

## **E-Assessment**

**Entwicklung und Güteprüfung von zwei internetbasierten  
Simulationsverfahren zur Messung der Planungs- und  
Problemlöseleistung von zukünftigen (pädagogischen)  
Führungskräften**

Dissertation  
zur Erlangung des Grades eines  
Doktors der Philosophie (Dr. phil.)

Leuphana Universität Lüneburg

## **Danksagung**

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich in der Zeit der Entstehung der Dissertation begleitet und unterstützt haben. Ohne das Zutun dieser Personen wäre diese Arbeit nicht zu Stande gekommen.

Da sind zuallererst meine Professoren und Gutachter Bernhard Sieland, Lutz Schumacher und Johannes Mayr zu nennen. Ihnen gilt besonderer Dank, da sie mich über die Jahre in vollem Umfang bei Planung und Durchführung der Arbeit unterstützt haben. Bei Professor Werner Sarges möchte ich mich für die fachliche Unterstützung und die inhaltlichen Anregungen bedanken.

Meinen Freunden und Kollegen Thilo Bergmann und Arno Gruber danke ich dafür, dass sie in unzähligen Stunden die Verfahren gestaltet und programmiert haben, so dass die durchgeführten Evaluationsstudien überhaupt möglich wurden.

Nikolina Kopping, Joachim Diercks und Ramin Mirhachemzadeh möchte ich für ihre Hilfe danken, insb. für die Möglichkeit, umfangreiche Stichproben nutzen zu können.

Dank gilt meinen Lüneburger Kolleginnen und Kollegen Birgit Nieskens, Thomas Petzel und Sylwia Neidhardt, die mich bei der Generierung von Stichproben tatkräftig unterstützt haben. Bedanken möchte ich mich auch bei den über 500 Testpersonen.

Des Weiteren möchte ich mich bei Petra Winkler und Helmut Riesche bedanken, die sich auch noch ganz zum Schluss bereit erklärt haben, die Arbeit orthografisch zu prüfen.

Großer Dank gebührt auch meinen Eltern Karin und Klaus, die mir nicht nur die Möglichkeit gegeben haben, eine Dissertation überhaupt verfassen zu dürfen, sondern mich auch tatkräftig bei der Entstehung dieser Arbeit unterstützt haben.

Und zu guter Letzt möchte ich mich bei meiner Familie, bei meinem Sohn Jonas Levin und meiner Frau Ines, für die unendliche Unterstützung in allen Phasen der Erstellung der Dissertation bedanken. Ohne den familiären Rückhalt wäre die Arbeit nicht möglich gewesen.



# Inhalt

1 Einleitung .....	1
2 Forschungsstand Planen und Problemlösen .....	5
2.1 Planungsfähigkeit und Problemlösefähigkeit als eignungsrelevante Personmerkmale von Führungskräften .....	7
2.1.1 Schulleitung als Führungstätigkeit - Schulleitungsforschung .....	12
2.1.2 Bedeutung von Planungs-, Organisations- und Verwaltungsaufgaben im Berufsalltag von Schulleitungen .....	14
2.1.3 Implikationen für die vorliegende Arbeit .....	23
2.2 Grundlagen Planen und Problemlösen .....	25
2.2.1 Planen und Problemlösen in der Tradition der Denkpsychologie .....	25
2.2.2 Berliner Intelligenzstrukturmodell .....	34
2.2.3 Definition Problemlösen .....	36
2.2.4 Einflussgrößen der Problemlöseleistung .....	37
2.2.5 Empirischer Forschungsstand zum Problemlösen .....	39
2.2.5.1 Aufgabenmerkmale als Einflussgrößen der Problemlöseleistung .....	39
2.2.5.2 Personmerkmale als Einflussgrößen der Problemlöseleistung .....	42
2.2.5.2.1 Intelligenz und Problemlösen .....	43
2.2.5.2.2 Bereichsspezifität von Problemlöseleistungen .....	53
2.2.6 Definition Planen .....	55
2.2.7 Einflussgrößen der Planungsleistung .....	58
2.2.8 Empirischer Forschungsstand zum Planen .....	59
2.2.8.1 Aufgabenmerkmale als Einflussgrößen der Planungsleistung .....	60
2.2.8.2 Personmerkmale als Einflussgrößen der Planungsleistung .....	62
2.2.9 Implikationen für die vorliegende Arbeit .....	64
3 Internetgestützte Berufseignungsdiagnostik und Forschungs- stand der Simulationsverfahren zum Planen und Problemlösen .....	67
3.1 E-Assessment und Blended-Assessment .....	70
3.1.1 Bedeutung von E-Assessment-Verfahren .....	70
3.1.2 Definition E-Assessment-Verfahren und E-Assessment .....	71

---

3.1.3 Blended-Assessment - eine Klassifikation eignungsdiagnostischer Verfahren unter Einbeziehung des Mediums Internet .....	73
3.1.4 Online-Simulationsverfahren.....	74
3.2 Ziele, Vorteile und Risiken.....	77
3.2.1 Praktische Bedeutung von Blended-Assessment .....	77
3.2.2 Ziele, Strategien und mögliche Vorteile des Einsatzes von E-Assessments .....	80
3.2.3 Vorteile und Risiken hinsichtlich der Testgüte von E-Assessment-Verfahren.....	82
3.3 Güte und Nutzen eignungsdiagnostischer Verfahren .....	89
3.3.1 Allgemeine Gütekriterien.....	89
3.3.2 Nutzen eignungsdiagnostischer Verfahren .....	97
3.4 Forschungsstand der Simulationsverfahren zum Planen und Problemlösen .....	100
3.4.1 Postkorbverfahren .....	100
3.4.1.1 Objektivität und Reliabilität.....	102
3.4.1.2 Validität und Nebengütekriterien .....	105
3.4.1.3 Implikationen für die vorliegende Arbeit.....	112
3.4.2 Computergestützte Szenarios.....	114
3.4.2.1 Objektivität und Reliabilität.....	116
3.4.2.2 Validität und Nebengütekriterien .....	121
3.4.2.3 Implikationen für die vorliegende Arbeit.....	126
4 Entwicklung der Verfahren <i>webPostkorb</i> und <i>webPlanungsaufgabe</i> .....	129
4.1 Rohformentwicklung.....	131
4.2 Vorprüfung der Rohformen an Studierenden.....	132
4.3 Entwicklung als internetbasierte E-Assessment-Verfahren .....	133
4.4 Problemlösen im <i>webPostkorb</i> .....	134
4.4.1 Szenario und Einweisung .....	136
4.4.2 Inhaltliche Merkmale .....	136
4.4.3 Operationalisierungen.....	137
4.4.4 Innere Struktur .....	139
4.4.5 Benutzerfreundlichkeit der Mensch-Maschine-Interaktion .....	144
4.5 Planen in der <i>webPlanungsaufgabe</i> .....	145
4.5.1 Szenario und Einweisung .....	145
4.5.2 Inhaltliche Merkmale .....	147
4.5.3 Operationalisierungen.....	147

---

5 Fragestellung und Hypothesen, Messinstrumente, Evaluationsvorgehen .....	151
5.1 Fragestellung und Hypothesen.....	151
5.2 Messinstrumente .....	153
5.2.1 Intelligenztestverfahren.....	154
5.2.2 Handlungs- und Lageorientierung.....	156
5.2.3 Computererfahrung und Computereinstellung .....	157
5.2.4 Eigener Fragebogen .....	158
5.3 Evaluationsvorgehen .....	159
6 Ergebnisse der Voruntersuchung.....	163
6.1 Stichproben .....	163
6.2 Itemkennwerte .....	164
6.3 Skalenkennwerte .....	168
6.4 Benutzerfreundlichkeit, Akzeptanz und Fairness.....	173
6.5 Erste Befunde zur Validität .....	179
6.6 Zusammenfassung und Implikationen für die Hauptuntersuchung .....	183
7 Ergebnisse der Hauptuntersuchung.....	187
7.1 Stichproben .....	187
7.2 Itemkennwerte .....	189
7.3 Objektivität, Reliabilität und Skalenkennwerte.....	192
7.4 Inhaltsvalidität.....	209
7.5 Kriteriumsvalidität .....	210
7.6 Konstruktvalidität .....	217
7.6.1 Dimensionsanalysen des <i>webPostkorbs</i> („innere“ Konstruktvalidität) .....	218
7.6.2 Zusammenhang mit dem Personmerkmal Handlungs- und Lageorientierung .....	223
7.6.3 Einsatz im Rahmen von Personalauswahlverfahren.....	227
7.6.4 Zusammenhang mit kognitiven Leistungsverfahren.....	229
7.7 Akzeptanz und Fairness .....	237
8 Diskussion und Ausblick .....	247
9 Literatur.....	259





## Tabellen

Tabelle 1: Die 26 Personmerkmale der AT&T Studie (nach Bray & Grant, 1966) .....	8
Tabelle 2: Beschreibung der beiden Bereiche Problemlösen und (Arbeitsplatz-) Organisation (nach Schuler et al., 1995).....	11
Tabelle 3: Durchschnittlicher monatlicher Stundenaufwand für verschiedene Schulleitertätigkeiten (nach Wissinger, 2002, S. 54) .....	17
Tabelle 4: Zeitaufwand pro Monat in Stunden für die Kategorie „Interne Verwaltungstätigkeit“ im Ländervergleich (nach Wissinger, 2002, S. 56) .....	18
Tabelle 5: Wettquotienten zur Vorhersage der Zugehörigkeit zur Gruppe der verbesserungsbedürftigen bzw. guten Schulen (nach Bensen, 2005) .....	21
Tabelle 6: Typische Aufgaben von Schulleitungen .....	22
Tabelle 7: Klassifikation von Barrieretypen nach den Dimensionen „Bekanntheitsgrad der Mittel“ und „Klarheit der Zielkriterien“ (nach Dörner, 1979, S. 14).....	26
Tabelle 8: Eigenschaften komplexer Probleme nach Dörner et al. (1983) .....	29
Tabelle 9: Vergleich eigenschafts- und simulationsorientierter Verfahrenskonzepte (vgl. Höft & U. Funke, 2001, S. 136, erweitert um den Aspekt der Validierungslogik) .....	69
Tabelle 10: Zusammenhänge zwischen zwei „Bonner-Postkorb- Modulen“ (nach Musch & Lieberei, 1997).....	107
Tabelle 11: Einfluss der Computervorerfahrung in der PISA- Vorbereitungsstudie (nach Klieme et al., 2001).....	125
Tabelle 12: Überblick über die Schritte der Verfahrenskonstruktion bis zur Programmierung der Erstversionen .....	130
Tabelle 13: Überblick über die Konzeption des <i>webPostkorbs</i> zum Problemlösen .....	138
Tabelle 14: Überblick über die Konzeption der <i>webPlanungsaufgabe</i> zum Planen .....	149
Tabelle 15: Überblick über die Evaluationsstudien zur Entwicklung der <i>webPlanungsaufgabe</i> und des <i>webPostkorbs</i> .....	161
Tabelle 16: Itemkennwerte des <i>webPostkorbs</i> (Voruntersuchung) .....	166
Tabelle 17: Deskriptive Statistiken des <i>webPostkorbs</i> und der <i>webPlanungsaufgabe</i> (Voruntersuchung) .....	168
Tabelle 18: Korrelationen zwischen der <i>webPlanungsaufgabe</i> sowie dem <i>webPostkorb</i> und konstruktfernen Variablen Ziel: $r < .3$ (Voruntersuchung) .....	179

Tabelle 19: Korrelationen zwischen dem <i>webPostkorb</i> sowie der <i>webPlanungsaufgabe</i> und Aufgaben des I-S-T 70 (Voruntersuchung) .....	181
Tabelle 20: Überblick über die wichtigsten Implikationen für die weitere Verfahrensentwicklung des <i>webPostkorbs</i> .....	184
Tabelle 21: Überblick über die wichtigsten Implikationen für die Hauptuntersuchung .....	185
Tabelle 22: Itemkennwerte des <i>webPostkorbs</i> (Hauptuntersuchung) .....	190
Tabelle 23: Deskriptive Statistiken und Cronbachs-Alpha-Werte des <i>webPostkorbs</i> (Hauptuntersuchung) .....	194
Tabelle 24: Interkorrelationen der Subskalen des <i>webPostkorbs</i> (Hauptuntersuchung) .....	197
Tabelle 25: Deskriptive Statistiken der <i>webPlanungsaufgabe</i> (Schulleitungs- und Wirtschaftsversion, Hauptuntersuchung) ..	198
Tabelle 26: Deskriptive Statistiken der Rohwerte der sechs BIS-K-S Aufgaben und des zusammengefassten Summenscores (Hauptuntersuchung) .....	202
Tabelle 27: Interkorrelationen der Aufgaben mit dem Gesamtscore BIS-K-S / Vergleich mit Werten des BIS-K der Normstichprobe (Jäger et al., 1997).....	204
Tabelle 28: Häufigkeiten HAKEMP-90 .....	206
Tabelle 29: Iteminterkorrelationen und interne Konsistenzen HAKEMP-90.....	207
Tabelle 30: Interkorrelationen der HAKEMP-90-Skalen.....	208
Tabelle 31: Korrelationen der Abitur- und der Zwischenprüfungsnote mit den Ergebnissen in den Leistungsverfahren (Hauptuntersuchung, Stichprobe Studierende) .....	211
Tabelle 32: Hierarchische Regression von Verarbeitungskapazität, Planungsleistung und Problemlöseleistung auf das Kriterium Studienerfolg (Note der Zwischenprüfung) .....	212
Tabelle 33: Korrelationen der Abitur- und der Zwischenprüfungsnote mit den Leistungsverfahren der Führungsnachwuchskräfte-Stichprobe .....	213
Tabelle 34: Hierarchische Regression von Abiturnote, kognitiver Leistungsfähigkeit und Planungsleistung auf das Kriterium Studienerfolg (Note der Diplomprüfung).....	214
Tabelle 35: Zusammenhang zwischen der <i>webPlanungsaufgabe</i> und den Bausteinen des Personalauswahlverfahrens .....	216
Tabelle 36: Extremgruppenvergleich zwischen Bewerbern und bewährten Führungsnachwuchskräften in der <i>webPlanungsaufgabe</i> (Wirtschaftsversion) .....	217
Tabelle 37: Dimensionsanalysen des <i>webPostkorbs</i> (Hauptuntersuchung) .....	219

---

Tabelle 38: Kennwerte für die Anpassungsgüte verschiedener faktorenanalytischer Modelle (Hauptuntersuchung) .....	221
Tabelle 39: Korrelationen zwischen HAKEMP-90-Skalen und erhobenen Leistungswerten .....	225
Tabelle 40: Korrelationen zwischen der HOP-Skala und den eingesetzten Leistungsskalen .....	226
Tabelle 41: Faktorenanalyse der berufsrelevanten Personmerkmale eines Personalauswahlprozesses .....	228
Tabelle 42: Korrelationen der Simulationsverfahren mit Intelligenztestleistungen (Hauptuntersuchung) .....	230
Tabelle 43: Kennwerte für die Anpassungsgüte verschiedener faktorenanalytischer Modelle (Hauptuntersuchung) .....	233
Tabelle 44: Konstrukt fremde Einflüsse (Korrelationen mit den Testergebnissen) auf die Leistungen im <i>webPostkorb</i> und in der <i>webPlanungsaufgabe</i> (Hauptuntersuchung) .....	245



## Abbildungen

Abbildung 1: Hintergrund der Verfahrenskonstruktion der <i>webPlanungsaufgabe</i> und des <i>webPostkorbs</i> .....	5
Abbildung 2: Klassische Managementfunktionen (nach Grunwald, 1995) .....	10
Abbildung 3: „Turm von Hanoi“-Problem mit drei Scheiben.....	27
Abbildung 4: Verschiebepproblem (nach J. R. Anderson, 1996).....	28
Abbildung 5: TOTE-Einheit nach Miller, Galanter und Pribram (1960) .....	31
Abbildung 6: Modell des Handlungsablaufs nach Dörner (1989, S. 67) .....	33
Abbildung 7: Berliner Intelligenzstrukturmodell (Jäger, Süß & Beauducel, 1997, S. 5) .....	35
Abbildung 8: Klassifikation der Einflussgrößen der Problemlöseleistung (vgl. Hussy, 1984; Vollmeyer & J. Funke, 1999).....	38
Abbildung 9: Interkorrelationen der Kompetenzfaktoren im normierten zweidimensionalen Raum (nach Klieme et al., 2001, S. 196) .....	52
Abbildung 10: Vorhersage der Problemlöseleistungen durch Verarbeitungskapazität (BIS-K), Systemwissen und Vorwissen (nach Süß, 1999, S. 225) .....	54
Abbildung 11: Klassifikation der Einflussgrößen der Planungsleistung .....	59
Abbildung 12: Blended-Assessment / Blended-Eignungsdiagnostik (angelehnt an Schuler & Höft, 2001, erweitert um das Medium Internet) .....	73
Abbildung 13: Gütekriterien eignungsdiagnostischer Verfahren .....	90
Abbildung 14: Oberfläche des „PC-Postkorbs“ (Demoversion) von Scharley und Partner (1998).....	110
Abbildung 15: Benutzeroberfläche des „PC-Postkorbs“ von J. Funke und Rasche (1988) .....	112
Abbildung 16: Bearbeitungsbildschirm des Szenarios „Textilfabrik“ von Hasselmann und Strauß (1995).....	118
Abbildung 17: Mindeststandards der Online-Teilnahme an den Verfahren.....	134
Abbildung 18: Beispiele für die Einweisungsseiten des <i>webPostkorbs</i> .....	135
Abbildung 19: Innere Struktur des <i>webPostkorbs</i> (Endversion).....	140
Abbildung 20: Operationalisierungsbeispiele der Aufgabengruppen <i>Wochenplan</i> und <i>Diagrammverständnis</i> .....	141
Abbildung 21: Operationalisierungsbeispiele der Aufgabengruppen <i>Textverständnis</i> und <i>Tabellenverständnis</i> .....	142
Abbildung 22: Operationalisierungsbeispiele der Aufgabengruppe <i>Umgang mit Störungen</i> .....	143

Abbildung 23: Übungsaufgabe und geöffnetes Instruktionfenster während der Verfahrensbearbeitung der <i>webPlanungsaufgabe</i> .....	146
Abbildung 24: Validierungslogik der verschiedenen eignungsdiagnostischen Ansätze (aus Schuler & Höft, 2001, S. 95, erweitert um die roten Markierungen).....	159
Abbildung 25: Boxplots der korrigierten Trennschärfen und Schwierigkeiten des <i>webPostkorbs</i> (Voruntersuchung, N = 38) .....	167
Abbildung 26: Histogramm der Punkterohwerte und Boxplot der benötigten Zeit <i>webPlanungsaufgabe</i> (Voruntersuchung, N = 35) .....	171
Abbildung 27: Rohwerteverteilungen der vier I-S-T 70 Aufgaben (Voruntersuchung).....	173
Abbildung 28: Bewertung der Bedienungseinweisung der <i>webPlanungsaufgabe</i> (Voruntersuchung) .....	175
Abbildung 29: Bewertung der Bedienungsfreundlichkeit der <i>webPlanungsaufgabe</i> (Voruntersuchung) .....	175
Abbildung 30: Bewertung der Gestaltung der <i>webPlanungsaufgabe</i> (Voruntersuchung).....	176
Abbildung 31: Bewertung der Zufriedenheit bei der Bearbeitung der <i>webPlanungsaufgabe</i> (Voruntersuchung) .....	177
Abbildung 32: Bewertung der Realitätsnähe der <i>webPlanungsaufgabe</i> (Voruntersuchung).....	180
Abbildung 33: Boxplots der Schwierigkeiten mit Inangriffnahmekorrektur und der part-whole korrigierten Trennschärfen <i>webPostkorb</i> (Hauptuntersuchung, N = 66).....	191
Abbildung 34: Streudiagramm der Schwierigkeiten mit Inangriffnahmekorrektur und der Standardabweichungen für die Items des <i>webPostkorbs</i> (Hauptuntersuchung).....	192
Abbildung 35: Histogramm für den Gesamtscore des <i>webPostkorbs</i> (Hauptuntersuchung) .....	195
Abbildung 36: Histogramm und Boxplot für die Punkterohwerte der <i>webPlanungsaufgabe</i> (Stichprobe Studierende, N = 45).....	199
Abbildung 37: Histogramm und Boxplot für die benötigte Zeit zur Bearbeitung der <i>webPlanungsaufgabe</i> (Stichprobe Studierende, N = 45) .....	200
Abbildung 38: Histogramme für die <i>webPlanungsaufgabe</i> (Stichprobe Führungsnachwuchskräfte, N = 369).....	201
Abbildung 39: Histogramme des Summenrohwerths und des Standardwerts des BIS-K-S (Stichprobe Studierende) .....	203
Abbildung 40: Bewertung der Realitätsnähe des <i>webPostkorbs</i> (Hauptuntersuchung) .....	210

---

Abbildung 41: Screeplot der Faktorenanalyse <i>webPostkorb</i> (Hauptuntersuchung) .....	220
Abbildung 42: Pfaddiagramm des Zwei-Faktoren-Modells mit geschätzten standardisierten Regressionsgewichten ( <i>webPostkorb</i> , Hauptuntersuchung) .....	223
Abbildung 43: Pfaddiagramm des Drei-Faktoren-Modells mit geschätzten standardisierten Regressionsgewichten (Hauptuntersuchung) .....	234
Abbildung 44: Bewertung der Bedienungseinweisung des <i>webPostkorbs</i> (Hauptuntersuchung) .....	238
Abbildung 45: Bewertung der Bedienungsfreundlichkeit des <i>webPostkorbs</i> (Hauptuntersuchung) .....	238
Abbildung 46: Bewertung der Gestaltung des <i>webPostkorbs</i> (Hauptuntersuchung) .....	239
Abbildung 47: Bewertung der Zufriedenheit bei der Bearbeitung des <i>webPostkorbs</i> (Hauptuntersuchung) .....	240
Abbildung 48: Bewertung der Anregung zur Auseinandersetzung mit der beruflichen Entwicklung (Hauptuntersuchung) .....	241
Abbildung 49: Vergleich der beiden Leistungsverfahren BIS-K-S und <i>webPostkorb</i> hinsichtlich verschiedener Akzeptanzaspekte (Hauptuntersuchung) .....	242





# 1 Einleitung

Der Einsatz von internetbasierten, eignungsdiagnostischen Verfahren - so genannten E-Assessment-Verfahren - ist ein relativ junges Unterfangen. Trotz zunehmender Bedeutung und Verbreitung von E- und Self-Assessments zum Zwecke der Personalauswahl, Personalentwicklung, Karriereberatung oder des Personalmarketings (vgl. Kupka, Diercks & Kopping, 2004) nähert sich die Wissenschaft erst langsam diesem Themengebiet an. Nur wenige empirische Studien haben sich bisher mit internetgestützter Eignungsdiagnostik bzw. E-Assessments befasst. Eignungsdiagnostische Arbeiten zur Entwicklung und Evaluation von arbeitsprobenartigen Online-Simulationsverfahren sind bisher kaum zu finden, obwohl sich insbesondere die simulativen Arbeitsproben als höchst valide zur Vorhersage von Berufserfolg erwiesen haben (Schmidt & Hunter, 1998, 2000).

Im Rahmen der Eignungsdiagnostikliteratur ist die Bedeutung der Personmerkmale „Planungsfähigkeit“ und „Problemlösefähigkeit“ für den Berufserfolg unumstritten. Fähigkeiten im Planen und Problemlösen gelten als erfolgskritische, eignungsrelevante Personmerkmale von Führungskräften. Die Schulleitungsforschung belegt, dass Schulleiterinnen und Schulleiter<sup>1</sup> als pädagogische Führungskräfte ein entscheidendes Element der Qualität und Gesundheit von Schulen darstellen (u. a. Bonsen, Gathen & Pfeiffer, 2002; S. G. Huber, 1999a, 1999b). Im Arbeitsalltag von Schulleitungen nehmen umfangreichen Untersuchungen zufolge Planungs-, Organisations- und Verwaltungsaufgaben *die* zentrale Rolle ein (Bonsen & Pfeiffer, 1998). Zu deren erfolgreicher Bewältigung sind Fähigkeiten im Planen und Problemlösen notwendig und entscheidend. An geeigneten, eignungsdiagnostischen Instrumenten zur Bestimmung der individuellen Planungs- und Problemlöseleistung mangelt es jedoch (vgl. J. Funke & Fritz, 1995).

---

<sup>1</sup> Der Lesbarkeit halber werden in dieser Arbeit teilweise lediglich die männlichen Begriffsformen verwendet. Diese schließen vom Sinngehalt die weiblichen stets mit ein.

Ziel dieser Arbeit ist daher die theoriegeleitete Entwicklung und Güteprüfung von zwei internetbasierten Simulationsverfahren zur Messung der Planungs- und Problemlöseleistung von zukünftigen (pädagogischen) Führungskräften. Den theoretischen und methodischen Hintergrund der Verfahrensentwicklung bilden zum einen die Auseinandersetzung mit dem Forschungsstand zu den Konstrukten „Planungsfähigkeit“ und „Problemlösefähigkeit“ sowie die Berücksichtigung bisheriger empirischer Tätigkeitsanalysen von pädagogischen Führungskräften. Zum anderen stehen die Probleme und Möglichkeiten des eignungsdiagnostischen Einsatzes bisheriger Simulationsverfahren im Mittelpunkt der Betrachtung. Neben der Darstellung der Entwicklung und Evaluation der Verfahren wird auch auf die grundsätzliche Bedeutung internetgestützter Eignungsdiagnostik eingegangen und eine Klassifikation zum „Blended-Assessment“ vorgestellt.

Die beiden verhaltensbasierten E-Assessment-Verfahren *webPostkorb* und *webPlanungsaufgabe* werden vor dem Hintergrund der Vorteile der Internetdarbietung als webbasierte, interaktive und auswertungsobjektive Verfahren konzipiert, die typische Planungs- und Problemlöseaufgaben in simulierten Arbeitssituationen abbilden. Mit der vorliegenden Arbeit soll aufgezeigt werden, dass sich interaktive Simulationsverfahren internetbasiert abbilden und hinsichtlich der Testgüte prüfen lassen. Dabei wird versucht, die Vorteile von zwei eignungsdiagnostischen Ansätzen zu verbinden: dem Simulations- und dem Konstruktansatz.

Die Herausforderung besteht darin, zielgruppengerechte und simulative Verfahren zu entwickeln, die sowohl hohe Akzeptanz und Arbeitsplatznähe als auch hohe Testgüte vereinigen und dies im Medium Internet auf eine benutzerfreundliche Art darstellen. Die Güteprüfung der *webPlanungsaufgabe* und des *webPostkorbs* steht dabei unter der Leitfrage, inwieweit die hier entwickelten eignungsdiagnostischen Simulationsverfahren den allgemeinen Standards genügen (zu den Standards siehe American Educational Research Association, 1999; Testkuratorium, 1986).

### *Zum Aufbau der Arbeit*

Einleitend wird im theoretischen Teil der Arbeit ein kurzer Überblick über die Bedeutung der „Planungsfähigkeit“ und „Problemlösefähigkeit“ als eignungsrelevante Personmerkmale für Führungskräfte gegeben. Empirische Studien verdeutlichen im Anschluss die typischen Aufgaben und Anforderungen von Schulleiterinnen und Schulleitern und die Relevanz von planerischen und problemlösenden Fähigkeiten für deren erfolgreiche Bewältigung. Die hieraus abzuleitenden Implikationen für die Verfahrensentwicklung schließen den ersten Theorieteil ab.

Nach einem kurzen Abriss über die Tradition der Denkpsychologie zu den Konstrukten „Planungsfähigkeit“ und „Problemlösefähigkeit“ wird auf das Begriffsverständnis und den empirischen Forschungsstand hinsichtlich der beiden Personmerkmale eingegangen. Aufgrund der Bedeutung für die vorliegende Arbeit werden dabei auch die wichtigsten Aspekte des Berliner Intelligenzstrukturmodells von Jäger (1982, 1984) dargestellt, bevor wiederum die Implikationen für die Verfahrensentwicklung den Abschnitt über die Grundlagen der psychologischen Forschung zu Fähigkeiten im Planen und Problemlösen abschließen.

Den theoretischen und methodischen Hintergrund der Arbeit rundet das dritte Kapitel mit der Vorstellung des Bereichs der internetgestützten Eignungsdiagnostik ab. Dieser Abschnitt thematisiert - neben der Begriffsbestimmung von E- und Blended-Assessment - Ziele, Vorteile und Risiken, die allgemeinen Gütekriterien sowie den Forschungsstand zu Simulationsverfahren zum Planen und Problemlösen. Das Augenmerk liegt dabei auf der Testgüte der beiden bisherigen Verfahrensklassen „Postkorbverfahren“ und „computergestützte Szenarios“.

Die Vorstellung der Entwicklung der simulativen E-Assessment-Verfahren *webPostkorb* und *webPlanungsaufgabe* steht im Mittelpunkt des vierten Kapitels. Dabei werden die einzelnen Entwicklungsschritte und die genaue Ausgestaltung der Verfahren näher betrachtet.

Das fünfte Kapitel thematisiert die Fragestellung und Hypothesen, das Evaluationsvorgehen und die im Rahmen der Güteuntersuchungen eingesetzten Messinstrumente.

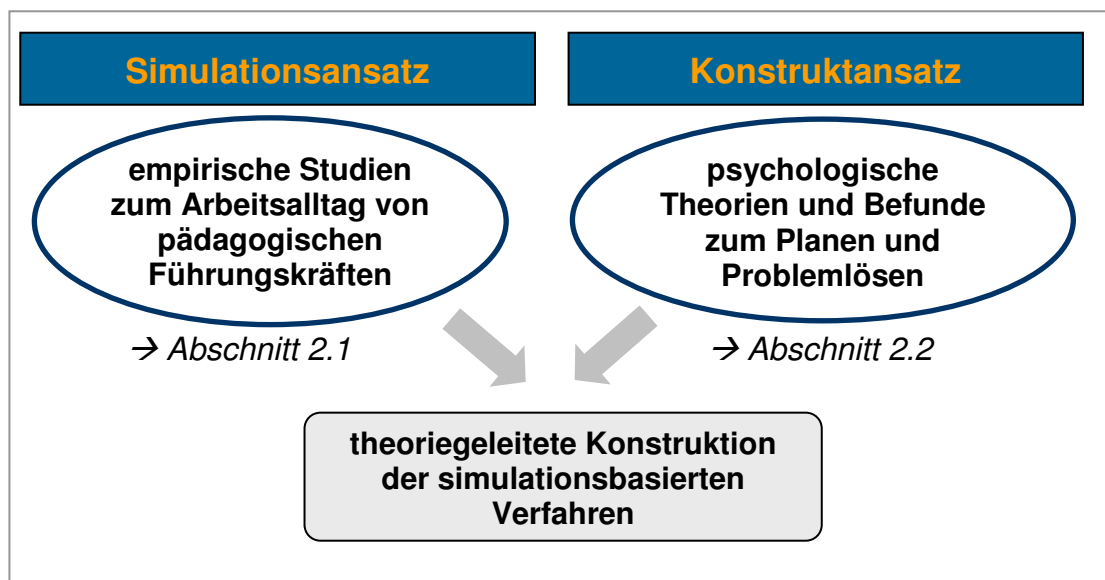
Die beiden genannten Simulationsverfahren wurden im Verlauf eines umfangreichen Entwicklungsprozesses optimiert und an Stichproben von 104 Studierenden sowie 19 bewährten und 369 potenziellen Führungsnachwuchskräften evaluiert. Eine Besonderheit ist, dass die Verfahren nicht nur im Rahmen eines Self-Assessments an studentischen Stichproben, sondern auch unter realen Auswahlbedingungen bei der Personalauswahl von Führungskräften zum Einsatz kamen.

Die Ergebnisse der Voruntersuchung und die daraus abzuleitenden Implikationen für die weitere Entwicklung der Verfahren sind Mittelpunkt der Darstellung in Kapitel sechs, während die Befunde der Hauptuntersuchung den Inhalt des siebten Kapitels bestimmen. Dabei wird auf Aspekte der Objektivität, Reliabilität, Validität sowie Akzeptanz und Fairness eingegangen. Besonderes Augenmerk liegt auf den Ergebnissen zur Konstruktvalidität, insbesondere zu korrelativen Zusammenhängen mit Intelligenztestverfahren.

Den Abschluss bildet das achte Kapitel mit der Diskussion und der Bewertung der Befunde sowie einem Ausblick auf zukünftige Entwicklungen.

## 2 Forschungsstand Planen und Problemlösen

Dieses und das folgende Kapitel widmen sich den theoretischen Grundlagen der vorliegenden Arbeit. Die Vorgehensweise bei der Verfahrenskonstruktion wird im Folgenden kurz vorgestellt, da sich der Aufbau dieses Kapitels danach richtet. Leitidee der Verfahrensentwicklung ist, Erkenntnisse der psychologischen Grundlagenforschung umfassend zu nutzen und die Instrumente somit theoriegeleitet zu entwickeln. Dabei wird versucht, die Vorteile von zwei methodischen Ansätzen zu verbinden. Die Entwicklungsstrategie der Kombination von Simulations- und Konstruktansatz ist in Abbildung 1 illustriert (zur genaueren Darstellung der Verfahrenskonstruktion siehe Kapitel 4).



**Abbildung 1:** Hintergrund der Verfahrenskonstruktion der *webPlanungsaufgabe* und des *webPostkorbs*

Ziel dieses Kapitels ist es, die Bedeutung der Planungs- und Problemlösefähigkeit für die erfolgreiche Bewältigung von Führungskräfteaufgaben - insbesondere im Schulleitungskontext - herauszuarbeiten und empirische Studien zum Arbeitsalltag von Schulleitungen vorzustellen (Abschnitt 2.1). Desweiteren werden die Erkenntnisse der psychologischen Grundlagenforschung zu den Konstrukten „Planungsfähigkeit“ und „Problemlösefähigkeit“ aus eignungsdiagnostischer

Perspektive betrachtet und die Implikationen für die vorliegende Arbeit deutlich gemacht (Abschnitt 2.2). Im Mittelpunkt dieses Abschnitts stehen sowohl die denkpsychologische Tradition als auch der aktuelle Forschungsstand zum Planen und Problemlösen. Aufgrund der Bedeutung für die vorliegende Arbeit wird darüber hinaus das Berliner Intelligenzstrukturmodell von Jäger (1982, 1984) vorgestellt. In diesem Zusammenhang gilt es auch die Frage zu erörtern, inwieweit von der Intelligenz unterscheidbare Fähigkeiten notwendig sind, um die Leistungen in Planungs- und Problemlösesituationen zu erklären.

Ziel der hier verfolgten Verfahrensentwicklungen ist es, mit Hilfe der Simulationsverfahren *webPlanungsaufgabe* und *webPostkorb* Leistungen im Planen und Problemlösen in einem berufstypischen Kontext zu diagnostizieren. Es wird dabei angenommen, dass eine generellere Planungs- bzw. Problemlösefähigkeit das Teilnehmerverhalten unter den spezifischen situativen Anforderungen der Simulationsverfahren erklärt. Daher ist es von Bedeutung, zwischen in den Verfahren gezeigten planerischen und problemlösenden Leistungen und der Planungs- und Problemlösefähigkeit zu differenzieren und die Fähigkeitskonstrukte von anderen möglichen Erklärungsvariablen des gezeigten Verhaltens - insbesondere der Intelligenz - abzugrenzen.

In der Literatur finden sich kaum Arbeiten, die trennscharf unterscheiden zwischen Planen, Planungsleistung und Planungsfähigkeit bzw. Problemlösen, Problemlöseleistung und Problemlösefähigkeit. Im Rahmen dieser Arbeit wird begrifflich - soweit möglich - die Leistung von der Fähigkeit abgegrenzt. Die Begriffe „Planen“ und „Problemlösen“ finden Verwendung, wenn beide Aspekte - „Leistung“ und „Fähigkeit“ - angesprochen bzw. wenn Unterscheidungen nicht möglich sind - beispielsweise aufgrund fehlender Angaben in den vorgestellten Studien.

Zunächst geht es im folgenden Abschnitt 2.1 um den Anforderungsbezug der Verfahrensentwicklungen. Dabei werden zuerst die Bedeutungen der

Planungs- und der Problemlösefähigkeit als eignungsrelevante Personmerkmale von Führungskräften skizziert. Nach einem Überblick über das aktuelle Verständnis von Schulleitungen als pädagogische Führungskräfte werden die spezifischen beruflichen Anforderungen durch die Vorstellung empirischer Arbeits- und Tätigkeitsanalysen von Schulleitungsmitgliedern dokumentiert. Diese Arbeits- und Tätigkeitsanalysen gilt es, bei den Verfahrensentwicklungen zu berücksichtigen.

## **2.1 Planungsfähigkeit und Problemlösefähigkeit als eignungsrelevante Personmerkmale von Führungskräften**

Die Auseinandersetzung mit den spezifischen Anforderungen von Führungsaufgaben und den eignungsrelevanten Personmerkmalen von Führungskräften hat eine lange Tradition (s. u.). In der Eignungsdiagnostik ist dabei zwischen Anforderungen, die die berufliche Arbeit an Führungskräfte stellt, und erfolgs- und eignungsrelevanten Personmerkmalen zu unterscheiden. Unter eignungsrelevanten Personmerkmalen sind alle Fähigkeiten, Fertigkeiten und Kenntnisse einer Person zu verstehen, die für die Leistung und Zufriedenheit in einem Beruf oder einer Ausbildung entscheidend sind (Maukisch, 1978).

Bereits 1948 identifizierte Stogdill in einem viel zitierten Überblicksartikel auf der Grundlage von über 100 Studien zentrale erfolgsrelevante Eigenschaften im Führungsbereich. Er unterschied dabei hinsichtlich unterschiedlicher Klassifikationskriterien wie Fähigkeiten, Leistungen, Verantwortung, Status und Situation. Der Versuch der Identifizierung bestimmter Führungseigenschaften, die deutlich und konsistent über verschiedene Situationen hinweg Führungskräfte von Nicht-Führungskräften unterscheiden können, blieb aber letztlich erfolglos (vgl. Neuberger, 1986). Als Ausgangsbasis für Listen eignungsrelevanter Personmerkmale von Führungskräften gilt der Katalog der „Management Progress Study“ (Bray & Grant, 1966, s. Tabelle 1). Innerhalb dieser Langzeitstudie (1956-1960)

wurden 422 Führungsnachwuchskräfte des amerikanischen Kommunikationsunternehmens AT&T untersucht (Bray, 1964; Bray & Grant, 1966). Zur Beurteilung von 26 eignungsrelevanten Personmerkmalen durchliefen die Kandidaten ein Assessment-Center, das auf dem Auswahlverfahren zur Rekrutierung von Geheimdienstagenten des Office of Strategic Services (OSS) aufbaute (Domsch & Jochum, 1989).

**Tabelle 1: Die 26 Personmerkmale der AT&T Studie (nach Bray & Grant, 1966)**

<ul style="list-style-type: none"> <li>• <b>Organization and planning</b></li> <li>• <b>Decision making</b></li> <li>• <b>Human relations skills</b></li> <li>• <b>Behavior flexibility</b></li> <li>• <b>Personal impact</b></li> <li>• <b>Tolerance of uncertainty</b></li> <li>• <b>Resistance to stress</b></li> <li>• <b>Scholastic aptitude</b></li> <li>• <b>Range of interests</b></li> <li>• <b>Primacy of work</b></li> <li>• <b>Inner work standards</b></li> <li>• <b>Ability to delay gratification</b></li> <li>• <b>Need for security</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Need for advancement</b></li> <li>• <b>Need for superior approval</b></li> <li>• <b>Need for peer approval</b></li> <li>• <b>Perception of social cues</b></li> <li>• <b>Goal flexibility</b></li> <li>• <b>Social objectivity</b></li> <li>• <b>Self-objectivity</b></li> <li>• <b>Energy</b></li> <li>• <b>Realism of expectations</b></li> <li>• <b>Bell System value orientation</b></li> <li>• <b>Staff prediction</b></li> <li>• <b>Oral communications skills</b></li> <li>• <b>Creativity</b></li> </ul>
---	--

Bereits Bray und Grant (1966) nennen in ihrer Liste das Personmerkmal „Organisation und Planen“ (organization and planning) als wichtige Größe. Das Personmerkmal „Problemlösefähigkeit“ ist nicht explizit zu finden, wengleich einige Kategorien, wie „decision making“, „scholastic aptitude“ und „goal flexibility“, zu diesem Merkmal große Schnittmengen haben dürften. Das Fehlen verwundert jedoch nicht weiter, da das Konstrukt der „Problemlösefähigkeit“ im Rahmen der Eignungsdiagnostik erst später eingeführt wurde (zur denkpsychologischen Tradition s. Abschnitt 2.2.1). Wengleich Listen, wie die von Bray und Grant (1966) oder Jeserich (1981), berechtigterweise kritisiert werden hinsichtlich mangelnder Breite (nicht alle relevanten Merkmale sind einbezogen), Tiefe (zu alltagssprachlich formuliert und operationalisiert) und fehlenden Bezugs zu Befunden der psychologischen Grundlagenforschung (vgl. Sarges, 1995), so zeigen sie doch ein mögliches Feld eignungsrelevanter Personmerkmale auf.



Sarges (1995) teilt die Merkmale zur besseren Orientierung in die drei Bereiche Kognition (Intelligenz, geistige Fähigkeiten), Motivation (Vitalität und Antrieb) und soziale Interaktion (Umgang mit Menschen) ein. In seinem Herausgeberwerk zur Management-Diagnostik wird den Bereichen Planen / Organisieren<sup>2</sup> und Problemlösen jeweils ein eigenes Kapitel gewidmet. Allgemein gelten Planen und Problemlösen in der eignungsdiagnostischen Literatur als wichtige Personmerkmale für Führungskräfte (vgl. Jeserich, 1981; Soworka & Sarges, 1995).

Neben der Bestimmung von Personmerkmalen für die erfolgreiche Bewältigung der heutigen Führungskräfteaufgaben haben u. a. Grunwald (1995) und Hofmann (2000) versucht, zukünftige Anforderungen an Führungskräfte zu ergründen. Es zeigt sich, dass die Autoren zur Bewältigung zukünftiger Anforderungen neben kommunikativen Kompetenzen auch vernetztes Denken im Sinne von Problemlösefähigkeit und Organisationsfähigkeit als entscheidende Personmerkmale identifizieren. Die Bedeutung der Personmerkmale „Planungsfähigkeit“ und „Problemlösefähigkeit“ spiegelt sich aber nicht nur in der psychologischen eignungsdiagnostischen Literatur wider, sondern wird auch im Rahmen der wirtschaftswissenschaftlichen Managementliteratur deutlich. Als klassische Managementfunktionen gelten zumeist die in der Abbildung 2 dargestellten Aspekte.

---

<sup>2</sup> Jeserich (1995) spricht von „Planung / Kontrolle“, Sarges (1995) von „Planung, Organisation und Kontrolle“. Die genaue Begriffsbestimmung im Rahmen dieser Arbeit findet sich in Kapitel 4. Im Folgenden wird der Übersichtlichkeit halber zumeist lediglich der Begriff „Planen“ verwendet.



