluative research: Principle and practice in public service and social action programs. New York: Russel Sage Foundation.

Sujan, H., Sujan, M. & Bettman, J. R. (1988). Knowledge Structure Differences Between More Effective and Less Effective Salespeople. Journal of Marketing Research, 25, 81-86.

Sujan, H., Sujan, M. & Bettman, J. R. (1991). The Practical Know-how of selling: Differences in knowledge content between more-effective and less effective performers. Marketing Letters, 2(4), 367-378.

Süß, H.-M., Oberauer, K., Wittmann, W. W., et al. (2002). Working-memory capacity explains reasoning ability - and a little bit more. Intelligence, 30, 261-288.

Szymanski, D. M. (1988). Determinants of Selling Effectiveness: The importance of declarative knowledge to the personal selling concept. Journal of Marketing, 52, 64-77.

Tack, W. H. (1980). Zur Theorie psychometrischer Verfahren - Formalisierung der Erfassung von Situationsunabhängigkeit und Veränderung. Zeitschrift für Differentielle und Diagnostische Psychologie, 1, 87-106.

Taylor, H. C. & Russel, J. T. (1939). The relationship between validity coefficients to the practical effectiveness in selection. Journal of Applied Psychology, 23, 565-578.

Tenopyr, M. L. (1994). Big Five, structural modeling, and item response theory. In Stokes, G., Mumford, M. & Owens, W. Biodata handbook: theory, research, and use of biographical information in selection and performance prediction, 519-534. Palo Alto: CPP Books.

Tett, R. P., Jackson, D. N. & Rothstein, M. G. (1991). Personality measures as predictors of job performance: a meta analytic review. Personnel Psychology, 44, 703-742.

Tett, R. P., Jackson, D. N., Rothstein, M. G., et al. (1994). Meta-analysis of personality-job performance relations: A reply to Ones, Mount, Barrick and Hunter (1994). Personnel Psychology, 47, 157-168.

Thorndike, R. L. (1949). Personnel selection. Test and measurement techniques. New York: Wiley.

Thorndike, R. L. (1986). The role of general ability in prediction. Journal of Vocational Behaviour, 29, 332-339.

Thornton, G. C., Gaugler, B. B., Rosenthal, D. B., et al. (1987). Die prädiktive Validität des Assessment Centers - eine Metaanalyse. In Schuler, H. & Stehle, W. (Hrsg.), Assessment Center als Methode der Personalentwicklung. Göttingen: Hogrefe/Verlag für Allgemeine Psychologie.

Thurstone, L. L. (1931). Multiple factor analysis. Psychological Review, 38, 406-427.

Thurstone, L. L. (1935). The vectors of mind: multiple factor analysis for the isolation of primary traits. Chicago: University of Chicago Press.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In Sternberg, R. L. (Ed.), The nature of creativity, 43-75. Cambridge: Cambridge University Press.

Trost, G. (1996). Interview. In Pawlik, K. (Hrsg.), Enzyklopädie der Psychologie. Grundlagen und Methoden der Differenziellen Psychologie (C/VIII/1). Göttingen: Hogrefe.

Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hursch, Hammond and Hursch, and by Hammond, Hursch and Todd. Psychological Review, 71, 528-530.

Tullar, W. L., Mullins, T. W. & Caldwell, S. A. (1979). Effects of interview length and applicant quality on interview decision time. Journal of Applied Psychology, 64, 669-674.

Tziner, A., Meir, E. I., Dahan, M., et al. (1994). An investigation of the predictive validity and economic utility of the assessment center for the high-management level. Canadian Journal of Behavioural Science, 26, 228-245.

Ulrich, L. & Trumbo, D. (1965). The selection interview since 1949. Psychological Bulletin, 63, 100-116.

Undheim, O. J. & Horn, J. L. (1977). Critical evaluation of Guilford´s structure-of-intellect-theory. Intelligence, 1, 65-81.