**Abbildung 2: Klassische Managementfunktionen (nach Grunwald, 1995)**

Führungskräfte sollen demnach verschiedenste Funktionen bedienen - angefangen beim Erkennen und Analysieren von Problemen über die Zielfindung und Planung von Lösungsalternativen bis hin zum Organisieren, Kontrollieren und Bewerten. Kommunikation und Kooperation verbinden die einzelnen Aspekte. Obwohl diese Managementfunktionen nicht gleichzusetzen sind mit eignungs- und erfolgsrelevanten Personmerkmalen, wird doch die grundsätzliche Bedeutung von Planen, Organisieren und Problemlösen im Managementbereich dokumentiert. Was im Rahmen der kognitionspsychologischen Forschungstradition unter Planen und Problemlösen verstanden wird, ist Inhalt der Abschnitte 2.2.3 und 2.2.6.

Die Bedeutung von Planen / Organisation und Problemlösen im Arbeitsalltag von Führungskräften unterschiedlicher Tätigkeitsfelder ist auch durch empirische Arbeits- und Anforderungsanalysen<sup>3</sup> belegt (vgl. u. a. Hasselmann, 1993; Ueckert, 1995). Die Analyse der Arbeitsanforderungen ist dabei wesentliche Grundlage für eine inhaltssvalide Konstruktion von eignungsdiagnostischen Verfahren (vgl. von Rosenstiel, 2000; Schuler, 2001; Wottawa, 1995). Exemplarisch wird hier die Studie von Schuler, U. Funke, Moser und Donat (1995) zu beruflichen Führungstätigkeiten im Forschungs- und Entwicklungsbereich (F&E) herangezogen. Schuler et al. (1995) untersuchen die beruflichen Anforderungen von etwa 140 Führungskräften in verschiedenen Hochtechnologieunternehmen, in dem sie typische Aufgaben aus dem Arbeitsalltag der Führungskräfte sammeln. Die Autoren bestimmen aus den Daten faktorenanalytisch dreizehn entscheidende Aufgabenbereiche, die als berufliche Anforderungen an potenzielle Nachwuchskräfte angesehen werden können. Trotz der unterschiedlichen Gewichtung der Aufgabenbereiche für verschiedene Arbeitsplätze zeigt sich die generelle Bedeutung der Aufgabenbereiche Problemlösen und (Arbeitsplatz-) Organisation. In Tabelle 2 wird das Verständnis der beiden Bereiche in der Untersuchung von Schuler et al. (1995) beschrieben.

**Tabelle 2: Beschreibung der beiden Bereiche Problemlösen und (Arbeitsplatz-) Organisation (nach Schuler et al., 1995)**

<b>Beschreibung der Aufgabenbereiche</b>	
<b>Problemlösen</b>	Forschungs- und Denkstrategien anwenden, z. B. zum Erkennen komplexer Zusammenhänge und flexibel mit Änderungen und Unsicherheit umgehen
<b>(Arbeitsplatz-) Organisation</b>	Verwaltungsarbeiten und Selbstmanagement

<sup>3</sup> Zur Klassifikation von Arbeits- und Anforderungsanalysen siehe Sonntag (1999) und Fleishman und Quaintance (1984).

Zusammenfassend lässt sich festhalten, dass Planungs- und Problemlösefähigkeit als generelle, eignungsrelevante Personmerkmale von Führungskräften eingeschätzt werden. Auch wenn jedes einzelne Unternehmen, jeder einzelne Arbeitsplatz unterschiedliche spezifische Anforderungen an Personen stellen dürfte, wird dennoch in empirischen Arbeits- und Anforderungsanalysen deutlich, dass die Aufgabenbereiche des Lösens von Problemen sowie des Organisierens, Verwaltens und Planens typisch für Führungskräfteaktivitäten sind (vgl. Schuler et al., 1995).

### ***2.1.1 Schulleitung als Führungstätigkeit - Schulleitungsforschung***

Schulleitung als Gegenstand der Forschung und Qualifizierung ist ein vergleichsweise junges Feld (Wissinger & S. G. Huber, 2002). Erst in den letzten Jahren hat sich eine neue Sichtweise auf Schulleitung durchgesetzt: Schulleitung als Management und der Schulleiter als Führungskraft. Das heisst, dass Schulleitungsmitglieder nicht mehr ausschließlich als Lehrer mit Verwaltungsfunktionen, sondern vermehrt als Führungskräfte oder Manager von Organisationen mit verschiedensten Funktionen und Rollen angesehen werden. Nicht zuletzt an der Fülle neuerer Bücher zu Führung und Management der Schule (vgl. u. a. die Bände von Ackermann & Wissinger, 1998; Dubs, 1994; Fischer & Schratz, 1999; Münch, 1999; Thiemermann, 1998), an der Kompetenzerweiterung bei Schulleitungen (Wissinger & S. G. Huber, 2002) und an der in den letzten Jahren beobachtbaren Personalauswahl und -qualifizierung von Schulleitern wird diese Tendenz deutlich. So berichten einige Autoren beispielsweise von ersten Einsätzen von Assessment-Center-Verfahren zur Auswahl und Entwicklung von potenziellen Schulleiterinnen und Schulleitern in der Schweiz, Österreich und einigen Bundesländern in Deutschland (vgl. Birri & Sonderegger, 2000; Glatzer & Roth, 2002; Wingert, 1999). Assessment-Center sind Verfahren, die sonst üblicherweise im Rahmen der Führungskräfte-Rekrutierung in der Wirtschaft Verwendung finden (s. a. Kapitel 3). Die unterschiedlichen (internationalen) Vorgehensweisen bei der Qualifizierung von Schulleitungen

beschreiben Rosenbusch und S. G. Huber (2002) und S. G. Huber (2002, 2003) näher.

Zwei Diskurse haben diese Entwicklung der Schulleitungsrolle und -funktion unterstützt (Wissinger, 2002): der Autonomie- und der Qualitätsdiskurs. Zum einen wird argumentiert, dass Schule, wo sie bislang enger staatlicher Aufsicht und Kontrolle unterliegt, erweiterter Eigenverantwortung bedarf. Die Schulleitungsfunktion durchläuft demnach einen Wandel von einer verengten Verwaltungs- und Kontrollfunktion hin zu einer Management- und Führungsfunktion (Wissinger & S. G. Huber, 2002). Dies zeigt sich bildungspolitisch durch eine deutliche Ausdehnung der Aufgaben der Schulleitung sowie die Aufwertung des Schulleiters als Führungskraft (ASD, 1999). Zum anderen wird angesichts veränderter ökonomischer, gesellschaftlicher, sozialer und individueller Anforderungen eine Führungs- und Koordinierungsfunktion der Schulleitungen betont, um im Hinblick auf die Schulqualität erfolgreich sein zu können (vgl. Pfeiffer, 2002; Wissinger & S. G. Huber, 2002). In diesem Zusammenhang sind die umfangreichen empirischen Untersuchungen der Schulwirksamkeitsforschung (school effectiveness) ebenso wie aktuelle Studien zur Schulentwicklung (school improvement) zu nennen, die hauptsächlich im englischsprachigen Raum, in den Niederlanden und den skandinavischen Ländern durchgeführt wurden. Sie zeigen laut S. G. Huber (1999a, 1999b) durchweg eines: die Bedeutung des Managements und der Führung durch die Schulleitung. Die innerschulische Steuerung gehört demnach zu den wichtigsten Qualitäts- und Einflussfaktoren für die Sicherung und Entwicklung der Qualität einer Schule (vgl. a. Bonsen, Gathen & Pfeiffer, 2002). Die Ergebnisse unterstreichen, dass erfolgreiche, als „gut“ eingeschätzte Schulen über eine kompetente und gute Schulleitung verfügen (für einen Überblick über die Untersuchungen siehe S. G. Huber, 1999a).

Beide Diskurse haben letztlich zu einer Betonung der Bedeutung der Schulleitung beigetragen. Trotz der recht jungen Tradition der Schulleitungsforschung gilt es als empirisch abgesichert, dass

Schulleitungen eine entscheidende Rolle innerhalb von Schulen zukommt (Bonsen & Pfeiffer, 1998; Wissinger & S. G. Huber, 2002). „Schulleitungen bekommen damit verstärkt Management- und Führungsaufgaben, sie müssen ihren Blick weg von der vorrangig pädagogischen Orientierung hin auf die Ebene der Institution richten“ (Pfeiffer, 2002, S. 22). De facto hat die Schulleitung somit heutzutage nicht nur pädagogische, sondern in gestiegenem Umfang auch Managementfunktionen zu erfüllen.

Nun stellt sich für die vorliegende Arbeit neben der herausgestellten Bedeutung von Schulleitungen auch die Frage nach der genauen Gewichtung verschiedener Tätigkeiten innerhalb des Schulleiteralltags und den entscheidenden eignungsrelevanten Personmerkmalen von Schulleiterinnen und Schulleitern. Wie bereits erwähnt, sind Arbeits- und Anforderungsanalysen entscheidend für eine valide, berufstypische Abbildung von Aufgaben der Schulleitung. Diese sollte nicht nur auf allgemeinen Vorstellungen, sondern idealerweise auf umfassenden empirischen Daten beruhen. Im nächsten Abschnitt werden daher Befunde zur Bedeutung von Planungs-, Organisations- und Problemlöseaufgaben im Berufsalltag von Schulleitungen vorgestellt.

### ***2.1.2 Bedeutung von Planungs-, Organisations- und Verwaltungsaufgaben im Berufsalltag von Schulleitungen***

Dass die Aufgabenbereiche Problemlösen und (Arbeitsplatz-) Organisation für die Arbeit von Führungskräften im Bereich Forschung und Entwicklung entscheidend sind, haben u. a. Schuler et al. (1995) zeigen können. Inwieweit dies auch für Schulleitungstätigkeiten gilt, soll hier erörtert werden. Dabei steht zunächst die Frage im Vordergrund, für welche Tätigkeiten Schulleitungen wie viel Zeit aufwenden. Die Schulleitungsforschung hat hierzu einige interessante Ergebnisse geliefert: So berichtet Rolff bereits 1980 von einer Befragung von Schulleitungen und Lehrkörpern an 102 hessischen Schulen unterschiedlicher Schulformen der ersten

Sekundarstufe. Dabei geht Rolff (1980) der Frage nach, wie viel Prozent des Zeitbudgets Schulleitungsmitglieder für bestimmte Tätigkeiten tatsächlich aufbringen bzw. im Idealzustand aufbringen wollen. Ergebnis ist, dass 37 Prozent der Befragten angeben, für administrative Tätigkeiten mehr als 30 Prozent der Arbeitszeit aufzuwenden.

In einer Beobachtung von Schmitz (1980), der einen Schulleiter eines Gymnasiums eine Woche lang begleitete, wird erstmals der absolute Vorrang der organisatorischen, verwaltenden Tätigkeiten im Rahmen des Berufsalltags deutlich. In dieser Einzelfallstudie zeigt sich, dass bis zu 90 Prozent der Tätigkeit des (beobachteten) Schulleiters aus Organisationsaufgaben besteht.

Eine der ersten umfassenden deutschsprachigen Untersuchungen zu den Tätigkeiten von Schulleitern führte Wolfmeyer (1981) durch. Dabei wurden 278 Schulleitungen zu ihrer täglichen Arbeit im Rahmen der Bereiche Personal, Organisation, Schülerverwaltung, Öffentlichkeitsarbeit, Haushalt und Finanzen, Beschaffung und Sachverwaltung befragt. Wolfmeyer (1981) stellt fest, dass die Schulleitungen von Gymnasien mehr als 70 Prozent ihrer Arbeitszeit für Verwaltungstätigkeiten aufwenden. Allerdings kritisieren Bosen und Pfeiffer (1998) die definitorische Eingrenzung der Tätigkeitsbereiche, insbesondere hinsichtlich der Bereiche „Verwaltungstätigkeiten“ und „Organisation“. Demnach subsumiert Wolfmeyer (1981) unter der Kategorie „Verwaltungstätigkeiten“ auch Vorgänge, die als pädagogische Tätigkeiten bezeichnet werden könnten.

Zwei weitere Studien liefern Hinweise zu Arbeitsanforderungen von Schulleitungen. Kmetz und Willower (1982) ermitteln in ihrer Studie, dass an Primarschulen bzw. an Sekundarschulen „ungeplante Treffen“ mit 32.5 bzw. 27.5 Prozent, „Schreibtischtätigkeiten“ mit 18.6 bzw. 16 Prozent und „geplante Treffen und Konferenzen“ mit 10.3 bzw. 17.3 Prozent die höchsten Zeitanteile am Zeitbudget der befragten Schulleiter aufweisen. Krüger (1983) findet bei der Beobachtung der Wochenarbeitszeiten von fünf Schulleitern

heraus, dass der Schulleiteralltag von Hektik und permanenter Anspannung geprägt ist (vgl. a. Baumert & Leschinsky, 1986).

Insgesamt kristallisiert sich als Fazit zu diesen älteren Untersuchungen - trotz der teilweise geäußerten Kritik an der begrifflichen Zuordnung - die Bedeutung der organisatorischen und verwaltenden Tätigkeiten heraus. Dies ist umso bemerkenswerter, weil das Lehrdeputat von Schulleitungen zu der Zeit im Allgemeinen noch sehr hoch war.

Die Gewichtung der heutigen Schulleitertätigkeiten ist Gegenstand weiterer aktueller Untersuchungen. Wissinger (2002) stellt die Ergebnisse der Schulleiterbefragung vor, die im Rahmen der internationalen Untersuchung „Third International Mathematics and Science Study (TIMSS)“ stattfand. Dabei wurden insgesamt 1006 Schulleitungen aus sechs verschiedenen Ländern befragt. Zwei Fragen stehen im Vordergrund: Erstens, worin bestehen Rolle und Aufgaben der Schulleitung und zweitens, was tun Schulleiterinnen und Schulleiter tagtäglich? Gerade die Ergebnisse zur zweiten Frage liefern eine quantitative Tätigkeitsanalyse, die wertvolle Hinweise für die vorliegende Arbeit geben kann. Die Schulleitungen wurden gefragt, wie viele Stunden sie monatlich mit einer Reihe von vorgegebenen Tätigkeiten verbringen. Die Ergebnisse sind in Tabelle 3 dargestellt.



**Tabelle 3: Durchschnittlicher monatlicher Stundenaufwand für verschiedene Schulleitertätigkeiten (nach Wissinger, 2002, S. 54)**

<b>Tätigkeit</b>	<b>Mittelwert (Stunden/Monat)</b>
interne Verwaltungsaufgaben	39.15
andere Aktivitäten	19.12
Elterngespräche führen	13.20
Schüler beraten und Behandlung von Disziplinproblemen	12.88
Unterrichten einschließlich Vorbereitung	12.59
Anfragen der Schulverwaltung beantworten	11.85
mit Lehrer/innen über Erziehungsaufgaben sprechen	11.42
curriculare Revisionen oder Planungen in die Wege leiten	8.78
berufliche Weiterbildung	7.07
die Schule bei offiziellen Anlässen repräsentieren	6.59
die Schule in der Gemeinde vertreten	6.16
Lehrer anleiten und fortbilden	3.64
Lehrer einstellen	2.15
Demonstrationsstunden halten	0.34

Anmerkung: N = 1006 Schulleiterinnen und Schulleiter.

Durchschnittlich nehmen interne Verwaltungstätigkeiten im Arbeitsalltag der befragten Schulleitungen den mit deutlichem Abstand größten Zeitanteil ein. Dazu gehören insbesondere auch typische Schreibtischtätigkeiten (M. Huber, 1997). Der Kategorie „andere Aktivitäten“ (Zusammenfassung einer Reihe von nicht näher definierten Tätigkeiten) weisen die Schulleiter mit ca. 19 Stunden Zeitaufwand pro Monat den zweitgrößten Wert zu. Die dem Bereich „Management und Führung“ zugeordneten Tätigkeitsbereiche „Elterngespräche führen“ und „mit Lehrer/innen über Erziehungsaufgaben

sprechen“, die zur „Pädagogik“ subsumierte „Schülerberatung bei Disziplinproblemen“ sowie „Schulverwaltungsanfragen“ (Bereich „Verwaltung“) bestimmen mit jeweils ca. 11 bis 13 Stunden pro Monat auch noch recht stark die Leitungstätigkeit.

Die internen Verwaltungsaufgaben nehmen in allen Ländern nominell den größten Anteil ein, wobei in den deutschsprachigen Ländern Österreich (48.13 Stunden/Monat) und Deutschland (41.1 Stunden/Monat) der Zeitaufwand für diesen Tätigkeitsbereich relativ deutlich über dem internationalen Mittelwert liegt (s. Tabelle 4). Eine Besonderheit ist, dass in Deutschland und Österreich die Schulleitungen zumeist auch selber unterrichten.

**Tabelle 4: Zeitaufwand pro Monat in Stunden für die Kategorie „Interne Verwaltungstätigkeit“ im Ländervergleich (nach Wissinger, 2002, S. 56)**

Tätigkeiten / Land	AUT	DNK	FRA	DEU	SWE	USA	Gesamt
interne Verwaltungstätigkeit	48.13	39.45	39.29	41.10	42.09	26.09	39.15

Wissinger (2002) schließt aus den Daten der Studie, dass in Staaten mit eher zentraler Ausrichtung der schulischen Organisationssysteme, wie Österreich, Frankreich und Deutschland, Verwaltungstätigkeiten den größten Zeitanteil ausmachen, während in Staaten mit dezentraler Ausrichtung, wie den Vereinigten Staaten, Schweden und Dänemark, für soziale Aspekte der Führung und des Managements - wie Eltern- und Lehrergespräche und Repräsentationsaufgaben - die meiste Zeit aufgewendet wird. Trotz der unterschiedlichen Gewichtung der Tätigkeiten in den einzelnen Ländern zeigt sich in dieser Studie wie auch in weiteren skandinavischen und angloamerikanischen Untersuchungen (vgl. Hultman, 1989; Wissinger, 1996), dass administrative, verwaltende Arbeit typisch für den Schulleitungsalltag in den untersuchten Ländern ist.

Welche Tätigkeiten deutsche Schulleitungen im Detail ausführen, ist Gegenstand einer weiteren Untersuchung. Wissinger (1996) baute auf der Untersuchung von Wolfmeyer (1981) auf und befragte insgesamt 196 Schulleitungen in Bayern zu ihren typischen Leitungstätigkeiten. 89.7 Prozent (höchster Wert) der befragten Schulleiter geben an, dass „Organisation“ einen hohen oder sehr hohen Anteil an ihrer Leitungstätigkeit hat. Dazu gehören explizit:

- Organisation der Klassen- und Kursbildung
- Organisation der Unterrichtsverteilung
- Organisation der Raumverteilung
- Gesamtstundenplan erstellen

Weiter stimmen 79 Prozent (zweithöchster Wert) der Befragten zu, dass „Personalaufgaben“ einen hohen oder sehr hohen Anteil an der Leitungstätigkeit einnehmen. Dazu gehören unter anderem die Personalplanung und die Personalbedarfsplanung. Darüber hinaus werden noch „Schülerverwaltung“ und „Öffentlichkeitsarbeit“ als Hauptbestandteile der Arbeit genannt, darunter auch die Auseinandersetzung mit Statistiken (s. a. Wolfmeyer, 1981). Der zeitliche Anteil, nicht aber die Bedeutung organisatorischer Tätigkeiten, hängt laut Wissinger (1996) auch von der Schulform ab. Insgesamt zeigt die Studie, dass Schulleitungen den Organisations- und Personal(verwaltungs)-Aufgaben einen herausragenden Stellenwert beimessen, unabhängig von Schulform und -größe (Wissinger, 1996, S. 93).

Bonsen (2002, 2003, 2005) und Bonsen, Gathen, Iglhaut und Pfeiffer (2002) haben in einer aktuellen Untersuchung an 30 zufällig ausgewählten Schulen versucht, führungsrelevante Merkmale von Schulleitungen zu identifizieren und Hinweise auf die Führungspraxis von Schulleitungen an besonders guten Schulen (s. u.) zu liefern. Ausgangspunkt der Analysen ist die Hypothese, dass „hinter einer erfolgreichen Schule eine fähige Schulleitung steht und umgekehrt Entwicklungsdefizite häufig mit Leitungsdefiziten zusammenhängen“ (Bonsen, 2002, S. 68). Die auf empirischem Wege

vorgenommene Unterteilung in „gute“ und „verbesserungsbedürftige“ Schulen erfolgt anhand von Kriterien, die aus der Schulqualitätsforschung abgeleitet wurden. Dabei berücksichtigt Bensen (2002) die Einschätzungen auf dreizehn Skalen von insgesamt 730 Lehrern, 382 Eltern und 2146 Schülern. Die Zuordnung der Schulen erfolgt über die Zuteilung von Rangplätzen und der anschließenden Drittelung der Rangreihe. Das obere Drittel bildet demnach die Gruppe der besonders guten Schulen, während die verbesserungswürdigen Schulen das untere Drittel der Rangreihe darstellen. Bei der Ergebnisbetrachtung zeigen sich deutliche Unterschiede in der Lehrerbeurteilung führungsbezogener Handlungsdimensionen der Schulleitungen zwischen „guten“ und „verbesserungsbedürftigen“ Schulen (bezogen auf den Mittelwert der eingesetzten Skalen). Bensen (2003) bestimmt Wettquotienten zur Vorhersage der Zugehörigkeit zur Gruppe der verbesserungsbedürftigen vs. der guten Schulen (s. Tabelle 5). Dabei erweisen sich die Dimensionen „zielbezogene Führung“, „Innovationsbereitschaft“ sowie „wahrgenommene Organisationskompetenz des Schulleiters“ als bedeutsame Merkmale von Schulleitungen an guten Schulen. Diese drei Dimensionen unterscheiden am besten zwischen guten und verbesserungsbedürftigen Schulen (Bensen, 2003). Hinsichtlich der Dimensionen „Förderung der Mitbestimmung“ und „Management sozialer Beziehungen“ hingegen erzielen alle Schulleiterinnen und Schulleiter hohe Einschätzungen. Somit eignet sich diese Handlungsdimension offensichtlich schlechter, um besonders qualifizierte Schulleitungen zu identifizieren.

Diese Ergebnisse haben auch Implikationen für das Personalmanagement. Wenn zielbezogene Führung, Innovationsbereitschaft und Organisationskompetenz erfolgskritische Personmerkmale sind, sollten diese auch in der Personalauswahl und -entwicklung von Schulleitungen berücksichtigt werden. Die Untersuchung hat zwar lediglich die von den Lehrern wahrgenommene Organisationskompetenz erfasst, es ist jedoch anzunehmen, dass ein hoher Zusammenhang zwischen tatsächlicher und wahrgenommener Organisationskompetenz vorhanden sein dürfte.

**Tabelle 5: Wettquotienten zur Vorhersage der Zugehörigkeit zur Gruppe der verbesserungsbedürftigen bzw. guten Schulen (nach Bensen, 2005)**

<b>Prädiktor</b>	<b>odds-ratio</b>	<b>p</b>
zielbezogene Führung	2.46	0.00
Innovationsbereitschaft	2.45	0.00
Förderung der Mitbestimmung	0.41	0.00
Organisationskompetenz	2.09	0.02
Visibilität	1.76	0.18
Management sozialer Beziehungen	0.07	0.93
individuelle Lehrerbegleitung	1.19	0.28
Förderung der Fortbildung	0.02	0.90
Belohnen von Lehrkräften	0.07	0.79
<i>-2 Log Likelihood</i>	<i>423.17</i>	
<i>Chi<sup>2</sup>(df)</i>	<i>148.82</i>	
<i>Nagelkerkes-R<sup>2</sup></i>	<i>.40</i>	

Anmerkung: N = 455 Lehrerinnen und Lehrer.

Bei den hier vorgestellten Studien zeigt sich allerdings, dass die Zuordnung von Tätigkeiten zu bestimmten Bereichen nicht einheitlich verläuft. So bemängeln auch Bensen und Pfeiffer (1998), dass beispielsweise Wolfmeyer (1981) unter der Kategorie „Verwaltungstätigkeiten“ Vorgänge subsumiert, die auch als pädagogische Tätigkeiten bezeichnet werden können (s. o.). Tätigkeiten, die Wolfmeyer als Verwaltungstätigkeiten versteht, decken sich nicht mit denen der TIMSS-Studie von Wissinger (2002). Ebenso fällt unter die Kategorie „Verwaltungstätigkeit“ im Rahmen der TIMSS-Studie eine ganze Reihe an Tätigkeiten, die in der Studie von Wissinger (1996) zu den „Organisationstätigkeiten“ gerechnet werden. Darüber hinaus wird in der TIMSS-Studie zwischen „Verwaltung“ und „Management / Führung“ unterschieden. Unter „Verwaltung“ sind in erster Linie Schreibtischtätigkeiten und unter „Management / Führung“ soziale Interaktionen subsumiert. Diese Unterscheidung irritiert, da in der Managementliteratur Verwaltungs-

tätigkeiten, wie das Organisieren und Planen, üblicherweise als Management- bzw. Führungstätigkeit bezeichnet werden (vgl. Sarges, 1995). Hier soll an der üblichen Terminologie festgehalten werden, nach der das Planen, Organisieren und Verwalten zu den Führungstätigkeiten gerechnet wird.

Trotz der uneinheitlichen Begriffsverwendung wird durch die Studien Folgendes deutlich: Schulleiterinnen und Schulleiter müssen tagtäglich ein breites Spektrum an Aufgaben bewältigen. Den größten Teil ihrer Leitungstätigkeit verbringen Schulleitungen mit Planungs-, Verwaltungs- und Organisationsaufgaben. Diesen Aufgaben weisen Schulleitungen auch selber eine sehr große Bedeutung für ihre Leitungsfunktion zu (vgl. u. a. Wissinger, 1996, 2002; Wissinger & S. G. Huber, 2002). In den vorgestellten empirischen Arbeiten werden u. a. folgende Aufgaben als exemplarisch für Planungs-, Verwaltungs- und Organisationsaufgaben von Schulleitungen genannt (s. Tabelle 6).

**Tabelle 6: Typische Aufgaben von Schulleitungen**

<b>Typische Aufgaben von Schulleitungen</b>
<b>Planen von Raum-, Aufsichts- und Kursverteilungen</b>
<b>Stundenplangestaltung</b>
<b>Vertretungen sichern</b>
<b>Konferenzen vorbereiten</b>
<b>Anfragen der Schulverwaltung bearbeiten</b>
<b>Statistiken bearbeiten</b>
<b>Personalplanungen</b>

Bonsen und Pfeiffer (1998, S. 16) ziehen in ihrem Überblick zur Schulleitungsforschung folgendes Fazit:

Die empirisch abgesicherten Forschungen zu Schulleitung in Deutschland haben einige wichtige Ergebnisse hervorgebracht:

- Schulleitung, insbesondere im Bereich der Sekundarschulen, ist ganz überwiegend eine männliche Domäne.
- Das Schulleitungshandeln ist von Organisations- und Managementtätigkeiten geprägt.
- Das Schulleitungshandeln unterliegt einem permanenten Zeitdruck sowie häufig und schnell wechselnden Sozialkontakten und Themenwechseln.
- Die Schulleitung hat keine eindeutig definierte Rolle nach innen und außen, sondern muß einen Rollenmix mit teilweise konfligierenden Zielen beherrschen.

Es bedarf einer ganzen Reihe an Fähigkeiten, um eine Schule erfolgreich zu leiten. Neben den sozialen, zwischenmenschlichen Aspekten von Führung und Management verdeutlichen die aufgeführten Studien, dass das kompetente Bearbeiten der Organisations- und Verwaltungsaufgaben entscheidend ist. Diese umfangreichen administrativen, organisatorischen Tätigkeiten (s. o.) stellen typischerweise Anforderungen an die Planungs- und Problemlösefähigkeit von Schulleitungen. Korrespondierend mit diesen Befunden weisen zahlreiche Autoren von Schulleitungs-Ratgebern darauf hin, dass Fähigkeiten im Planen und Problemlösen für Schulleitungen zu den entscheidenden erfolgs- und eignungsrelevanten Personmerkmalen zählen (u. a. Dubs, 1994; Fischer & Schratz, 1999; Herz & Lohmann, 1999; Thiemermann, 1998).

### ***2.1.3 Implikationen für die vorliegende Arbeit***

In diesem Abschnitt wurden die Bedeutung der Personmerkmale „Planungsfähigkeit“ und „Problemlösefähigkeit“ für Führungskräfte im Allgemeinen und Schulleitungen im Speziellen beleuchtet und typische Tätigkeiten und Anforderungen der Arbeit von Schulleitungen dokumentiert. Dabei ist es angesichts der Bedeutung von Planungs-, Verwaltungs- und Organisationstätigkeiten erstaunlich, dass die Erfassung der tatsächlichen

Fähigkeiten und Leistungen von Schulleitungen im Planen und Problemlösen solcher Aufgaben bisher nicht expliziter Gegenstand der Schulleitungsforschung geworden ist. Dies mag in erster Linie an dem grundsätzlichen Mangel an geeigneten eignungsdiagnostischen Instrumenten zum Planen und Problemlösen liegen (vgl. Ueckert, 1995). Internetbasierte eignungsdiagnostische Verfahren zum Planen und Problemlösen von Schulleiterinnen und Schulleitern wurden bisher noch gar nicht publiziert. Wottawa (1995, S. 180) betont für die Diagnostik von Personmerkmalen die Unterscheidung zwischen

1. der Bedeutung und
2. dem Ausprägungsgrad.

Das bedeutet, dass es Folgendes zu berücksichtigen gilt: Erstens, ist eine Fähigkeit bzw. ein Personmerkmal wie beispielsweise die „Problemlösefähigkeit“ wichtig und bedeutend für einen Arbeitsplatz? Zweitens, wie hoch sollte der Ausprägungsgrad sein, um die Arbeit erfolgreich durchführen zu können? Die vorgestellten Studien zeigen, dass für die Bewältigung der umfangreichen Planungs-, Organisations- und Verwaltungsaufgaben von Schulleitungen Fähigkeiten im Planen und Problemlösen bedeutend sind und der Ausprägungsgrad (möglichst) hoch sein sollte (vgl. Bonsen, 2002). Fähigkeiten im Planen und Problemlösen können somit als entscheidende personelle Leistungsvoraussetzungen zur erfolgreichen Bewältigung von Schulleitungsaufgaben verstanden werden (vgl. Schuler & U. Funke, 1995). Ziel dieser Arbeit ist es daher, geeignete Instrumente zur Erfassung von Planungs- und Problemlöseleistungen zu entwickeln. Im Sinne des Simulationsansatzes (s. Abbildung 1) sollen dabei die in diesem Abschnitt vorgestellten, aktuellen Studien zu den Tätigkeiten von Schulleitungen Berücksichtigung finden (zur Umsetzung und Operationalisierung s. Kapitel 4). Gemäß der Vorgehensweise der Verfahrensentwicklung nach dem Konstruktansatz gilt es darüber hinaus, die psychologischen Theorien und Befunde hinsichtlich der Merkmale „Planungsfähigkeit“ und „Problemlösefähigkeit“ zu betrachten. Diese konstruktorientierten Grundlagen thematisiert das folgende Kapitel.



## **2.2 Grundlagen Planen und Problemlösen**

Die Begriffe „Planen“ und „Problemlösen“ haben eine gemeinsame Tradition in der Denkpsychologie, die sich weit zurückverfolgen lässt. Im folgenden kurzen Überblick über die Historie werden vor allem Arbeiten betrachtet, die für das Verständnis beider Konstrukte einflussreich sind und anhand derer wichtige Ansätze der Denkpsychologie erläutert werden können. Die eignungsdiagnostisch relevante Definition von Problemlösen im Rahmen dieser Arbeit wird im Abschnitt 2.2.3 erörtert, um anschließend die empirische Befundlage zu dem Konstrukt zu beleuchten. Dabei interessiert insbesondere der vielfach untersuchte Zusammenhang zwischen Intelligenz und Problemlösen. Abschnitt 2.2.6 führt in das eignungsdiagnostische Verständnis von „Planen“ im Rahmen dieser Arbeit ein, während die Ausführungen über den aktuellen empirischen Stand der Planungsforschung die theoretischen Grundlagen abschließen.

### ***2.2.1 Planen und Problemlösen in der Tradition der Denkpsychologie***

Aktuelle Darstellungen zur Geschichte der Problemlöseforschung finden sich u. a. bei Bösel (2001), J. Funke (2003), Hussy (1998), Klauer (1995), zum Planungskonzept bei J. Funke und Fritz (1995). Daher soll hier nur auf wesentliche Aspekte eingegangen werden, die aus diagnostischer Perspektive von Bedeutung sind. Auf weitere - hier nicht näher erläuterte - einflussreiche Arbeiten, wie beispielsweise das opportunistische Planen von Hayes-Roth und Hayes-Roth (1979), Hackers Drei-Ebenen-Modell der Handlungsregulation (1986), Österreichs 5-Ebenen-Modell (1981) oder J. R. Andersons ACT-Theorie (1983, 1996) sei jedoch ausdrücklich hingewiesen.

#### *Probleme, Problemlösen, Problemtypen*

Forscher der gestaltpsychologischen Schule befassen sich bereits zu Beginn des 20. Jahrhunderts mit Denkprozessen und insbesondere dem

Problemlösen (vgl. Ach, 1905; Duncker, 1935; Köhler, 1917; Selz, 1913; Wertheimer, 1943). Eine der ersten psychologischen Definitionen für „Probleme“ wird von Duncker (1935) vorgestellt: „Ein Problem entsteht z. B. dann, wenn ein Lebewesen ein Ziel hat und nicht weiß, wie es dieses Ziel erreichen soll“ (S. 1). Spätere Definitionen weisen im Kern die gleichen Merkmale auf, wie beispielsweise die von Dörner (1979): „Ein Individuum steht einem Problem gegenüber, wenn es sich in einem inneren oder äußeren Zustand befindet, den es aus irgendwelchen Gründen nicht für wünschenswert hält, aber im Moment nicht über die Mittel verfügt, um den unerwünschten Zustand in den wünschenswerten Zustand zu überführen“ (S. 10). Das Verständnis von „Problemlösen“ ist daran angelehnt. Darunter wird allgemein der Vorgang verstanden, der einen gegebenen Ausgangszustand in einen gewünschten Zielzustand überführt und dabei eine Barriere überwindet (vgl. Dörner, 1979; Duncker, 1935; Edelmann, 2000; Hussy, 1984; Lüer & Spada, 1992).

Dörner (1979) betrachtet die zu überwindende Barriere genauer und klassifiziert Probleme nach verschiedenen Typen von Barrieren. In seiner viel zitierten Einteilung von Problemen unterscheidet Dörner (1979) vier verschiedene Barriere- bzw. Problemtypen in Abhängigkeit vom „Bekanntheitsgrad der Mittel“ und von der „Klarheit der Zielkriterien“ (s. Tabelle 7).

**Tabelle 7: Klassifikation von Barrieretypen nach den Dimensionen „Bekanntheitsgrad der Mittel“ und „Klarheit der Zielkriterien“ (nach Dörner, 1979, S. 14)**

		Klarheit der Zielkriterien	
		<i>hoch</i>	<i>gering</i>
Bekanntheitsgrad der Mittel	<i>hoch</i>	Interpolationsbarriere	dialektische Barriere
	<i>gering</i>	Synthesebarriere	dialektische und Synthesebarriere

Bei Problemen mit einer Interpolationsbarriere ist das Ziel bekannt und die Mittel zur Lösung sind gegeben. Es fehlt demnach die richtige Kombination oder die Festlegung einer zeitlichen Abfolge. Bei der Synthesebarriere ist das Ziel klar erfasst, jedoch sind wichtige Operationen zur Transformation nicht bekannt oder werden nicht in Betracht gezogen. Probleme mit einer dialektischen Barriere zeichnen sich dadurch aus, dass das Ziel nicht genau bekannt ist, während die Mittel transparent sind.

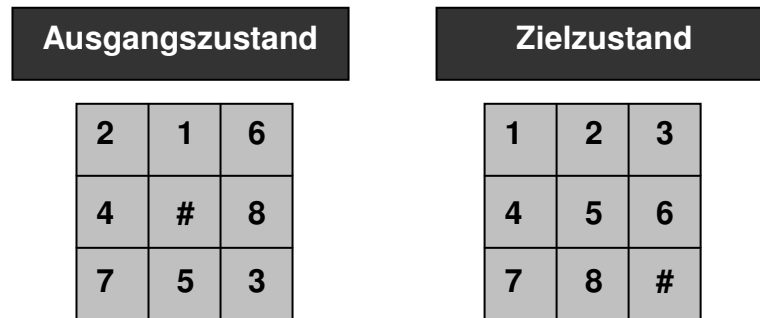
Die Problemlöseforschung befasste sich zuerst mit Problemen mit einer Interpolationsbarriere. Eines der bekanntesten Probleme ist der „Turm von Hanoi“. Die Abbildung 3 zeigt eine einfache Version der Aufgabenstellung. Ziel ist es, drei unterschiedlich große Scheiben durch Umlegen vom linken Stab auf den rechten zu überführen. Dabei sind jedoch nur bestimmte, klar definierte Handlungsmöglichkeiten erlaubt, die auch als Transformationen bezeichnet werden. Jede Scheibe darf von einem Stab auf einen beliebigen anderen gelegt werden, solange jeweils nur die oberste Scheibe umgelegt und keine größere Scheibe auf einer kleineren abgelegt wird. In sieben Zügen ist dieses Problem zu lösen.<sup>4</sup>



**Abbildung 3: „Turm von Hanoi“-Problem mit drei Scheiben**

Auch die in Abbildung 4 dargestellte Verschiebeaufgabe fällt in die Gruppe der Probleme mit einer Interpolationsbarriere. Die Aufgabenstellung lautet, die neun Felder der Matrix so zu verschieben, dass sie vom Ausgangszustand (linke Abbildung) in den Zielzustand (rechte Abbildung) gebracht werden. Dabei ist das mit der Raute (#) gekennzeichnete Feld frei und kann mit einem beliebigen anderen Feld aufgefüllt werden.

<sup>4</sup> Der Lösungsweg zu dem hier geschilderten „Turm von Hanoi“-Problem ist unter <http://www.mathematische-basteleien.de/hanoi.htm> dargestellt.



**Abbildung 4: Verschiebeprobem<sup>5</sup> (nach J. R. Anderson, 1996)**

Anhand dieser Interpolationsprobleme lassen sich einige Begriffe der Denkpsychologie erläutern. Im Verschiebeprobem ist die unerwünschte Ausgangslage der Felder im linken Zustand in die gewünschte richtige Reihenfolge (rechts) zu bringen. Dabei ist aber nicht auf Anhieb klar, wie dies zu bewältigen ist. Es liegt vor der direkten Lösung somit eine Hürde bzw. Barriere. Auf der Suche nach der Lösung durchforstet nach J. R. Anderson (1996) der Problemlösende den Problemraum, der aus verschiedenen Problemzuständen besteht. Der objektive Problemraum ist beim Verschiebeprobem in Abbildung 4 die Zahl der verschiedenen Zustände, die ein Problemlöser durch Verschieben erreichen kann. Der subjektive Problemraum ist das mentale Abbild der Aufgabenstruktur beim Problemlösenden. Gerade die Analyse der subjektiven und objektiven Anforderungen ist für diagnostische Fragestellungen von Interesse, wie im Abschnitt 2.2.4 erläutert wird.

Das „Turm von Hanoi“- und das Verschiebeprobem sind in ihren objektiven Anforderungen genau zu definieren und daher in experimentellen Studien vielfach verwendet worden, um Problemlöseprozesse (Anzai & Simon, 1979; Klix, 1971; Kotovsky, Hayes & Simon, 1985; VanLehn, 1991) und Planungsprozesse (Karat, 1982) zu betrachten.

<sup>5</sup> Der Lösungsweg findet sich bei J. R. Anderson (1996).

### *Komplexe Probleme*

Nach Auffassung einiger Autoren (vgl. Dörner, Kreuzig, Reither & Stäudel, 1983) unterscheiden sich diese Interpolations- oder „künstlichen“ Probleme von realen Alltagsproblemen. Daher stellen Dörner et al. (1983) diesen Problemen Probleme mit einer dialektischen bzw. synthetischen Barriere gegenüber. Die Forschung an Problemen mit geringer Bekanntheit der Mittel und geringer Klarheit der Ziele ist in Deutschland vor allem bekannt geworden durch die Arbeiten der Bamberger Gruppe um Dietrich Dörner (in England: Broadbent, 1977).

**Tabelle 8: Eigenschaften komplexer Probleme nach Dörner et al. (1983)**

<b>Komplexität</b>	Die Testpersonen müssen bei der Lösung eines Problems viele Aspekte berücksichtigen. Als Richtmaß wird vielfach die Zahl der vernetzten Variablen verwendet.
<b>Vernetztheit</b>	Die Variablen einer Problemsituation sind nicht unabhängig voneinander, sondern zeichnen sich durch einen hohen Grad an Abhängigkeit aus.
<b>Eigendynamik</b>	Ohne Eingriff von außen verändert das System eigenständig seinen Zustand.
<b>Intransparenz</b>	Für die Lösung relevante Informationen sind den Problemlösenden teilweise unbekannt.
<b>Polytelie</b>	Die Problemlösenden müssen gleichzeitig viele verschiedene, teilweise gegensätzliche Ziele verfolgen.
<b>Offenheit</b>	Die Ziele sind teilweise ungenau bestimmt und müssen von den Problemlösenden zunächst präzisiert werden.

Der unter dem Namen „komplexes Problemlösen“ eingeführte Gegenstandsbereich hat in den letzten Jahren beträchtliche

Forschungsaktivitäten angeregt. Dietrich Dörner und seine Mitstreiter entwickelten in den siebziger Jahren erste komplexe Probleme, die computergestützt dargeboten wurden.<sup>6</sup> Dabei mussten Problemlöser beispielsweise die Rolle des Bürgermeisters der fiktiven Kleinstadt Lohhausen einnehmen und Güter, Finanzen und Bürger verwalten. Nach Auffassung von Dörner et al. (1983) zeichnen sich gerade reale Probleme durch geringe Bekanntheit der Mittel und geringe Transparenz der Ziele aus. Komplexe Probleme lassen sich nach Dörner et al. (1983) anhand von sechs Merkmalen beschreiben (s. Tabelle 8).

#### *Phasen des Problemlösens, Planen als ein Teil des Problemlöseprozesses*

Neben der Betrachtung von Problemtypen oder von Eigenschaften von Problemen werden auch mögliche Phasen des Problemlösens bzw. Planens thematisiert. In diesen Arbeiten wird die enge Verzahnung von „Planen“ und „Problemlösen“ deutlich.

Polya hat bereits 1945 Phasen des Problemlösens vorgestellt. Dazu zählen seiner Ansicht nach folgende:

1. Bestimmung des Zieles
2. Analyse der Ausgangssituation und Aufbau einer mentalen Repräsentation, eines Situationsmodells
3. Bestimmung der Lösungsstrategie und Planung von Lösungsschritten
4. Ausführen des Lösungsplans, begleitende Kontrolle und gegebenenfalls Modifizierung der Lösung
5. Evaluation der Lösung

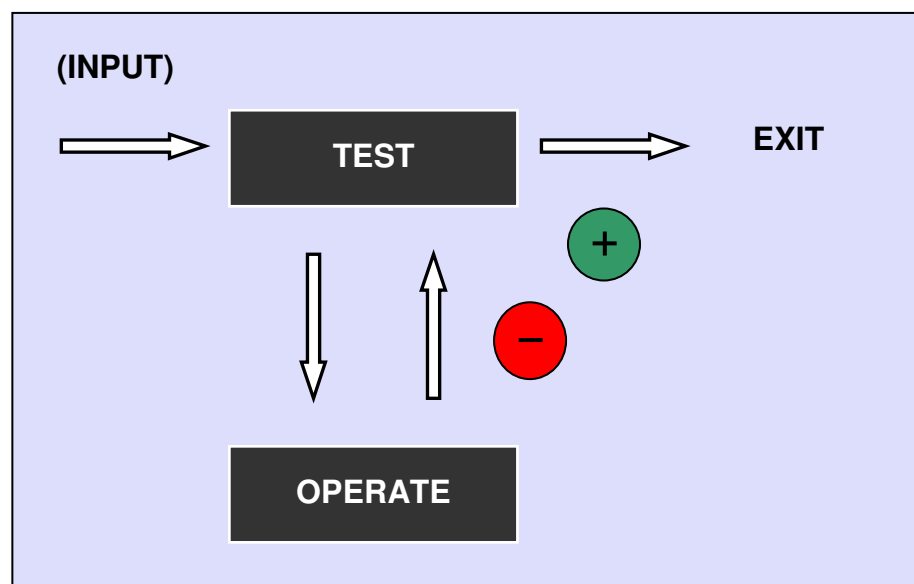
Planen wird von Polya (1945) und vielen anderen Autoren (u. a. Dörner, 1989, s. Abbildung 6) als *ein* Schritt innerhalb des Problemlöseprozesses verstanden. Handlungsschritte oder Phasen des Problemlösens werden seit der Arbeit Polyas (1945) zumeist in ähnlicher Weise beschrieben, wenngleich Newell und Simon (1972) den Problemlöseprozess deutlich kürzer darstellen. Angeregt durch die Nachrichten- und Computertechnologie entwickelten

---

<sup>6</sup> Die Verfahrensklasse der computergestützten Szenarios wird in Kapitel 3 näher vorgestellt.

Newell, Shaw und Simon (1960) das Computerprogramm „General Problem Solver“, das ihre methodologische Grundidee verdeutlicht: Wenn es gelänge, ein Computerprogramm zu schreiben, das Probleme genauso löst wie ein menschlicher Problemlöser, so könnte dieses Programm als ein hinreichendes Modell der menschlichen Denkprozesse verstanden werden. Es wird dabei von einer allgemeinen, umfassenden Problemlösefähigkeit ausgegangen. Newell und Simon (1972) verstehen das Problemlösen als einen Prozess aus lediglich zwei Phasen: erstens, dem Verstehen des Problems und zweitens, dem Suchen nach einer Lösung.

Eine andere klassische Arbeit von Miller, Galanter und Pribram (1960) hat die Entwicklung von Handlungstheorien und die Begriffsbildung entscheidend beeinflusst. Anstelle der bis dahin vorherrschenden Fokussierung auf die Außensteuerung des Verhaltens wie beim Modell des Reiz-Reaktion-Lernens beispielsweise, entwickeln die Autoren ein kybernetisches Modell mit einer Rückkopplungsschleife (s. Abbildung 5).



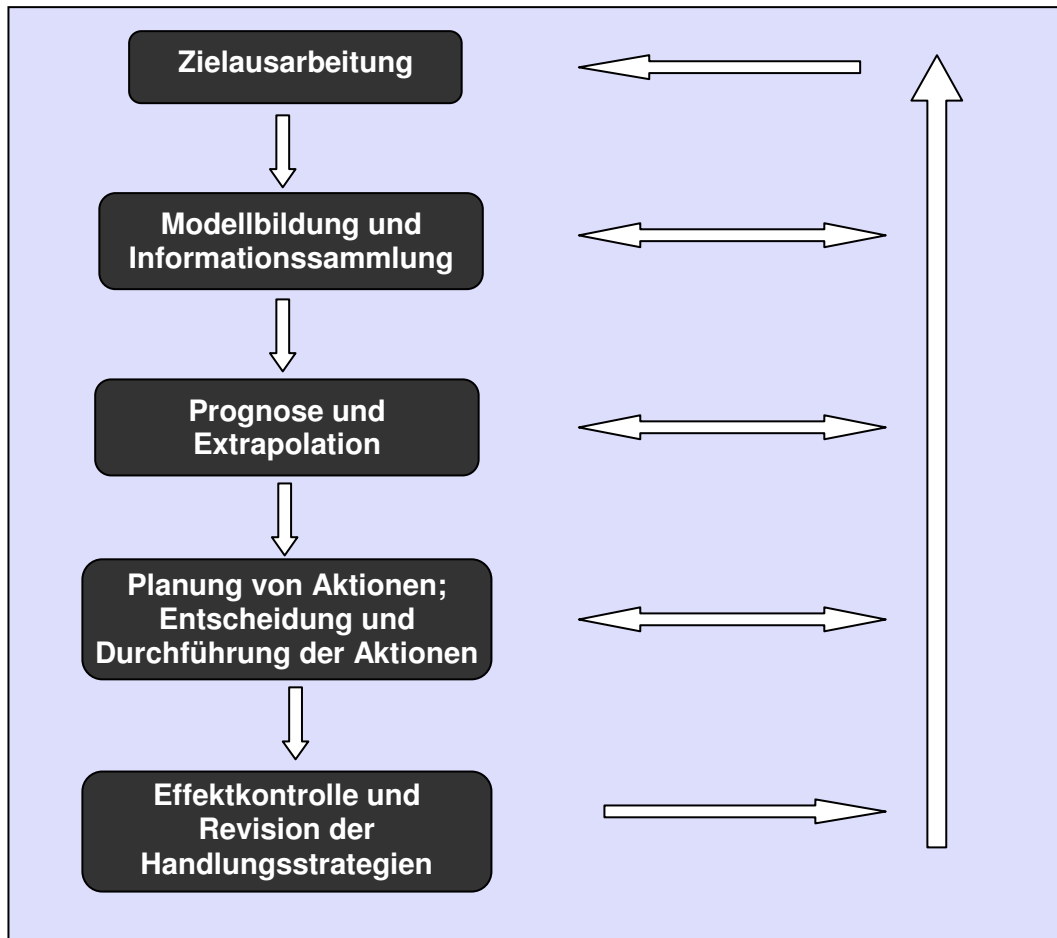
**Abbildung 5: TOTE-Einheit nach Miller, Galanter und Pribram (1960)**

Die Handlungsregulation wird demnach mittels TOTE-Einheiten (Test-Operate-Test-Exit) modelliert. Ziele können nach Miller et al. (1960) nur verwirklicht werden, wenn die „Test“-Einheit dem Handelnden bereits

während der Ausführung die Übereinstimmung der Handlung mit dem anvisierten Plan rückmeldet. Einlaufende Informationen (INPUT) werden auf eine Ist-Soll-Abweichung hin kontrolliert (TEST). Bei einer Abweichung (-) greift das System ein und versucht, die Abweichung auszugleichen (OPERATE). Dieser Prozess wird solange wiederholt, bis der Ist- dem Sollzustand entspricht (+). Dies führt dann zum Ende der TOTE-Einheit (EXIT). Das Modell wurde sowohl auf das Problemlösen als auch das Planen angewendet, was die Nähe dieser beiden Konzepte hinsichtlich ihrer theoretischen Fundierung verdeutlicht (vgl. J. Funke & Fritz, 1995; Hussy, 1998).

Im Rahmen des Forschungsansatzes „komplexes Problemlösen“ werden neben der Beschreibung von Problemeigenschaften auch Begriffsdefinitionen zum Planen vorgestellt, so beispielsweise von Dörner (1989), der die Planung, Entscheidung und Durchführung von Aktionen als einen Teil des Handlungsablaufs beim Bearbeiten eines komplexen Problems versteht (s. Abbildung 6). Planen ist in diesem Modell *ein* Schritt innerhalb des gesamten Ablaufs. Planen besteht nach Dörner (1989) darin, einzelne Aktionen auf ihre Konsequenzen hin zu untersuchen und diese Einzelaktionen probeweise zu Ketten zusammen zufügen, um wiederum die Konsequenz solcher Aktionsketten zu analysieren. Nach Dörner (1989) bestehen die Ketten aus drei Einheiten, wenn sie vollständig sind, „nämlich dem *Bedingungsteil*, dem eigentlichen *Aktionsteil* und dem *Ergebnisteil*“ (S. 235, Hervorhebungen im Original). Dadurch könnten verzweigte oder auch kreisförmige Planungssequenzen entstehen. Planen ist demnach „Probearbeiten“ bzw. das Überprüfen der Konsequenzen von Aktionen.





**Abbildung 6: Modell des Handlungsablaufs nach Dörner (1989, S. 67)**

Die geringe Trennschärfe zwischen den beiden Begriffen „Planen“ und „Problemlösen“ wird auch in den Folgearbeiten der Bamberger Gruppe deutlich. Von der Weth und Strohschneider (2002) übertragen gar alle Merkmale des komplexen Problemlösens auf - wie sie es nennen - Planungskonstellationen. Pläne könnten demnach die typischen, oben genannten Eigenschaften komplexer Probleme annehmen.

Bevor auf das eignungsdiagnostische Begriffsverständnis von Planen und Problemlösen für diese Arbeit näher eingegangen wird, sollen aufgrund der besonderen Bedeutung für die vorliegende Arbeit im Folgenden die wichtigsten Aspekte des Berliner Intelligenzstrukturmodells (Jäger, 1982, 1984) vorgestellt werden.

### **2.2.2 Berliner Intelligenzstrukturmodell**

Wie eingangs erläutert, ist für diese Arbeit die Frage bedeutend, ob Planungs- und Problemlösefähigkeit von der Intelligenz unterscheidbare Konstrukte sind bzw. inwieweit von der Intelligenz unterscheidbare Fähigkeiten notwendig sind, um die Leistungen in Planungs- und Problemlösesituationen zu erklären. Die Überschneidung der Definitionen von Problemlösen und Intelligenz wird nicht zuletzt deutlich durch die Aussage Sternbergs (1982, S. 225; Hervorhebung im Original):

Reasoning, problem solving, and intelligence are so closely interrelated that it is often difficult to tell them apart. ... Whatever *intelligence* may be, *reasoning* and *problem solving* have traditionally been viewed as important subsets of it. Almost without regard to how *intelligence* has been defined, *reasoning* and *problem solving* have been part of the definition.

Die folgende kurze Vorstellung des Berliner Intelligenzstrukturmodells (BIS) von Jäger (1982, 1984) geschieht zum einen aufgrund der Nähe der Konstrukte „Intelligenz“ und „Problemlösen“ sowie der Bedeutung des Modells im Rahmen der Problemlöseforschung. Zum anderen sollen Aspekte des Modells in die hier angestrebte Verfahrensgestaltung mit einfließen. Dabei wird eine Variation des Aufgabenmaterials des *webPostkorbs* im Sinne der unterschiedlichen Inhaltsbereiche des Berliner Intelligenzstrukturmodells anvisiert (s. Kapitel 4).

Das Berliner Intelligenzstrukturmodell steht in der Tradition der Strukturmodelle in der Intelligenzforschung und versucht, die bisherigen Ansätze von Thurstone (1938), Cattell (1971) und Vernon (1971) zu integrieren.

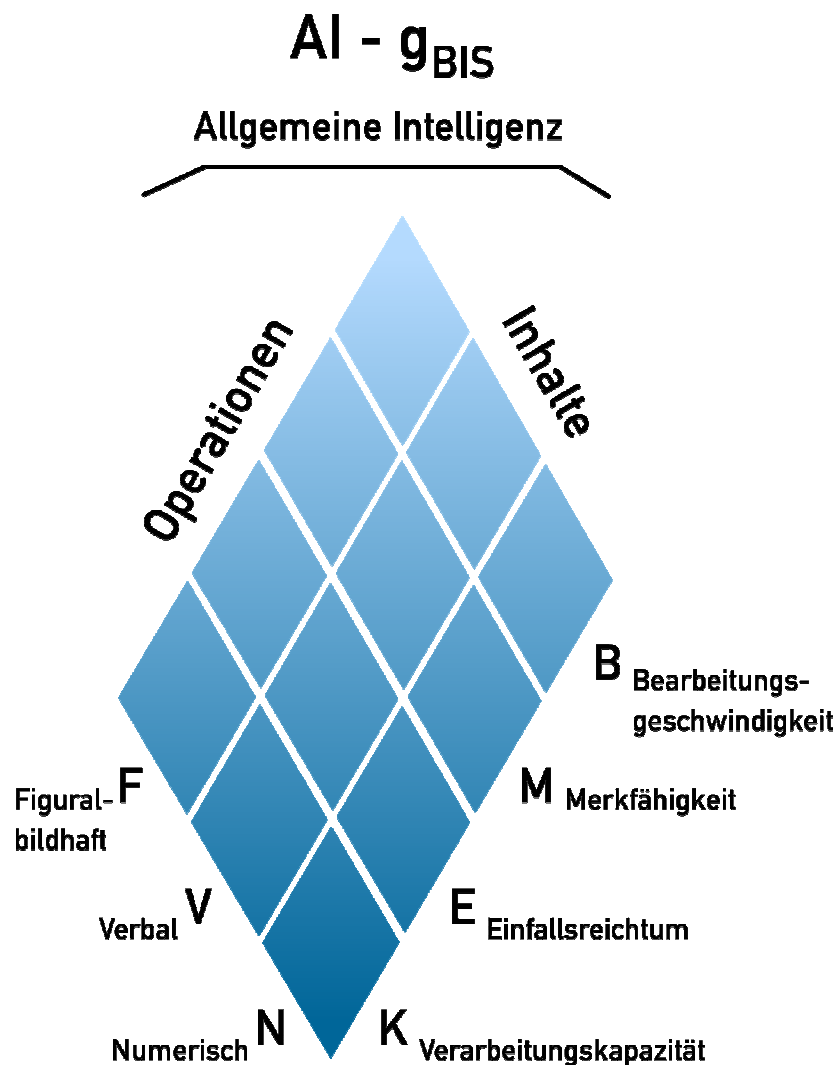


Abbildung 7: Berliner Intelligenzstrukturmodell (Jäger, Süß & Beauducel, 1997, S. 5)

Die Struktur kognitiver Leistungen ist nach Jäger (1984) - wie in Abbildung 7 zu sehen - hierarchisch und bimodal aufgebaut. Es werden sieben hochgradig generelle Hauptkomponenten (Fähigkeitsbündel) unterschieden, die den beiden Modalitäten „Inhalte“ und „Operationen“ zugeordnet werden. Modalitäten sind hier als unterschiedliche Aspekte zu verstehen, unter denen sich dieselben Intelligenzleistungen klassifizieren lassen. Unter die Modalität „Inhalte“ lassen sich die drei verschiedenen Fähigkeitsbündel figural-bildhaft (F), numerisch (N) und verbal (V) subsumieren (s. Abbildung 7). Einheitsstiftendes Merkmal ist jeweils die Gemeinsamkeit des Aufgabenmaterials. Dabei gilt der Grad ihrer Aneignung und Verfügbarkeit als mitbestimmend.

Auf der Ebene der „Operationen“ differenziert das Modell zwischen Einfallsreichtum (E), Merkfähigkeit (M), Bearbeitungsgeschwindigkeit (B) und Verarbeitungskapazität (K), wobei im Zusammenhang mit Planen und Problemlösen in erster Linie die Verarbeitungskapazität von Bedeutung ist (s. Abschnitt 2.2.5). Unter der Verarbeitungskapazität (K) versteht Jäger (1984) die „...Verarbeitung komplexer Informationen bei Aufgaben, die nicht auf Antriebe zu lösen sind, sondern Heranziehen, Verfügbarhalten, vielfältiges Beziehungsstiften, formallogisch exaktes Denken und sachgerechtes Beurteilen von Informationen erfordern“ (S. 31).

### **2.2.3 Definition Problemlösen**

Soll die Problemlöseleistung bzw. -fähigkeit einer Person diagnostiziert werden, gilt es zunächst zu klären, was unter Problemlöseleistung und insbesondere unter der Qualität von Problemlösen verstanden wird. Dies entspricht einer resultatsorientierten Sichtweise. Aufbauend auf der einschlägigen psychologischen Forschungstradition und in Anlehnung an das Begriffsverständnis im Rahmen der PISA-Studien (vgl. Klieme, J. Funke, Leutner, Reimann & Wirth, 2001) wird als Problemlösen im Rahmen dieser Arbeit Folgendes aufgefasst:

Zum Problemlösen ist zielorientiertes Denken und Handeln notwendig, um Situationen (Probleme) bewältigen zu können, für die keine routinierten Vorgehensweisen verfügbar sind. Die Problemlöser wissen dabei nicht unmittelbar, wie sie ein anvisiertes Ziel erreichen können. Problemlöseprozesse beinhalten das Verstehen der Problemsituation und deren schrittweise Veränderung bis zur möglichen Erreichung der Ziele. Dies erfolgt durch das Erkennen von entscheidenden Zusammenhängen, durch die Analyse der verfügbaren Informationen, durch analytisches und schlussfolgerndes Denken und durch das folgerichtige Entscheiden. Die Qualität des Problemlösens zeigt sich in erster Linie durch die Angemessenheit der erreichten Lösung, für die das Verständnis der

Problemsituation und die Denkprozesse bei der Problembearbeitung entscheidend sind.

Diese Definition lehnt sich an das allgemeine Verständnis vom Problemlösen als kognitive Leistung an. Ein solches Begriffsverständnis findet sich in fast allen eignungsdiagnostischen und auch pädagogisch-psychologischen Untersuchungen wieder (vgl. PISA, Klieme et al., 2001). Motivationale und emotionale Faktoren, Interesse oder der soziale Kontext können beim Problemlöseprozess ebenfalls eine Rolle spielen (vgl. Schaub, 2001; Vollmeyer & J. Funke, 1999). In dieser Arbeit geht es in erster Linie jedoch um die Erfassung von kognitiv-analytischem Problemlösen.

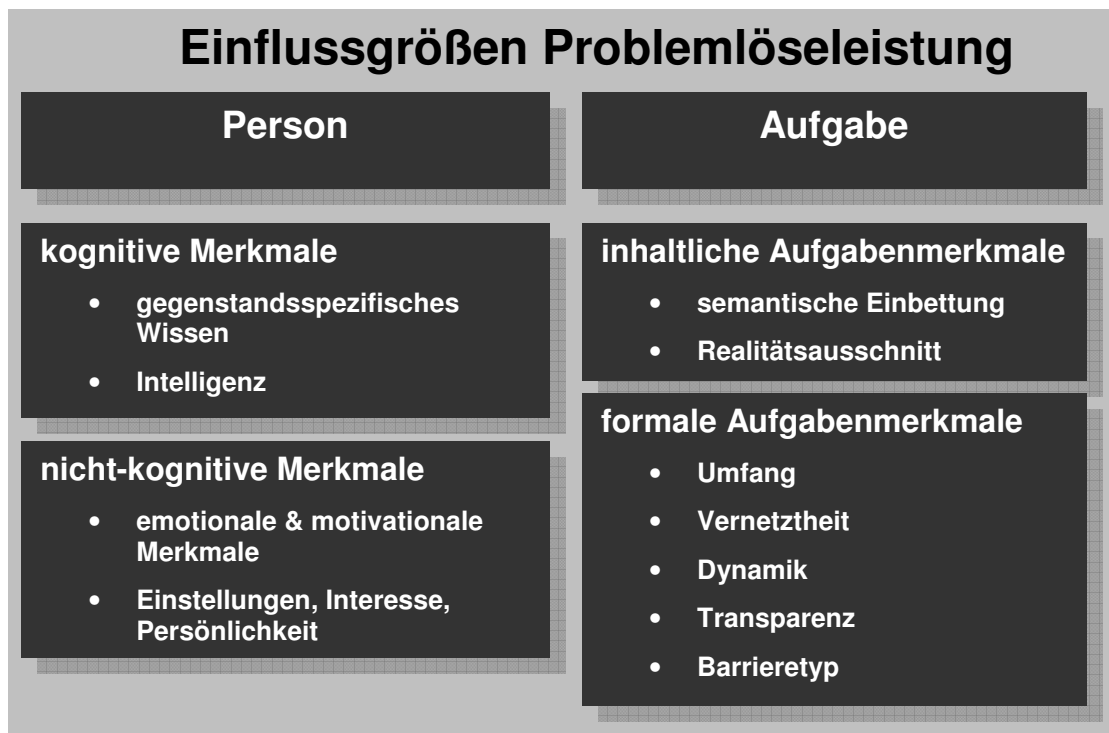
#### ***2.2.4 Einflussgrößen der Problemlöseleistung***

Für die Entwicklung eines simulationsbasierten Verfahrens zum Problemlösen sind insbesondere Faktoren zu berücksichtigen, die die Problemlöseleistung und die Schwierigkeit eines Problems beeinflussen. Verschiedene Autoren haben diesbezüglich Klassifikationen vorgestellt. Eine der bekanntesten Taxonomien liefert J. Funke (1990), später Vollmeyer und J. Funke (1999), die der von Hussy (1984) nah ist. Es wird dabei zumeist zwischen zwei Einflussfaktoren unterscheiden: den Personmerkmalen auf der einen und den Aufgabenmerkmalen auf der anderen Seite (Hussy, 1984, nennt letztere Problemmerkmale). J. Funke (1990) erweitert diese beiden Größen noch um die Situationsmerkmale, die sich aber nach M. Kersting (1999) auch den Aufgabenmerkmalen zurechnen lassen können.

Gemäß Hussy (1984) und J. Funke und Fritz (1995) entsteht erst durch das Zusammenspiel zwischen Person und Aufgabe, zwischen Problemlöser und Problem, die Schwierigkeit. Abbildung 8 veranschaulicht die beiden Einflussgrößen der Problemlöseleistung. Dabei lassen sich Personmerkmalen in kognitive und nicht-kognitive Merkmale unterscheiden. Die Aufgabenmerkmale sind hier weiter in inhaltliche und formale unterteilt. Diese Einteilung der Merkmale lässt sich sowohl auf Interpolations- oder „künstliche“ Probleme als auch auf komplexe Probleme anwenden.

Trotz Kritik von Süß (1996) oder Strohschneider (1991b) hat sich diese Klassifikation für diagnostische Fragestellungen als hilfreich erwiesen. Süß (1996) beispielsweise bemängelt, dass psychologische Anforderungen an den Problemlösenden zu wenig berücksichtigt würden. Er folgert, dass diese Einteilung primär dazu diene, mögliche Einflussgrößen der Problemlöseleistung zu strukturieren. Gerade diese Strukturierungsfunktion soll hier bewusst genutzt werden, weshalb an der Darstellung festgehalten wird.

Die Klassifikation dient der Unterteilung der verschiedenen empirischen Befunde zu Einflussgrößen der Problemlöseleistung. Es sollen somit zum einen schwierigkeitsvariierende Faktoren herauskristallisiert werden, die für die hier angestrebte Entwicklung eines Problemlöseverfahrens entscheidend sind. Zum anderen sollen die Forschungsergebnisse zu Einflussgrößen der Problemlöseleistung bei der grundsätzlichen Verfahrensentwicklung und der Verfahrensevaluation Berücksichtigung finden.



**Abbildung 8: Klassifikation der Einflussgrößen der Problemlöseleistung (vgl. Hussy, 1984; Vollmeyer & J. Funke, 1999)**

## **2.2.5 Empirischer Forschungsstand zum Problemlösen**

### *2.2.5.1 Aufgabenmerkmale als Einflussgrößen der Problemlöseleistung*

Zu den Aufgabenmerkmalen oder Problemmerkmalen, wie es Hussy (1984) nennt, gehören inhaltliche und formale Aspekte, die zusammen die genaue Ausgestaltung eines Problems konstituieren.

Inhaltliche Merkmale sind die semantische Einbettung und der gewählte Realitätsausschnitt (J. Funke, 1990). Dazu zählen beispielsweise die Rahmengeschichte, die Instruktion und die Namen der Variablen. Die inhaltlichen Aufgabenmerkmale korrespondieren mit dem Vorwissen des Problemlösers. Studien haben zeigen können, dass die semantische Einbettung formal identische Probleme unterschiedlich schwierig erscheinen lässt. So variieren Kotovsky, Hayes und Simon (1985) das künstliche Problem „Turm von Hanoi“ inhaltlich. Die semantische Einbettung hat deutliche Effekte auf die Schwierigkeit der Aufgabe. In einer anderen Untersuchung lässt J. Funke (1992) 80 Studierende das komplexe Problem „Altöl“ bearbeiten. Die Problemstellung ist zum einen in eine den Studierenden vertraute semantische Rahmengeschichte und zum anderen in einen unbekanntem Aufgabenhintergrund eingebaut. Die Problemstellung, die dem Vorwissen der Studierenden entspricht, wird durchgehend besser bearbeitet. Ähnliche Ergebnisse liefert auch Beckmann (1994), der Testpersonen ein formal identisches Problem als computergestützte Szenarios „Kirschbaum“ und „Maschine“ vorgibt. Es zeigen sich deutliche Unterschiede im Wissenserwerb, da nur beim Szenario „Maschine“ die Komplexität hilfreich reduziert werden kann.

Zu den formalen Aufgabenmerkmalen sind die eigentlichen, objektiven Eigenschaften des zu lösenden Problems zu rechnen. Allgemein wird dazu eine ganze Reihe von Einflussgrößen gezählt, die in Abbildung 8 aufgeführt sind (vgl. Dörner et al., 1983; J. Funke, 1990; Hussy, 1984). Die meisten Studien kommen aus dem Forschungsbereich „komplexes Problemlösen“.

Daher wird zumeist auch die dort verwendete Terminologie (s. Tabelle 8, Eigenschaften komplexer Probleme) verwendet. Neben den von allen Autoren genannten Faktoren Umfang, Vernetztheit, Transparenz und Dynamik führen Hussy (1984) und Dörner et al. (1983) darüber hinaus auch den Einflussfaktor Barrieretyp an. Dieser lässt sich unter die formalen Aufgabenmerkmale subsumieren. Wenngleich die Begrifflichkeiten und exakten Einordnungen bei den verschiedenen Autoren teilweise ein wenig voneinander abweichen, so sind diese Unterschiede für die hier interessierende Strukturierung möglicher Schwierigkeitsfaktoren nicht von Bedeutung.

Zahlreiche Studien haben den Einfluss der formalen Aufgabenmerkmale auf das Problemlösen untersucht. Die Analyse und Beschreibung der formalen Aspekte kann auch zur Formulierung einer optimalen Problemlösung führen (M. Kersting, 1999). Dies kann somit als ein wichtiger Vergleichsmaßstab zur Bewertung der Lösungsqualität dienen.

Unter Umfang werden allgemein die Anzahl der relevanten Variablen und die Zahl der Lösungsalternativen verstanden. Es zeigt sich in einigen Studien, dass der Umfang erwartungsgemäß einen direkten Einfluss auf die Schwierigkeit von Problemen hat. Je mehr Variablen beim Problemlösen zu berücksichtigen sind, desto höher werden die kognitiven Anforderungen an die Problemlösenden (vgl. J. Funke, 1990; Hussy, 1984). Aufgrund von Untersuchungen von Strauß (1993) ist dieser Zusammenhang noch differenzierter zu betrachten. So ist nicht nur die Zahl von Lösungsalternativen entscheidend, sondern auch der Anteil der richtigen Lösungen am Problemumfang. Strauß (1993) zeigt auf, dass eine unterschiedliche Zahl von Variablen nicht zwangsläufig zu einer unterschiedlichen Schwierigkeit führt, wenn der Anteil der richtigen Lösungen am Problemumfang konstant bleibt.

Neben dem Umfang ist auch die Vernetztheit der Variablen für die Schwierigkeit von Problemen bedeutsam. Reichert und Dörner (1988) finden heraus, dass es schwieriger ist, Variablen zu steuern, die zeitverzögert



reagieren. In einer Studie zur Problemschwierigkeit variiert Hussy (1985) die Art der Variablenvernetztheit. Die nichtlineare Vernetzung liefert dabei niedrigere Gruppenmittelwerte hinsichtlich der Problemlösegütemaße als die lineare. Die Aspekte Vernetztheit und Umfang sind nicht ganz unabhängig voneinander zu betrachten. Eine steigende Anzahl an Variablen bedeutet zumeist auch eine umfangreichere Vernetztheit.

Eine weitere Einflussgröße ist nach Dörner et al. (1983), Hussy (1984) und J. Funke (1990) der Aspekt der Eigendynamik. Das bedeutet, dass ein Problem, das sich ohne äußeres Einwirken von alleine (eigendynamisch) verändert, schwieriger zu lösen ist. In einer Studie können Gediga, Schöttke und Tücke (1983) zeigen, dass ein sonst identisches computergestütztes Szenario unter der Eigendynamikbedingung schwieriger zu bewältigen ist als ohne Dynamik.

Der Einfluss der Transparenz auf die Schwierigkeit von Problemen hat unterschiedliche Ergebnisse hervorgebracht. Während beispielsweise bei Hörmann und Thomas (1989) höhere Mittelwerte für das Systemwissen unter der Transparenzbedingung gefunden werden, zeigt sich bei Süß (1996), dass Transparenz keine Auswirkungen auf die Qualität der Systemsteuerung hat.

Nach Hussy (1984) und Dörner et al. (1983) hat auch der Barrieretyp Einfluss auf die Schwierigkeit von Problemen. Wie bereits weiter oben ausgeführt, unterscheidet Dörner (1979) zwischen vier verschiedenen Typen. Die Annahme lautet: Die Lösung eines Problems erweist sich als umso schwieriger, je geringer die Klarheit der Zielkriterien und der Bekanntheitsgrad der Mittel sind. Diese Aspekte haben eine große Verwandtschaft mit dem Merkmal der Transparenz. Allerdings wird - wie gerade dargestellt - bei der empirischen Überprüfung von uneinheitlichen Ergebnissen berichtet.

### *2.2.5.2 Personmerkmale als Einflussgrößen der Problemlöseleistung*

Zu den Personmerkmalen als Einflussfaktoren der Problemlöseleistung sind alle Eigenschaften, Fähigkeiten und Kenntnisse zu zählen, die eine Person in die Problemsituation mitbringt oder sie während der Situation erwirbt (J. Funke, 1990). Im Rahmen der Problemlöseforschung steht insbesondere der Einfluss kognitiver Merkmale im Mittelpunkt der Untersuchungen. Aber auch nicht-kognitive wie beispielsweise emotionale und motivationale Personmerkmale können möglicherweise die Schwierigkeit von Problemen beeinflussen. Nur vergleichsweise wenige Studien haben sich mit dem Zusammenhang von Problemlösen und nicht-kognitiven Personmerkmalen beschäftigt. Süß (1996) untersucht den Einfluss der Selbstwirksamkeitserwartung (Schwarzer & Jerusalem, 1995) und der Leistungsorientierung und Beanspruchung - mittels Freiburger Persönlichkeits-Inventar (Fahrenberg, Hampel & Selg, 1989) - auf das Szenario „Tomaten“. Die Korrelationen mit verschiedenen Leistungsmaßen zeigen zwar in die erwarteten Richtungen. Sie liegen allerdings nur im niedrigen bis mittleren Bereich. Vollmeyer und Rheinberg (1998) nehmen an, dass die Motivation während der Aufgabenbearbeitung davon abhängen kann, ob jemand Erfolg oder Misserfolg erlebt und dadurch seine Fähigkeit einschätzen kann. Nach ihrem kognitiv-motivationalen Prozessmodell ist die Motivation, die nach dem Lesen der Aufgabenbeschreibung und vor der Aufgabenbearbeitung wirkt, die relevante Motivation. Die Autoren können bei ersten Studien ihr Modell des Einflusses der Motivation auf die Problemlöseleistung für die lineare Aufgabe „Biology Lab“ bestätigen. Insgesamt wird jedoch in den meisten Untersuchungen von unbedeutenden oder nur sehr geringen Korrelationen zwischen Problemlösen und nicht-kognitiven Personmerkmalen berichtet (vgl. Schaub, 2001; Vogt, 1998).

Zu den kognitiven Merkmalen sind laut Süß, M. Kersting und Oberauer (1991) sowohl generelle intellektuelle Fähigkeiten als auch gegenstandsspezifisches Wissen zu zählen. Darunter ist Wissen zu verstehen, das für die Problemsituation von Bedeutung ist oder sein kann und somit die Bereichsspezifität von Problemlöseleistungen betrifft. Zur

genaueren Darstellung der teilweise widersprüchlichen Befundlage und aufgrund der besonderen Bedeutung für die vorliegende Arbeit werden die Aspekte Intelligenz und Problemlösen und Bereichsspezifität von Problemlöseleistungen in den nächsten Abschnitten gesondert vorgestellt.

#### 2.2.5.2.1 Intelligenz und Problemlösen

Hinsichtlich des Zusammenhangs zwischen Problemlösen und Intelligenz lassen sich zwei unterschiedliche Bereiche unterscheiden. Zum einen geht es um den Einfluss der Intelligenz auf die Problemlöseleistung bei künstlichen oder einfachen Problemen und zum anderen um den Einfluss auf die Problemlöseleistung in computergestützten Szenarios zum komplexen Problemlösen.

Die theoretische Nähe der Konzepte „Intelligenz“ und (künstliches) „Problemlösen“ wurde bereits im Abschnitt 2.2.2 in der Aussage von Sternberg (1982) deutlich. Aufgaben aus Intelligenztests und einfache oder künstliche Probleme sind demnach verwandt und stellen den Bearbeiter vor ähnliche kognitive Anforderungen.

Süß (1996) kann in seiner Untersuchung belegen, dass die Problemlöseleistung bei einem künstlichen Problem hoch mit der Intelligenztestleistung zusammenhängt. In zwei Studien ( $N = 30$  und  $N = 25$ ) wird der Zusammenhang der operativen Fähigkeit Verarbeitungskapazität des Berliner Intelligenzstruktur-Tests (Jäger et al., 1997) mit der Problemlöseleistung beim chinesischen Puzzle Tangram untersucht. Die Versuchspersonen absolvieren dabei mehrere Aufgabenstellungen. Die Leistungen in den verschiedenen Durchgängen aggregiert Süß (1996) zu Summenscores. Es können enge Zusammenhänge zwischen der Problemlöseleistung und der Verarbeitungskapazität ermittelt und repliziert werden, auf dem höchsten Aggregationsniveau in Höhe von  $r = .60$  und  $.84$ . Diese Werte liegen im Bereich der Korrelationshöhe zwischen verschiedenen Intelligenztests. Die Verlaufskurven der Korrelationen folgen dabei laut

Süß (1996) der Spearman-Brown-Formel für die Steigerung der Reliabilität durch Testverlängerung. Süß (1996) schließt aus den Daten, dass die Enge des Zusammenhangs demnach eine Funktion der Reliabilität der Problemlöseleistung ist und auf nur einer unabhängigen Messung beruhende Kriterien bei Einsichtsproblemen völlig untauglich seien, um reliable Daten zu erhalten.

Diese Studie zeigt beispielhaft, dass die Problemlöseleistung bei künstlichen oder einfachen Problemen stark mit der Intelligenztestleistung zusammenhängt, vorausgesetzt, dass die Problemlöseleistung reliabel und valide erfasst wird.

Hinsichtlich des Einflusses der Intelligenz auf die Problemlöseleistung in computergestützten, komplexen Szenarios sind seit den frühen Arbeiten zum komplexen Problemlösen eine ganze Reihe an teilweise kontradiktorischen Untersuchungen publiziert worden. Bereits 1981 untersuchte beispielsweise Putz-Osterloh (1981) den Einfluss von Intelligenztestleistungen auf das komplexe Problemlösen im Szenario „Schneiderwerkstatt“. Putz-Osterloh (1981) ermittelte Rangkorrelationen zu Leistungen im Intelligenztest APM (Advanced Progressive Matrices) von Raven (1962) in Höhe von  $r = -.01$  (N = 13) bzw.  $r = -.22$  (N = 36). Weitere Studien lieferten ebenfalls größtenteils Nullkorrelationen zwischen Intelligenz und komplexem Problemlösen (vgl. Kluwe, Misiak & Haider, 1991). Dörner (1986) und die Barmberger Gruppe werteten dies als Hinweis für die Unfruchtbarkeit des gesamten Zweigs der Intelligenzforschung. Diese Publikationen riefen ein großes Echo in der Forschungsgemeinschaft hervor. Es folgten zahlreiche Diskussionen, theoretische Auseinandersetzungen und weitere Studien zum Zusammenhang zwischen Leistungen in komplexen Computersimulationen und Intelligenztests (u. a. J. Funke, 1983; Hussy, 1989, 1991a; Jäger, 1986; J. Funke, Buchner, Dörner, Süß & Vollmeyer, 1999). Bei näherer Betrachtung der Studien werden die anfänglich recht kontroversen Ergebnisse klarer.

Es zeigen sich dabei insbesondere drei Aspekte, die von Bedeutung sind:

- methodische Probleme
- Einfluss der Aufgabenmerkmale
- verwendetes Intelligenzmaß

### *Methodische Probleme*

Die unterschiedlichen Zielsetzungen der Untersuchungen werden in den gewählten Designs deutlich. Während es Dörner und seinen Mitarbeitenden in erster Linie um die Aufdeckung von Problemlöseprozessen - also um allgemeine psychische Prozesse - ging, haben andere Autoren den differenziell-diagnostischen Nutzen der computergestützten komplexen Problemlöseszenarios überprüfen wollen (vgl. J. Funke, 1983; Hasselmann, 1993; Jäger, 1984). Aus Sicht der Diagnostik werden neben der im nächsten Kapitel thematisierten mangelnden Testgüte einiger Szenarios (insbesondere bei der Verwendung von Problemlösegutemaßen) vor allem auch methodische Probleme als ausschlaggebend für nicht gefundene Zusammenhänge zwischen Intelligenz und Problemlöseleistung in computergestützten Szenarios genannt. So führt beispielsweise Süß (1999) aus, dass die bisherigen Ergebnisse der Forschung einen Zusammenhang zwischen Verarbeitungskapazität und komplexer Problemlöseleistung von  $r = .30$  realistisch erscheinen lassen. Um einen Effekt dieser Größenordnung mit einer statistischen Power von  $.80$  bei einer Irrtumswahrscheinlichkeit von  $p < .05$  (einseitige Testung) aufdecken zu können, ist eine Stichprobengröße von  $N = 64$  erforderlich. Tatsächlich haben jedoch die meisten Studien weniger als 64 Personen untersucht, so dass sich Interpretationen über nicht gefundene signifikante Zusammenhänge erübrigen, weil bei diesen Stichprobengrößen überzufällige Korrelationen in mittlerer Höhe gar nicht zu erwarten sind. Untersuchungen, die jedoch auf eine genügend große Stichprobe zurückgreifen können (Hussy, 1989; M. Kersting, 1999; Süß, 1996), haben Zusammenhänge mit zumindest mittlerer Effektstärke ausgemacht. Weiterführende Literatur zur Methodenkritik findet sich bei Süß (1996, 1999), Tent (1984), Hörmann und Thomas (1989) und Jäger (1986, 1991).

### *Einfluss der Aufgabenmerkmale*

Die verschiedenen Studien unterscheiden sich zum Teil auch erheblich aufgrund der Unterschiedlichkeit der eingesetzten computergestützten Szenarios. So kann beispielsweise die Zahl der Variablen von ca. zehn bis zu zweitausend („Lohhausen“) variieren. Kluwe, Misiak und Haider (1991) halten daher Metaanalysen für nicht durchführbar. Strohschneider (1991a) hat das abstrakte Szenario „Vektor“ mit dem semantisch reichen Entwicklungshelferszenario „Moro“ verglichen. Die Leistungen der Probanden in den beiden Systemen hängen nicht signifikant zusammen, was die Uneinheitlichkeit der Operationalisierung von komplexen Problemlöseleistungen verdeutlicht. Die Ergebnisse einiger Untersuchungen weisen darauf hin, dass der Zusammenhang zwischen Intelligenz und Problemlöseleistung umso niedriger wird, je komplexer (vernetzter, intransparenter, dynamischer, umfangreicher) ein Szenario ist. Für kleinere, weniger komplexe computergestützte Szenarios, die zumeist auf nachvollziehbaren Verknüpfungen der Variablen beruhen, konnten höhere Korrelationen mit Intelligenztestleistungen gefunden werden als für mittelgroße und große Computersimulationen. So zeigen sich beispielsweise für die kleineren Szenarios „ZAP“ (Hussy, 1989) und „Vektor“ (Strohschneider, 1991a) substantielle Zusammenhänge zu Intelligenztestleistungen in Höhe von  $r = .31$  bis  $.58$ .

Die Annahme, dass die Anforderungen an die Testpersonen in kleineren computergestützten Szenarios große Schnittbereiche haben zu denen in Intelligenztests, kann Kröner (2001) in seiner Arbeit belegen. Kröner (2001) entwickelt in mehreren Schritten die Computersimulation „Multiflux“ mit dem Ziel der Konstruktion eines simulationsbasierten Intelligenzdiagnostikums. In einer Studie an 101 Schüler/innen zeigen sich Korrelationen zwischen einem kombinierten Wert aus „Multiflux“-Systemwissen und -Steuerungsleistung und der Verarbeitungskapazität des Berliner Intelligenzstruktur-Tests (Jäger et al., 1997) in Höhe von  $r = .65$  (minderungskorrigiert  $r = .71$ ). Dies entspricht der Stärke des Zusammenhangs zwischen unterschiedlichen

Intelligenztests. Kröner (2001) folgert, dass mit Hilfe von „Multiflux“ Intelligenz gemessen wird.

Strohschneider (1991a) variiert in seinen Studien die Transparenz der Zielvorgaben. Während zwanzig Studierende das System „Moro“ mit klarer Sollwertvorgabe steuern, werden fünfundzwanzig anderen keine Vorgaben gemacht. Bei der Gruppe mit Zielspezifikation zeigen sich deutlich höhere Korrelationen ( $r = .50$  bis  $.77$ ) zur Intelligenztestleistung im BIS-Test (Jäger et al., 1997) gegenüber der anderen Gruppe ( $r = -.05$  bis  $.14$ ). Die Befunde zum Einfluss der Transparenzbedingung auf den Zusammenhang von Intelligenz- und komplexer Problemlöseleistung sind widersprüchlich. Putz-Osterloh und Lüer (1981) sowie Hörmann und Thomas (1989) berichten von höheren Korrelationen unter transparenten Bedingungen als unter intransparenten. Transparenz wird dabei durch die Vorgabe einer Vernetzungsgrafik, die alle Variablen und Variablenrelationen veranschaulicht, operationalisiert. Süß (1996) ermittelt hingegen für die Gruppe unter der Transparenzbedingung ähnliche Zusammenhänge wie für die Kontrollgruppe.