Van Iddekinge, C. H., Raymark, P. H., Eidson, C. E., et al. (2004). What Do Structured Selection Interviews Really Measure? The Construct Validity of Behaviour Description Interviews. Human Performance, 17(1), 71-93.

Vance, R. J. & Colella, A. (1990). The Utility Analysis. Human Performance, 3(2), 123-139.

Vinchur, A. J., Schippmann, J. S., Switzer, F. S., et al. (1998). A meta analytic review of predictors of job performance for salespeople. Journal of Applied Psychology, 83 (4), 586-597.

Viswesvaran, C., Ones, D. S. & Schmidt, F. L. (1996). Comparative Analysis of the Reliability of Job Performance Ratings. Journal of Applied Psychology, 81 (5), 557-574.

Viswesvaran, C., Schmidt, F. L. & Ones, D. S. (2002). The Moderating Influence of Job Performance Dimensions on Convergence of Supervisory and Peer Ratings of Job Performance: Unconfounding Construct-Level Convergence and Rating Difficulty. Journal of Applied Psychology, 87 (2), 345-354.

Viswesvaran, C., Schmidt, F. L. & Ones, D. S. (2005). Is There a General Factor in Ratings of Job Performance? A Meta-Analytic Framework for Disentangling Substantive and Error Influences. Journal of Applied Psychology, 90(1), 108-131.

Wagner, R. (1949). The employment interview: A critical summary. Personnel Psychology, 2, 17-46.

Wainer, H. A. & Rubin, I. M. (1971). Motivation of research and development entrepreneurs: Determinants of company success. In Kolb, D. A., Rubin, I. M. & McIntire, J. (Eds.), Organizational Psychology, 131-139. Englewood Cliffs, NJ: Prentice-Hall.

Walger, G., Achatzi, H. J. & Benkert, W. (1999). Formen der Unternehmensberatung. In (Hrsg.). Köln: Schmidt.

Weatherly, L. A. (2003). The Value of People. SHRM Research Quartely, 3, 1-10.

Webb, R. M., Lubinski, D. & Benbow, C. P. (2002). Mathematically Facile Adolescents With Math-Science Aspirations: New Perspectives on Their Educational and Vocational Development. Journal of Educational Psychology, 94,4, 785-794.

Weekley, J. A., O'Connor, E. J., Frank, B., et al. (1985). A comparison of three methods of estimating the standard deviation of performance in dollars. Journal of Applied Psychology, 79, 122-126.

Weinert, A. B. (1992). Lehrbuch der Organisationspsychologie (3. Aufl.). Weinheim: Psychologie Verlags Union.

Weiss, C. H. (2001). Evaluation. Upper Saddle River: Prentice Hall.

Weitz, B. A., Sujan, H. & Sujan, M. (1986). Knowledge, Motivation, and Adaptive Behaviour: A Framework for Improving Selling Effectiveness. Journal of Marketing, 50, 174-191.

Wentling, T. L. & Lawson, T. E. (1975). Evaluating occupational education and training programs (2nd ed.). Boston, MA: Allyn & Bacon.

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

Westhoff, K., Hellfritsch, L. J., Hornke, L. F., et al. (Hrsg.) (2004). Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430. Lengerich, Berlin, Bremen, Miami, Riga, Viernheim, Wien, Zagreb: Pabst Science Publishers.

Westhoff, K. & Kluck, M. L. (2003). Psychologische Gutachten schreiben und beurteilen (4. vollst. überarb. und erw. Auflage). Berlin: Springer.

Whyte, G. & Latham, G. P. (1997). The Futility of Utility Analysis Revisited: When Even an Expert Fails. Personnel Psychology, 50, 601-610.

Wickel-Kirsch, K. (2002). Personalauswahl mit der Unterstützung von Personalkennziffern. Personal, 7, 50-55.

Wiesner, W. H. & Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure an the validity of the employment interview. Journal of Occupational Psychology, 61, 275-290.