#### *Verwendetes Intelligenzmaß*

Kritik an den Untersuchungen zum Zusammenhang zwischen Intelligenz und Problemlösen bezieht sich auch auf die verwendeten Intelligenzmaße. So bemängelt Hussy (1989) an den meisten Studien die Praxis, den Gesamtscore von Intelligenztests in die Zusammenhangsanalyse eingehen zu lassen und somit nicht differenzierteren Intelligenzmodellen mit Faktorenstruktur gerecht zu werden. Gerade das Berliner Intelligenzstrukturmodell (Jäger, 1982, s. Abbildung 7) mit seiner bimodalen, hierarchischen Struktur eigne sich nach Ansicht einiger Autoren in besonderer Weise zum Vergleich mit Problemlösemodellen (Hussy, 1989; M. Kersting, 1999). Beim Einsatz verschiedener Versionen des Berliner Intelligenzstruktur-Tests (Jäger et al., 1997) können zumeist substantielle Zusammenhänge zwischen den Intelligenztestscores und verschiedenen computergestützten Problemlöseszenarios erfasst werden. Wenngleich in den zwei Untersuchungen von Reichert und Dörner (1988) und

J. Funke (1985) keine signifikanten Korrelationen ermittelt werden, so weisen eine ganze Reihe von weiteren kleineren Studien insbesondere für die operativen Fähigkeiten Verarbeitungskapazität und Bearbeitungsgeschwindigkeit mittlere bis sehr hohe Zusammenhänge in Höhe von  $r = .27$  bis  $.81$  aus. Dabei zeigt sich bei Hussy (1989, 1991a, 1991b), Schoppek (1996) und Strohschneider (1991a) die Verarbeitungskapazität als bester Einzelprädiktor für die Problemlöseleistung in den Szenarios „Jogi“, „ZAP“, „Moro“ und „Vektor“. Hörmann und Thomas (1989) und Putz-Osterloh (1985) finden die höchsten Korrelationen hingegen für den Zusammenhang zwischen der Problemlöseleistung und der Bearbeitungsgeschwindigkeit (Szenarios „Moro“ und „Schneiderwerkstatt“).

Süß (1996, 1999) kann schließlich in mehreren umfangreichen Studien zeigen, dass die Problemlöseleistung in reliablen und validen computergestützten Szenarios mittelstark mit der Verarbeitungskapazität zusammenhängt. Die Berliner Untersuchungen zum Zusammenhang von Intelligenz, Wissen und komplexen Problemlösen (Süß, 1993, 1996; Süß et al., 1991) können die Bedeutung der Verarbeitungskapazität für die Problemlöseleistung in der „Schneiderwerkstatt“ belegen. An 137 Testpersonen werden in der Erstuntersuchung Korrelationen in Höhe von  $r = .34$  zwischen der Verarbeitungskapazität des BIS-Tests und der Steuerungsleistung ermittelt. Süß (1996) aggregiert danach die Daten aus der Erst- und Wiederholungsuntersuchung über alle fünf Bearbeitungsdurchgänge, um ein reliableres Problemlösegutemaß zu erhalten. Sowohl die Verarbeitungskapazität als auch die zahlengebundenen kognitiven Fähigkeiten erweisen sich demnach als gute Prädiktoren, wobei die Korrelationen zur Verarbeitungskapazität insgesamt am höchsten ausfallen ( $r = .47$  bzw.  $r = .49$ ; minderungskorrigiert:  $r = .55$ ). Der Berliner Untersuchungsansatz wurde in den Mannheimer Studien von Wittmann, Süß und Oberauer (1996), Wittmann und Süß (1999) und Süß (1999) weiterverfolgt. Dabei setzen die Wissenschaftler zur Messung der Problemlöseleistung drei verschiedene computergestützte Szenarios („Schneiderwerkstatt“, „PowerPlant“ und „Learn“) und zur Intelligenzmessung



wiederum den Berliner Intelligenzstruktur-Test von Jäger et al. (1997) ein. Anhand der Daten von 136 Studierenden kann dabei bestätigt werden, dass die operative Fähigkeit Verarbeitungskapazität entscheidend für die Zusammenhänge ist. Die Einzelkorrelationen liegen zwischen  $r = .35$  und  $.46$ . Beim aggregierten Problemlösegutemaß aller drei Szenarios lässt sich der Zusammenhang auf  $r = .52$  steigern.

Die empirischen Belege zum Zusammenhang zwischen Verarbeitungskapazität und Problemlösen werden auch theoretisch begründet. Nach Auffassung vieler Autoren sind Aufgaben zum komplexen Problemlösen als eine Erweiterung der herkömmlichen Intelligenzdiagnostik zu verstehen (M. Kersting, 1999; Kröner, 2001; Putz-Osterloh, 1985). Demnach müsste die Stärke des Zusammenhangs in etwa vergleichbar mit der Höhe der Korrelationen zwischen Aufgaben zur Messung verschiedener Dimensionen der Intelligenz oder zwischen unterschiedlichen Intelligenztests sein. Hinsichtlich der Zusammenhänge zwischen verschiedenen Subskalen des BIS-4-Tests berichten Jäger et al. (1997) von Koeffizienten im Streubereich zwischen  $r = .25$  bis  $r = .47$ . Diese sind vergleichbar mit den berichteten empirischen Ergebnissen zum Zusammenhang zwischen der Verarbeitungskapazität und der Problemlöseleistung in computergestützten Szenarios. Dabei ist die Stärke des Zusammenhangs abhängig von der in jedem einzelnen Untersuchungsfall realisierten Ähnlichkeit der Anforderungen und Aufgaben (M. Kersting, 1999, Süß, 1996). Süß (1996) argumentiert entsprechend mit der Gleichartigkeit der Anforderungen von computergestützten Szenarios und Aufgaben zum schlussfolgernden Denken bzw. zur Verarbeitungskapazität in Intelligenztests. Er führt aus, dass die Lösung eines Problems, unabhängig davon, ob es sich um ein künstliches oder ein realitätsnahes, komplexes Problem handelt, das Arbeitsgedächtnis beansprucht, „weil simultanes Speichern und Prozessieren von Informationen erforderlich ist und Routineabläufe beim Denken und Handeln gesteuert und überwacht werden müssen“ (S. 49). Die Ansicht, dass Problemlösen das Arbeitsgedächtnis fordert, teilen praktisch alle Autoren (zum Beispiel J. R. Anderson, 1983; Hussy, 1984; Klauer, 1995; Newell &

Simon, 1972). Die operative Fähigkeit Verarbeitungskapazität (K) gilt gemäß dem Berliner Intelligenzstrukturmodell (Jäger, 1982, s. Abschnitt 2.2.2) als eine hochgradig generelle, leistungsbegrenzende, kognitive Ressource und kann nach Süß (1996) als Kapazität des Arbeitsgedächtnisses interpretiert werden. Süß (1996) kommt daher zu dem Schluss, dass bei Problemen, die das Arbeitsgedächtnis belasten, die Verarbeitungskapazität ein substantieller Prädiktor sein sollte.

### *Strukturanalytisches Ergebnis*

Neben diesen Studien aus der Kognitionspsychologie und Eignungsdiagnostik hat eine Studie aus einem anderen Feld Hinweise auf die Struktur des Konstrukts „Problemlösefähigkeit“ geliefert: die PISA-Vorbereitungsstudie. Im Rahmen des von der OECD getragenen „Program for International Student Assessment“ (PISA) werden alle drei Jahre weltweit die Lese-, Mathematik- und Naturwissenschaftskompetenz von fünfzehnjährigen Schülerinnen und Schülern untersucht (Baumert et al., 2001). Neben diesen auf bestimmte Lernbereiche zugeschnittenen Leistungsmessungen werden auch „Cross-curricular Competencies (CCC)“ (OECD, 1997) ermittelt. Darunter werden fächerübergreifende Kompetenzen verstanden, die junge Menschen in die Lage versetzen, „eine positive Rolle als Bürger in der Gesellschaft zu spielen“ (Klieme et al., 2001, S. 181). Problemlösefähigkeit wird als eine fächerübergreifende Kompetenz verstanden und wurde in einem vorbereitenden PISA-Feldtest untersucht. Im Sommer 1999 konnten dabei insgesamt 650 Schülerinnen und Schüler aus 18 Schulen in vier Bundesländern getestet werden (Klieme et al., 2001). Um möglichst die Breite des Konstrukts abzudecken, kam eine ganze Reihe an Problemlöseinstrumenten (drei computergestützte und drei schriftliche) zum Einsatz. Die Validitätsprüfung liefert einige interessante Ergebnisse. So finden Klieme et al. (2001) keine Hinweise dafür, dass Problemlösefähigkeit ein homogenes Konstrukt ist. Vielmehr folgern die Autoren aus den Daten: „Es gelingt also nicht, eine allgemeine Problemlösekompetenz zu identifizieren und von anderen Konstrukten zu trennen“ (Klieme et al., 2001, S. 195). Zur weiteren Aufklärung sollten mit Hilfe von

Strukturgleichungsmodellen Teilfaktoren des Konstrukts erfasst werden, wobei in die Analyse auch die im Rahmen der PISA-Studie erhobenen Konstrukte „Lesekompetenz“, „mathematische Kompetenz“ und „naturwissenschaftliche Kompetenz“ mit aufgenommen wurden. Die beste Passung an die gewonnenen Daten zeigt ein Modell mit sieben Kompetenzfaktoren.<sup>7</sup> Im Einzelnen sind das:

- Reasoning bzw. kognitive Grundfähigkeit (Reason): Dabei werden Aufgaben zum figuralen und verbalen schlussfolgernden Denken gestellt.
- Lesekompetenz (Reading)
- Mathematische Kompetenz (Math)
- Naturwissenschaftliche Kompetenz (Science)
- drei getrennte Problemlösefaktoren:
  - Problemlösekompetenz bei schriftlichen Aufgaben (PS-Paper)
  - Wissenserwerb in computergestützten Systemen (PS-Know)
  - Strategien beim Explorieren und Steuern eines komplexen Systems (PS-Strat)

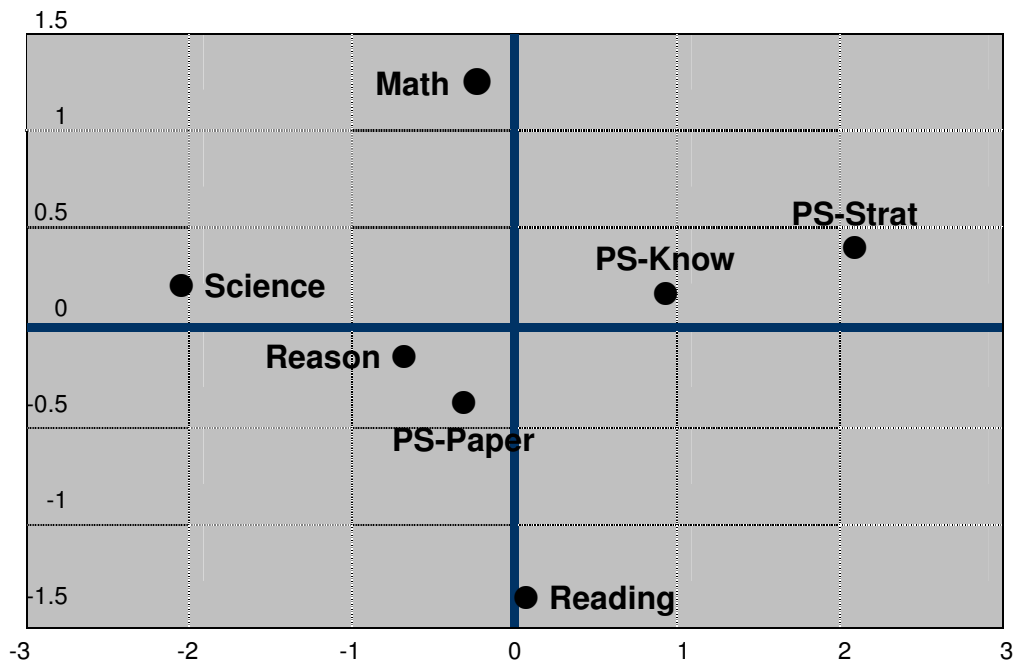
Die Abbildung 9 stellt das Ergebnis einer nichtmetrischen mehrdimensionalen Skalierung dar. Je näher zwei Variablen grafisch angeordnet sind, desto höher ist der im Strukturgleichungsmodell geschätzte Zusammenhang. Besonderheit dieses Ergebnisses ist die Unterscheidung von drei Problemlösefaktoren sowie die zentrale Positionierung des Faktors „Reasoning“. Die Variable „Problemlösekompetenz, erfasst in schriftlichen Verfahren“ (PS-Paper) liegt sehr nahe am Faktor der kognitiven Grundfähigkeit (Reason), während die beiden Faktoren, die auf Computerverfahren basieren (PS-Know und PS-Strat) weiter davon entfernt platziert sind.

---

<sup>7</sup> In Klammern stehen die jeweiligen PISA-Abkürzungen.

Wenngleich zu kritisieren ist, dass die Bezeichnung für die Problemlösefaktoren nicht einheitlich erfolgt<sup>8</sup>, so zeigt diese Studie doch zweierlei: Problemlösekompetenz ist erstens nicht als ein eindimensionales Konstrukt zu verstehen, sondern besteht aus verschiedenen Facetten. Zweitens wird in den schriftlichen Problemlöseverfahren, obwohl sie laut Klieme et al. (2001) durchaus komplex und realitätsnah ausgestaltet sind, zu einem erheblichen Teil Problemlösekompetenz im Sinne von schlussfolgerndem Denken (Reasoning) erfasst.

Diese Ergebnisse könnten möglicherweise auch die teilweise widersprüchlichen Befunde aus der Forschung zum Zusammenhang zwischen Problemlöseleistungen in komplexen Computersimulationen und Intelligenztestleistungen erklären. Das würde Folgendes bedeuten: Der Zusammenhang mit Intelligenztestleistungen bzw. der Verarbeitungskapazität ist vermutlich umso stärker, je näher die (komplexe) Problemlöseleistung hinsichtlich Anforderungen an das schlussfolgernde Denken operationalisiert wird.



**Abbildung 9: Interkorrelationen der Kompetenzfaktoren im normierten zweidimensionalen Raum (nach Klieme et al., 2001, S. 196)**

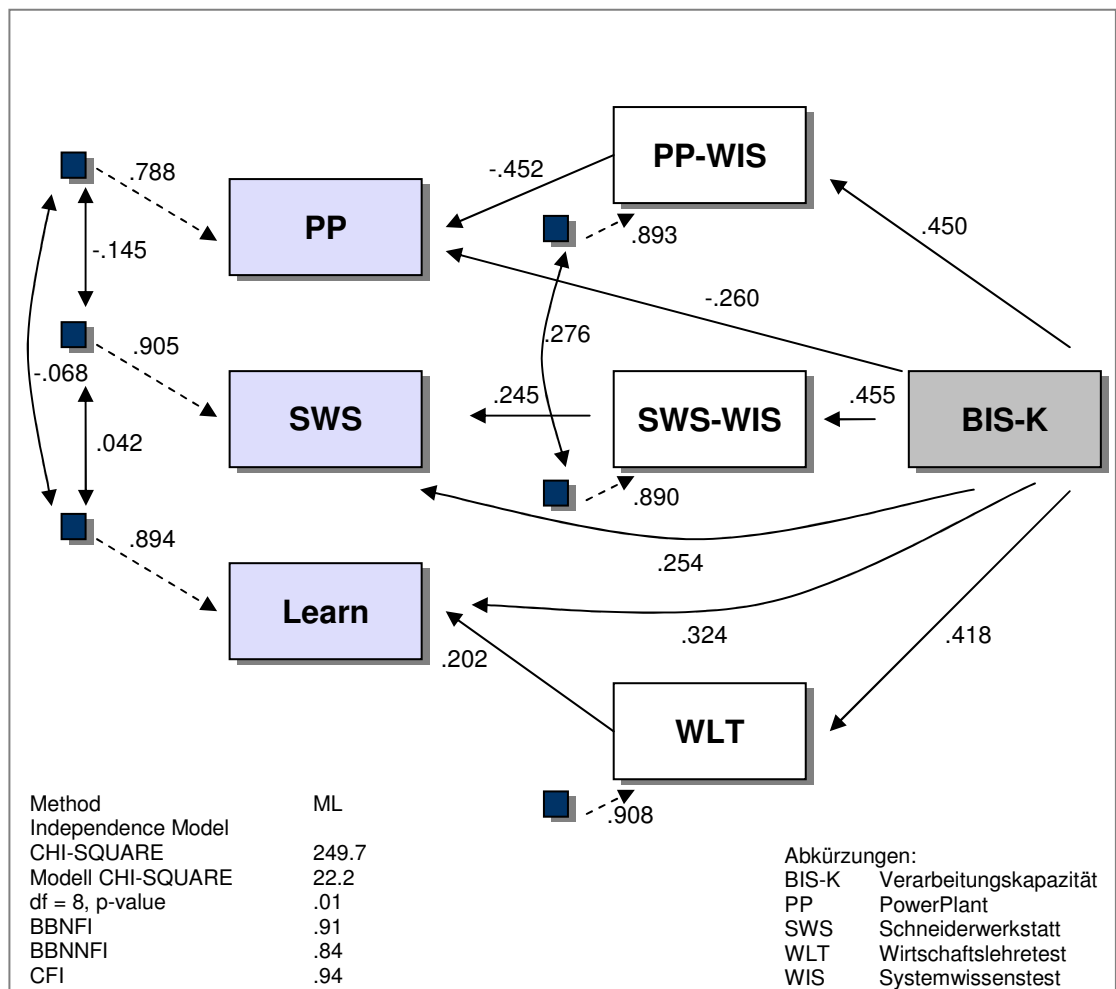
<sup>8</sup> Es wird nicht klar, inwieweit die Problemlösefaktoren hinsichtlich der Form der Darbietung (schriftlich vs. computergestützt) oder hinsichtlich der Art der Beschreibung (Kompetenzen, Strategien und Erwerb von Wissen) unterschieden werden.

### 2.2.5.2.2 Bereichsspezifität von Problemlöseleistungen

Der Einfluss des Vorwissens auf Problemlöseleistungen wurde im Rahmen der komplexen Computersimulationen nur selten direkt untersucht. Die Ergebnisse der durchgeführten Untersuchungen deuten jedoch darauf hin, dass die Problemlöseleistung sowohl von der Intelligenz der Person als auch vom Wissen in der jeweiligen Domäne, dem Vorwissen, abhängig ist. Die bereits zitierte Studie von J. Funke (1992) zur semantischen Einbettung liefert hierzu Hinweise. J. Funke (1992) findet heraus, dass die Version des „Altöl“-Szenarios, die dem Vorwissen der Testpersonen entspricht, erfolgreicher bearbeitet wird als eine den Testpersonen fremdere Version. J. Funke (1992) erfasst dabei aber nicht explizit das Vorwissen für die zwei verschiedenen Versionen. Es kann somit nur implizit geschlossen werden, dass das Vorwissen die Problemlöseleistung beeinflusst hat.

Auch Ergebnisse von Experten-Novizen-Studien sind in diesem Zusammenhang zu nennen. Diese Studien werden im Abschnitt über die Kriteriumsvaliditäten bisheriger Verfahren detaillierter dargestellt. Wenngleich die Befundlage nicht eindeutig ist, so konnte doch in einigen Studien festgestellt werden, dass Experten die besseren Leistungen gegenüber Novizen liefern. Ein Problem dieser Studien ist jedoch, dass das Vorwissen zumeist nicht explizit erfasst, sondern nur über die Berufszugehörigkeit implizit geschlossen wird. Süß (1996, 1999) hingegen erfasst das Vorwissen in seinen Studien mit Hilfe eines allgemeinen wirtschaftswissenschaftlichen Tests. Es ergeben sich jeweils substantielle Zusammenhänge zwischen den Problemlöseleistungen in den drei eingesetzten Szenarios „PowerPlant“, „Schneiderwerkstatt“ und „Learn“ und dem Wirtschaftswissen. Die Höhe der Korrelationen liegt zwischen  $r = .28$  und  $r = .40$ . Dabei kann das bereichsspezifische Wissen in den durchgeführten hierarchischen Regressionsanalysen inkrementelle Varianz neben der Verarbeitungskapazität (BIS-K) erklären. In einem um die Computererfahrung erweiterten Prognosemodell können schrumpfungskorrigiert insgesamt 51.5 Prozent der Kriteriumsvarianz des über drei computergestützte Szenarios aggregierten Problemlösegußmaßes

aufgeklärt werden (s. Abbildung 10). Süß (1996) sieht seine Hypothese durch die Befunde gestützt, dass das allgemeine wirtschaftswissenschaftliche Vorwissen ein guter Prädiktor für den Problemlöseerfolg ist. Neben dem Aspekt des Vorwissens hat Süß (1999) auch das systemspezifische Wissen mit Hilfe eines Systemwissenstests untersucht. Es zeigt sich, dass auch das Systemwissen die Problemlöseleistung beeinflusst (s. Abb. 10).



**Abbildung 10: Vorhersage der Problemlöseleistungen durch Verarbeitungskapazität (BIS-K), Systemwissen und Vorwissen (nach Süß, 1999, S. 225)**

In der PISA-Vorbereitungsstudie zum Problemlösen von Klieme et al. (2001) finden sich zumeist auch bedeutsame Korrelationen zwischen dem Vorwissen und den Leistungen in verschiedenen Problemlöseverfahren. Allerdings wird das Vorwissen nicht über einen Test erhoben, sondern die Schüler bewerten ihr Vorwissen in einer Domäne selbständig. Das selbst

eingeschätzte Vorwissen korreliert dabei mit den Leistungen in sechs von acht Verfahren signifikant. Die Höhe der sechs Werte liegt im Bereich von  $r = .16$  bis  $r = .31$ . Klieme et al. (2001) kommen zu der Einschätzung, dass die Bestimmung einer allgemeinen, bereichsunabhängigen Kompetenz des Problemlösens in computergestützten Szenarios nicht möglich ist. Die oben geschilderten Vergleiche der Leistungen in verschiedenen computergestützten Szenarios unterstützen diese Einschätzung. Demnach ist für die Diagnose von Problemlöseleistung die Bereichsspezifität von Bedeutung. Klieme et al. (2001) folgern: „Testaufgaben müssen daher sorgfältig auf die Zielpopulation ausgerichtet sein“ (S. 189).

### **2.2.6 Definition Planen**

Die Fähigkeit des Planens wird von Sydow (1990) als grundlegende Kulturtechnik bezeichnet. Planen ist nach Sydow (1990) von entscheidender Bedeutung, da es die Effektivität kognitiver Leistungen steigert. Den Begriff „Planen“ haben J. Funke und Fritz (1995, S. 29) aus diagnostischer Sicht folgendermaßen beschrieben:

Planen bedeutet: gedanklicher Entwurf einer zielgerichteten Aktionsfolge, der auf unterschiedlichen Auflösungs niveaus erfolgen kann, unter Beachtung von einschränkenden Randbedingungen räumlicher, zeitlicher, materieller und logischer Art, und bei einem jeweils gegebenen aktuellen Kenntnis- und Fertigkeitenstand. Planen bedeutet auch die Überwachung des erstellten Plans bei dessen Ausführung mit den Optionen der Revision oder des Abbruchs.

Dieser Planungsbegriff baut auf neueren Ansätzen zur Handlungsregulation auf. Nach J. Funke und Fritz (1995) sind fünf Elemente der Definition von zentraler Bedeutung.

(1) Das Basiselement des Planens ist die Aktion, die strukturiert und geordnet, also zielgerichtet erfolgt. Die Aktionsfolge wird mit dem Begriff des Denkens als „gedanklichem Entwurf“ verknüpft, um die kognitive Leistung des Planenden zu unterstreichen.

(2) Mit dem Begriff des Auflösungs niveaus soll laut J. Funke und Fritz (1995) der Ansatz der TOTE-Einheiten von Miller et al. (1960) aufgegriffen werden (s. Abbildung 5). Pläne können demnach auf unterschiedlichen Ebenen entwickelt werden.

(3) Die einschränkenden Randbedingungen machen das Planen erst zu einem Problem. Planen ohne behindernde Randbedingungen ist demnach eine Koordinationsaufgabe.

(4) Planende können mit unterschiedlichem Hintergrundwissen in eine Planungssituation kommen bzw. Kenntnisse während der Bearbeitung unterschiedlich erwerben. Dies hat zur Folge, dass die gleiche Planungssituation verschieden schwierig wahrgenommen werden kann.

(5) Planen bedeutet nicht nur die Planerstellung, sondern auch die Überwachung der Planausführung, die erst mit der Zielerreichung oder dem Abbruch beendet ist (vgl. von der Weth & Strohschneider, 2002).

J. Funke und Glodowski (1990) beschreiben in ihrem logisch begründeten Modell Basiskompetenzen des Planens. Je nach Planungsaufgabe werden verschiedene Kompetenzen in unterschiedlichem Maße beansprucht. Demnach ist die gesamte Planungsleistung in Teilleistungen zu unterteilen, die wiederum zwei grundlegenden Phasen zugeordnet sind - der Planerstellung und der Planausführung. Unter „Planerstellung“ werden nach J. Funke und Glodowski (1990) fünf Teilleistungen verstanden:

1. Abfolgen erkennen

Planende müssen die angemessene zeitliche Abfolge von Teilschritten erkennen.

2. Randbedingungen erkennen

Planende müssen einschränkende räumliche, zeitliche, materielle und logische Faktoren berücksichtigen.

3. Zwischenzielbildung

Planende müssen einen Gesamtplan in Teilziele aufgliedern können.

4. Verfügbarkeit von Alternativen

Planende müssen im Störfall Alternativen abrufen können.

5. Angemessenheit der Auflösung

Planende müssen Teilziele in handhabbarer Weise erstellen können.



Unter „Planausführung“ werden demnach vier Teilleistungen unterschieden:

1. Planüberwachung

Planende müssen kontinuierlich die Übereinstimmung von Planung und Ausführung kontrollieren.

2. Fehlerdiagnostik

Planende müssen im Bedarfsfall Faktoren identifizieren, die die Planausführung beeinträchtigen.

3. Planrevision

Planende müssen Planungskomponenten ersetzen bzw. die Planungsebene wechseln können.

4. Planverwerfung

Planende müssen die Unmöglichkeit der Plansausführung erkennen können.

Planen ist nach J. Funke und Glodowski (1990) also nicht nur handlungsvorbereitend, sondern auch -begleitend.

Zur Bestimmung der Planungsleistung von Personen ist es von besonderer Bedeutung, was unter der Qualität von Planen verstanden wird. Nach J. Funke und Krüger (1995) ist die Qualität des Planens in erster Linie durch die Güte der Aufgabenbearbeitung bestimmt. Dabei sind die Güte der Teilleistungen während der Planerstellung und der Planausführung entscheidend. Das Planungskonzept von J. Funke und Fritz (1995) baut auf einem kognitiven Planungsbegriff auf, wie es in fast allen diagnostischen Untersuchungen der Fall ist (vgl. für den personaldiagnostischen Bereich: J. Funke und Krüger, 1995; Pascha, Schöppe & Hacker, 2001; Ueckert, 1995; für den entwicklungspsychologischen Bereich: Fritz & Hussy, 1995; für den allgemeinspsychologischen und neuropsychologischen Bereich: Dreher & Oerter, 1985).

In dieser Arbeit geht es in erster Linie um kognitiv-analytisches Planen, das eine gewisse Nähe zum Problemlösen vorweist, sich aber davon unterscheiden lässt. Der hier verfolgte Planungsbegriff grenzt sich laut J. Funke und Fritz (1995) vom Problemlösen in erster Linie durch drei Aspekte

ab. Erstens, Planen besteht aus der kognitiven Sequenz von Schritten, während Problemlösen noch ganz andere Formen annehmen kann. Zweitens, Problemlösen kann im Gegensatz zum Planen, das sich spezifisch auf eine zukünftige Handlung bezieht, Themen der Vergangenheit aufgreifen. Drittens, Planen zielt auf die *Art* der Leistung - also auf die Frage nach dem *wie* ab. Problemlösen richtet sich dagegen auf die Ursachen oder Ergebnisse - also auf die Frage nach dem *warum* (J. Funke & Fritz, 1995; Kreitler & Kreitler, 1987). Laut J. Funke und Glodowski (1990) kann Planen Problemlösen implizieren, muss es aber nicht. Daher erscheint eine Gleichsetzung der Konstrukte nicht sinnvoll (vgl. J. Funke & Fritz, 1995).

Neben der eignungs- und leistungsdiagnostischen Erfassung von Planen sind auch eine Reihe von Fragebogenverfahren zur Beschreibung von Planungskompetenz als Persönlichkeitsausprägung publiziert worden, die allerdings für diese Arbeit nicht weiter von Bedeutung sind. Eine Übersicht hierzu findet sich bei Wiesner (1995).

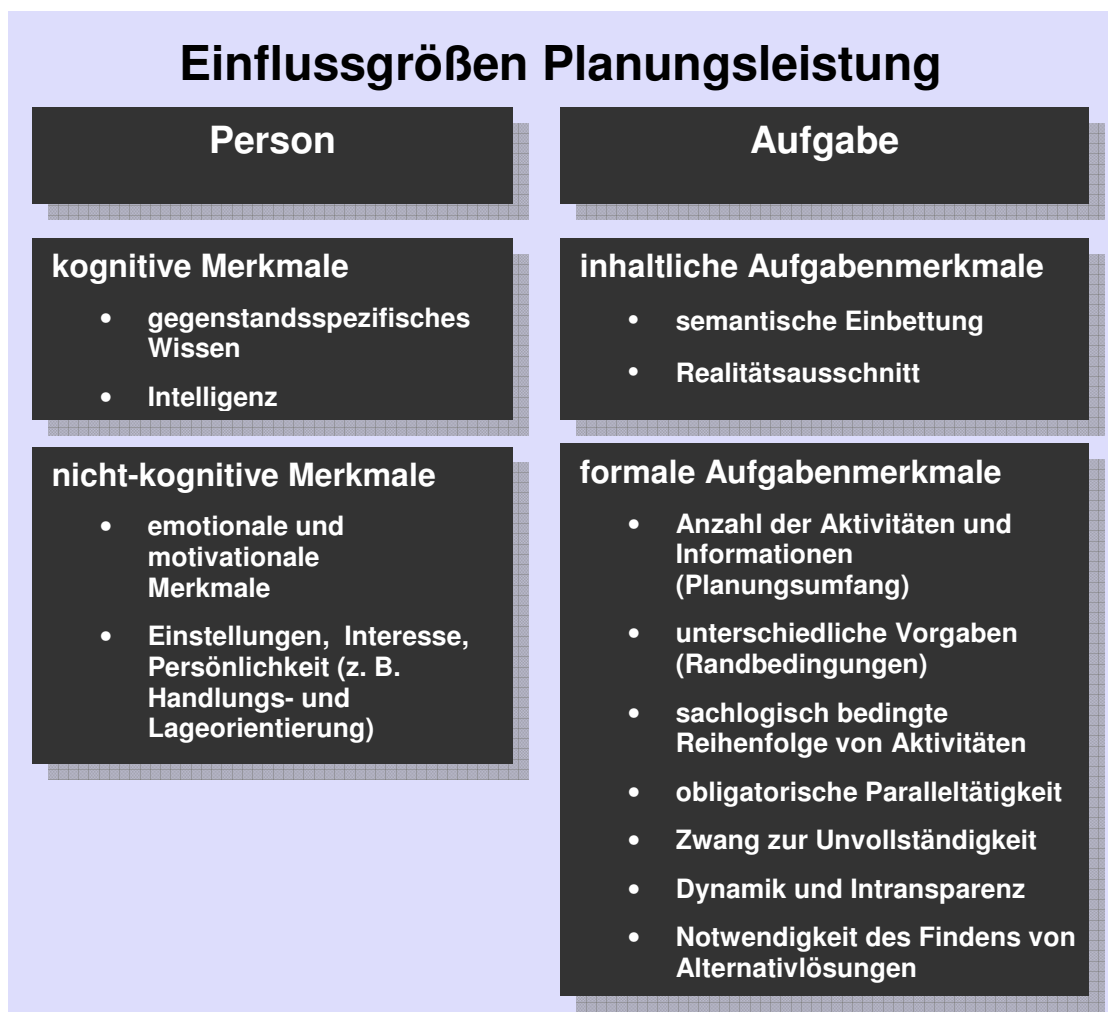
### **2.2.7 Einflussgrößen der Planungsleistung**

Analog zum Vorgehen beim Abschnitt Problemlösen (s. o.) gilt es bei der Verfahrensentwicklung zur Diagnose der Planungsleistung insbesondere die schwierigkeitsvariierenden Faktoren zu berücksichtigen. Abbildung 11 veranschaulicht die beiden Einflussgrößen für Planungsleistungen. Die Einteilung basiert auf der oben vorgestellten Klassifikation zum Problemlösen. Diese Klassifikation ist bereits in der Forschung zum Planungskonzept eingeführt worden und hat sich als hilfreich für diagnostische Fragestellungen herausgestellt (vgl. Fritz & J. Funke, 1995).<sup>9</sup> Dabei lassen sich kognitive von nicht-kognitiven Personmerkmalen unterscheiden. Den Aufgabenmerkmalen sind sowohl inhaltliche als auch

---

<sup>9</sup> J. Funke und Fritz (1995) nennen neben den Faktoren Person und Aufgaben zusätzlich den Aspekt der Zeit und den Bereich der Situation. Der Bereich der Situation lässt sich den Aufgabenmerkmalen zuordnen. Der Aspekt der Zeit ist für die Aufgabenentwicklung im Rahmen dieser Arbeit nicht entscheidend.

formale Aspekte zuzuordnen. Die Schwierigkeit der Aufgabe ergibt sich laut Hussy (1984) und J. Funke und Fritz (1995) aus der Wechselwirkung zwischen Person- und Aufgabenmerkmalen, wobei insbesondere die formalen Aufgabenmerkmale für die Verfahrensentwicklung entscheidend sind. Diese formalen Merkmale können bei der Aufgabenkonstruktion so berücksichtigt werden, dass die Schwierigkeit der Aufgaben variiert.



**Abbildung 11: Klassifikation der Einflussgrößen der Planungsleistung**

### **2.2.8 Empirischer Forschungsstand zum Planen**

Die Forschung hat bisher erst recht wenige Ergebnisse zu den Einflussgrößen der Planungsleistung (s. Abbildung 11) hervorgebracht, die

für die Entwicklung einer simulationsbasierten Planungsaufgabe von Bedeutung sein können. Im Folgenden sollen die entscheidenden Befunde vorgestellt werden.

#### *2.2.8.1 Aufgabenmerkmale als Einflussgrößen der Planungsleistung*

Zu den Aufgabenmerkmalen gehören inhaltliche und formale Aspekte. Sie bestimmen die eigentliche Planungsaufgabe. Wie bereits im Abschnitt über das Problemlösen vorgestellt, geht es bei inhaltlichen Aufgabenmerkmalen primär um den gewählten Realitätsausschnitt bzw. um die semantische Einbettung, die das Vorwissen des Planenden ansprechen. Pascha et al. (2001) haben in zwei Studien an 240 Psychologie-Studierenden untersucht, wie verschiedene Faktoren die Lösungsgüte, den Zeitbedarf und die subjektiv wahrgenommene Schwierigkeit bei der Bearbeitung von Planungsaufgaben beeinflussen. Dabei variieren die Autoren verschiedene Einflussgrößen, unter anderem auch die semantische Einbettung. Die Studierenden sollen einerseits einen Handwerkertag und andererseits einen Familienumzugstag planen. Pascha et al. (2001) vermuten, dass der Tätigkeitsbereich „Beruf“ - operationalisiert durch die Semantik des Handwerkertags - den Studierenden weniger vertraut ist. Die Ergebnisse weisen tatsächlich eine höhere wahrgenommene Schwierigkeit für die Planung des Handwerkertags aus. Im Rahmen der Problemlöseforschung wurde der Einfluss des dargebotenen Inhalts ebenfalls untersucht. Wie weiter oben ausgeführt, zeigt sich, dass die semantische Einbettung zumeist die Schwierigkeit einer Aufgabe beeinflusst.

Die eigentlichen, objektiven Eigenschaften von Planungsaufgaben sind unter den formalen Aufgabenmerkmalen zusammengefasst. Die Befundlage ist diesbezüglich recht spärlich, so dass zumeist aus benachbarten Bereichen - wie der Problemlöseforschung - Ergebnisse herangezogen werden. Eine Ausnahme bilden die Studien von Pascha et al. (2001). Die Autoren können erwartungsgemäß belegen, dass der Planungsumfang die Lösungsgüte, den Zeitbedarf und die subjektive Schwierigkeit beeinflusst. Je mehr Aktivitäten

die Testpersonen in einen Tagesplan einbauen sollen, desto schwieriger gestaltet sich die Bewältigung der Aufgabe. Oerter und Dreher (1998) beschreiben eine Untersuchung zum Planungsverhalten bei Organisationsaufgaben, die der Planungsaufgabe von J. Funke und Krüger (1995) sehr ähnlich sind. Dabei wird das Planungsverhalten von Erwachsenen und Kindern verglichen. Kinder haben größere Schwierigkeiten, alle Aspekte einer Aufgabe zu berücksichtigen. Es zeigt sich, dass eine Planungsaufgabe umso schwieriger zu lösen ist, je mehr Informationen vorab verarbeitet werden müssen. Der Einfluss der Anzahl zu bearbeitender Aufgaben bzw. zu berücksichtigender Informationen auf die Aufgabenschwierigkeit kann auch in einigen anderen Studien zum Planen belegt werden (u. a. J. Funke & Krüger, 1995). Pascha et al. (2001) erklären die Befunde damit, dass die Arbeitsgedächtniskapazität umso stärker belastet wird, je mehr Einzelheiten einer Planungsaufgabe es mental zu repräsentieren gilt.

Pascha et al. (2001) haben auch den Einfluss unterschiedlicher Vorgaben (Randbedingungen) untersucht. Dabei finden sie bei einer Variation der zeitlichen Angaben keine eindeutigen Ergebnisse. Während bei Zeitdauerangaben die Aufgabe objektiv und subjektiv schwieriger wird, steigt bei Fixzeitpunktangaben lediglich die subjektiv wahrgenommene Aufgabenschwierigkeit. Die objektiven Größen (Lösungsgüte und Zeitbedarf) ergeben, dass Zeitpunktangaben die Aufgabe vereinfachen.

J. Funke und Krüger (1995) und Pascha et al. (2001) haben darüber hinaus weitere plausible Einflussgrößen für Planungsaufgaben genannt, deren Einfluss jedoch noch nicht empirisch belegt wurde. Darunter fällt der Zwang zur Unvollständigkeit im Test „Plan-A-Day“ von J. Funke und Krüger (1995). Die Autoren argumentieren, dass die Schwierigkeit und die ökologische Validität von Planungsaufgaben gesteigert würden, wenn nicht alle Aufträge erledigt werden können. Pascha et al. (2001) nennen in diesem Zusammenhang die Notwendigkeit des Findens von Alternativlösungen. Die Autoren nehmen an, dass eine Planungsaufgabe schwieriger wird, sobald

eine ursprünglich angestrebte Lösung im Laufe der Bearbeitung revidiert werden muss.

Laut Pascha et al. (2001) haben auch die aus der Problemlöseforschung bekannten Aspekte Transparenz und Dynamik einen Einfluss auf die Schwierigkeit von Planungsaufgaben. Inwieweit die Eigenschaften des komplexen Problemlösens grundsätzlich auf Planungskonstellationen zu übertragen sind, wird von den verschiedenen Autoren allerdings kontrovers diskutiert. Während beispielsweise von der Weth und Strohschneider (2002) Planungskonstellationen mit komplexen Problemen gleichstellen, kritisieren J. Funke und Fritz (1995) ein solches Vorgehen und stellen dem ein sachlogisches Begriffsverständnis entgegen, das auch die Grundlage für diese Arbeit bildet (s. o.).

#### *2.2.8.2 Personmerkmale als Einflussgrößen der Planungsleistung*

Unter Personmerkmalen als Einflussfaktoren der Aufgabenschwierigkeit sind alle Eigenschaften, Fähigkeiten und Kenntnisse des Planenden zu verstehen, die für die Lösung der Planungsaufgabe von Bedeutung sein können. Grob lassen sich die Personmerkmale wiederum in kognitive und nicht-kognitive unterscheiden. Zu letzterem gehören neben emotionalen und motivationalen Merkmalen auch individuelle Voraussetzungen in Form von Werten, Einstellungen und Interessen (vgl. M. Kersting, 1999). Nur wenige Studien haben sich mit dem Zusammenhang von Planungsleistung und Personmerkmalen beschäftigt.

Die Definition des Planens nach J. Funke und Glodowski (1990, s. o.) lässt die begriffliche Nähe zum kognitiven Personmerkmal des schlussfolgernden Denkens bzw. zur allgemeinen Intelligenz (vgl. Abschnitt 2.2.2) erkennen. M. Kersting (1999, S. 23 ff.) bemängelt bei der Betrachtung der Ergebnisse der Planungsforschung die fehlende Auseinandersetzung mit den Ergebnissen der Intelligenzforschung - insbesondere hinsichtlich der Abgrenzung zwischen Planen, Intelligenz und Wissen. Gerade die theoretischen Ausführungen von J. Funke und Fritz (1995), J. Funke und Glodowski (1990) deuten auf eine Nähe der Konstrukte. Planende werden in vielen

Planungssituationen hinsichtlich ihrer intellektuellen Fähigkeiten konfrontiert, wie beispielsweise beim Erkennen und Berücksichtigen der Randbedingungen (M. Kersting, 1999). Es ist daher umso erstaunlicher, dass es bisher nur wenige Untersuchungen zum Zusammenhang zwischen Planen und Intelligenz sowie Planen und Problemlösen gibt. Eine korrelative Studie liefern Fritz und Hussy (1995). Sie untersuchen die Planungsleistung von Grundschulkindern mit Hilfe des „Zoo-Spiels“. Beim „Zoo-Spiel“ geht es darum, sechs Tiere unter Beachtung von einschränkenden Randbedingungen in den Zoo zu bringen. Fritz und Hussy (1995) ermitteln neben der Planungsleistung im „Zoo-Spiel“ die Intelligenztestleistung mit Hilfe des HAWIK-R (Tewes, 1984). Dabei zeigen sich insgesamt nur geringe Korrelationen der Leistungsindizes mit den Intelligenztestwerten. Drei von acht Korrelationen erweisen sich als signifikant mit Werten um  $r = .30$ , alle anderen Zusammenhänge sind nicht substantiell. Die Autoren folgern, dass Intelligenz möglicherweise mit Leistungsaspekten des Zoo-Spiels kovariiert. Allerdings lassen sich aufgrund dieser Ergebnisse noch keine Schlüsse für einen generellen Zusammenhang zwischen Planen und Intelligenz ziehen.

Planen ist auch in einigen computergestützten Problemlöseszenarios erfasst worden, wenngleich nur als Nebenprodukt. Eine umfassende oder eigenständige Diagnose von Planungsleistung innerhalb der Szenarios findet sich nicht (Fritz und J. Funke, 1995).

Der Einfluss von nicht-kognitiven Personmerkmalen auf die Planungsleistung ist ebenfalls nur selten untersucht worden. Fritz und Hussy (1995) setzen im Rahmen ihrer Studie auch einen Planungsfragebogen nach Kreitler und Kreitler (1987) ein, der Planungsfähigkeit als Persönlichkeitskonstrukt erfasst. Es zeigen sich keinerlei signifikante Korrelationen zwischen den im Fragebogen ermittelten Persönlichkeitsaspekten und den Leistungsmaßen im „Zoo-Spiel“ (Fritz und Hussy, 1995).

J. Funke und Krüger (1995) führen ihr Verfahren „Plan-A-Day“ im Rahmen eines Assessment-Centers durch. Es können auch hier keine bedeutsamen Zusammenhänge zwischen der Planungsleistung und den im Assessment-Center gemessenen Dimensionen wie Teamfähigkeit, Kundenorientierung und Führungsfähigkeit gefunden werden. Auch die im Assessment-Center

erfassten kognitiven Dimensionen wie Entscheidungsvermögen und Organisationsfähigkeit hängen nicht mit der Planungsleistung zusammen.

Es zeigt sich abschließend, dass der Zusammenhang zwischen Planen und kognitiven sowie nicht-kognitiven Personmerkmalen bisher nur sehr selten empirisch überprüft wurde. Dies liegt sicherlich auch an dem grundsätzlichen Mangel geeigneter diagnostischer Verfahren zur Erfassung von Planungsleistungen (s. Kapitel drei). Die begriffliche Nähe der Konstrukte „Planen“ und „Intelligenz“ lässt zwar einen Zusammenhang vermuten (vgl. M. Kersting, 1999), dieser konnte in den bisherigen Studien aber nicht überzeugend empirisch bestätigt werden.

### ***2.2.9 Implikationen für die vorliegende Arbeit***

In diesem Abschnitt wurden die Erkenntnisse der psychologischen Grundlagenforschung zum Planen und Problemlösen aus eignungsdiagnostischer Perspektive betrachtet und das Begriffsverständnis der beiden Personmerkmale, das für diese Arbeit entscheidend ist, vorgestellt. Der Darstellung des Forschungsstands hinsichtlich entscheidender Einflussgrößen der Leistungen im Planen und Problemlösen galt dabei besonderes Augenmerk, da sich hieraus Implikationen für diese Arbeit ergeben. So sollen die Befunde zu den formalen Aufgabenmerkmalen (s. Abbildungen 8 und 11) bei der Konstruktion der Testaufgaben Berücksichtigung finden. Die Ergebnisse zum Einfluss von kognitiven und nicht-kognitiven Personmerkmalen auf Leistungen im Planen und Problemlösen gilt es, insbesondere bei der Prüfung der Konstruktvalidität der Verfahren, einzubeziehen.

Die empirischen Befunde zeigen auf, dass ein Problem umso schwieriger zu lösen ist, je größer sich der Umfang, die Vernetztheit und die Eigendynamik darstellen. Diese Einflussfaktoren sind daher bei der Aufgabenkonstruktion von Simulationsverfahren zu berücksichtigen, um a priori möglichst verschiedene Schwierigkeitsgrade zu erzeugen. Hinsichtlich weiterer



Aspekte wie der Transparenz liefert die empirische Befundlage keine eindeutigen Ergebnisse.

Es zeigt sich, dass sich die Problemlösefähigkeit als ein heterogenes Konstrukt mit verschiedenen Facetten darstellt (Klieme et al., 2001). Dabei gilt es komplexes, strategisches Problemlösen in dynamischen Situationen innerhalb computergestützter Szenarios von Problemlösen in weniger komplexen bzw. so genannten „einfachen“ Problemsituationen zu unterscheiden. Letztere Facette des Problemlösens hängt offensichtlich stark mit schlussfolgerndem Denken und daher Intelligenztestleistungen zusammen. Der Einfluss der Intelligenz - insbesondere der Verarbeitungskapazität - auf das komplexe Problemlösen in computergestützten Szenarios ist ebenfalls substantiell, wenngleich nur mittlerer Größenordnung von  $r = .30$  bis  $.40$  (Süß, 1996).

Für die Verfahrensentwicklung und -evaluation bedeutet dies, dass die verschiedenen Facetten des Personmerkmals „Problemlösefähigkeit“ berücksichtigt und der Einfluss der Intelligenz überprüft werden sollen. Dabei sind die methodischen Kritikpunkte an früheren Studien zu beachten und daher geeignete Intelligenzmaße sowie eine angemessene Stichprobengröße zu verwenden.

Studien zum Zusammenhang von Problemlösen mit nicht-kognitiven Personmerkmalen haben teilweise zwar erwartungsgemäße, jedoch nur geringe Korrelationen nachweisen können. Hinsichtlich des Einflusses des Vorwissens zeigt sich, dass Problemlöseleistungen zumeist bereichsspezifisch sind. Das bedeutet für die Entwicklung von eignungsdiagnostischen Verfahren, dass es zu berücksichtigen gilt, inwieweit die Bewältigung bestimmter Aufgaben für den jeweiligen Berufs- oder Arbeitsbereich erfolgsrelevant ist. Dies geschieht typischerweise mit Hilfe von Arbeits- und Anforderungsanalysen, die in Abschnitt 2.1 vorgestellt wurden.

Die empirische Befundlage zu Einflussgrößen auf Leistungen im Planen erscheint bisher als recht karg. Lediglich der Einfluss des Planungsumfangs auf die Aufgabenschwierigkeit kann dabei belegt werden (Pascha et al., 2001).

Die diagnostisch relevante Definition von Planen im Rahmen dieser Arbeit basiert in erster Linie auf den Arbeiten von J. Funke und Glodowski (1990), J. Funke und Fritz (1995) sowie J. Funke und Krüger (1995). Planen gliedert sich laut J. Funke und Glodowski (1990) in die Planerstellung und die Planausführung mit jeweils mehreren Teilleistungen. Die Qualität der Planungsleistung wird nach J. Funke und Krüger (1995) durch die Güte der Aufgabenbearbeitung bestimmt. Dabei sind die Güte der Teilleistungen während der Planerstellung und der Planausführung entscheidend.

Eignungsdiagnostische Instrumente sollten diese Aspekte sowie weitere schwierigkeitsvariierende Faktoren berücksichtigen. Die begriffliche Nähe zwischen Planen und schlussfolgerndem Denken lässt einen Zusammenhang mit Intelligenz erwarten. Der Einfluss von Intelligenztestleistungen auf die Planungsleistung in Simulationsverfahren wurde jedoch bisher noch nicht umfassend untersucht.

### 3 Internetgestützte Berufseignungsdiagnostik und Forschungsstand der Simulationsverfahren zum Planen und Problemlösen

Die internetgestützte<sup>10</sup> Berufseignungsdiagnostik befasst sich mit der Methodologie der Entwicklung, Prüfung und Anwendung psychologischer Verfahren zum Ziele eignungsbezogener Erfolgsprognosen und Entscheidungshilfen im beruflichen Kontext mit Hilfe des Mediums Internet (vgl. Schuler & Höft, 2001). Als Eignung wird dabei das Ausmaß der Übereinstimmung von beruflichen Anforderungen des Arbeitsplatzes und der weiteren Arbeitsumgebung mit den Potenzialen und Kompetenzen einer Person verstanden. Dahinter steht die Annahme, dass der Grad der Übereinstimmung die Erfolgswahrscheinlichkeit für eine Berufs- oder Ausbildungstätigkeit mitbestimmt (Schuler & U. Funke, 1995). Daher ist für die Entwicklung von eignungsdiagnostischen Verfahren neben der Betrachtung der Person die Untersuchung der Situation im Rahmen der Arbeits- und Anforderungsanalyse von entscheidender Bedeutung (zur Problematik der Bestimmung zukünftiger Anforderungen s. N. Anderson, Lievens, Van Dam & Ryan, 2004).

Dieser arbeits- und anforderungsanalytische Bezug der Verfahren *webPlanungsaufgabe* und *webPostkorb* wurde im zweiten Kapitel mit Hilfe empirischer Tätigkeitsanalysen von Schulleitungen dargelegt. Die Dokumentation der Bedeutung planerischer und problemlösender Fähigkeiten für die erfolgreiche Bewältigung der Arbeit von Schulleitungen und Führungskräften sowie der Forschungsstand zu den Konstrukten komplettierten das zweite Kapitel.

In diesem Kapitel stehen die bisherigen methodischen Ansätze zur Diagnose von Planungs- und Problemlöseleistung im Mittelpunkt. Bevor jedoch näher auf den Aspekt „internetbasiert“ im Rahmen der Begriffsbestimmung „E-Assessment“ und „Blended-Assessment“ eingegangen wird und dabei mögliche Ziele, Vorteile und Risiken der internetgestützten Eignungsdiagnostik erörtert werden, soll zunächst eine Klassifikation von

---

<sup>10</sup> Die Begriffe „internetgestützt“, „internetbasiert“ und „online“ werden im Rahmen dieser Arbeit synonym verwendet.

Schuler und Höft (2001) die Ausführungen einleiten. Ihr „Prinzip der trimodalen Eignungsdiagnostik“ (vgl. Abbildung 17) ist hilfreich zum Verständnis des methodischen Hintergrunds dieser Arbeit. Die Autoren unterscheiden drei grundsätzliche Herangehensweisen eignungsdiagnostischer Verfahren, und zwar

- den Eigenschafts- oder Konstruktansatz,
- den Simulationsansatz und
- den biografischen Ansatz.

Hinter jedem Ansatz stehen Schuler und Höft (2001) zufolge eine teilweise eigenständige Validierungslogik sowie spezifische Methoden der Erfassung der jeweiligen Personmerkmale (vgl. Abbildung 17). Zielsetzung des Eigenschafts- oder Konstruktansatzes ist eine Erfassung von Merkmalen, die als zeitlich relativ stabil angenommen werden. Typische Messverfahren, um Konstrukte möglichst homogen zu diagnostizieren, stellen psychologische Tests dar, wie beispielsweise Intelligenz- oder Persönlichkeitstests. Das zugehörige Validitätsprinzip ist das der Überprüfung der Konstruktvalidität.

Der Simulationsansatz zielt hingegen auf eine möglichst gute Abbildung der Realität beruflicher Anforderungen ab, um erfolgsrelevantes Verhalten zu erfassen. Inwieweit das Ziel erreicht wurde, gilt es mittels der inhaltlichen Repräsentativität oder Inhaltsvalidität zu kontrollieren. Simulationen, Arbeitsproben oder Assessment-Center-Übungen sind nach Schuler und Höft (2001) typische Erhebungsformen.

Die dritte Konstruktionsstrategie eignungsdiagnostischer Verfahren repräsentiert nach Schuler und Höft (2001) der biografische Ansatz. Typischerweise erfolgt die Ermittlung biografischer Daten mit Hilfe von Fragebögen, Bewerbungsunterlagen oder Interviews. Dem biografischen Ansatz liegt die prognostische Validität als Validierungslogik zu Grunde.

Das „Prinzip der trimodalen Eignungsdiagnostik“ geht zum Teil auf Wernimont und Campbell zurück, die bereits 1968 die Unterschiede zwischen „signs“ und „samples“ betonen. „Signs“ bilden demnach Testverfahren, die allgemeine psychologische Dimensionen messen,

„samples“ hingegen stellen Arbeitsbeispiele oder Simulationen von typischen Situationen der Arbeitswelt dar.

**Tabelle 9: Vergleich eigenschafts- und simulationsorientierter Verfahrenskonzepte (vgl. Höft & U. Funke, 2001, S. 136, erweitert um den Aspekt der Validierungslogik)**

	<b>Konstruktansatz (eigenschaftsorientiert)</b>	<b>Simulationsansatz (simulationsorientiert)</b>
Bezug zum Arbeitsplatz	geringer (eigenschaftsbezog. Items = signs)	höher (tätigkeitsbezog. Aufgabe = sample)
Anforderungsbereich	eigenschaftshomogen	eigenschaftsheterogen
Ablaufstruktur	elementaristisch (Einzelitems)	ganzheitlich (Gesamtaufgabe mit Planungsanforderungen)
Rückmeldung & Kontrolle	gering	Ablauf-Feedback Ergebnis-Feedback
Verlaufscharakter	statisch	dynamisch (reaktionsabhängige Aufgabenentwicklung)
Validierungslogik	Konstruktvalidierung	Inhaltsvalidierung
typische Beispiele	Intelligenztest, Persönlichkeitsverfahren	Gruppendiskussion, Postkorb, diagnostisches Rollenspiel

U. Funke (1993) hat die entscheidenden Charakteristika der beiden eignungsdiagnostischen Ansätze „Konstrukt“ und „Simulation“ verglichen. Dabei zeigen sich hinsichtlich der Kriterien Bezug zum Arbeitsplatz,

Anforderungsbereich, Ablaufstruktur, Rückmeldung und Kontrolle, Verlaufscharakter sowie Validierungslogik deutliche Unterschiede (s. Tabelle 9). Die unterschiedlichen Kennzeichen weisen auch auf typische Vor- und Nachteile der Ansätze hin (Höft & U. Funke, 2001). So wird als Vorteil der Testverfahren, die den Konstruktansatz verfolgen, zumeist die gut nachvollziehbare Konstruktion und die umfangreichen Möglichkeiten der Güteprüfung beschrieben. Problematisch erscheinen die zumeist niedrige inhaltliche Validität und die häufig fehlende Akzeptanz bei den Testpersonen aufgrund des geringen Bezugs zum Arbeitsplatz. Als Vorteile der Verfahren, die simulativ ausgerichtet sind, gelten demgegenüber die postulierte höhere Akzeptanz und Inhaltsvalidität. Diese Verfahren werden allerdings wegen mangelnder oder gar nicht erfasster Testgüte vielfach kritisiert. In dieser Arbeit soll die Kombination des Konstrukt- und Simulationsansatzes im Sinne eines Hybridvorgehens verfolgt werden, um die Vorteile beider Herangehensweisen zu nutzen.

### **3.1 E-Assessment und Blended-Assessment**

#### ***3.1.1 Bedeutung von E-Assessment-Verfahren***

Die praktische Bedeutung der internetgestützten Eignungsdiagnostik steigt (Hünninghausen, 2002). Der Studie „Recruiting Trends 2005“ der Universität Frankfurt und des Jobportals Monster zufolge, finden mittlerweile etwa 43 Prozent aller Bewerbungen online statt (Keim, König, von Westarp, Wendt & Weitzel, 2005). Gleichermaßen gewinnt die internetgestützte Berufs-, Ausbildungs- und Studienberatung mit Hilfe von umfassenden Informationen und eignungsdiagnostischen Verfahren deutlich an Bedeutung (Kupka, 2005, in Vorbereitung). Arbeits-, Ausbildungs- und Weiterbildungsinteressierte nutzen Intra- und Internet zunehmend für ihre Zwecke. Dabei wird das Internet gemäß einer bereits 2001 von TNS Emnid im Auftrag von Jobware durchgeführten Studie von mehr als der Hälfte der Bewerber zur Stellensuche genutzt. Bei den Gruppen der Abiturienten und

Hochschulabsolventen, Führungskräften und Selbständigen bzw. den 18- bis 29-Jährigen stellt das Online-Medium gar die bevorzugte Informationsquelle dar (TNS Emnid, 2001).

Auch Unternehmen nutzen die Möglichkeiten des Internets für ihre Personalarbeit und setzen alle Formen eignungsdiagnostischer Verfahren ein (vgl. Kupka, Diercks & Kopping, 2004). Zahlreiche Firmen bieten mittlerweile zumindest webbasierte Bewerbungsformulare an. Die Wissenschaft nähert sich erst in jüngster Zeit vermehrt dem Themengebiet. So erschienen kürzlich erste deutschsprachige Überblickswerke (Hertel & Konradt, 2004; Kirbach, Montel, Oenning & Wottawa, 2004; Konradt & Sarges, 2003).

Für die eignungsdiagnostischen Verfahren gilt Folgendes: Während sich die Übertragung von Fragebögen und klassischen Testverfahren auf das Internet relativ einfach gestaltet und somit auch die Mehrzahl der Online-Verfahren ausmacht (vgl. Lievens & Harris, 2003; Konradt, Lehmann, Böhm-Rupprecht & Hertel, 2003), erfolgt der Einsatz von internetbasierten, interaktiven Instrumenten im Sinne des Simulationsansatzes noch sehr selten (Eichstaedt, 2005; Konradt & Hertel, 2004). Simulationsbasierte Verfahren existieren bisher zwar computergestützt, jedoch kaum internetbasiert. Gezielte Untersuchungen zur Testgüte von interaktiven Online-Simulationsverfahren finden sich nur sehr wenige.

### **3.1.2 Definition E-Assessment-Verfahren und E-Assessment**

Unter Online- oder E-Assessment-Verfahren sind internetgestützte Instrumente zur Beurteilung und Vorhersage beruflich relevanter biografischer und psychologischer Variablen zur Abschätzung der Eignung zu verstehen (Konradt & Sarges, 2003). Sowohl die Durchführungen als auch die Rückmeldungen verlaufen typischerweise online. Übersetzt ins Deutsche bedeutet „Assessment“ Festsetzung, Beurteilung, Einschätzung (Langenscheidt, 2005). Demnach ist ein E-Assessment-Verfahren ein elektronisches, internetgestütztes Instrument zur (Eignungs-) Beurteilung einer Person. Es handelt sich bei E-Assessment-Verfahren um eine neue Klasse eignungsdiagnostischer Instrumente, die online ablaufen und somit

eine Reihe von Eigenschaften, wie Orts- und Zeitunabhängigkeit und Kosteneinsparpotenziale, aber auch Manipulationsgefahren, verbindet (s. Abschnitt 3.2).

E-Assessment-Verfahren können im Rahmen der Personalauswahl, des Personalmarketings, der Personaleinschätzung, der Personalförderung oder der Berufs- und Ausbildungsberatung zum Einsatz kommen. Instrumente, die dem Zweck der Selbsteinschätzung und Selbstauswahl dienen, werden zumeist als Self-Assessment- oder Selbsterkundungs-Verfahren bezeichnet<sup>11</sup> (vgl. Mayr & Nieskens, 2004). Bei einem weiten Begriffsverständnis werden alle internetgestützten Instrumente der Eignungsdiagnostik als E-Assessment-Verfahren bezeichnet, wie in Abbildung 12 dargestellt.<sup>12</sup> Einige Autoren (vgl. Konradt & Sarges, 2003) verwenden eine engere Definition, wonach ausschließlich die verhaltensorientierten, interaktiven Simulationsverfahren und nicht Bewerbungsformulare oder die direkte Adaptation der klassischen psychologischen Tests als E-Assessment-Instrumente gelten.

Die systematische Kombination verschiedener E-Assessment-Verfahren lässt sich als E-Assessment (teilweise auch Online-Assessment-Center) bzw. Self-Assessment bezeichnen. Dieser Einsatz mehrerer Einzelverfahren dient typischerweise der umfassenden Eignungsbeurteilung von Testpersonen. E-Assessments sind zu unterscheiden von Assessment-Centern und von computergestützten Verfahren. Der Begriff „Assessment-Center“ steht im Allgemeinen für eine ein- bis mehrtägige Veranstaltung. Im Rahmen dieser physischen Zusammenkunft bewerten einige Beobachter/innen mehrere Bewerber/innen in unterschiedlichen, meist simulativen Übungen hinsichtlich ihrer Berufseignung (Thornton & Byham, 1982). Demgegenüber stellen computergestützte Verfahren eine Gruppe von Tests oder Simulationen dar, die auf dem Medium Computer, aber nicht intergestützt ablaufen (Kubinger, 1993).

---

<sup>11</sup> Im Folgenden wird der Übersichtlichkeit halber zumeist lediglich der Begriff „E-Assessment-Verfahren“ verwendet.

<sup>12</sup> Hier wird das weite Begriffsverständnis angewendet.



### 3.1.3 Blended-Assessment - eine Klassifikation eignungsdiagnostischer Verfahren unter Einbeziehung des Mediums Internet

Im Zusammenhang mit der Begriffsklärung von „E-Assessment“ wird im Folgenden das Modell des „Blended-Assessments“ eingeführt: Die in Abbildung 12 abgebildete Klassifikation eignungsdiagnostischer Ansätze erweitert das „Prinzip der trimodalen Eignungsdiagnostik“ von Schuler und Höft (2001) um den Aspekt des Internets.

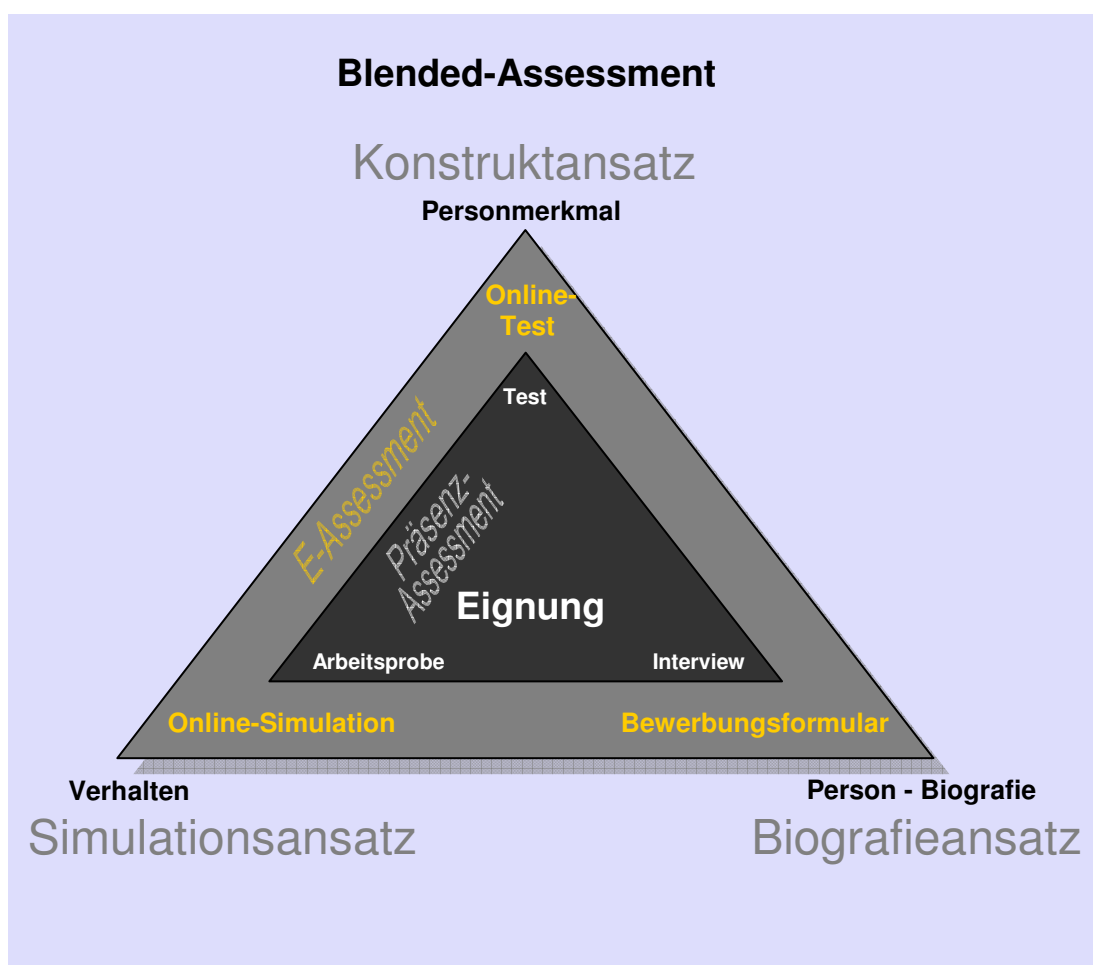


Abbildung 12: Blended-Assessment / Blended-Eignungsdiagnostik (angelehnt an Schuler & Höft, 2001, erweitert um das Medium Internet)

Das Modell des Blended-Assessments kann zum einen der Beschreibung verschiedener Herangehensweisen und typischer Verfahrensarten dienen und zum anderen darstellen, wie Online- und Offline-Eignungsdiagnostik in der Praxis sinnvoll zu verbinden sind (s. Abschnitt 3.2). Dabei verwendet die

Klassifikation teilweise neue Begriffe gegenüber dem Ausgangsmodell von Schuler und Höft (2001). So steht anstelle von „Eigenschaften“ der oben bereits eingeführte Begriff „Personmerkmal“, um zu verdeutlichen, dass innerhalb des Konstruktansatzes sämtliche eignungsrelevanten personenbezogenen Merkmale Berücksichtigung finden. Dazu zählen neben den Eigenschaften auch die Interessen oder das Wissen.

Alle Ansätze der herkömmlichen Eignungsdiagnostik nach Schuler und Höft (2001), die der letzte Abschnitt thematisierte, sind aufgrund der erweiterten Möglichkeiten der Computer- und Netztechniken auch im Rahmen von E-Assessments umsetzbar. Jeder der drei eignungsrelevanten Diagnosebereiche Verhalten, Biografie und Personmerkmale wird - wie bereits oben beschrieben - typischerweise mit speziellen (Offline-) Methoden erhoben. Gleiches gilt für die internetgestützte Bestimmung der Eignung. So lassen sich biografische Informationen über die Bewerber mit Hilfe von Bewerbungsformularen mit Upload-Funktionen für Zeugnisse, Fotos etc. erfassen. Dies ist nicht mit der einfachen Versendung von Bewerbungsunterlagen per E-Mail zu verwechseln, wodurch zumeist ein innerbetrieblicher Mehraufwand entsteht (vgl. Wottawa, Kirbach, Montel & Oenning, 2005). Auch mittels internetgestützter Interviews können biografische Informationen erhoben werden.

Im Rahmen des Konstruktansatzes kommen Online-Tests zum Einsatz. Dies dient typischerweise der Diagnose entscheidender Personmerkmale.

Die Eignungsdiagnostik hinsichtlich des dritten Merkmalsbereichs beinhaltet die Auseinandersetzung mit berufsrelevantem Verhalten, das in ähnlicher Form am Arbeits- bzw. Ausbildungsplatz gefordert wird. Als typische Online-Methoden kommen Simulationen, Planspiele oder auch Gruppendiskussion in Chats in Frage.

### **3.1.4 Online-Simulationsverfahren**

Aufgrund der Bedeutung für die vorliegende Arbeit soll im Folgenden die Verfahrensklasse der Online-Verfahren, die dem Simulationsansatz zuzuordnen sind, näher betrachtet werden. Online-Simulationsverfahren oder

Online-Simulationen sind interaktive Verfahren, die die Möglichkeiten der Computertechnik bewusst internetgestützt einsetzen, um dynamische und möglichst realistische Szenarios zu erzeugen. Konradt und Sarges (2003) betonen die zukünftig weiter steigende Bedeutung dieses Bereichs und konstatieren, dass die Nutzung der interaktiven und multimedialen Möglichkeiten des Mediums „Internet“ sowie die damit verbundene Erfassung zusätzlicher Verhaltensparameter allerdings erst in Ansätzen erkennbar seien. Gerade diese Verfahrensklasse eröffnet jedoch neue Möglichkeiten für die Eignungsdiagnostik (vgl. Hornke und M. Kersting, 2006). Trotzdem sind in der eignungsdiagnostischen Literatur zu diesem Themengebiet bislang kaum tiefer gehende Arbeiten zu finden.

Internetbasierte Simulationsverfahren können sich in ihrer Ausgestaltung stark unterscheiden. Dabei ist der Aspekt der Interaktivität entscheidend. Aus dem Forschungsbereich des E-Learnings liefert Schulmeister (2002) hierzu eine Taxonomie, die sich auch für die unterschiedliche Ausgestaltung von E-Assessment-Verfahren als hilfreich erweisen kann und daher kurz vorgestellt wird. Schulmeister (2002, S. 194 ff.) unterscheidet zwischen sechs Stufen der Modellierung von Multimedia-Komponenten, die nach dem Interaktivitätsniveau differenzieren:

- Stufe I:        Objekte betrachten und rezipieren
- Stufe II:       multiple Darstellungen betrachten und rezipieren
- Stufe III:      die Repräsentationsform variieren
- Stufe IV:      den Inhalt der Komponente modifizieren
- Stufe V:       das Objekt bzw. den Inhalt der Repräsentation konstruieren
- Stufe VI:      den Gegenstand bzw. Inhalt der Repräsentation konstruieren  
                  und durch manipulierende Handlungen intelligente  
                  Rückmeldung vom System erhalten

Mit steigender Stufe nimmt demnach die Handlungsfreiheit der Probanden zu. Während die Stufe I noch keine wirkliche Interaktivität beinhaltet und Multimediainhalte lediglich betrachtet werden können, ist es auf

Stufe II zumindest möglich, unterschiedliche Varianten auszuwählen. So vermag der Proband beispielsweise durch Klicken oder Auswählen eines Multimedia-Objekts Inhalte, Bilder oder Animationen auszutauschen. Auf Stufe III ist es dem Benutzer möglich, durch direkte Manipulation Grafiken oder Objekte zu skalieren und zu bewegen oder durch Klick auf Objekte in Filmen zu anderen Abschnitten des Filmes zu verzweigen. Diese Stufe der Interaktivität bietet dem Probanden somit einen aktiven Einfluss auf die Repräsentation des Multimedia-Objekts. Dabei bleiben das Objekt oder der Film selbst unverändert und die Benutzerhandlungen verändern nur die Repräsentationsform und nicht den Inhalt. Diese Stufe ist laut Schulmeister (2002) bedeutsam für die Motivation der Benutzer, was Implikationen für die Akzeptanz von E-Assessment-Verfahren haben dürfte. Auf der Stufe IV ist der Inhalt der Multimedia-Komponenten nicht vorgefertigt, sondern wird erst durch die Interaktion mit dem Probanden gebildet. Der Benutzer kann durch Eingabe von Daten oder Variieren von Parametern innerhalb eines gesetzten Rahmens andere Darstellungen erzeugen oder andere Verknüpfungen visualisieren. Typisch für die Stufe V ist, dass Probanden mit Hilfe von interaktiven Werkzeugen im Stande sind, Objekte selbst zu kreieren, Ideen zu visualisieren oder Modelle zu entwerfen. Für die letzte Stufe der Interaktivität gilt nach Schulmeister (2002), dass der Anwendung oder dem Programm Objekte bzw. Aktionen geschickt werden, die das Programm versteht und auf die es mit entsprechend bedeutungsvollen Rückmeldungen antwortet. Dies ist zwar nicht mit menschlicher Kommunikation oder sozialer Interaktion gleichzusetzen, erlaubt allerdings schon einen Austausch symbolischer Inhalte in einer restringierten Domäne. Bisherige internetbasierte Simulationsverfahren erreichen typischerweise erst die unteren Stufen der möglichen Interaktivitätsskala. Die eignungsdiagnostische Forschung hierzu steht dabei erst am Anfang. Interessant wird sein, inwieweit umfangreichere und hoch interaktive Verfahren in Zukunft weitere Verhaltensaspekte ermöglichen und bewerten können. Allerdings bleibt für Testentwickler die Schwierigkeit, trotz steigender Interaktivität die Vergleichbarkeit und Standardisierung eignungsdiagnostischer Informationen zu gewährleisten.

### **3.2 Ziele, Vorteile und Risiken**

Der folgende Abschnitt thematisiert Möglichkeiten und Grenzen der internetgestützten Eignungsdiagnostik. Dabei werden die praktische Bedeutung des oben eingeführten Modells des Blended-Assessments im Rahmen einer Verbindung von Online- und Präsenz-Assessments diskutiert, Ziele, Strategien und mögliche Vorteile des Einsatzes von E-Assessments für Organisationen vorgestellt und abschließend die Besonderheiten von E-Assessment-Verfahren im Hinblick auf Testgütekriterien behandelt.

#### ***3.2.1 Praktische Bedeutung von Blended-Assessment***

Hinsichtlich der praktischen Umsetzung bedeutet das Modell des Blended-Assessments, dass für komplexe eignungsdiagnostische Entscheidungen - wie die der Personalauswahl, Personalentwicklung oder auch der Berufswahl - eine multiple, vielschichtige Herangehensweise angemessen erscheint, da psychologische Diagnosen teilweise auch von der Erfassungsmethode abhängig sind (Schuler & Höft, 2001). Erst die systematische Kombination unterschiedlicher Verfahren gewährleistet laut Schuler und Höft (2001), dass alle relevanten Facetten eines Merkmalsbereichs berücksichtigt werden. Neben den von Schuler und Höft (2001) genannten eignungsdiagnostischen Grundsätzen der

- Trimodalität (drei eignungsrelevante Merkmalsbereiche) und
- Multimethodalität (verschiedene Verfahren) ist auch eine
- Multimedialität (verschiedene Medien, Kommunikationsformen) zu fordern.

Mit Hilfe einer Vorgehensweise, die diese drei Aspekte berücksichtigt, kann eine große Spannweite möglicher Erhebungsformen abgedeckt werden, wie beispielsweise verhaltens-, biografie- und personmerkmalbezogene Diagnosen oder Selbst- und Fremdeinschätzungen (Schuler & Höft, 2001). Laut Schuler und Höft (2001) zieht ein multimethodales und trimodales

Vorgehen eine erhöhte Generalisierbarkeit der eignungsdiagnostischen Ergebnisse auf zukünftige berufliche Leistungen nach sich.

Durch den Einsatz verschiedener Medien bzw. der Nutzung unterschiedlicher Kommunikationsformen ist es möglich, Stimulus- und Antwortformen zu variieren - wie beispielsweise schriftliche, mündliche und verhaltensbasierte oder computervermittelte und Face-to-Face. So können ggf. vorhandene Präferenzen von Testpersonen hinsichtlich einer Kommunikationsform ausgeglichen werden.

Es ist anzunehmen, dass der berufliche Alltag von Führungskräften typischerweise durch die Arbeit mit verschiedenen Medien bzw. mittels unterschiedlicher Kommunikationsformen geprägt ist. Durch multimediale Erfassungsmethoden sind daher ggf. weitere berufsrelevante Aspekte diagnostizierbar, die bei einer Beschränkung auf nur ein Medium möglicherweise nicht erfasst werden können (vgl. Döring, 1997).

Durch eine theoriegestützte Systematik der Eignungsdiagnose minimiert sich letztlich das Risiko gravierender Folgen von unpassenden Karriere- und Personalauswahlentscheidungen (vgl. Kupka, 2005).

Wie die systematische Kombination verschiedener Verfahren zu höheren prognostischen Validitäten führen kann, zeigen nicht zuletzt die Ergebnisse der Metaanalyse von Schmidt und Hunter (1998, 2000). Die Autoren ermitteln, dass die Qualität der Prognose von Berufserfolg, gemessen mit Hilfe des Validitätskoeffizienten, durch die Kombination von Intelligenztests und Arbeitsproben von  $r = .51$  auf  $r = .63$  steigt. Dies entspricht einem Validitätszuwachs von 24 Prozent hinsichtlich der Berufsleistung gegenüber der alleinigen Prädiktorleistung der Intelligenztests. Andere Verfahrenskombinationen ergeben wiederum nur geringen oder keinen Validitätszuwachs (zur genauen Darstellung der inkrementellen Validität verschiedener Verfahren gegenüber Intelligenztests s. Schmidt & Hunter, 2000).

Neben möglichen Vorzügen für die Qualität der Eignungsdiagnostik durch einen theoriegeleiteten Einsatz von Blended-Assessments sprechen auch

realisierbare Kosten- und Zeitersparnisse für den praktischen Nutzen des Modells (vgl. Finke & Eckl, 2001; Lievens & Harris, 2003).

Dabei ist umfassende Eignungsdiagnostik nach dem hier vorgeschlagenen Ansatz des Blended-Assessments zweistufig aufgebaut, um die Möglichkeiten und Vorteile von Präsenz- und Online-Verfahren systematisch zu verbinden. Im Rahmen eines vorgeschalteten E-Assessment lässt sich bereits eine ganze Reihe von Aspekten untersuchen, die es sonst aufwändiger in Präsenzveranstaltungen bzw. nicht elektronisch zu erheben gilt. So ist es möglich, die Papierbewerbungen durch Bewerbungsformulare mit Upload-Funktionalität zu ersetzen. Bewerbungen liegen dabei sofort digital vor mit den verbundenen positiven Effekten für den innerbetrieblichen Workflow (s. u., vgl. Wottawa et al., 2005). Tests und auch Simulationen können zumeist schneller sowie kosten- und aufwandsgeringer online durchgeführt werden, so dass in der zweiten Stufe des Blended-Assessments, dem Offline- oder Präsenz-Assessment, vermehrt die Aspekte im Mittelpunkt stehen können, die sich internetgestützt gar nicht oder nur sehr schwierig ermitteln lassen. Dazu gehören insbesondere die Bereiche der sozialen Interaktion, wie sie typischerweise in Assessment-Centern oder Interviews beleuchtet werden. Durch ein solches Vorgehen ist es möglich, die kostbare Offline-Zeit gezielter zu nutzen. E- und Präsenz-Assessments stellen in diesem Modell sinnvolle Bestandteile eines umfassenden Blended-Assessments dar.

### **3.2.2 Ziele, Strategien und mögliche Vorteile des Einsatzes von E-Assessments**

Der Einsatz von E-Assessments im Kontext der Personalarbeit kann nach Konradt und Sarges (2003, S. 4) im Rahmen von vier Strategien erfolgen:

- Personalmarketing
- Personalauswahl
- Personalentwicklung
- Workflow-Optimierung

Neben diesen auf ein Unternehmen bezogenen Zielrichtungen werden E-Assessments auch zu folgenden Zwecken eingesetzt:

- Berufs-, Ausbildungs- und Studienberatung
- Anregung zur Selbstauswahl

Die Perspektive des Personalmarketings zielt auf die Bindung von potenziellen Mitarbeitern an das Unternehmen und den Aufbau und die Pflege eines positiven Images und dadurch indirekt auf die Gewinnung neuer Mitarbeiter auf dem Personalmarkt ab. Bei der Personalauswahl wird die Vorauswahl (Screening), Auswahl und Platzierung von externen und internen Bewerbern anvisiert. Bei der Personalentwicklung ist die Potenzialanalyse und Personalförderung durch zielgerechte Informationen und Instrumente entscheidend.

Die Workflow-Optimierung lässt sich als eine übergeordnete Zielsetzung verstehen, weil es ebenfalls als Zusatznutzen bei Erreichung der anderen Ziele anvisiert werden kann (Konradt & Sarges, 2003). Dabei steht der Gedanke im Zentrum, durch effizientes und rationales E-Assessment Kosten- und Zeiteinsparungen bei gleichzeitiger Informationsverbreitung zu erreichen (vgl. K. Kersting, 2004). Durch eine vollständige, elektronische Bewerberadministration (Datenerfassung und Datenverarbeitung) sind zusätzliche Rationalisierungspotenziale möglich (Konradt & Sarges 2003).



Bei der Berufs-, Ausbildungs- oder Studienberatung werden Ratsuchenden in Self-Assessments Informationen und Verfahren zur Verfügung gestellt, die Karriereentscheidungen verbessern helfen oder weitere Maßnahmen - wie Karrieregespräche - sinnvoll vorbereiten und ergänzen können (vgl. Mayr, 2001; Mayr & Nieskens, 2004; Sieland, 2002, 2003; Sieland & Nieskens, 2001; Zimmerhofer & Hornke, 2005).

Das Ziel der Anregung zur Selbstauswahl beim E-Assessment-Einsatz überschneidet sich mit den Aspekten des Personalmarketings, der Personalauswahl und der Beratung. Dabei gilt es eine Prüfung der Passung zwischen Unternehmen bzw. Studiengang oder Ausbildung und Individuum beim potenziellen Bewerber mit dem Ziel zu initiieren, geeignete Personen „anzuziehen“ und ungeeignete „abzuschrecken“. Wenn mit Hilfe von E-Assessments die Anzahl geeigneter Bewerbungen an der Gesamtzahl der Bewerbungen steigt, zeugt das von effektivem Personalmarketing bzw. effektiver Beratung und hat positive Auswirkungen auf den Erfolg der gesamten Personalauswahl (s. Abschnitt 3.3).

Abgesehen von der weiten Verbreitung von internetfähigen Computern (vgl. Bartram, 2000) und den inhaltlichen Vorteilen, die innerhalb der sechs Zielrichtungen beim Einsatz von E-Assessments anvisiert werden, wird die wachsende Bedeutung von E-Assessments sicherlich auch durch Kostenargumente befördert (vgl. Naglieri et al., 2004). So rechnen u. a. Kirbach et al. (2004) und Konradt und Sarges (2003) beispielhaft vor, dass sich deutliche Kosteneinsparpotenziale durch die internetgestützte Personalarbeit ergeben. Lievens und Harris (2003) und Finke und Eckl (2001) berichten von Fallstudien, bei denen sich durch die Einführung von internetgestützter Personalauswahl alleine kurzfristige Einsparungen in Höhe von fünfzig Prozent bei den mit der Einstellung anfallenden Kosten und von rund sechzig Prozent bei der aufgewendeten Zeit realisieren ließen. Einige Unternehmen akzeptieren aufgrund des Mehraufwands bei herkömmlichen Papierbewerbungen seit kurzer Zeit nur noch digitale Bewerbungen (Diercks & Kupka, 2004). Für die Konstruktion und Durchführung von herkömmlichen

Tests, Interviews oder Assessment-Centern ergeben sich Kosten, die sich möglicherweise deutlich reduzieren, wenn geeignete E-Assessments als (Vor-) Selektion der Bewerber zur Verfügung stehen. Der höhere Aufwand durch den E-Assessment-Einsatz ist normalerweise deutlich geringer als die eingesparten Kosten (vgl. Kopping, Diercks & Kupka, 2007). E-Assessments beinhalten darüber hinaus sämtliche Vorteile des Mediums „Internet“, wie Unabhängigkeit von Ort und Zeit und eine mögliche automatisierte Kommunikation per E-Mail.

In der Praxis kommen E-Assessments im Rahmen aller sechs oben genannten Bereiche zum Einsatz. Oft werden dabei verschiedene Ziele verfolgt, um Synergieeffekte zu nutzen, wie beispielsweise bei der Kombination von Personalmarketing, Personalauswahl und Workflow-Optimierung im Rahmen von „Recruitment“-Anwendungen. So lassen sich die Pflege eines positiven Unternehmensimages, die eignungsdiagnostische Vorauswahl von Bewerbern im Internet und eine Zeit- und Kosteneinsparung durch Optimierung von Arbeitsabläufen verbinden (Kupka et al., 2004).

### ***3.2.3 Vorteile und Risiken hinsichtlich der Testgüte von E-Assessment-Verfahren***

Kubinger (1993) zieht in seinem Vergleich zwischen herkömmlichen und computergestützten Testverfahren das Fazit, dass bei computergestützten Verfahren keine grundsätzlichen Unterschiede gegenüber Papier-Bleistift-Verfahren in Bezug auf Reliabilität, Validität, Interpretationseindeutigkeit und Verfälschbarkeit zu erwarten sind. Es ist anzunehmen, dass dies auch für internetgestützte Verfahren gilt. Der Vergleich ist lediglich in Bezug auf die Verfälschbarkeit einer internetgestützten Testdurchführung einzuschränken, da Verfälschungen bei einem Einsatz ohne Aufsicht, beispielsweise von zu Hause oder einem öffentlichen Internetzugang aus, möglich sind (s. u.).

Gegenüber der Papier-Bleistift-Variante - aber auch gegenüber computergestützten Verfahren - bieten E-Assessment-Verfahren einige Vorteile, die sich vor allem durch die Möglichkeiten des Internets bzw. der

Computertechnologie ergeben. Allerdings liegen diesbezüglich insgesamt noch recht wenige empirische Studien vor.

Der wesentlichste testtheoretische Vorteil von computergestützten Verfahren gegenüber Papier-Bleistift-Verfahren liegt nach Kubinger (1993) insbesondere in der Nützlichkeit der Verfahren. Diese Nützlichkeit zeige sich insbesondere bei der Diagnose von Realverhalten in Simulationsverfahren, da sich Realverhalten in Papier-Bleistift-Verfahren kaum ermitteln lasse. Ein Vorteil von computer- und internetgestützten Simulationsverfahren besteht womöglich auch in der einfachen Speicherung der gesamten Testverläufe und -ergebnisse und der damit verbundenen Vorteile hinsichtlich einer späteren bzw. nachträglichen Nutzung - beispielsweise im Rahmen einer Überprüfung der Testgüte.