Wiggins, J. S. (1973). Personality and prediction: principles of personality assessment. Reading, Mass.: Addison-Wesley.

Wilhelm, O., Schulze, R., Schmiedek, F., et al. (2003). Interindividuelle Unterschiede im typischen intellektuellen Engagement. Diagnostica, 49, 49-60.

Wittmann, W. W. (1985). Evaluationsforschung - Aufgaben, Probleme und Anwendungen. Heidelberg: Springer.

Wittmann, W. W. (1987). Grundlagen erfolgreicher Forschung in der Psychologie: Multimodale Diagnostik, Multiplismus, multivariate Reliabilitäts- und Validitätstheorie. Diagnostica, 33, 209-226.

Wittmann, W. W. (1988). Multivariate reliability theory. Principles of symmetry and successful validation strategies. In Nesselroade, J. R. & Cattell, R. B. (Eds.), Handbook of multivariate experimental psychology, (2nd ed.). New York: Plenum.

Wittmann, W. W. (1990). Brunswik-Symmetrie und die Konzeption der fünf Datenboxen - Ein Rahmenkonzept für umfassende Evaluationsforschung. Zeitschrift für Pädagogische Psychologie, 4, 241-251.

Wittmann, W. W. & Matt, G. E. (1986). Aggregation und Symmetrie. Grundlagen einer multivariaten Reliabilitäts- und Validitätstheorie, dargestellt am Beispiel der differentiellen Validität des Berliner Intelligenzstrukturmodells. Diagnostica, 32, 309-329.

Wöhe, G. (1990). Einführung in die klassische Betriebswirtschaftslehre. München: Vahlen.

Wortmann, P. M. (1983). Evaluation Research: A Methodological Perspective. Annual Review of Psychology, 34, 223-260.

Wotschke, A. (1995). Motivatorische, lebenslauf- und erziehungsbedingte Determinanten des beruflichen Erfolges von Führungskräften: Ein diagnostischer Dreigruppenvergleich von Managern aus dem Profit- und Nonprofitsektor sowie dem Selbständigenbereich. Hamburg: Universität der Bundeswehr Hamburg.

Wottawa, H. (1990). Einige Überlegungen zu (Fehl-)Entwicklungen der psychologischen Methodenlehre. Psychologische Rundschau, 41(2), 84-97.

Wottawa, H. (1991). Entwicklungstrends psychologischer Eignungsdiagnostik. In Schuler, H. & Funke, U. (Hrsg.), Eignungsdiagnostik in Forschung und Praxis, 10, 1-5. Stuttgart: Verlag für angewandte Psychologie.

Wottawa, H. (1993). Computergestützte Diagnostik [Themenheft]. Zeitschrift für Arbeits- und Organisationspsychologie, 37(3),

Wottawa, H. & Hossiep, R. (1987). Grundlagen psychologischer Diagnostik. Göttingen: Verlag für Psychologie.

Wottawa, H. & Thierau, H. (2003). Lehrbuch Evaluation (3. korr. Auflage). Bern: Huber.

Wright, P. M., Lichtenfels, P. A. & Pursell, E. D. (1989). The structured interview: Additional studies and meta-analysis. Journal of Occupational Psychology, 62, 191-199.

Wroten, S. P. (1984). Overcoming the futilities of utility applications: Measures, models, and management. Paper presented at the annual meeting of the American Psychological Association, Toronto,

Zeithaml, V. A. (1981). How Consumer Evaluation Process Differ Between Goods and Services. In Donelly, J. H. & George, W. R. (Eds.), Marketing of Services, 186-190. Chicago: American Marketing Association.

Zimmermann, H. & Schuler, H. (1991). Persönlichkeitskonstrukte als Urteilseinheiten im Assessment Center. In Schuler, H. & Funke, U. (Hrsg.), Eignungsdiagnostik in Forschung und Praxis, Band 10, 139-143. Stuttgart: Verlag für Angewandte Psychologie.