Auch hinsichtlich der variablen Einsatzmöglichkeiten im Rahmen von E- oder Self-Assessments erscheinen internetgestützte Verfahren gegenüber der herkömmlichen Variante als vorteilhaft, da die Durchführung und Auswertung unabhängig von einem real anwesenden Testleiter ist. So wäre es denkbar, je nach anvisierter Einsatzstrategie ein Verfahren zur Verbesserung der Grundquote oder zur Verbesserung der Validität einzusetzen (siehe genauer Abschnitt 3.3). Ebenso ist eine Verbindung von Testergebnissen mit weiterführenden Informationen zur diagnosegeleiteten, individuellen Beratung möglich (s. Mayr, 2001; Mayr & Nieskens, 2004; Sieland & Nieskens, 2001).

Ein wesentlicher Aspekt von E-Assessment-Verfahren ist die Orts- und Zeitunabhängigkeit der Durchführung sowie der Testadministration. Dadurch erweist sich die Testdurchführung und Administration als bequemer, einfacher und flexibler gegenüber nicht internetfähigen Varianten (Barak & Buchanan, 2004). So kann der Testleiter die Administration vollkommen alleine durchführen, angefangen bei dem Einladen von Personengruppen, über die Kontrolle der Testdurchführung bis hin zur Testauswertung, dem Überarbeiten, Erweitern oder Normieren (Bartram, 2005).

Durch die internetbasierte Durchführung bietet sich gerade im Zusammenhang mit Selbsttestungen der Vorteil, Tests durchzuführen, wenn das Bedürfnis beim Ratsuchenden entsteht. So ist nicht nur eine ggf. wirksamere Beratung durch die Verbindung von spontanen Änderungsimpulsen (Bedürfnissen) mit dem Änderungsbedarf, der sich aus Testverfahren ergibt, möglich (Sieland, 2002). Barak und Buchanan (2004) zufolge beeinflusst die Bereitschaft des Probanden, sich testen zu wollen, auch die Messqualität des eingesetzten Testverfahrens positiv. Empirische Studien, die diesen Aspekt internetbasierter Verfahren untersuchen, sind allerdings bisher nicht zu finden.

Drasgow und Olson-Buchanan (1999) weisen darauf hin, dass sich computer- und internetgestütztes Testen gegenüber der herkömmlichen Variante als psychometrisch vorteilhaft erweisen kann. So ist die menschliche Fehlerquelle, wie beispielsweise durch falsches Ablesen oder Übertragen von Ergebnissen durch den Testleiter, bei einem auswertungsobjektiven, programmierten Bewertungsablauf - vorausgesetzt die Programmierung ist in sich fehlerfrei - beseitigt (vgl. Drasgow & Olson-Buchanan, 1999). Barak und Buchanan (2004) zufolge zeigen sich beispielsweise in der computergestützten Version des Self-Directed Search (SDS, Holland et al., 1994) deutlich weniger Auswertungsfehler als in der herkömmlichen Version. Auch die einfachere Handhabung hinsichtlich Überarbeitungen und Normierungen kann sich nach Barak und Buchanan (2004) positiv auf die Messqualität auswirken.

Einem Überblick von Booth (1998) zufolge stehen die meisten Probanden computergestützten Verfahren positiv gegenüber und ziehen häufig elektronische Testverfahren der Papierform vor. Booth (1998) zieht das Fazit, dass eine Reihe von Aspekten zu berücksichtigen sind, um eine möglichst hohe Akzeptanz zu erzielen. So sollten die Bewerber vor der Testbearbeitung über die Zielsetzung und die Verfahrensweise informiert werden. Die Anwendung sollte leicht zu bedienen, fehlertolerant, die Aufgaben angemessen und erwartungskonform sein und möglichst auch

Spaß machen (Booth, 1998). Diese Aspekte sind vermutlich auch bei einer internetbasierten Testdurchführung von Bedeutung.

In einigen Studien zeigt sich, dass sich auch internetbasiertes Testen bei den meisten Probanden einer hohen Akzeptanz erfreut und der Papier-Bleistift-Durchführung vorgezogen wird (vgl. Barak & L. Cohen, 2002; Lievens & Harris, 2003). Die am häufigsten genannten Gründe sind dabei die Möglichkeit, die Test orts- und zeitunabhängig durchzuführen sowie die schnelle Darbietung der Ergebnisse (vgl. N. Anderson, 2003). Allerdings ist ein positiver Effekt für das Employer-Branding (im Sinne des Personalmarketings) auf Dauer kaum mit der einfachen Übertragung von Multiple-Choice-Tests oder biografischen Fragen auf das Internet zu erwarten. Gerade die simulativen oder auch spielerischen Elemente mit aufwändigen Spielgeschichten, Animationen, Grafiken und interaktiven Objekten versprechen einen Imagegewinn und eine erhöhte Zufriedenheit der Benutzer. So wird im Rahmen der Personalwahl insbesondere simulationsorientierten E-Assessment-Verfahren eine hohe Akzeptanz bei Bewerbern mit einem verbundenen Marketing-Effekt konzidiert (Heuke, 2005; Kirbach et al., 2004).

Für E-Assessments zum Zwecke der Berufs-, Ausbildungs- und Studienberatung gelten interaktive und spielerische Ausgestaltungen ebenfalls als sinnvoll (vgl. Schulmeister, 2002). Hier wird der Schnittbereich des E-Assessments mit dem E-Learning deutlich. Auch bei letzterem ist die einfache Übertragung von textbasierten Buchinhalten auf das Internet wenig Erfolg versprechend (Schulmeister, 2002). Insbesondere der Aspekt der Interaktivität ist bei der Online-Eignungsdiagnostik wie auch beim Lernen im Internet von besonderer Bedeutung und ermöglicht neue Lern- und Testumgebungen. Es ist anzunehmen, dass mit ansteigendem Grad an Interaktivität die Akzeptanz von E-Assessment-Verfahren zunimmt.

Die Einbettung in spielerische Szenarios zieht nach Weber und Busch (2002) wiederum ein weniger sozial erwünschtes Antwortverhalten nach sich, da dadurch der Testcharakter reduziert wird. Bereits für einfache psychologische Testverfahren mit Multiple-Choice-Logik findet

Joinson (1999) in einer Studie heraus, dass bei internetgestützten Testversionen weniger sozial erwünscht geantwortet wird als bei den Papier-Bleistift-Varianten.

Als Probleme und Risiken von E-Assessment-Verfahren sind in erster Linie folgende zu nennen (s. a. Barak & Buchanan, 2004; Naglieri et al., 2004):

- technische Faktoren - wie die mangelnde Berücksichtigung der Systemvoraussetzungen, Verbindungsabbruch, Browserabhängigkeit
- Probleme, die bedingt sind durch die Mensch-Maschine-Interaktion
- psychologische Effekte, die sich aus der Testsituation ergeben, insbesondere hinsichtlich der Übertragbarkeit von Papier-Bleistift- auf Online-Versionen
- mangelnde Kontrolle und Standardisierung der Testsituation, insbesondere hinsichtlich der Manipulation der Testergebnisse und Sicherstellung der Teilnehmer-Identität

Kubinger (1993) weist darüber hinaus auf eine mögliche mangelnde Fairness computergestützter Verfahren in Bezug auf auditive Wahrnehmungstypen hin, da Instruktionen und Beispiele nur schriftlich dargeboten würden. Dieser Umstand gilt auch für die meisten E-Assessment-Verfahren. Es ist jedoch einzuwenden, dass eine standardisierte, akustische Instruktion auch bei E-Assessment-Verfahren sehr wohl technisch möglich ist. Damit ist allerdings zumeist die Problematik verbunden, dass lediglich bei einer Durchführung in kontrollierten Computertesträumen sichergestellt sein kann, dass Kopfhörer oder Lautsprecher für alle Testpersonen vorhanden und einsatzfähig sind.

Aus den möglichen Problemen und Risiken leiten sich auch Faktoren ab, die es bei der Konstruktion und beim Einsatz von E-Assessments zu beachten gilt. Dazu zählen technische Aspekte - wie die fehlerfreie Programmierung und die Transparenz hinsichtlich der Systemanforderungen. Es sollte dabei gewährleistet sein, dass nur Teilnehmer das E-Assessment bearbeiten, deren Computer und Internetanbindungen die Voraussetzungen zur fehlerfreien Durchführung erfüllen.

Um Probleme zu minimieren, die sich aus der Mensch-Maschine-Interaktion ergeben können, sind Aspekte der Benutzerfreundlichkeit (Usability) und der Einfluss der Computererfahrung bei der Gestaltung zu berücksichtigen und zu überprüfen (vgl. M. Kersting, 1999; Nielsen, 2001). Das bedeutet, dass computererfahrene Personen nicht systematisch bevorteilt werden sollten. Solche Effekte lassen sich laut einer Studie von Hergovich (1992, zitiert nach Kubinger, 1993) mittels Einweisungsseiten nivellieren. In dieser Untersuchung zeigt sich, dass selbst bei der Benutzung einer Maus zum Zeichnen am Bildschirm keine signifikanten Niveauunterschiede hinsichtlich der Computererfahrung von Personen ermittelt werden, wenn dem eigentlichen Verfahren eine Lerneinheit vorausgeht.

Als weitere Risiken von E-Assessment-Verfahren sind psychologische Effekte zu nennen, die sich aus der computergestützten Online-Testsituation ergeben können. Etliche Arbeiten haben sich daher mit der Übertragbarkeit von einfachen Papier-Bleistift-Verfahren auf das Medium Computer bzw. Internet befasst (vgl. Barak & English, 2002; Buchanan et al., 2005; Epstein, Klinkenberg, Wiley & McKinley, 2001; Hertel, Naumann, Konradt & Batinic, 2002; Potosky & Bobko, 2004). Insgesamt erweisen sich die Online-Versionen als äquivalent, in einigen Arbeiten sogar der Papier-Bleistift-Version überlegen hinsichtlich der internen Konsistenzen (Barak & L. Cohen, 2002) oder Akzeptanz (Salgado & Moscoso, 2003).

Einige Autoren werten die computergestützte Testdurchführung Bezugnehmend zur Forschung computergestützter Kommunikation gar als vorteilhaft hinsichtlich der Validität der psychologischen Bewertungen, weil ehrlicher und offener geantwortet werde (u. a. Barak & Buchanan, 2004). Laut Döring (1999) ist hinsichtlich der computervermittelten Kommunikation im Internet allerdings weder ein Identitätsverlust oder ein „Verstecken hinter Schein-Identitäten“ zu erwarten, noch dass das Internet generell Selbstexploration und Selbstentwicklung fördere und damit therapeutisch wirke. Es ist also nicht davon auszugehen, dass alleine der Umstand, dass

die Kommunikation im Internet computervermittelt verläuft, zu offeneren und ehrlicheren Antworten in internetgestützten Testverfahren führt.

Zu den wesentlichen Problemen mangelnder Kontrolle und Standardisierung der Testsituation bei einer Online-Durchführung zählt das Risiko einer möglichen Manipulation von Testergebnissen. Wenngleich eine völlige Kontrolle bei der Online-Durchführung kaum möglich sein dürfte, wird die reale Gefahr von Fälschungsversuchen jedoch laut Bartram (2005) zumeist überschätzt. Bartram und Brown (2004) legen in einer Untersuchung dar, dass lediglich zwei von 54 anonym befragten Kandidaten für einige Items weitere Personen zur Hilfe genommen haben, um einen Online-Fragebogen im Rahmen der Bewerberauswahl zu bearbeiten.

Dem Risiko lässt sich mit Hilfe einer Reihe von Strategien begegnen. So ist es beispielsweise möglich, eine Freischaltung für die E-Assessment-Verfahren erst nach einer bestandenen Überprüfung der individuellen Voraussetzungen anhand von biografischen Daten zu gestatten oder Teile der E-Assessment-Verfahren innerhalb einer nachfolgenden Präsenzveranstaltung wiederholt darzubieten, um mit Hilfe von Äquidistanz-Maßen Manipulationsversuche zu identifizieren. Segall (2001, zitiert nach Naglieri et al., 2004) stellt in einer Studie dar, dass sich Verfälschungen effektiv identifizieren lassen, indem ein Testteil mit hoch diskriminierenden Items nach der Online-Bearbeitung innerhalb eines kontrollierten Settings erneut bearbeitet wird. Auch eine zielgerechte Kommunikation über die Folgen von Täuschungsversuchen kann hilfreich sein, um Manipulationen vorzubeugen (vgl. Naglieri et al., 2004).

Bei einem Einsatz im Sinne des Modells des Blended-Assessments ist die Gefahr der erfolgreichen Testmanipulation ebenfalls vermindert. Im Rahmen des oben vorgestellten Blended-Assessments kommen verschiedene Verfahren zum Einsatz, wobei das E-Assessment typischerweise der Vorauswahl dient. Dabei steht zumeist die Strategie im Fokus, nicht die geeigneten Bewerber, sondern zuallererst die offensichtlich ungeeigneten Personen zu identifizieren. Folglich hat bei einem mehrstufigen



Personalauswahlverfahren derjenige, der unter Zuhilfenahme von Manipulationen das E-Assessment erfolgreich absolviert, lediglich eine Stufe übersprungen. Allerdings erfolgt die Einstellung typischerweise nicht aufgrund der internetgestützten Daten, sondern nach Ablauf weiterer Präsenz-Assessments, die nicht der Problematik der Online-Verfälschbarkeit unterliegen.

Zusammenfassend erscheint es trotz einiger Gefahren angesichts der Vorteile von E-Assessment-Verfahren nur eine Frage der Zeit zu sein, wann der Computer bzw. das Internet das Standardmedium für die Durchführung eignungsdiagnostischer Testverfahren darstellt (Bartram, 2000).

### **3.3 Güte und Nutzen eignungsdiagnostischer Verfahren**

#### ***3.3.1 Allgemeine Gütekriterien***

Die Qualitätsüberprüfung von E-Assessment-Verfahren erfolgt typischerweise analog zum Vorgehen bei herkömmlichen eignungsdiagnostischen Instrumenten anhand der allgemeinen Gütekriterien.<sup>13</sup> Die allgemeinen Haupt- und Nebengütekriterien (vgl. Amelang & Zielinski, 2002; American Educational Research Association, 1999; Lienert & Raatz, 1998; Testkuratorium, 1986) sind in Abbildung 13 dargestellt. Als Hauptkriterien gelten die Objektivität, die Reliabilität und die Validität. Dabei kommt der Validität die höchste Bedeutung zu. Die Objektivität und Reliabilität sind der Validität vorgeordnet. Das bedeutet, dass ohne ein Mindestmaß an Objektivität keine hinlängliche Reliabilität und ohne ein Mindestmaß an Reliabilität keine hinlängliche Validität möglich ist. Zu den Nebengütekriterien werden im Allgemeinen die Normierung, Nützlichkeit, Vergleichbarkeit, Ökonomie, Akzeptanz und Fairness gezählt. Die wechselseitige Beeinflussung der Gütekriterien beschreiben ausführlich Lienert (1969) und Amelang und Zielinski (2002).

---

<sup>13</sup> Die Gütekriterien werden häufig auch als Testgütekriterien bezeichnet. Sie gelten allerdings nicht nur für Tests im Sinne der trimodalen Eignungsdiagnostik, sondern für alle eignungsdiagnostischen Verfahren (s. Abbildung 24).

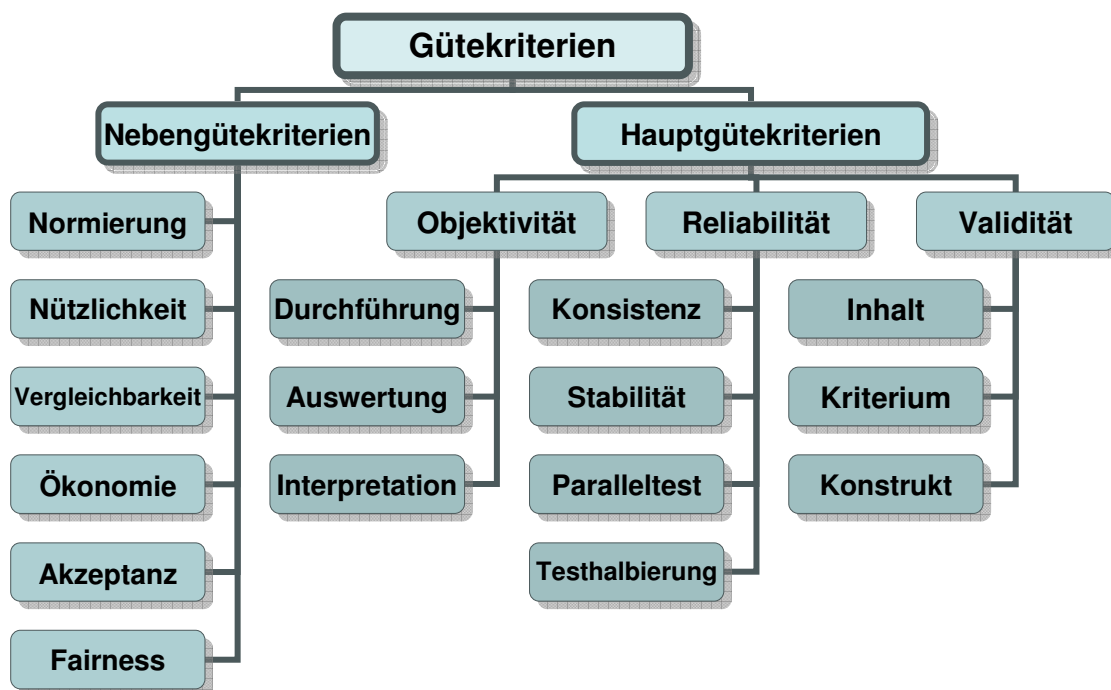


Abbildung 13: Gütekriterien eignungsdiagnostischer Verfahren

Unter Objektivität verstehen Lienert & Raatz (1998) den „Grad, mit dem die Ergebnisse eines Tests unabhängig vom Untersucher“ (S. 7) und somit auf die Testperson zurückzuführen sind. Dabei gilt es, drei Aspekte zu berücksichtigen: die Untersucherunabhängigkeit hinsichtlich der Durchführung, der Auswertung und der Interpretation (Amelang & Zielinski, 2002). Die Durchführungsobjektivität bezeichnet dabei das Ausmaß, in dem Variationen im Verhalten des Testleiters und der von diesem hergestellten Durchführungsbedingungen zu Veränderungen im Verhalten der Testpersonen führen. Die Auswertungsobjektivität beleuchtet, inwieweit die Auswertung des Tests unabhängig von personenbedingten oder apparativen Störquellen ist. Interpretationsobjektivität ist „das Ausmaß, in dem aus gleichen Scores verschiedener Probanden identische Schlüsse gezogen werden“ (Amelang & Zielinski, 2002, S. 151).

Unter Reliabilität oder Zuverlässigkeit eines Verfahrens ist das Ausmaß der Messgenauigkeit zu verstehen. Lienert & Raatz (1998) beschreiben die Reliabilität eines eignungsdiagnostischen Tests daher als „den Grad der Genauigkeit, mit dem er ein bestimmtes Persönlichkeits- oder

Verhaltensmerkmal mißt, gleichgültig, ob er dieses Merkmal auch zu messen beansprucht“ (S. 9). Dabei wird zumeist unterschieden hinsichtlich der Testwiederholung (Stabilität), der Paralleltestreliabilität, der Testhalbierung und der inneren Konsistenz.

Die Validität eines Tests sagt aus, wie gut der Test in der Lage ist, genau das zu messen, was er zu messen vorgibt (Bortz & Döring, 1995). Lienert (1969) definiert Validität folgendermaßen: „Die Validität eines Testes gibt den Grad der Genauigkeit an, mit dem dieser Test dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise, das (die) er messen soll oder zu messen vorgibt, tatsächlich mißt“ (S. 16). Es wird allgemein zwischen Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität differenziert, unter der Berücksichtigung, dass die unterschiedlichen Aspekte jedoch als Facetten einer Gesamtvalidität nicht unabhängig voneinander sind, sondern lediglich verschiedene Ansätze bzw. Herangehensweisen zur Validitätsprüfung darstellen (vgl. Amelang & Zielinski, 2002).

Die Inhaltsvalidität betrifft das Ausmaß, inwieweit die Aufgaben eines Tests das zu messende Merkmal auch wirklich bzw. hinreichend repräsentieren. Inhaltsvalidität ist dann gegeben, wenn sich aus dem gezeigten Verhalten innerhalb der Testsituation auf das zu ermittelnde Zielverhalten im Sinne eines Repräsentationsschlusses folgern lässt (Michel und Conrad, 1982). Als Augenscheinvalidität („Face Validity“) bezeichnen einige Autoren, wenn selbst Laien unmittelbar den Zusammenhang zwischen Testinhalten und zu messendem Verhalten erkennen. Nach Bortz und Döring (1995) liegt Inhaltsvalidität vor, „wenn der Inhalt der Test-Items das zu messende Konstrukt in seinen wichtigsten Aspekten erschöpfend erfaßt“ (S. 185). Die Inhaltsvalidität wird in der Regel nicht numerisch anhand eines Kennwerts, sondern aufgrund „logischer und fachlicher Überlegungen“ (Michel & Conrad, 1982, S. 52), einer umfassenden Verfahrenskonstruktion oder „durch ein Rating von Experten“ (Lienert, 1969, S. 17) bestimmt. Dabei gilt, dass die Inhaltsvalidität in erster Linie durch eine sorgfältige Verfahrensentwicklung

gewährleistet wird. Die Einschätzung von Experten oder Testpersonen unterstützt diese lediglich.

Bei der Kriteriumsvalidität wird typischerweise untersucht, in welchem Ausmaß die im Verfahren gemessenen Werte mit einem Kriterium zusammenhängen, mit dem das Verfahren aufgrund seines Messanspruchs korrelieren sollte. Dabei lassen sich die prognostische - oder prädiktive - und die konkurrente Validität unterscheiden. Erstere bezieht sich auf Kriterien, die in der Zukunft liegen, wie beispielsweise bei der Vorhersage von Berufserfolg mit Hilfe von Personalauswahlverfahren. Bei der konkurrenten Validität wird der Vergleich zwischen Testwerten und Kriterien, die zeitgleich erhoben werden, vorgenommen. Einige Autoren grenzen davon noch die retrograde Validität ab, die zeitlich zurückliegende Kriterien betrachtet. Bei der Bestimmung der Kriteriumsvalidität bedarf es somit eines Korrelationsschlusses, um von den Testergebnissen auf das interessierende Zielmerkmal oder das vorherzusagende Verhalten schließen zu können (Amelang & Zielinski, 2002).

Die Konstruktvalidität betrifft die Überprüfung der Frage, inwieweit ein Test ein Personmerkmal, das er messen soll, auch wirklich misst. Den Begriff „Konstruktvalidität“ diskutieren erstmals umfassend Cronbach und Meehl (1955). Demnach sind Konstrukte nicht beobachtbare Eigenschaften oder Merkmale von Personen (latente Variablen). Sie stehen zu anderen latenten aber auch zu beobachtbaren Variablen in Beziehung und werden durch dieses „nomologische Netzwerk“ implizit definiert. Verfahren zur Überprüfung der Konstruktvalidität versuchen, das Netzwerk aufzuklären und konstruktnahe (konvergente) und konstruktferne (diskriminante) Beziehungen zu untersuchen. „Ein Test ist konstruktvalid, wenn aus dem zu messenden Zielkonstrukt Hypothesen ableitbar sind, die anhand der Testwerte bestätigt werden können“ (Bortz und Döring, 1995, S. 186).

Neben diesen Hauptgütekriterien sind eine Reihe von Nebengütekriterien für die Qualität von eignungsdiagnostischen Verfahren von Bedeutung.

Die Normierung dient als Bezugssystem für die Einordnung der individuellen Testergebnisse. Angaben erfolgen zumeist hinsichtlich einer Vergleichsgruppe.

Ein Verfahren gilt dann als nützlich, wenn es ein Merkmal misst, für dessen Untersuchung ein praktisches Bedürfnis besteht. Die Vergleichbarkeit behandelt den Umstand, inwieweit für einen Test eine oder mehrere Parallelformen existieren. Ein ökonomischer Test ist innerhalb kurzer Zeit durchzuführen, verbraucht wenig Material, ist einfach zu handhaben, als Gruppentest durchführbar und schnell und bequem auszuwerten. Die Fairness bezieht sich auf den Aspekt, ob konstrukt fremde Variablen wie Alter, Geschlecht, Computererfahrung und ethnische Gruppierung das Abschneiden in einem Test beeinflussen. Die Akzeptanz eines Tests beschreibt das Ausmaß, inwieweit die Testpersonen mit dem Verfahren zufrieden sind und inwieweit sie die Durchführung, Auswertung und Interpretation akzeptieren. Dabei erweisen sich die Aspekte Akzeptanz und Fairness nicht immer als trennscharf und werden häufig gemeinsam betrachtet. Die thematisierten Anforderungen an die Qualität psychologischer Tests sind vielfältig und in der Praxis kaum alle gleichzeitig zu erreichen, da einige Zielgrößen nur begrenzt miteinander vereinbar sind. Der Katalog der Kriterien ist nicht verbindlich definiert. Gleichwohl beinhaltet die hier vorgestellte Auflistung die zumeist genannten Gütekriterien.

Bei der Beurteilung eines eignungsdiagnostischen Verfahrens ist in Abhängigkeit von den vorliegenden Bedingungen und Zielsetzungen teilweise auch eine Gewichtung der Gütekriterien vorzunehmen. So sind hinsichtlich der Bewertung der Qualität computer- oder internetbasierter Verfahren insbesondere die Aspekte der Akzeptanz und Fairness neben weiteren rein technischen Qualitätskriterien bedeutsam, um eventuelle Effekte, die sich ggf. aus der computergestützten Durchführung ergeben, zu kontrollieren.

Das aus dem Forschungsbereich der Human Computer Interaction kommende Konzept der „Usability“ oder „Benutzerfreundlichkeit“ ist in diesem Zusammenhang hilfreich. Es lässt sich aus testpsychologischer Sicht

den Gütekriterien der Fairness und Akzeptanz zuordnen und befasst sich mit der Kommunikation zwischen Mensch und Maschine, die für internet- und computergestützte Verfahren konstituierend ist. „Usability“ kann mit "Benutzerfreundlichkeit", "Benutzbarkeit" oder „Gebrauchstauglichkeit“ übersetzt werden (Langenscheidt, 2005). Die Benutzerfreundlichkeit steht dabei für die Bewertung eines interaktiven Systems aus Sicht des Teilnehmers. Interaktive Systeme sind Systeme, die eine Benutzerschnittstelle aufweisen und auf Benutzeraktionen reagieren. Dazu gehören einfache Maschinen ebenso wie komplexe Softwareanwendungen oder auch computer- und internetgestützte Testverfahren. Es gibt eine Reihe von Definitionen für Benutzerfreundlichkeit, die meist aus dem angloamerikanischen Raum kommen. Eine viel zitierte Beschreibung liefert die ISO 9241-11 (1995). Danach wird Usability als "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (S. 2) verstanden. Daten zur Überprüfung der Usability eines Programms können aus verschiedenen Quellen gewonnen werden, wobei insbesondere folgende drei zu nennen sind:

- Trackingverfahren und Analyse der Protokolldateien (Logfiles)
- Expertenbefragung
- Probandenbefragung

Mit Hilfe von Trackingverfahren - zumeist basierend auf Zählpixeln (kleine Bilddatei) - lassen sich relevante Informationen über die IT-Umgebung des Nutzers und das genaue Benutzerverhalten registrieren. So ist es möglich, Abbrüche, typische Verläufe und sämtliche Klickpfade exakt zu identifizieren. Ein Logfile hingegen ist eine automatisch generierte Protokolldatei, die alle Anfragen an einen Webserver und deren Ergebnisse aufzeichnet. In einem Logfile können zu jeder Anfrage die IP-Nummer des zugreifenden Rechners, Datum und Uhrzeit des Zugriffs, die angeforderten Dateien und das Ergebnis der Übertragung erfasst werden. Auch die zuvor besuchte Adresse sowie Betriebssystem und Browser des Benutzers lassen sich mit Hilfe der Protokolldatei ermitteln. Trackingverfahren bzw. Logfiles eignen sich somit

für eine Bewertung und Optimierung einer webbasierten Anwendung in Bezug auf Effizienz und Nutzerverhalten und finden beispielsweise auch bei der Beurteilung von Unternehmenswebseiten Anwendung (Nielsen, 1993).

Bei der Darbietung im Internet oder Intranet sind neben den psychologischen Gesichtspunkten einer Verfahrenskonstruktion auch eine Reihe von technischen Mindestanforderungen zu gewährleisten, die zumeist von Experten bewertet werden. Dazu gehören unter anderem:

- Sicherstellung der Lauffähigkeit des Programms auf allen Rechnern,
- Browserkompatibilität,
- korrekte Verlinkung,
- Optimierung von Ladezeiten
- Eindeutigkeit der Sicherung der Daten

Diese Aspekte sind sorgfältig zu prüfen und bereits früh zu berücksichtigen, um die Möglichkeit von Datenverlusten zu minimieren (vgl. die Untersuchungen von Süß, 1996, bei der Datensätze aufgrund technischer Pannen verloren gingen).

Eine Befragung von Experten ist in verschiedener Hinsicht denkbar: Zum einen können Personen befragt werden, die eine Expertise auf dem Gebiet der benutzerfreundlichen Webgestaltung oder der Gestaltung von E-Assessment-Verfahren erworben haben. Zum anderen lassen sich Einschätzungen zur inhaltlichen Relevanz und Realitätsnähe von Experten gewinnen, die sich in dem Bereich sehr gut auskennen, der den Inhalt einer Anwendung ausmacht. Letztere Expertenbefragung dient in erster Linie der Inhaltsvalidierung, kann allerdings auch Auswirkungen auf die Akzeptanz und Fairness haben.

Die wirklichen Anwender eines Systems stellen die eigentliche Zielgruppe der Überprüfung der Benutzerfreundlichkeit dar. Es erscheint daher als sinnvoll, die Probanden bereits in der Entwicklungsphase mit einzubeziehen, um eine möglichst optimale Bedienbarkeit und Benutzerfreundlichkeit zu

erzielen. Die Nähe des Konzepts der „Benutzerfreundlichkeit“ zu psychologischen Kriterien wird u. a. bei Dumas & Redish (1994) deutlich. Die Autoren konstatieren einem System oder einem Produkt eine hohe Benutzerfreundlichkeit, wenn

- es den Benutzenden im Fokus hat,
- der Benutzende einen Nutzen durch das Produkt erhält,
- der Benutzende die Benutzerfreundlichkeit eines Produkts bewertet.

Die Überprüfung der Benutzerfreundlichkeit steht somit in enger Verwandtschaft mit dem Konzept der „sozialen Validität“, das sich den Nebengütekriterien Akzeptanz und Fairness zuordnen lässt. Das von Schuler und Stehle (1983) eingeführte und von Schuler (1990) weiterentwickelte Konzept der „sozialen Validität“ steht als Sammelbegriff für das, „was die eignungsdiagnostische Situation zu einer akzeptablen sozialen Situation macht“ (Schuler & Stehle, 1983, S. 35). Darunter sind im Wesentlichen vier Aspekte zu verstehen, die bei der Konstruktion und beim Einsatz eignungsdiagnostischer Verfahren zu berücksichtigen sind:

1. Mitteilung relevanter Informationen über wichtige Charakteristika von Arbeitsplatz und Organisation sowie über erfolgskritische Anforderungen
2. Partizipation an der Entwicklung und Anwendung eignungsdiagnostischer Instrumente
3. Transparenz des Verfahrens und der Schlussfolgerungen
4. Kommunikation bzw. Feedback in rücksichtsvoller, verständlicher Form  
(vgl. Schuler, 1990)

Dass die Akzeptanz und wahrgenommene Fairness sowohl für die Benutzer als auch für die Testanwender (Unternehmen, Hochschulen etc.) von entscheidender Bedeutung ist, zeigt u. a. M. Kerstings (1998, S. 62) Darstellung möglicher Folgen auf. Demnach kann sich die Akzeptanz der Personalauswahlverfahren auf das allgemeine Bewerbungsverhalten, auf die



Entscheidung der ausgewählten Personen für oder gegen die Organisation sowie auf das zukünftige Erleben und Verhalten der ausgewählten Personen in der Ausbildung und/oder im Beruf ebenso auswirken wie auf das zukünftige Verhalten der abgelehnten Personen, auf das allgemeine Image der Organisation und auf die Anzahl von Rechtsstreitigkeiten infolge von Auswahlverfahren.

Im Rahmen dieser Arbeit stehen auf Grund des gewählten Hybridansatzes insbesondere die Aspekte der Konstrukt- und Inhaltsvalidierung sowie der Objektivität, Reliabilität, Akzeptanz und Fairness im Vordergrund (s. a. Kapitel 5). Dabei stellt sich gerade vor dem Hintergrund der Probleme bisheriger Verfahren (s. Abschnitt 3.4) die Frage, wie sich bereits bei der Verfahrensentwicklung die Problembereiche berücksichtigen und eine möglichst optimale Testgüte gewährleisten lassen.

Aufgrund der zunehmenden Bedeutung internetgestützter Testungen wurden in jüngster Vergangenheit Anstrengungen unternommen, Qualitätsstandards für den Einsatz von Online-Testverfahren auszuarbeiten. Die vom amerikanischen Psychologenverband (APA) eingesetzte „Task Force“ und die „International Commission of Test (ITC)“ haben kürzlich wesentliche Aspekte verfasst, die beim internetgestützten Einsatz zu berücksichtigen sind (International Test Commission, 2006; Naglieri et al., 2004). Auf diese und weitere wichtige Aspekte im Zusammenhang mit der Verfahrensentwicklung fokussiert das Kapitel 4.

### ***3.3.2 Nutzen eignungsdiagnostischer Verfahren***

Zur Bestimmung des Nutzens eignungsdiagnostischer Verfahren wurden einige Modelle entwickelt. Eines der bekanntesten ist das von Taylor und Russell (1939), wonach für die Praxis eignungsdiagnostischer Auswahlentscheidungen die Wahrscheinlichkeit entscheidend ist, mit der ein als geeignet identifizierter Bewerber tatsächlich erfolgreich sein wird. Dieser

Zusammenhang - als Erfolgswahrscheinlichkeit oder Trefferquote bezeichnet  
- wird von folgenden drei Faktoren beeinflusst:

- Anteil der Geeigneten unter den Bewerbern (Basisrate oder Grundquote)
- Anteil der Personen, die eingestellt werden im Verhältnis zur gesamten Bewerbergruppe (Selektionsquote)
- Validitätskoeffizient (Prognosegüte) der eignungsdiagnostischen Verfahren

Mit Hilfe des Taylor-Russell-Modells<sup>14</sup> lässt sich zeigen, dass die Qualität der Personauswahl nicht nur durch die Validität eignungsdiagnostischer Instrumente bedingt ist. So führen beispielsweise auch Maßnahmen, die den Anteil der grundsätzlich Geeigneten in der Bewerbergruppe erhöhen, zu einer Verbesserung der gesamten Auswahl. Self-Assessments oder ausführliche Beschreibungen der Anforderungen des Berufs können nützlich sein, um die Selbstausswahl unter den potenziellen Bewerbern anzuregen und somit die Basisrate günstig zu beeinflussen (vgl. Diercks, Kupka & Jägeler, in Druck; Kupka & Diercks, in Druck).

Die Validität eines Verfahrens ist umso bedeutender, je niedriger die Grundquote und je größer die Selektionsquote ist. Umgekehrt hat der Validitätskoeffizient umso weniger Einfluss auf die Trefferquote, je größer sich die Grundquote und je niedriger sich die Selektionsquote darstellen.

Gerade im Zusammenhang mit Auswahltests, die zusätzlich zu bereits nutzbaren Prädiktoren eingeführt werden sollen - wie beispielsweise im Rahmen der Studierendenauswahl bei einer möglichen Einführung von Studierfähigkeitstests ergänzend zur Berücksichtigung der Abiturnote - wird deutlich, dass nicht nur die Validität der eignungsdiagnostischen Verfahren die Erfolgswahrscheinlichkeit verbessern kann. Unter gegebenen Bedingungen ist der zusätzliche Nutzen, der sich durch zielführendes

---

<sup>14</sup> Im Internet steht unter <http://www.hr-guide.com/data/G906.htm> eine Software zur Verfügung, mit deren Hilfe sich die Erfolgswahrscheinlichkeiten nach den Annahmen von Taylor und Russell (1939) berechnen lassen.

Personalmarketing oder Selbstauswahl-Maßnahmen (beispielsweise mittels Self-Assessment) für die Qualität der gesamten Auswahl ergibt, ggf. deutlich größer (vgl. Kupka & Diercks, in Druck).

Dies ist insofern von praktischer Bedeutung, als dass die Art des Verfahrenseinsatzes dabei möglicherweise deutlich unterschiedliche Effekte nach sich zieht. So kann dasselbe Verfahren statt zur Fremdauswahl im Rahmen eines E-Assessments ggf. zur Anregung zur Selbstauswahl im Rahmen eines Self-Assessments eingesetzt werden. Die Testergebnisse würden dabei nicht einer Institution - wie der Hochschule oder einem Unternehmen zur Verfügung gestellt - sondern dem Testkandidaten, der dies als Unterstützung zur Berufs- oder Ausbildungsentscheidung nutzen kann (s. ausführlicher zu dieser Problematik Kupka, in Vorbereitung).

Zur Qualitätssicherung des praktischen Einsatzes berufseignungsdiagnostischer Verfahren ist in jüngerer Vergangenheit die DIN-Norm 33430 entstanden (Normausschuss Gebrauchstauglichkeit und Dienstleistungen, 2002). Die DIN 33430 beschreibt Qualitätsanforderungen an Verfahren und Zielgrößen bei deren Einsatz innerhalb berufsbezogener Eignungsbeurteilungen. Dabei werden alle Phasen der Personalentscheidung fokussiert - von der Anforderungsanalyse über die verwendeten Verfahren bis zur abschließenden Beurteilung. Auch die Gütekriterien finden in der Norm Berücksichtigung. Die DIN-Norm 33430 formuliert Kriterien, um Verfahren vor dem eigentlichen Einsatz bereits auf deren Qualität hin zu prüfen und die Qualität von berufsbezogenen Eignungsbeurteilungen insgesamt zu gewährleisten (Hornke & M. Kersting, 2004; Westhoff et al., 2005). Die Qualitätssicherung eines praktischen Einsatzes der im Rahmen dieser Arbeit entwickelten Verfahren ist der Prüfung der Gütekriterien allerdings nachgeordnet und steht nicht im Fokus dieser Arbeit. Daher soll hier nicht weiter auf die DIN-Norm 33430 eingegangen werden, Näheres über die Norm findet sich u. a. bei Westhoff et al. (2005).

### **3.4 Forschungsstand der Simulationsverfahren zum Planen und Problemlösen**

Die Prinzipien unterschiedlicher Ansätze zur Erfassung eignungsrelevanter Personmerkmale wurden im ersten Abschnitt dieses Kapitels verdeutlicht. In der simulativen Eignungsdiagnostik haben sich zur Messung der Konstrukte „Planungsfähigkeit“ und „Problemlösefähigkeit“ insbesondere zwei Verfahrensarten hervorgetan:

1. Postkorbverfahren
2. computergestützte Szenarios (komplexe Computersimulationen)

Ergebnisse zu den Gütekriterien der beiden Verfahrensgruppen werden nachfolgend thematisiert.

#### **3.4.1 Postkorbverfahren**

Beim klassischen Postkorbverfahren werden von der Testperson in Einzelarbeit verschiedene schriftliche Materialien bearbeitet, die einen typischen Postkorb einer Führungskraft abbilden sollen (Höft & U. Funke, 2001). Erste Ansätze dieser Methode finden sich bereits im Offiziersauswahlverfahren der deutschen Wehrmacht, das Mitte der 20er Jahre entwickelt und bis zum Ende des 2. Weltkrieges als Regelform der Auswahl zum Einsatz kam (Domsch & Jochum, 1989). Im Rahmen der so genannten „charakterologischen Komplexprüfung“ wurden potenzielle Offiziersanwärter mit Hilfe verschiedenster Verfahren untersucht (Simoneit, 1933). Eines dieser Verfahren war die "Befehlsreihe", die als Vorläufer der heutigen Postkorbverfahren gilt (Domsch & Jochum, 1989).

Frederiksen, Saunders und Wand (1957) führten im englischsprachigen Raum das erste standardisierte Postkorbverfahren („In-Basket-Test“) zur Erfassung von Verhaltensdimensionen wie dem Organisieren und Ordnen von Abläufen ein. Eine gewisse Relevanz für die heutige Praxis der Verfahren hat der Postkorb von Jeserich (1981) erlangt, der - im Gegensatz

zu den meisten anderen - komplett inhaltlich publiziert wurde. Er dient als Grundlage für einige aktuelle Instrumente - wie beispielsweise für das Dispositionsproblem „Plan-A-Day“ von J. Funke und Krüger (1993). In Jeserichs Verfahren sollen Testpersonen in der Rolle von Hans Schnell insgesamt fünfzehn berufliche und private Vorgänge unter Zeitdruck erledigen. Darunter finden sich verschiedene Aufgaben, die in Form von Briefen, Notizen oder Rundschreiben den Kandidaten vorgelegt werden.

Postkorbverfahren werden zumeist als Arbeitsproben verstanden (vgl. Dommel, 1995; Fritz und J. Funke, 1995; J. Funke, 1993; Musch & Lieberei, 1997). Sie gehören zu den am meisten verwendeten Methoden in Assessment-Centern und zählen zu den Standardverfahren der Eignungsdiagnostik. Laut Thornton und Byham (1982) finden sie in 95 Prozent aller Assessment-Center Anwendung. Postkorbverfahren gelten als standardisierte Verfahren zur Erhebung inhaltsvalider Stichproben des eignungsrelevanten Verhaltens (J. Funke, 1993; Jeserich, 1981). Mit der steigenden Verbreitung von Computern in Beruf und Freizeit und aufgrund möglicher Vorteile der computergestützten Diagnostik (s. Abschnitt 3.2; Kubinger, 1993) haben auch rechnergestützte Verfahren vermehrt Einzug in die Eignungsdiagnostik gefunden. So wurden in den letzten Jahren eine Reihe von computergestützten Postkorbverfahren publiziert, beispielsweise „Mailbox ´90“ (Roest, Scherzer, Urban, Gangl & Brandstätter, 1989), „PC-Office“ (Fennekels, 1995), „PC-Postkorb Seeblick“ (Scharley & Partner, 1991), „PC-Postkorb“ (J. Funke & Rasche, 1988), „Plan-A-Day“ (J. Funke & Krüger, 1993), die „Bonner-Postkorb-Module“ (Musch & Lieberei, 2001) und die „Managementarbeitsprobe“ (Etzel & Küppers, 2000). Typischerweise lassen sich folgende Personmerkmale mit Hilfe der Verfahren erfassen:

- Analysieren
- Planen und Organisieren
- Entscheiden
- Problemlösen

Postkorbverfahren sind zumeist situativ angelegt (Fritz & J. Funke, 1995). Sie unterscheiden sich daher hinsichtlich Inhalt, Präsentationsbedingung und

Auswertung. Mangels Einheitlichkeit der Postkorb-Varianten ist es bisher kaum möglich, generelle Gütebefunde sowohl für die klassische als auch für die computergestützte Variante zu berichten (vgl. u. a. Dommel, 1995; Höft & U. Funke, 2001; Schippmann, Prien & Katz, 1990). Im Folgenden werden daher zentrale empirische Ergebnisse einzelner Arbeiten zur Testgüte von Postkorbverfahren referiert. Dabei steht - aufgrund der Bedeutung für die vorliegende Arbeit - der Stand der Forschung zu computergestützten Postkorbverfahren im Mittelpunkt der Betrachtung. Da hierzu allerdings erst wenige Studien vorliegen, werden im Folgenden auch wichtige Ergebnisse zu klassischen Postkorbverfahren referiert. Überblicke über Studien zu Gütekriterien bieten Schippmann et al. (1990) hinsichtlich klassischer Postkorbverfahren sowie Riediger und Rolfs (1998) für die drei computergestützten Verfahren „Mailbox '90“, „PC-Office“ und „PC-Postkorb Seeblick“.

#### *3.4.1.1 Objektivität und Reliabilität*

In Bezug auf klassische Postkorbverfahren kritisieren Musch und Lieberei (1997) und auch Fritz und J. Funke (1995) die häufig anzutreffende Praxis, Verfahren unstandardisiert und nicht objektiv einzusetzen. Dabei würden Testkandidaten problematischerweise im Rahmen einer Präsentation ihrer Ergebnisse bewertet und nicht bei der eigentlichen Postkorbbearbeitung. Die Bewertungen und Auswertungen basierten zumeist auf freier Eindrucksbildung der Beobachter hinsichtlich vorgegebener Eigenschaftsdimensionen (Musch & Lieberei, 1997). Musch und Lieberei (1997) kritisieren die Vorgehensweise, Postkorbleistungen im Zuge einer globalen Eindrucksbildung auf vorgegebenen Eigenschaftsdimensionen zu bewerten. Dies könne zu Beobachterfehlern und Urteilsverzerrungen führen. Die Möglichkeit einer Standardisierung hinsichtlich Beobachtungs- und Auswertungsdimensionen werde so bei vielen Verfahren nicht genutzt, was sich negativ auf die Objektivität auswirke. Die Feststellung und Bewertung der Güte der Aufgabenbearbeitung sollte daher hinsichtlich a priori festzulegender Kriterien bzw. Verhaltensweisen erfolgen (Musch &

Lieberei, 1997). Nur so könnten diese Verhaltensweisen als "sample" von Handlungen im Sinne von Wernimont und Campbell (s. o., 1968) verstanden werden, die durch eine gewisse, dahinter liegende Fähigkeit bestimmt werden (Musch und Lieberei, 1997).

Gleiches gilt im Grunde auch für computergestützte Postkorbverfahren. Rechnergestützte Postkorbverfahren bieten auf Grund der programmgesteuerten Durchführung und Auswertung grundsätzlich die Möglichkeit des objektiven und auch standardisierten Einsatzes wie beispielsweise bei „Plan-A-Day“, den „Bonner-Postkorb-Modulen“ oder der „Managementarbeitsprobe“ (vgl. Ueckert, 1995).

Allerdings erscheinen die bisherigen computergestützten Postkorbverfahren diesbezüglich zumeist verbesserungswürdig: Riediger & Rolfs (1998) ziehen in ihrem Review das Fazit, dass die genaue Auswertung und Interpretation der erbrachten Leistungen in allen drei beschriebenen Computer-Postkörben (s. o.) entweder gar nicht dargelegt oder nur wenig überzeugend seien.

J. Funke (1993) hat in diesem Zusammenhang auch die teilweise beliebig wirkenden Indikatoren und Operationalisierungen kritisiert: „Warum sollte die Häufigkeit, mit der eine Umblättern-Funktion genutzt wird, etwas mit Aktivität im Sinne einer Führungseigenschaft zu tun haben?“ (S. 124).

Nicht zuletzt angesichts der teils mangelnden Objektivität ist die Streubreite der beobachteten Interrater-Reliabilitäten bei klassischen Postkorbverfahren kaum verwunderlich. Schippmann et al. (1990) berichten von Werten zwischen  $r = -.20$  bis  $.97$  für klassische Postkorbverfahren. Es werden dabei allerdings höhere Beobachter-Übereinstimmungen für die Gesamtbewertungen der Übungen („overall ratings“) berichtet als für die Bewertungen einzelner Sub-Dimensionen. Ein Beobachtertraining kann außerdem den Grad der Übereinstimmung erhöhen. Dies zeigt eine Studie von Richards und Jaffee (1972). Die Autoren haben dabei untrainierte und trainierte Beobachter hinsichtlich der Beurteiler-Übereinstimmung verglichen. Bei untrainierten Beobachtern zeigen sich Interrater-Reliabilitätskoeffizienten von  $r = .46$  und  $r = .58$  für die Bewertungen auf zwei Dimensionen, während bei trainierten Beobachtern Werte von  $r = .78$  und  $r = .90$  ermittelt werden.

Angaben zur Reliabilität der Skalen computergestützter Postkorbverfahren liegen laut Riediger und Rolfs (1998) für die drei vorgestellten Instrumente im niedrigen bis mittleren Bereich. Dabei werden für das Verfahren „Mailbox ´90“ Cronbachs-Alpha-Werte zwischen  $\alpha = .50$  bis  $.79$  und für den Postkorb „PC-Office“ zwischen  $\alpha = .62$  bis  $.83$  genannt (Fiedler, 2002). Die Reliabilitätsschätzungen durch Testhalbierung (Split-Half) liegen für den Postkorb „PC-Office“ zwischen  $\alpha = .64$  und  $.84$ . Für das computergestützte Verfahren zur Erfassung von Planungsfähigkeit „Plan-A-Day“ (PAD) von J. Funke und Krüger (1993) wird von Halbierungsreliabilitäten von  $r_{tt} = .23$  und  $.55$  berichtet (J. Funke & Krüger, 1995). Die Autoren werten dies als nicht zufrieden stellend. Sie weisen allerdings darauf hin, dass die im Rahmen einer Diplomarbeit von Evers (1995) ermittelten Werte durch die nicht streng standardisierten Durchführungsbedingungen sowie durch die unsystematische Auswahl der Versuchspersonen bedingt sein könnten.

Für zwei neuere Verfahren werden höhere Reliabilitäten referiert, zum einen für die computergestützte „Managementarbeitsprobe“ von Etzel und Küppers (2000) und zum anderen für die „Bonner-Postkorb-Module“ (BPM) von Musch und Lieberei (2001). Musch und Lieberei (2001) haben versucht, die teststatistischen Probleme vieler Postkorbverfahren zu berücksichtigen, in dem sie vier auswertungsobjektive Postkörbe konstruierten, die alle auf dem gleichen Konstruktionsprinzip beruhen. Die Bewertung der Teilnehmerleistung im Postkorb ist dabei standardisiert aufgrund von Musterlösungen möglich. Angaben zur Reliabilität liegen für zwei Verfahren vor. Für den Postkorb „CaterTrans“ ermitteln die Autoren Cronbachs-Alpha-Werte in Höhe von  $\alpha = .91$  ( $N = 47$ ) und für den Postkorb „Chronos“ in Höhe von  $\alpha = .82$  ( $N = 54$ ). Die Split-Halft-Reliabilität wird lediglich für den Postkorb Chronos berichtet, diese liegt in Höhe von  $r = .83$ . Die „Managementarbeitsprobe“ (MAP) von Etzel und Küppers (2000) ist ein computergestütztes Verfahren. Der Subtest zur Erfassung von Planungsfähigkeit ähnelt computergestützten Postkorbverfahren. Dabei müssen die Testpersonen berufliche Termine unter Zeitdruck koordinieren. Die Autoren berichten von guten Reliabilitäten in den



verschiedenen Untersuchungen mit insgesamt annähernd eintausend Testpersonen zwischen  $\alpha = .78$  und  $.94$  (Cronbachs-Alpha).

#### *3.4.1.2 Validität und Nebengütekriterien*

Ein Grund für den häufigen Einsatz von Postkorbverfahren in Assessment-Centern liegt wahrscheinlich in der hohen Augenscheinvalidität. Die Relevanz der dargebotenen Aufgaben für den späteren Arbeitsplatz erscheint offensichtlich (Höft & U. Funke, 2001). Es ist jedoch anzumerken, dass der bloße augenscheinliche Bezug zu berufsrelevanten Aufgaben nicht als Beleg für die Inhaltsvalidität eines Verfahrens ausreicht (Schippmann et al., 1990). So kritisieren beispielsweise Höft und U. Funke (2001) die Verwendung von Original-Briefpapier, das Anbringen von Falzkanten etc. als Vorspielung von Authentizität. Riediger und Rolfs (1998) weisen darauf hin, dass die Testentwickler der drei bewerteten computergestützten Postkorbverfahren „mehr oder weniger ‘stillschweigend’ die Inhaltsvalidität ihrer Übungen“ (S. 49) annehmen. Nachweise der Inhaltsvalidität von computergestützten Postkorbverfahren lassen sich nur schwer finden. Dabei ist auf jeden Fall die Beschreibung der Verfahrensentwicklung hinsichtlich der anvisierten Sicherung der Inhaltsvalidität zu fordern (Schuler, 2001; Wottawa, 1995).

Einige Kriteriumsvaliditätsuntersuchungen aus dem englischen Sprachraum zeigen trotz der unterschiedlichen Konstruktionen von Postkorbverfahren und teilweise nicht vergleichbaren Anforderungsdimensionen zumeist positive Zusammenhänge der Postkorbleistungen mit externen Kriterien der Berufs- oder Ausbildungsleistung auf. Die größte Zahl der 22 von Schippmann et al. (1990) zitierten Studien zeugen davon, dass klassische Postkorbverfahren relativ hoch mit Kriterien wie Verkaufszahlen, Gehaltserhöhungen, Examensnoten und Vorgesetztenbeurteilungen korrelieren. Wenngleich in einigen Studien auch negative Werte bis zu  $r = -.45$  ermittelt wurden, so hängt die große Mehrheit der erfassten beruflichen Leistungskriterien substantiell positiv mit den Leistungen in den Postkorbverfahren zusammen ( $r_{\max} = .76$ ). Allerdings sind nur wenige der vorgestellten Untersuchungen

prognostisch angelegt, die meisten liefern Angaben zur konkurrenten Validität.

Auch für computergestützte Postkorbverfahren kann von ersten Hinweisen zur Kriteriumsvalidität berichtet werden. Für „Plan-A-Day“ ermittelte J. Funke und Krüger (1995) einen bedeutsamen Mittelwertsunterschied zwischen Führungskräften und Kontrollpersonen bei jedoch recht kleinen Stichproben ( $N = 22$  und  $N = 16$ ). Scharley und Partner (1991) berichten für ihren „PC-Postkorb Seeblick“ von einer signifikanten Korrelation in Höhe von  $r = .33$  zwischen der Postkorbleistung und einer Vorgesetztenbeurteilung ( $N = 54$ ). Allerdings sind die Angaben und Ergebnisse insgesamt sehr spärlich, so dass Riediger und Rolfs (1998) schlussfolgern, dass ein endgültiger Nachweis der psychometrischen Güte und prognostischen Aussagekraft der von ihnen betrachteten computergestützten Verfahren noch ausstehe.

Studien zur Überprüfung der Konstruktvalidität von computergestützten Postkorbverfahren mit Hilfe von psychologischen Testverfahren finden sich nur wenige. Die meisten Studien haben zur Prüfung der Konstruktvalidität von computergestützten Postkorbverfahren lediglich Zusammenhänge zu Assessment-Center-Übungen hergestellt. Dabei zeigen sich uneinheitliche Befunde, wenngleich sich zumindest in einer Reihe von Untersuchungen mittlere Korrelationen zu verschiedensten Assessment-Center-Übungen und zum Gesamtergebnis ergeben (z. B. Musch & Lieberei, 2001; Scharley & Partner, 1991; Thornton & Byham, 1982). Die Uneinheitlichkeit wird auch hinsichtlich der konstruktionsgleichen Bonner-Postkorb-Module „CaterTrans“ und „Chronos“ (Musch & Lieberei, 2001) deutlich. Während für das Modul „Chronos“ durchgehend positive mittelstarke Zusammenhänge zu Leistungen in AC-Übungen ermittelt wurden, liefert eine andere Studie negative Zusammenhänge hinsichtlich des Moduls „CaterTrans“ (s. Tabelle 10). Der Postkorb „CaterTrans“ erweist sich nach Einschätzung der Autoren als Suppressorvariable mit negativem Betagewicht, die durch Bindung von Varianzanteilen in den anderen AC-Übungen (Gruppendiskussion, Mitarbeitergespräch, Präsentation und schriftliche Fallstudie) zur Erhöhung der konkurrenten Kriteriumsvalidität beiträgt. Inwieweit die Schlussfolgerung

der Autoren, dass der Postkorb "CaterTrans" - im Gegensatz zum Postkorb "Chronos" - spezifische, von anderen AC-Verfahren nicht abgedeckte Varianzanteile erfasse, treffend ist, bleibt zu bezweifeln angesichts der teilweise auch recht starken, negativen Korrelationen zwischen dem Postkorb „CaterTrans“ und der AC-Übung „Präsentation“ in Höhe von  $r = -.27$ .

**Tabelle 10: Zusammenhänge zwischen zwei „Bonner-Postkorb-Modulen“ (nach Musch & Lieberei, 1997)**

	Gruppen- diskussion	Mitarbeiter- gespräch	Präsentation	Fallstudie
Postkorb „CaterTrans“ N = 39	-.19	-.07	-.27	-.22
Postkorb „Chronos“ N = 54	.43	.30	.51	.31

J. Funke und Krüger (1995) untersuchen in einer Studie, inwieweit die im computergestützten Verfahren „Plan-A-Day“ erfasste Planungsfähigkeit mit anderen Bewertungen innerhalb eines Assessment-Centers zusammenhängt. Dabei können die Autoren keinerlei signifikante Korrelationen zwischen der im „Plan-A-Day“ (PAD) erhobenen Planungsfähigkeit und anderen im Assessment-Center erhobenen Variablen finden. Eine Faktorenanalyse der gesamten Bewertungen des Assessment-Centers erbringt, dass einer von zwei Faktoren mit Eigenwert größer als 1 durch den PAD markiert ist. Die durch diesen Faktor aufgeklärte Varianz liegt bei dreizehn Prozent. Ob dieser Befund als Hinweis auf Konstruktvalidität im Sinne diskriminanter Validität zu verstehen ist, bleiben die Autoren allerdings schuldig.

Einen Großteil ihrer prädiktiven Validität - so vermuten Goldstein, Yusko, Bravermann, Smith und Chung (1998) - verdanken Postkorbverfahren ihrer Nähe zu kognitiven Fähigkeitstests. Nicht zuletzt Schmidt und Hunter (1998, 2000) können die Bedeutung von kognitiven Fähigkeiten als Prädiktoren für berufliche Leistungen aufzeigen. In der Untersuchung von Goldstein et al.

(1998) finden sich für den eingesetzten Offline-Postkorb die höchsten Korrelationen zu einem allgemeinen Intelligenztest ( $r = .29$ ). Alle anderen Assessment-Center-Übungen wie Rollenspiel, Gruppendiskussion und Präsentationsübung hängen niedriger zusammen ( $N = 85$ ). Dieses Ergebnis ist allerdings nicht besonders überraschend, da Gruppendiskussionen und Rollenspiele zumeist eher soziale und kommunikative Fähigkeiten erfassen sollen. Nach Goldstein et al. (1998) können Postkorbverfahren, die berufliche Anforderungen simulieren, möglicherweise als bereichsspezifische Intelligenztests verstanden werden.

Von einem weiteren Ergebnis, das diese These unterstreichen könnte, berichten Scharley und Partner (1991). Für ihren „PC-Postkorb Seeblick“ ermitteln sie den Korrelationswert ( $r = .29$ ) zwischen Postkorbergebnis und Intelligenztestresultat bei einer Untersuchung an 117 Mitarbeitern aus dem Gesundheitswesen. Allerdings stellen Scholz und Schuler (1993) in einer Metaanalyse niedrigere Zusammenhänge für klassische Postkorbverfahren fest. Die Autoren finden heraus, dass die allgemeine Intelligenz - erhoben mittels Intelligenztests - zwar für die Gesamtleistung von Assessment-Center-Teilnehmern der beste Prädiktor ist (korrigiertes  $r = .43$ ;  $N = 17373$ ), sich jedoch der Zusammenhang zwischen klassischen Postkorbleistungen und allgemeiner Intelligenz als schwach mit lediglich  $r = .18$  (korrigiert) bei einem  $N$  von 877 erweist. Den Zusammenhang bewerten Scholz und Schuler (1993) selber als auffallend niedrig und erklären dies mit möglichen Stichprobeneffekten.

Eine testtheoretisch andere Herangehensweise wählen die Autoren der computergestützten „Managementarbeitsprobe“ (Etzel & Küppers, 2000). Sie überprüfen die Konstruktvalidität mit Hilfe eines probabilistischen Modells. Laut Einschätzung der Autoren deuten die durchgeführten Rasch-Analysen daraufhin, dass die Kompetenzen eindimensional und konstruktvalid erfasst werden.

Scholz und Schuler (1993) haben Studien zusammengefasst, die Ergebnisse in Persönlichkeitstests und Leistungen in klassischen Postkorbverfahren vergleichen. Dabei stellen sich für die meisten Variablen nur unbedeutende

Korrelationen heraus, so beispielsweise für Neurotizismus, Offenheit, Verträglichkeit und Leistungsmotivation. Selbstvertrauen und Dominanz zeigen die höchsten Werte mit  $r = .24$  bzw.  $r = .23$ . Die Autoren bewerten das Ergebnis als verwunderlich, da Postkorbverfahren klassischerweise als Einzelverfahren konzipiert werden und Variablen, die Interaktionsverhalten mitbestimmen, somit in dieser Aufgabe keine wesentliche Rolle spielen sollten. Sie führen als möglichen Grund für diese Zusammenhänge auf, dass der Bearbeitung eines Postkorbs häufig ein Gespräch zwischen den Beobachtern und der Testperson folge, in dem die Interaktion eine wichtige Rolle spiele. Dies könne den Zusammenhang erklären, dass für Interaktionssituationen typische Erfolgsfaktoren wie Selbstvertrauen und Dominanz mit dem Postkorbergebnis korrelieren. Eine solche Einschätzung spricht auch für den von Musch und Lieberei (2001) in den „Bonner-Postkorb-Modulen“ umgesetzten auswertungsobjektiven Verfahrenseinsatz. Die mittels Lösungsskizze ermittelte Postkorbleistung ist somit von unbeabsichtigten Effekten - wie sie Scholz und Schuler (1993) beschreiben - nicht beeinflusst.

Hinsichtlich der Nebengütekriterien Fairness und Akzeptanz werden positive Ergebnisse berichtet. Klassische Postkorbverfahren erfreuen sich laut Schippmann et al. (1990) einer hohen Akzeptanz bei Testpersonen und gelten als fair. Das bedeutet, dass sich hinsichtlich des Geschlechts oder der Zugehörigkeit zu einer bestimmten ethnischen Gruppe keine Verzerrungen ergeben (Thornton, Hollenshead & Larsh, 1997). Musch und Lieberei (2001) können für die Bonner-Postkorb-Module „CaterTrans“ und „Chronos“ keinerlei signifikante Korrelationen zwischen Alter, Geschlecht und den Ergebnissen in den Postkorbverfahren ermitteln. In einer Studie von Borman (1982) finden sich Korrelationen nahe Null zwischen dem eingesetzten klassischen Postkorb und den beiden Variablen „Erster Eindruck“ und „Physische Attraktivität“ - Ergebnisse, die auch für die Fairness des verwendeten Postkorbverfahrens sprechen.

Postkorbverfahren werden allgemein von Testpersonen als stark berufsbezogen und somit positiver eingeschätzt als beispielsweise

Fähigkeitstests oder biografische Fragebögen (Smither, Millsap, Stoffey, Reilly & Pearlmann, 1996). Für den „PC-Postkorb Seeblick“ von Scharley und Partner (1991) wird im Handbuch von einer Untersuchung der Tilburg Universität in den Niederlanden berichtet, wonach die befragten 35 Führungskräfte das Programm mit durchschnittlich 3.4 Punkten auf einer 5-Punkte-Skala zur Zufriedenheit bewerten.

Nr.	Vorgang	Person	Problemfeld	Priorität	Termin	Notiz
01	Demobrief 1)	Bayer	Einweihung	Prio. 3	Mi. 08-12	Demo1
02	Demobrief 2)	Heider	PR/Werbung	Prio. 2	Di. 08-13	
04	Demobrief 4)	Reich	PR/Werbung	Prio. 3	Mi. 08-12	Demo
05	Demobrief 5)	Staufner	Sonstiges	Prio. 1	Mi. 08-10	
03	Demobrief 3)					
06	Hotel-Führer, Magazin					
07	Brandschutzmassnahmen					
08	Hoteldirektor, Besuch					
09	Wochenend-Gäste					
10	Brandschutzprüfung					
11	Vortragsvorbereitung					
12	Parkplatznutzung					
13	Mittagessen, Einladung					
14	Abnahme Bauarbeiten					
15	Beiratsitzung					
16	Gerätebestandsliste					
17	Fototermin					
18	EDV-Präsentation					
19	Neue Bewerbungen					
20	Parkplatzvergrößerung					
21	Konzertpianist					
22	Lokalpresse, Interview					

**Abbildung 14: Oberfläche des „PC-Postkorbs“ (Demoversion) von Scharley und Partner (1998)**

Ein Problem, das sowohl für die Papier-Bleistift als auch die computergestützten Versionen gilt, ist testtheoretischer Art: Werden einzelne Schriftstücke als Items verwendet, so ist die „lokale stochastische Unabhängigkeit“ dieser Items verletzt, wenn die Bearbeitung der einzelnen Schriftstücke voneinander abhängt. Somit sind Verfahren der klassischen oder probabilistischen Testtheorie auf Itemebene nur bedingt bzw. gar nicht anwendbar (vgl. Kubinger, 1993). Ein solches Verfahren liefert im Grunde nur eine einzige unabhängige Messung, deren Reliabilität und - damit verbundene - Vorhersagekraft möglicherweise eingeschränkt ist. Neuere Ansätze - wie das Beispiel der „Bonner-Postkorb-Module“ (Musch & Lieberei,

2001) - versuchen daher, über einen objektiven Kriterienkatalog hinsichtlich der Güte der Aufgabenbearbeitung eine Standardisierung auf Itemstruktur zu erlangen. Die oben referierten Befunde deuten auf die Wirksamkeit dieser Strategie hin - zumindest hinsichtlich der Objektivität und Reliabilität.

Für computer- und internetbasierte Verfahren ist neben den üblichen Anforderungen an die Testkonstruktion (vgl. American Educational Research Association, 1999; Normausschuss Gebrauchstauglichkeit und Dienstleistungen, 2002; Testkuratorium, 1986) auch eine sinnvolle Einbindung des Mediums Computer zu fordern (vgl. u. a. Bartram, 2005; International Test Commission, 2006; Naglieri et al., 2004). Darunter sind Aspekte zu verstehen, die die Benutzerfreundlichkeit betreffen (s. Abschnitt 3.3 Gütekriterien). Es ist zu erwarten, dass der Fortschritt der Computer- und Netzwerktechnologien auch realistischere und leichter verständliche Gestaltungen der Benutzeroberflächen und Navigationen möglich macht. In diesem Zusammenhang verwundern die bisher recht dürftigen Angaben bzw. die geringe Anzahl von Studien zur Benutzerfreundlichkeit der computergestützten Verfahren. Studien, inwieweit beispielsweise die Navigation und Bedienung der computergestützten Postkorbverfahren von J. Funke und Rasche (1988) sowie Scharley und Partner (1991) oder die Oberfläche des computergestützten Szenarios von Hasselmann und Strauß (1995) (s. Abbildungen 14, 15 und 16) von der Zielgruppe verstanden und als benutzerfreundlich bewertet werden, sind bisher Mangelware. Auch die Rezensenten der drei computergestützten Postkorbverfahren Riediger und Rolfs (1998) berücksichtigen diesen Aspekt nicht durch Einbeziehung empirischer Daten, sondern lediglich durch persönliche Einschätzung. So berichten sie, dass der Postkorb „Seeblick“ sich gegenüber „Mailbox ´90“ durch eine größere Bedienungsfreundlichkeit auszeichne. Empirische Ergebnisse zur Untermauerung dieses Eindrucks bleiben die Autoren jedoch schuldig.



Abbildung 15: Benutzeroberfläche des „PC-Postkorbs“ von J. Funke und Rasche (1988)

### 3.4.1.3 Implikationen für die vorliegende Arbeit

In diesem Abschnitt wurden relevante Arbeiten zur Testgüte von Postkorbverfahren vorgestellt. Dabei interessierten insbesondere neuere Entwicklungen hinsichtlich der computergestützten Darbietung und mögliche problematische Aspekte in Bezug auf den eignungsdiagnostischen Einsatz. Es lässt sich konstatieren, dass trotz der häufigen Anwendung von Postkorbverfahren die bisherige Befundlage zu den Testgütekriterien recht spärlich ist, gerade auch im Hinblick auf die computergestützte Darbietung (Riediger & Rolfs, 1998; Schippmann et al., 1990).



Hinsichtlich eines eignungsdiagnostischen Einsatzes zeigen sich dabei folgende Problemkreise älterer Arbeiten:

- Mangelnde Standardisierung sowie Auswertungs- und Bewertungsobjektivität
- Abhängigkeit der Schriftstücke (Items)
- noch nicht ausreichende Belege zur Testgüte
- häufig unzureichende theoretische Fundierung

Werden diese Aspekte berücksichtigt (vgl. die Arbeiten von Musch & Lieberei, 2001; Etzel & Küppers, 2000), können Postkorbverfahren sehr wohl als eine sinnvolle Ergänzung zu anderen eignungsdiagnostischen Instrumenten angesehen werden (Dommel, 1995; Höft & U. Funke, 2001; Riediger & Rolfs, 1998; Ueckert, 1995). Gerade standardisierte Verfahren wie die „Bonner-Postkorb-Module“ (Musch & Lieberei, 2001) oder die „Managementarbeitsprobe“ (Etzel & Küppers, 2000) mit einer objektiven Auswertung auf Basis einer Itemstruktur sind viel versprechend und deuten auf die mögliche Testgüte von computergestützten Postkorbverfahren hin. Dommel (1995) und Höft und U. Funke (2001) kommen in ihren Übersichten hinsichtlich des aktuellen Forschungsstands zu dem Schluss, dass standardisierte Postkorbverfahren einen sinnvollen Bestandteil des Methodenrepertoires der angewandten Eignungsdiagnostik darstellen.

Auch Ueckert (1995) weist (computergestützten) Postkorbverfahren beim Einsatz als Diagnoseinstrument ein großes Potenzial im Bereich der Erfassung der Personmerkmale „Planen und Organisieren“ zu. Ueckert (1995) zieht in seinem Überblick über die berufsrelevanten Konstrukte „Planung, Organisation und Kontrolle“ das Fazit, dass diese Schlüsselmerkmale des Führungshandelns nicht „...mit den herkömmlichen Methoden der Diagnostik beizukommen ist“ (S. 262). Als Erfolg versprechender ordnet er zum einen das Postkorbverfahren und zum anderen computergestützte Szenarios ein, die im nächsten Abschnitt behandelt werden.

### **3.4.2 Computergestützte Szenarios**

Unter dem Begriff „computergestützte Szenarios“ werden Verfahren verstanden, bei denen die Testperson ein Computerprogramm steuert, das eine bestimmte sich dynamisch verändernde Problemsituation simuliert (Strauß & Kleinmann, 1995b). Diese Problemsituation ist zumeist in ein mehr oder weniger realistisches Rahmenszenario eingebettet. Dies kann beispielsweise das Steuern eines Unternehmens, wie beim Verfahren „DISKo“ von U. Funke (1992) oder das Regieren einer Kleinstadt - wie bei „Lohhausen“ von Dörner et al. (1983) - sein. Die Begriffe „Computersimulation“, „Computerszenario“ „komplexes Problemszenario“ und „computersimuliertes“ bzw. „computergestütztes Szenario“ sind synonym zu verstehen.<sup>15</sup> In diesem Kontext interessieren Verfahren und Befunde, die für die Diagnostik von Planen und Problemlösen relevant sind. Der Ursprung dieser Verfahren liegt nicht in der angewandten Psychologie, vielmehr werden computergestützte Szenarios seit Anfang der siebziger Jahre als Forschungsinstrumente in der kognitiven Psychologie eingesetzt (vgl. Broadbent, 1977; Dörner & Reither, 1978; Kleiter, 1970). Computergestützte Szenarios fanden zuerst in der wissenschaftlichen Kognitions- und Problemlösepsychologie Verwendung. Gerade die Arbeiten der Bamberger Gruppe um Dietrich Dörner (s. a. Kapitel 2) initiierten im deutschsprachigen Raum eine rege Forschungstätigkeit. Unter dem Begriff „komplexes Problemlösen“ ist eine Forschungsrichtung entstanden, die „sich allgemein mit dem Denken und Handeln von Menschen in unbestimmten, dynamischen und komplexen Situationen beschäftigt“ (Dörner, Schaub & Strohschneider, 1999, S. 198). Dörner und seine Mitstreiter suchten nach der Erweiterung des Gegenstandsbereichs der Denkpsychologie von den klassischen, linearen, „einfachen“ Problemen wie dem „Turm von Hanoi“ hin zu komplexen, dynamischen, intransparenten und polytelischen Problemstellungen. Die genaue Erläuterung der Eigenschaften komplexer

---

<sup>15</sup> Einige Autoren grenzen diese Begriffe noch weiter ein. So sieht Kluwe (1995) eine Verwendung des Simulationsbegriffs in diesem Kontext als problematisch an. Strauß (2000) unterscheidet computergestützte Planspiele von computergestützten Szenarios. Beide lassen sich nach dem Autor zu den computergestützten Simulationsverfahren zählen (zur genauen Unterscheidung s. Strauß, 2000).

Problemszenarios, die als psychologisch relevante Anforderungen an die Testperson zu verstehen sind, liefert Abschnitt 2.2.1.

Diese Computerprogramme fanden schnell Verwendung außerhalb der Grundlagenforschung, vor allem auf dem Gebiet der Eignungsdiagnostik, aber auch im pädagogischen Kontext (s. J. Funke, 1999; Klieme et al., 2001). Hintergrund des Einsatzes ist die Annahme, dass die komplexen Anforderungen der Realität im Grundsatz die gleichen sind wie sie an Problemlöser in computergestützten Szenarios gestellt werden. Die Anwendung der Szenarios würde demnach ermöglichen, Konstrukte wie „komplexes Problemlösen“, „strategische Kompetenzen“, „strategisches Denken“ oder „operative, praktische Intelligenz“ von Testpersonen realistisch einzuschätzen (vgl. Strauß & Kleinmann, 1995b). Dabei ist die Diagnostik von Planungsfähigkeit - wenn überhaupt - nur ein Randprodukt (vgl. Fritz und J. Funke, 1995). Trotz einiger Kritik (vgl. Fritz & J. Funke, 1995; M. Kersting, 1999; Süß, 1996) kommen Computersimulationen seit einigen Jahren beispielsweise im Rahmen von Assessment-Centern oder auch im Rahmen internationaler Längsschnittstudien wie der PISA-Studie (s. o.) zum Einsatz. Die insgesamt gewachsene Bedeutung der computergestützten Szenarios für die Eignungsdiagnostik wird neben einigen Herausgeberbänden (u. a. Frensch & J. Funke, 1995; Geilhardt & Mühlbradt, 1995; Strauß & Kleinmann, 1995a) auch durch die gestiegene Anzahl an Szenarios dokumentiert. Bereits 1992 hat Graf ca. 140 Verfahren (inkl. Planspiele) in einer Marktübersicht aufgezählt, die sich u. a. in Rahmengeschichte, Bearbeitungsmodus, Darbietung, Anzahl und Art der Variablen und Anforderung an die Problemlöser unterscheiden. Die Zahl der Szenarios dürfte heute noch deutlich größer sein. Die Unterschiedlichkeit der Verfahren hat naturgemäß auch diagnostische Implikationen. Das bedeutet, dass die einzelnen Szenarios sich teilweise deutlich hinsichtlich ihrer Testgüte unterscheiden können. Es ist somit vor einem eignungsdiagnostischen Einsatz die Validierung für jede Verfahrensfamilie von computergestützten Szenarios zu fordern. Fritz und J. Funke (1995) kritisieren in diesem Zusammenhang die mangelnde Überprüfung der Testgüte der Instrumente. So vergleichen sie die Anzahl der Verfahren, über die Angaben zu Reliabilität

und Validität publiziert wurden, mit der Anzahl der Verfahren, die generell verfügbar sind. Sie ziehen das Fazit: „...vermutlich wird auf 100 Simulationssysteme *ein* untersuchtes und nach eignungsdiagnostischen Kriterien überprüfetes System kommen“ (S. 72, Hervorhebung im Original). Aufgrund der zahlreichen jüngeren, umfangreichen Handbücher zu diesem Themenkomplex (u. a. J. Funke, 2003; Strauß & Kleinmann, 1995a) sowie aktueller Überblicksartikel (z. B. U. Funke, 1995a, 1995b; Höft & U. Funke, 2001; Kleinmann & Strauß, 1998) soll auf eine detaillierte Darstellung einzelner Instrumente verzichtet und stattdessen zusammenfassend auf relevante Aspekte hinsichtlich der Testgütekriterien eingegangen werden.

#### *3.4.2.1 Objektivität und Reliabilität*

Grundsätzlich ist durch den Einsatz von Computern eine verbesserte Objektivität in Durchführung und Auswertung gegenüber Papier-Bleistift-Tests möglich (Kubinger, 1993). Die Objektivität von computergestützten Szenarios wurde jedoch wiederholt kritisiert (vgl. u. a. J. Funke, 1983; M. Kersting, 1999; Süß, 1996). So können durch die Eigendynamik vieler Computerszenarios die Bedingungen für die Testpersonen bei der Durchführung unterschiedlich und somit nicht objektiv sein. Dabei führen Szenariozustände, die auf Basis der vorangegangenen Entscheidungen der Testperson berechnet werden, möglicherweise zu unterschiedlich schwierigen Problemkonstellationen (vgl. U. Funke, 1995a). Bei einigen Szenarios ist es auch möglich, in eine ausweglose „Sackgasse“ zu gelangen. Das bedeutet, dass bei einer bestimmten Konstellation die Testperson das computergestützte Szenario nicht mehr positiv beeinflussen kann. Eine vergleichende Bewertung und Interpretation der erzielten Leistungen von Probanden ist dann fraglich. Nach Süß (1996) ist für Computersimulationen, die Leistungsmessungen vornehmen wollen, die Vorgabe eindeutiger, präzise formulierter Ziele Voraussetzung für die Objektivität, da sich nur dann Fähigkeits- oder Eignungsaussagen im Vergleich zu anderen treffen lassen, wenn alle Testpersonen ein Ziel anstreben, nach dem ihre Leistung bewertet wird. In einigen Szenarios geben die Autoren bewusst keine Hinweise zu

Handlungsmöglichkeiten oder Zielen. Dies ist für den Einsatz in der Personalarbeit problematisch und kaum mit der Forderung nach Transparenz von eignungsdiagnostischen Verfahren zu vereinbaren (vgl. Testkuratorium, 1986).

Die Eigendynamik vieler computergestützter Szenarios und die Abhängigkeit der einzelnen Gütemaße ist auch hinsichtlich der Bestimmung von itembasierten Kennwerten problematisch (vgl. Kröner, 2001). Dadurch, dass die Bearbeitungsschritte voneinander abhängen, liefert ein Szenariodurchlauf im Grunde nur eine einzige unabhängige Messung, deren Reliabilität und damit verbundene Vorhersagbarkeit dementsprechend eingeschränkt ist. Da computergestützte Szenarios in der Praxis jedoch häufig über mehrere Stunden bearbeitet werden müssen, spricht dieser Aspekt nicht für eine hohe Testökonomie und Zumutbarkeit.

M. Kersting (1999) weist auf die Bedeutung der Reliabilität im eignungsdiagnostischen Kontext hin. Die Bestimmung der Reliabilität ist Voraussetzung für die Überprüfung von Zusammenhangsannahmen, wie beispielsweise zwischen der Leistung in einem Szenario als Prädiktor und Berufserfolg als Kriterium. Angesichts dieser zentralen Bedeutung bewertet M. Kersting (1999) die Anzahl und teilweise auch Qualität der vorhandenen empirischen Studien als unzureichend. Die meisten Untersuchungen stammen dabei aus dem grundlagenwissenschaftlichen Bereich. Bis 1999 lagen nach M. Kersting lediglich für zwei explizit anwendungsorientierte Szenarios („DISKo“ und Szenarios aus der Verfahrensfamilie „Textilfabrik“) überhaupt Reliabilitätsdaten vor. Hasselmann (1993) berichtet in einer Retest-Reliabilitätsstudie des Szenarios „Textilfabrik“ von Werten zwischen  $r_{tt} = .47$  bis  $r_{tt} = .74$  für verschiedene Gütemaße. Es wurden die Leistungen von 52 Studierenden untersucht, die das Szenario im Abstand von 14 Tagen steuerten. Für den erreichten Kapitalendwert ergeben sich dabei Werte von  $r_{tt} = .47$ , während sich für das Problemlösegütemaß, das die Entwicklung von vier wichtigen Systemvariablen (Produktion, Arbeitszufriedenheit, Maschinenzustand und Zahl verkaufter Hemden) berücksichtigt, der höchste Wert in Höhe von  $r_{tt} = .74$  liegt. Für das Szenario „DISKo“ ermittelt U. Funke (1992) Retest-Reliabilitäten zwischen  $r_{tt} = .53$  und  $.63$ . Die Werte liegen somit

insgesamt in ähnlicher Höhe wie in den Überblicksarbeiten von U. Funke (1993, 1995a) und Kleinmann und Strauß (1998). Dort wird von Retest-Reliabilitäten zwischen  $r_{tt} = .40$  und  $.83$  bei insgesamt zehn grundlagenorientierten Studien berichtet.

Zustand der Firma		Ihre Entscheidungen	
Arbeitszufriedenheit	> 0.98	Werbeausgaben	2800.00
Maschinenzustand	> 47.04	Instandhalt.-Ausgab.	1200.00
Produktion	> 403	Sozialausgaben/Arb.	50.00
Rohstoffe im Lager	> 16	Lohn pro Arb.	1080.00
Verkaufte Hemden	> 407	Preis pro Hemd	52.00
Hemden im Lager	> 80	<b>Rohstoffe kaufen</b>	<b>0</b>
Anzahl Arbeiter an XR-3	> 8	neue Arb. an XR-3	
Anzahl Arbeiter an XR-6	> 0	neue Arb. an XR-6	
Anzahl Maschinen XR-3	> 10	Arb. entlassen(XR-3)	
Anzahl Maschinen XR-6	> 0	Arb. entlassen(XR-6)	
Anzahl Lieferwagen	> 1	XR-3 kaufen	
Kassenstand	> 15774.66	XR-6 kaufen	
Kapitalwert	> 101120.64	XR-3 verkaufen	
		XR-6 verkaufen	
		Lieferwagen kaufen	
		Lieferwagen verkaufen	
Naechster Monat : F5		<b>Eingabe-Hinweis:</b>	
Dies ist der 1. Monat		Rohmaterial pro Einheit kostet: 2.96	

Abbildung 16: Bearbeitungsbildschirm des Szenarios „Textilfabrik“ von Hasselmann und Strauß (1995)

Die Ergebnisse für zwei neuere Szenarios liegen insgesamt im zufriedenstellenden Bereich. Dabei wurden die Simulationen „FSYS“ und „Multiflux“ untersucht, die explizit diagnostischen Zwecken dienen sollen. So berichtet Wagener (2001) für das Szenario „FSYS“ von Retest-Reliabilitäten hinsichtlich des Gesamtergebnisses von  $r_{tt} = .81$  - jedoch bei einem N von lediglich 21. Die Testpersonen bearbeiteten dabei nach etwa einem Monat die Parallelversion „FSYS-Farm“. Für die Detailskalen zeigen sich laut Wagener (2001) erwartungsgemäß weniger reliable Werte im Bereich von  $r_{tt} = .50$  bis  $.70$ . Auch Kröner (2001) ermittelt gute Reliabilitäten für das

Szenario „Multiflux“, das als Intelligenzdiagnostikum konzipiert wurde. Die interne Konsistenz der Testscores liegt zwischen .82 und .93 (Cronbachs-Alpha), die Werte der Retest-Reliabilität im Bereich von  $r_{tt} = .60$  bis .65.

Wesentlich für den Einsatz computergestützter Szenarios im eignungsdiagnostischen Kontext sind die Qualität und die Relevanz der Auswertungsergebnisse. Bei der Bewertung der erbrachten Leistungen lassen sich laut Höft und U. Funke (2001, S. 140) drei Arten von Problemlösegütemaßen unterscheiden:

- Kennzahlen des Endergebnisses oder des Verlaufs bestimmter Systemvariablen bilden die „Steuerleistung“ einer Testperson ab. Zumeist werden dabei aus den Rohwerten des Szenarios semantisch eingekleidete Koeffizienten gebildet - wie beispielsweise das Firmenskapital oder die Anzahl der Beschäftigten.
- Bei „kognitiven Kriterien“ wird die Richtigkeit des erworbenen Wissens, der mentalen Modelle und der Denk- und Planungsprozesse bezüglich des Szenarios erfasst.
- „Verhaltensmaße“ zum Umgang mit einem computergestützten Szenario werden beispielsweise mit Hilfe von Indikatoren für das Frage- und Entscheidungsverhalten erhoben.

Inwieweit mit Hilfe dieser Problemlösegütemaße Problemlösefähigkeit erfasst wird bzw. inwieweit die Problemlösegütemaße eignungsdiagnostischen Ansprüchen gerecht werden, ist von vielen Autoren diskutiert worden (vgl. J. Funke, 1983; M. Kersting, 1999; Kröner, 2001; Süß, 1996). So werden die Gütemaße als zumeist nicht vergleichbar, nicht standardisiert, unreliabel und verfälschbar kritisiert. Viele Szenarios besitzen keine optimale Lösung, die als Grundlage für die Bemessung der Qualität der Lösung dienen kann. Süß, M. Kersting und Oberauer (1991, 1993) haben für die „Schneiderwerkstatt“ durch eine Aufgabenanalyse zeigen können, dass die Kennwerte für die Problemlöseleistung nicht reliabel sind. Die bisherige Operationalisierung von Erfolg kann demnach zu paradoxen Ergebnissen führen. Diese

Überlegungen führten zu der Bildung eines neuen, geeigneten Problemlösegütemaßes.

M. Kersting (1999) und Süß (1996) zeigen in zwei unabhängigen Studien, dass für Personen, die mit der Bearbeitung der Simulationen überfordert sind, die herkömmlichen Problemgütemaße keine interne Validität besitzen. Die Überforderung von Probanden ist bei vielen Studien berichtet worden (s. M. Kersting, 1999), so dass dieser Befund vermutlich auch für andere Szenarios Relevanz hat. Kleinmann und Strauß (1998) argumentieren, dass für eine geeignete Messung der Problemlöseleistung die computergestützten Szenarios auf Basis formaler Modelle beschreibbar sein müssen (vgl. J. Funke, 1986; J. Funke & Buchner, 1992). Buchner (1999) stellt dar, wie komplexes Problemlösen vor dem Hintergrund der Theorie finiter Automaten zu betrachten wäre und welche Konsequenzen sich daraus für die Konstruktion von computergestützten Szenarios ergeben. Allerdings ist die Praxis der bisherigen im Personalbereich eingesetzten Verfahren hinsichtlich dieses Aspekts wenig ermutigend, denn keines der in der Personalarbeit eingesetzten Szenarios lässt sich laut Kleinmann und Strauß (1998) exakt formal beschreiben (zur Theorie finiter Automaten bzw. formaler Modelle in Bezug auf computergestützte Szenarios s. genauer Buchner, 1999; Strauß & Kleinmann, 1995c).

Einen Ausweg aus der problematischen Bestimmung der Problemlösemaße hat Kröner (2001) in seiner Arbeit untersucht. Testpersonen müssen dabei im Anschluss an eine Explorationsphase, wobei das Szenario „Multiflux“ systematisch zu erkunden ist, einen simulationsbasierten Fragebogen beantworten. Dabei handelt es sich um einzelne Fragen zum Systemwissen und zur Steuerung des Szenarios, die unabhängig voneinander sind. Dieser Aufbau mit Itemstruktur hat u. a. den Vorteil, dass sich itembasierte Kennwerte berechnen lassen und die oben genannten methodischen Probleme größtenteils gelöst werden. Bei einer Itemstruktur ist jede zu bearbeitende und später bewertete Aufgabe von den anderen unabhängig und somit als einzelnes Item zu verstehen, wo hingegen bei vielen computergestützten Szenarios die Aufgaben voneinander abhängen. Das



bedeutet, dass sich die Art und Weise, wie eine Person das computergestützte Szenario bearbeitet, direkt auf den weiteren Verlauf der Szenario-Durchführung auswirkt. Die Ergebnisse der Untersuchung Kröners (2001) weisen darauf hin, dass die gebildeten Problemlösegütemaße reliable und valide sind (s. o.).

#### *3.4.2.2 Validität und Nebengütekriterien*

Eines der Hauptargumente für den Einsatz von computergestützten Szenarios ist laut Dörner (1989) die hohe Übereinstimmung der komplexen Anforderungen der Realität und der Anforderungen innerhalb von komplexen Computersimulationen. Diese Annahme inhaltlicher Validität ließ sich allerdings nicht endgültig nachweisen (M. Kersting, 1999). So wurden laut Kleinmann und Strauß (1998) bisher auch keine Studien zur Gleichheit von Managementaufgaben und Computerszenarios durchgeführt.

U. Funke (1995a) kommt in einer Übersicht von vier Studien zur Kriteriumsvalidität zu dem Schluss, dass insgesamt realistischerweise von Validitätskoeffizienten für die Vorhersage von Berufserfolg in der Größenordnung von .30 bis .40 auszugehen sei. Dies ordnet er in etwa dem Bereich anderer simulationsorientierter Verfahren zu. Allerdings kritisiert er die methodischen Probleme und die geringe Anzahl bisheriger Studien. Hinsichtlich der Kriteriumsvalidität bezüglich der Vorhersage von Berufs- oder Ausbildungserfolg finden sich in einigen Studien Hinweise dahingehend, dass Experten die Szenarios teilweise besser steuern als Novizen, wengleich andere Untersuchungen hierzu keinerlei Unterschiede feststellen. Für das Szenario „Textilfabrik“ berechnen Hasselmann und Strauß (1995) einen Extremgruppenvergleich zwischen 41 Studenten und 17 Führungsnachwuchskräften einer Bank. Die Gruppe der Führungsnachwuchskräfte zeigt dabei hinsichtlich aller Gütemaße des Systems bessere Leistungen als die Studentengruppe nicht-wirtschaftswissenschaftlicher Fachrichtungen. Im Vergleich zwischen den elf Wirtschaftsstudenten und den Führungskräften erweisen sich die

Führungskräfte jedoch lediglich hinsichtlich eines Gütemaßes als signifikant besser. Leistungsunterschiede zwischen Experten und Novizen lassen sich auch für andere Szenarios außerhalb des eignungsdiagnostischen Einsatzfeldes finden (vgl. u. a. Putz-Osterloh, 1987; Putz-Osterloh & Lemme, 1987), während wiederum Schaub und Strohschneider (1992) keinen Leistungsvorteil von Managern gegenüber Studierenden beim ökonomischen Szenario „Manutex“ feststellen können. Streufert, Pogash und Piasecki (1988) berichten von teilweise signifikanten Zusammenhängen zwischen der Zahl unterstellter Mitarbeiter, der Beförderung, der erreichten Position und dem Einkommen auf der einen Seite und den Verhaltensvariablen im Szenario „SMS“ auf der anderen Seite. Dabei ergeben sich bei einer Stichprobe von 111 Personen des mittleren Managements Korrelationen zwischen  $r = -.09$  und  $r = .31$ .

Hinsichtlich der prädiktiven Validität finden sich nur sehr wenige Studien. Hasselmann (1993) vergleicht die Daten von 21 Führungsnachwuchskräften im Szenario „Textilfabrik“ mit Kriterien der beruflichen Bewährung, die ca. zwei Jahre später erhoben wurden. Die höchsten Zusammenhänge mit den Szenario-Leistungen finden sich für Karriereprognosen bzw. Potenzialeinschätzungen durch Vorgesetzte ( $r = .15$  bis  $r = .49$ ).

Hasselmann (1993) ermittelt in diesem Zusammenhang auch für Karriereindizes wie die berufliche Position, das Gehalt und die Leistungsbeurteilungen durch die Personalbetreuer mittelstarke Zusammenhänge mit der Leistung in Computersimulationen. Eine jüngere, prädiktive Studie liefert M. Kersting (1999), der Polizisten durch ihre Vorgesetzten hinsichtlich ihrer beruflichen Problemlösefähigkeit und Intelligenz einschätzen ließ. Die Daten der Vorgesetztenurteile vergleicht M. Kersting (1999) mit der Leistung im Szenario „DISKo“, die ein Jahr zuvor erhoben wurde. Die Korrelationen belaufen sich bei einer Stichprobengröße von 72 auf  $r = .38$  bzw.  $r = .31$ . Wenngleich in einigen Studien positive Hinweise auf Kriteriumsvalidität bezüglich der Vorhersage von Berufserfolg gefunden wurden, so lässt sich aufgrund der geringen Anzahl an Studien und der teilweise niedrigen Korrelationen noch kein abschließendes Fazit ziehen.

Zur Konstruktvalidität bzw. konvergenten und diskriminanten Validität von Computersimulationen wurden Zusammenhänge zu Assessment-Center Daten, zu psychologischen Tests (u. a. Persönlichkeits- und Intelligenztests) und zu anderen Szenarios ermittelt. Die Korrelationen zu den Leistungen in Assessment-Center-Übungen sind zumeist nicht bedeutend. So erweisen sich beispielsweise bei Hasselmann (1993) von den 24 gefunden Werten lediglich fünf als signifikant ( $N = 21$ ). Diese Korrelationen ergeben sich zwischen dem Endkapitalwert als Gütemaß und den im AC beobachteten Kriterien Problemlösefähigkeit ( $r = .45$ ), Entscheidungsfähigkeit ( $r = .39$ ), Belastbarkeit ( $r = .38$ ) und Führungspotenzial ( $r = .38$ ). Inwieweit dies als Hinweis zur Konstruktvalidität des Verfahrens zu werten ist, bleibt allerdings unklar.

Der Zusammenhang zwischen den Leistungen in computergestützten Szenarios und Intelligenztestleistungen wurde kontrovers diskutiert (s. Kapitel 2). Umfassende Untersuchungen (u. a. Süß, 1996, 1999) haben die methodischen Probleme vergangener Studien aufgedeckt und den Einfluss neu untersucht. Es gilt demnach als gesichert, dass Intelligenz bzw. Verarbeitungskapazität aus dem Berliner Intelligenzstrukturmodell (Jäger, 1982) mittelstark mit der Problemlöseleistung in computergestützten Szenarios zusammenhängt. Dies lässt sich als Beleg für konvergente Validität interpretieren. Im Abschnitt 2.2.5.2.1 wird die Diskussion und der aktuelle Forschungsstand zum Zusammenhang zwischen den Leistungen in computergestützten Szenarios und Intelligenztests genauer vorgestellt.

Im Hinblick auf die Prüfung der Konstruktvalidität wurden auch Leistungen in verschiedenen Simulationen verglichen. Es zeigen sich zumeist keine oder nur uneinheitliche Zusammenhänge wie beispielsweise bei Wagener (2001). Er setzt das Szenario „FSYS“ zusammen mit der Schneiderwerkstatt-Version "Hemdenfabrik" von Süß, Oberauer und M. Kersting (1993) ein. Dabei ergeben sich in drei Untersuchungen deutlich unterschiedliche Korrelationen in Höhe von  $r = -.35$ ,  $r = .20$  und  $r = .36$ , was ggf. auf methodische Probleme bei der Bestimmung der Problemlösegütemaße (s. o.) oder möglicherweise auf die Unterschiedlichkeit der Szenarios hinweist.

Einige Studien haben sich mit dem Zusammenhang der Leistungen in computergestützten Szenarios und nicht-kognitiven Personmerkmalen beschäftigt. Übersichten finden sich dazu bei Hasselmann (1993), Vogt (1998), M. Kersting (1999) und Schaub (2001). Die meisten Untersuchungen berichten von unbedeutenden oder nur sehr geringen Korrelationen. In Studien, in denen ein Zusammenhang aufgedeckt wurde, handelt es sich zumeist um erwartungsgemäße, jedoch nur kleine Effekte, die bislang nicht überzeugend repliziert werden konnten. Die Leistungen in computergestützten Szenarios werden somit zumeist nicht systematisch von nicht-kognitiven Personmerkmalen beeinflusst. Inwieweit dies als diskriminante Validität zu interpretieren ist, lässt sich den Studien nicht entnehmen.

Nebengütekriterien wie Akzeptanz und Fairness wurden ebenfalls untersucht. Hinsichtlich des Einflusses der Computererfahrung auf die Leistungen in computergestützten Szenarios ist die Befundlage insgesamt uneinheitlich. Während beispielsweise U. Funke (1992) und Hasselmann (1993) keinerlei Zusammenhang finden konnten, weisen die Ergebnisse von Süß (1996) im Rahmen der Berliner Untersuchung mit dem Szenario „Schneiderwerkstatt“ auf einen substanziellen Einfluss hin. Dabei korreliert das Ausmaß der Computererfahrung zum ersten Messzeitpunkt zu  $r = .41$  mit dem Aggregat der drei folgenden Bearbeitungen. Jedoch ist die Computererfahrung auch mit den Intelligenz- und Wissensmaßen konfundiert.

M. Kersting (1999) untersucht den Einfluss der Computererfahrung und Computereinstellung auf die Problemlöseleistung im Szenario „Schneiderwerkstatt“. Dabei zeigt sich für die Computererfahrung ein schwacher positiver Zusammenhang von  $r = .19$  bei einer Stichprobengröße von 100 Personen.

Im Rahmen der PISA-Vorbereitungsstudie wurde der Einfluss der Computererfahrung auf die Erfassung der Problemlösefähigkeit für sechs verschiedene Verfahren untersucht (Klieme et al., 2001, s. Kapitel 2). Dabei ergeben sich für alle eingesetzten Verfahren positive signifikante

Zusammenhänge (s. Tabelle 11). Die Höhe der Korrelationen liegt zwischen  $r = .14$  und  $.36$ . Die Autoren haben sich das Ziel gesetzt, Korrelationen von niedriger als  $.3$  zu erreichen. Es lassen sich jedoch keine Angaben dazu finden, wie die Computererfahrung gemessen wurde - ob per Single-Item oder mit Hilfe einer umfangreicheren Erfassungsmethode.

**Tabelle 11: Einfluss der Computervorerfahrung in der PISA-Vorbereitungsstudie (nach Klieme et al., 2001)**

Verfahren	Skala	Computererfahrung
virtuelles Labor	Wissen	.21
ökologisches Planspiel	Wissen	.28
Raumfahrtspiel	Exploration	.26
	Wissen	.18
	Steuerung	.36
Projektaufgaben		.16
technisches Problemlösen		.22
analoges Problemlösen		.14

Anmerkung: Die ersten drei Verfahren sind computergestützt. Die letzten drei sind Papier-Bleistift-Versionen.

Der Einfluss von Computererfahrung und -einstellung ist bislang nur bei M. Kersting (1999) umfassend untersucht worden (s. o.). Bei den meisten anderen Studien sind nach M. Kersting (1999) Zweifel an der Qualität der Messung der Computererfahrung (teilweise nur per Single-Item) berechtigt. M. Kersting (1999) empfiehlt, zur Absicherung der mit Hilfe von computergestützten Szenarios gestellten Diagnosen stets auch die Computererfahrung und -einstellung zu erfassen. Dies ist jedoch bisher nur bei sehr wenigen Studien erfolgt.

Zur Akzeptanz der Szenarios bei Testpersonen und Assessment-Center-Beobachtern werden ebenfalls unterschiedliche Ergebnisse berichtet. Hartung und Schneider (1995) konstatieren zum Beispiel eine hohe

Akzeptanz für das Szenario „Utopia“ beim Einsatz im Rahmen eines Assessment-Centers, während M. Kersting (1999) keine grundsätzlich höhere Akzeptanz von computergestützten Szenarios gegenüber Intelligenztests ausmachen konnte. Insgesamt weisen computergestützte Szenarios jedoch laut Kleinmann und Strauß (1998) zumeist hohe Akzeptanz- und Zufriedenheitswerte bei Testpersonen auf.

#### *3.4.2.3 Implikationen für die vorliegende Arbeit*

Im Mittelpunkt dieses Abschnitts standen eignungsdiagnostisch relevante Befunde hinsichtlich der Testgüte von computergestützten Szenarios. Zusammenfassend lässt sich festhalten, dass zwar einige Hinweise zur befriedigenden eignungsdiagnostischen Testgüte von computergestützten Verfahren gefunden wurden, wie beispielsweise bei U. Funke (1992), Hasselmann (1993) und insbesondere in den neueren Arbeiten von Kröner (2001) und Wagener (2001), es für den eignungsdiagnostischen Einsatz jedoch einige Problembereiche zu berücksichtigen gilt:

Die Auswertungs-, Bewertungs- und Interpretationsobjektivität lässt sich - wie Kröner (2001) zeigt - durch eine standardisierte Durchführung mit Musterlösung und Itemstruktur erreichen. Durch eine itembasierte Struktur mit objektiven Maßen der Problemlöseleistung können die bisherigen Probleme der Qualität der Maße gelöst werden. Die Ergebnisse verschiedener Testpersonen sind dadurch vergleichbar und beschreibbar. Durch dieses Vorgehen ist eine reliable und valide Bestimmung der Lösungsqualität möglich. Auch das Methodenproblem der stochastischen Abhängigkeit der Messungen innerhalb der computergestützten Szenarios lässt sich mit Hilfe einer itembasierten Struktur mit unabhängigen Aufgaben auffangen.

Wie Wagener (2001) für das Szenario „FSYS“ darstellt, kann durch die transparente Vorgabe der wichtigsten Systemeigenschaften eine geeignete Bestimmung der Problemlöseleistung erfolgen. Dabei können Testpersonen während der Bearbeitung Informationstexte aufrufen. Die damit verbundene Einschränkung des Intransparenzprinzips erscheint als das kleinere Übel gegenüber der sonst drohenden Zufallsdetermination der

Problemlöseleistung (Wagener, 2001). Durch eine transparente Vorgabe der Ziele und Anforderungen lassen sich ggf. fairere und reliablere Bedingungen schaffen (vgl. Süß, 1996).

Der Aspekt der Eigendynamik hat sich - wie in Kapitel 2 dargelegt - als Einflussgröße der Aufgabenschwierigkeit erwiesen. Für einen eignungsdiagnostischen Einsatz ist allerdings die eigendynamische Entwicklung eines computergestützten Szenarios hin zu einem ausweglosen Zustand nicht sinnvoll, da dies negative Auswirkungen auf Objektivität und ggf. Testfairness haben kann (vgl. Süß, 1996).

Werden diese Aspekte berücksichtigt, so ist mit Hasselmann (1993), Kröner (2001) und Kleinmann und Strauß (1998) festzuhalten, dass computergestützte Szenarios einen wichtigen und Erfolg versprechenden Teil des methodischen Repertoires der angewandten Eignungsdiagnostik darstellen können.





## **4 Entwicklung der E-Assessment-Verfahren**

### ***webPostkorb* und *webPlanungsaufgabe***

Der theoretische und methodische Hintergrund der Verfahrensentwicklung und die daraus gezogenen Implikationen für die vorliegende Arbeit waren Mittelpunkt der Ausführungen der vorangegangenen Kapitel. Daraus abgeleitet wurden im Rahmen dieser Arbeit zwei E-Assessment-Verfahren zum Planen und Problemlösen von zukünftigen (pädagogischen) Führungskräften entwickelt: der *webPostkorb* und die *webPlanungsaufgabe*. Bevor die genannten Verfahren in den Abschnitten 4.4 und 4.5 näher betrachtet werden, sollen zunächst die Entwicklungsschritte bis zur Programmierung der Erstversionen vorgestellt werden. Tabelle 12 gibt einen Überblick über die wesentlichen Abschnitte der Verfahrenskonstruktion. Auf die weiteren Schritte zur Entwicklung der Endversionen und insbesondere auf den Ablauf der Evaluationsstudien wird im fünften Kapitel eingegangen.

Wie in Tabelle 12 dargestellt, bestand der erste Entwicklungsschritt aus umfangreichen Analysearbeiten. Dazu gehörten im Sinne des hier gewählten Hybridansatzes u. a. die Berücksichtigung empirischer Tätigkeitsanalysen von Schulleitungen zum Zwecke der Identifizierung berufstypischer Anforderungen und Aufgaben sowie die Bestimmung des Verhaltens bzw. der Fähigkeiten, die für die erfolgreiche Bewältigung dieser Anforderungen und Aufgaben entscheidend und notwendig sind. Darüber hinaus galt es, die Erkenntnisse der psychologischen Forschung zu den Konstrukten „Planen“ und „Problemlösen“ sowie die bisherigen methodischen Ansätze bei der Verfahrensentwicklung zu berücksichtigen. Dieser theoretische und methodische Hintergrund der Verfahrens- und Aufgabenentwicklung stand im Fokus der vorangegangenen Kapitel. Daher werden im Folgenden lediglich die wesentlichsten Implikationen für die Rohformentwicklung noch einmal kurz aufgeführt. Auf die weiteren Entwicklungsschritte bis zur Erstversion (s. Tabelle 12), wie die Rohformentwicklung, die Vorprüfung der Rohformen an Studierenden, die Programmierung der Verfahren als internetbasierte Instrumente und die konkreten Umsetzungen und Operationalisierungen wird darauffolgend in eigenen Abschnitten näher eingegangen.

**Tabelle 12: Überblick über die Schritte der Verfahrenskonstruktion bis zur Programmierung der Erstversionen**

Entwicklungsschritt	Inhalt	Ziele
Analysearbeiten und Implikationen für die Verfahrensentwicklung	<ul style="list-style-type: none"> <li>• Analyse der Anforderungen an Schulleitungen (s. Abschnitt 2.1)</li> <li>• Analyse der Literatur zu den Konstrukten „Planen“ und „Problemlösen“ (s. Abschnitt 2.2)</li> <li>• Analyse der bisherigen Methoden (s. Kapitel 3)</li> </ul>	<ul style="list-style-type: none"> <li>• anforderungsgeleitete Verfahrensentwicklung als Hybrid aus Konstrukt- und Simulationsverfahren</li> <li>• theoriegeleitete Entwicklung der Verfahren</li> <li>• Weiterentwicklung bisheriger Methoden</li> </ul>
Rohformentwicklung	<ul style="list-style-type: none"> <li>• Interviews mit Schulleitungsmitgliedern</li> <li>• Zusammenarbeit mit Schulleitungsmitgliedern</li> </ul>	<ul style="list-style-type: none"> <li>• anforderungsgeleitete Entwicklung</li> <li>• Entwicklung von alltagsnahen und berufstypischen Aufgaben</li> </ul>
	<ul style="list-style-type: none"> <li>• Zusammenarbeit mit Experten der Gestaltung von Benutzeroberflächen und Internet-Programmierungen</li> </ul>	<ul style="list-style-type: none"> <li>• Gestaltung von Aufgaben, die internetbasiert durchgeführt werden können</li> <li>• Rohformen (Offline-Versionen) der <i>webPlanungsaufgabe</i> und des <i>webPostkorbs</i></li> </ul>
Vorprüfung der Rohformen an Studierenden	<ul style="list-style-type: none"> <li>• Storyboard-Bearbeitung mit Thinking-Aloud</li> </ul>	<ul style="list-style-type: none"> <li>• benutzerorientierte Entwicklung</li> <li>• Verständnisprüfung</li> <li>• Prüfung der Zufriedenheit mit der Gestaltung</li> <li>• Überprüfung der Bearbeitungsdauer</li> <li>• Sicherung der Benutzerfreundlichkeit, Akzeptanz, sozialen Validität</li> </ul>
Programmierung der Erstversion des <i>webPostkorbs</i>	<ul style="list-style-type: none"> <li>• Programmierung der Erstversion</li> <li>• Beta-Testung</li> </ul>	<ul style="list-style-type: none"> <li>• lauffähige, fehlerfreie Erstversion des <i>webPostkorbs</i></li> </ul>
Programmierung der Erstversion der <i>webPlanungsaufgabe</i>	<ul style="list-style-type: none"> <li>• Programmierung der Erstversion</li> <li>• Beta-Testung</li> </ul>	<ul style="list-style-type: none"> <li>• lauffähige, fehlerfreie Erstversion der <i>webPlanungsaufgabe</i></li> </ul>

Aus den Analysearbeiten ergab sich, dass die Konzeption der E-Assessment-Verfahren an bestehende Methoden wie den Postkorbverfahren und den computergestützten Szenarios anknüpfen sollte, allerdings mit der Vorgabe, deren erkennbare Schwächen für den eignungsdiagnostischen Einsatz aufzuheben und durch verbesserte Lösungen zu ersetzen. Die Einweisung in die Programmbedienung sollte mit Hilfe von Übungsbeispielen illustriert werden, um eventuelle Vorteile von computererfahrenen Personen zu minimieren (vgl. Hervogich, 1992 zitiert nach Kubinger, 1993). Zur Steigerung der Durchführungsobjektivität wurde eine direkte Interaktion der Testpersonen mit dem Computer gewählt. Im Sinne von Booth (1998, s. o.) waren die Eingabemöglichkeiten für die Benutzer so zu gestalten, dass sie sich als einfach, eindeutig, erwartungskonform, kontrollierbar und korrigierbar erweisen. Um den Transparenzanforderungen (vgl. Testkuratorium, 1986) gerecht zu werden, waren die Handlungsmöglichkeiten zu erklären und offen zu legen sowie eindeutige und jederzeit aufrufbare, transparente Zielvorgaben mit Hilfe von Informationstexten zur Verfügung zu stellen (vgl. Wagener, 2001). Darüber hinaus wurden die Datenschutzerklärung und die Kontaktmöglichkeiten dargestellt. Die vier Aspekte der sozialen Validität nach Schuler (1990, s. o.) hinsichtlich der Mitteilung relevanter Informationen, der Partizipation an der Entwicklung und Anwendung, der Transparenz der Verfahren und der Schlussfolgerungen und des Feedbacks in rücksichtsvoller, verständlicher Form galt es bei der Verfahrensentwicklung zu berücksichtigen. Dabei sollten die Programmierung und Gestaltung benutzerfreundlich, aber auch ansprechend, simulativ und spielerisch sein, um eine möglichst hohe Akzeptanz zu erreichen.

#### **4.1 Rohformentwicklung**

Neben der Berücksichtigung der Erkenntnisse der Schulleitungsforschung (s. o.) wurden zur weiteren Absicherung des berufsbezogenen Anforderungsbezugs Interviews mit Schulleitungsmitgliedern geführt, um

berufsrelevante und alltagsnahe Aufgaben zu extrahieren. Mit Hilfe offener, explorativer Befragungen und anschließender Feedbackschleifen konnten so möglichst typische und realistische Aufgaben aus dem Alltag pädagogischer Führungskräfte entwickelt werden.

Die aus den Vorarbeiten abgeleiteten Aufgaben wurden dann in Zusammenarbeit mit Experten der benutzerorientierten Webgestaltung als Rohform erstellt. So sollte sichergestellt werden, dass die angedachten Aufgaben internetbasiert darstellbar sind und dass sich die Navigation und der Verfahrensaufbau als möglichst intuitiv und leicht erlernbar gestaltet.

#### **4.2 Vorprüfung der Rohformen an Studierenden**

Im Rahmen der Entwicklung der Verfahren war es ein Anliegen, Gesichtspunkte der Benutzerfreundlichkeit, Akzeptanz und Fairness zu berücksichtigen. Die Zielgruppe der Studierenden wurde dabei schon im Entwicklungsstadium mit in die Konstruktion und Gestaltung der Verfahren einbezogen. Hierbei wurde so vorgegangen, dass auf Basis der psychologischen Operationalisierungen grafische Storyboards entstanden. Diese Bildergeschichten setzten sich zusammen aus den bis dahin konstruierten und gestalteten Seiten der Verfahren. Die Storyboards wurden einzelnen Studierenden in Papierform vorgesetzt. Dabei gaben die Personen mit Hilfe der Thinking-Aloud-Methode ihre Eindrücke während des Bearbeitens wieder. So ließen sich bereits die Rohformen hinsichtlich der Verständlichkeit und der Zufriedenheit mit der Gestaltung als auch hinsichtlich der vermutlichen Bearbeitungsdauer überprüfen. Dieses Prozedere wurde für die verschiedenen Prototypen wiederholt. So konnten wichtige Hinweise und Erkenntnisse bereits im Entwicklungsstadium aufgenommen und mittels einiger Feedbackschleifen abgestimmt in die programmierten Erstversionen integriert werden.

### 4.3 Entwicklung als internetbasierte E-Assessment-Verfahren

Um die Möglichkeiten und Vorteile der Internetdarbietung nutzen zu können, erfolgte die Konstruktion der E-Assessment-Verfahren internetbasiert. Dabei galt es, eine ganze Reihe von Aspekten zu berücksichtigen, die sich in erster Linie durch den Einsatz im Internet ergeben. So ist es im Hinblick auf das Untersuchungsdesign (s. o.) von Bedeutung, nur solche Testteilnehmer zum Online-Teil der Untersuchung zuzulassen, deren Computer über gewisse Mindeststandards verfügten (vgl. International Test Commission, 2006). Dabei wurde u. a. die so genannte „Sniffer-Technologie“ von flash© (1995-2007 Macromedia, Inc.) eingesetzt, die die Einstellungen beim Benutzenden überprüft und gegebenenfalls die aktuelle Flash-Player-Version kostenlos installiert. Die Installation erfolgt dabei typischerweise automatisiert, so dass auch computerunerfahrene Personen die Voraussetzungen ohne eigenes Eingreifen in die Computereinstellungen erfüllen können. Die wesentlichen Vorteile der Flash-Programmierung sind neben der weiten Verbreitung und weitestgehenden Browserunabhängigkeit vor allem die Möglichkeiten der interaktiven Verfahrensgestaltung, die weit über die Möglichkeiten der html-Programmierung hinausgehen (zum Aspekt der Interaktivität s. genauer Abschnitt 3.1.4).

Darüber hinaus wurden allgemeine technische Aspekte, wie die Sicherstellung der Lauffähigkeit des Programms auf allen Rechnern, die Browserkompatibilität, die korrekte Verlinkung, die Optimierung von Ladezeiten sowie die Eindeutigkeit der Sicherung der Daten berücksichtigt und vielfach getestet, um die Gefahr von Datenverlusten zu minimieren.

Die Mindeststandards wurden darüber hinaus vor der Testbearbeitung kommuniziert (s. Abbildung 17) und sind bewusst auf niedrigem Niveau der Rechnerleistung festgesetzt, um möglichst alle Teilnehmer erreichen zu können.

Um eine reibungslose Teilnahme zu gewährleisten, muss Ihr Computer-System folgende Voraussetzungen erfüllen:	
CPU/Prozessor:	mindestens Pentium II (200 MHz)
Arbeitsspeicher/RAM:	mindestens 64 MB
Browser:	mindestens Internet Explorer 4.0 oder Netscape Navigator 4.6
Javascript:	aktiviert
Flashplayer:	Version 5 oder höher installiert
Cookies:	aktiviert (nur für temporäres PHP-Sessionhandling, es werden keine permanenten Cookies bei Ihnen gesetzt)
Internet-Verbindung:	alle Verbindungen möglich
Bitte nutzen Sie eine Maus.	

**Abbildung 17: Mindeststandards der Online-Teilnahme an den Verfahren**

Die Anwendung wurde so programmiert, dass es möglich ist, die Verbindung während der Bearbeitung zu unterbrechen und erst nach Abschluss wieder aufzubauen. Dadurch können eventuell entstehende Kosten möglichst niedrig gehalten werden. Auch die Schnelligkeit der Internetverbindung hat somit keinen Einfluss auf die Bearbeitung der Verfahren.

Die Mindeststandards waren jedoch nur bei der Durchführung des *webPostkorbs* der Voruntersuchung und der *webPlanungsaufgabe* der Hauptuntersuchung von Bedeutung, da alle anderen Verfahren an geeigneten Rechnern in den Computerräumen der Universität bzw. als Papier-Bleistift-Version durchgeführt wurden.

#### **4.4 Problemlösen im *webPostkorb***

Beim internetbasierten Simulationsverfahren *webPostkorb* wurde das Postkorbverfahren unter Berücksichtigung wichtiger Aspekte der computergestützten Szenarios weiterentwickelt und mit einer Musterlösung konzipiert (s. Tabelle 13). Der *webPostkorb* wurde itembasiert und auswertungsobjektiv konstruiert, so dass die Probleme hinsichtlich der Reliabilität und Validität von Problemlösemaßen bisheriger Verfahren gelöst werden konnten. Der gewählte Aufbau ermöglicht somit den sinnvollen Einsatz der Analyseverfahren der klassischen Testtheorie auf Itemniveau zur weiteren Optimierung des Verfahrens.

### Einweisung


Im folgenden webPostkorb geht es um Ihre Fähigkeit, komplexe Probleme zu analysieren, wichtige Zusammenhänge zu erkennen und in dynamischen Situationen zu handeln.

Damit für alle Teilnehmenden die Bedienung dieses Tests klar ist, werden wir Ihnen nun im folgenden den Aufbau der Oberfläche erklären und mit zwei Beispielen diese Einweisung abschließen.

Sie haben die Möglichkeit, sich die Einweisungs-Seiten so lange anzuschauen, bis Sie die Bedienung verstanden haben. Bei Bedarf können Sie zu den vorherigen Einweisungsseiten zurückkehren (über den "< rückwärts -Schalter, unten links).

---

**Szenario:**  
Für die folgenden Aufgaben sollen Sie bitte die Aufgaben des **Schulleiters Thomas Thomasen** übernehmen.  
Stellen Sie sich bitte vor, dass Sie vor dem Schreibtisch des Schulleiters sitzen und einige Poststücke und die Verteilung der Fachräume für Ihn in den nächsten Minuten übernehmen sollen.



Dies ist ein Bild der Schule.

Navigation: zurück < | fertig | Arbeitsplatz | > weiter

### Beispiel 1

**Fachraum-Plan** Mo | Di | Mi | Do | Fr    **Postfach** email verfassen    1. Aufgabe    2. Aufgabe    3. Aufgabe    **Telefon**

Hallo Thomas!

Ich bräuchte dringend noch für die 10b einen Bio-Fachraum mit Experimentiermöglichkeit; bitte in den letzten beiden Stunden am Montag.

Danke im voraus. Anke

**Klasse 10b (Bio)**  
Experimentieren  
2 Stunden;  
14:20-16:00

Zeit	Fachraum 1 Bio/ Chemie/ Physik Experimentiermöglichkeit	Fachraum 2 Bio/ Chemie/ Physik
8:00 - 8:45	7a (Bio)	7b (Bio) Exp.
8:50 - 9:35	2 Std./ vormittags	2 Std./ vormittags
10:00 - 10:45		
10:50 - 11:35	6a (Che) 10:00-10:45	
11:45 - 12:30		
Mittag	x	x
13:30 - 14:15	9b (Bio) 1 Std., nachm.	
14:20 - 15:05	10b (Bio)	
15:15 - 16:00	2 Std./ 14:20-16:00	

**Vorgehen:**  
Nun würden Sie einfach den Block 9b in die Stunde 13:30-14:15 und den Block 10b in die letzten beiden Stunden in Fachraum 1 ziehen.  
Klicken Sie "weiter >"

Um mit der Einweisung fortzufahren, bitte unten rechts auf "Beispiel verstanden >" klicken.

Navigation: zurück < | fertig | Arbeitsplatz | > weiter

Abbildung 18: Beispiele für die Einweisungsseiten des *webPostkorbs*

#### **4.4.1 Szenario und Einweisung**

Die semantische Einbettung des *webPostkorbs* ist durch den schulischen Rahmen bestimmt. Das Szenario ist so gewählt, dass einige Aufgaben des fiktiven Schulleiters Thomas Thomassen übernommen werden sollen. Dabei sitzt die Testperson vor dem virtuellen Schreibtisch des Schulleiters und ist angehalten, innerhalb einer bestimmten, standardisierten Zeit, die anfallenden Schriftstücke und den Wochenplan zu bearbeiten sowie ggf. E-Mails und Anrufe zu beantworten. Vor der eigentlichen Testbearbeitung erhalten die Testpersonen eine umfangreiche Einweisung in die Aufgabenstellung und in die Bedienung (s. Abbildung 18). Da der *webPostkorb* mit Freiheitsgraden konstruiert ist, kommt der Bedienungseinweisung besondere Bedeutung zu. Es müssen einige unbewertete Übungsaufgaben (s. Abbildung 23) durchlaufen werden. In kleinen Filmen und Animationen wird die Navigation und Bedienung erläutert. Analog zur *webPlanungsaufgabe* ist die Aufgabeninstruktion direkt vor der Testbearbeitung dargestellt, allerdings auch während des Tests immer aufrufbar, um einen möglichen Einfluss der Merkfähigkeit zu minimieren.

#### **4.4.2 Inhaltliche Merkmale**

Die abzubildende Realität, in diesem Falle also die Arbeit von Schulleitungen, wie im zweiten Kapitel beschrieben, bestimmt primär die inhaltlichen Aufgabenmerkmale. Hinsichtlich der semantischen Einbettung wurde der Schulleitungsalltag gewählt. Um die inhaltliche Validität zu sichern, wurden berufstypische Organisations- und Verwaltungsaufgaben genutzt. Oberflächlich handelt es sich dabei um die Bearbeitung von Tabellen, Schaubildern oder Schriftstücken, beispielsweise zur Vorbereitung der anstehenden Gesamtkonferenz. Es wird allerdings angenommen, dass auf der tieferen Ebene zur Lösung der unterschiedlichen Aufgaben eine generellere Problemlösefähigkeit notwendig ist. Die Aufgaben des *webPostkorbs* stellen sich dem Bearbeiter in einer Schreibtisch-Simulation



dar, die dem empirischen Befund des permanenten Zeitdrucks sowie häufig und schnellen Themenwechseln (Bonsen & Pfeiffer, 1998) entspricht.

#### **4.4.3 Operationalisierungen**

Es wurde versucht, die Aufgaben des *webPostkorbs* so zu gestalten, dass die typischen kognitiven Prozesse des Problemlösens aktiviert werden. Die rationale Aufgabenkonstruktion (vgl. Amelang und Zielinski, 2002) ist dabei regelgeleitet erfolgt. Problemlösen wurde innerhalb des *webPostkorbs* aufgrund der Heterogenität des Konstrukts zum einen als Informationsverarbeitung und schlussfolgerndes Denken (analytisches Problemlösen) und zum anderen als Handeln in komplexen, dynamischen Problemstellungen (dynamisches Problemlösen) operationalisiert.

Bei der ersten Operationalisierung handelt es sich um ein kognitiv-analytisches Problemlöseverständnis in der Forschungstradition der einfachen Probleme, wobei das Aufgabenmaterial vor dem Hintergrund des Berliner Intelligenzstrukturmodells (Jäger, 1982, 1984) und hinsichtlich schwierigkeitsbeeinflussender Merkmale systematisch variiert wurde.

Das Problemlösen als Handeln in komplexen, dynamischen Problemstellungen baut ebenfalls auf dem kognitiv-analytisch Problemlöseverständnis auf, jedoch erfolgte hierbei die Operationalisierung von Problemlösen durch verhaltensbasierte Aufgaben in dynamischen Problemsituationen. Dies steht in Bezug zur Forschungstradition des komplexen Problemlösens. Abbildungen 20 bis 22 zeigen Operationalisierungsbeispiele für die verschiedenen Bereiche.

**Tabelle 13: Überblick über die Konzeption des *webPostkorbs* zum Problemlösen**

### Konzeption des *webPostkorbs* zum Problemlösen

- Weiterentwicklung der Postkorbmethode als auswertungsobjektives, standardisiertes, internetbasiertes Verfahren
- Itemstruktur mit Musterlösung
- Schreibtisch-Simulation in Anlehnung an computergestützte Szenarios mit Zeitdruck und Themenwechseln
- semantische Einbettung in den Schulalltag
- Problemlösen als Informationsverarbeitung und schlussfolgerndes Denken
  - analytisches Problemlösen
  - systematische Variation des Aufgabenmaterials nach dem Berliner Intelligenzstrukturmodell von Jäger (1982, 1984) und hinsichtlich schwierigkeitsvariierender Aufgabenmerkmale
- Problemlösen als Handeln in komplexen, dynamischen Problemstellungen
  - dynamisches, komplexes, verhaltensbasiertes Problemlösen
  - systematische Variation der schwierigkeitsvariierenden formalen Aufgabenmerkmale

Die systematische Variation der formalen Aufgabenmerkmale soll ein möglichst breites Spektrum der Schwierigkeiten abbilden. Durch eine rationale Aufgabengestaltung wurde a priori versucht, die unterschiedlichen Schwierigkeiten der Aufgaben durch die Art der vorgegebenen Informationen und die erforderlichen Schlussfolgerungen zu bestimmen. Dabei galt es, die empirisch bestätigten Einflussgrößen des Umfangs und der Vernetztheit so

zu kombinieren, dass sich unterschiedlich schwierige Aufgaben ergeben (s. Abschnitt 2.2.5). Es wird angenommen, dass für die Bearbeitung dieser Aufgaben Fähigkeiten im Problemlösen notwendig und hilfreich sind.

Die Reihenfolge der Bearbeitung innerhalb des *webPostkorbs* ist dem Benutzer - wie typischerweise in der Realität auch - frei gestellt. Alle Aufgaben können während der Simulation überarbeitet, revidiert bzw. rückgängig gemacht werden.

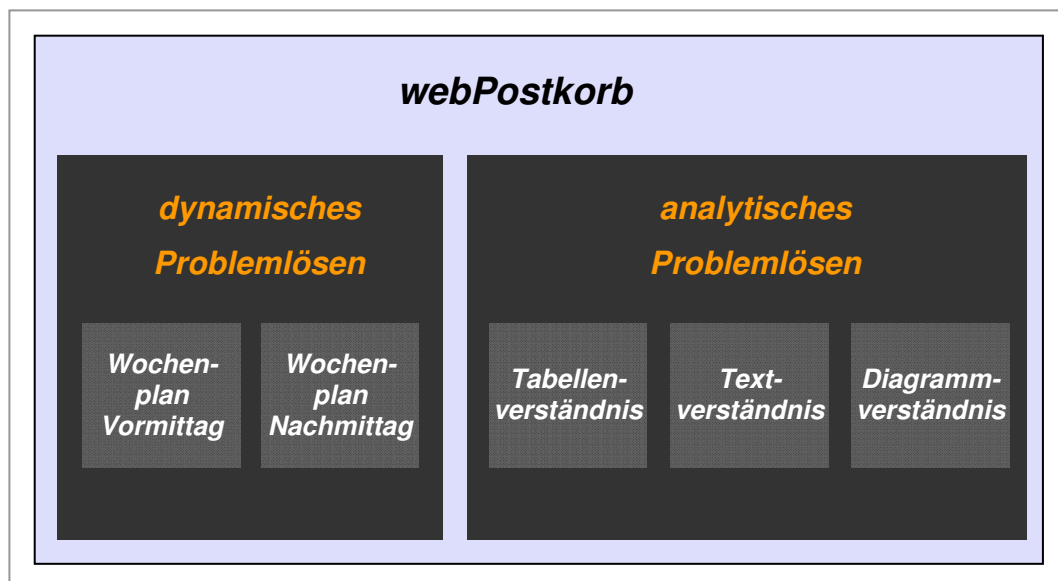
Die Güte der Aufgabenbearbeitung wird als Indikator der Problemlöseleistung verstanden. Die Auswertung erfolgt hinsichtlich der Qualität der jeweiligen Problemlöseleistung. Dies entspricht einer resultatsorientierten Sichtweise. Dabei wird jede richtig beantwortete Frage gewertet.

#### **4.4.4 Innere Struktur**

In der Schreibtisch-Simulation *webPostkorb* sollen aufgrund von inhaltlichen Gesichtspunkten (s. o.) verschiedene Aspekte von Problemlösefähigkeit abgebildet werden. Die innere Struktur des *webPostkorbs* (Endversion) ist in Abbildung 19 dargestellt<sup>16</sup>.

---

<sup>16</sup> Aufgrund der Ergebnisse der Voruntersuchung wurde die innere Struktur des Bereichs *dynamisches Problemlösen* überarbeitet. Die zuerst entwickelten Aufgabengruppen *Umgang mit Störungen* und *Wochenplan* wurden in der Endversion des *webPostkorbs* überführt in die Aufgabengruppen *Wochplan Vormittag* und *Wochenplan Nachmittag* (s. genauer Kapitel 6).



**Abbildung 19: Innere Struktur des *webPostkorbs* (Endversion)**

Der *webPostkorb* ist in fünf Aufgabengruppen unterteilt, die wiederum zu zwei Subskalen zusammengefasst werden. Die Aufgabengruppen *Wochenplan Vormittag* und *Wochenplan Nachmittag* sind inhaltlich der Subskala *dynamisches Problemlösen* zugeordnet, während unter *analytisches Problemlösen* die Aufgabengruppen *Tabellenverständnis*, *Textverständnis* und *Diagrammverständnis* fallen. Es ist anzunehmen, dass im Schulalltag Aufgaben aus allen drei Inhaltsbereichen des Berliner Intelligenzstrukturmodells zu bearbeiten sind (vgl. Bensen & Pfeiffer, 1998), so dass die Inhaltsbereiche bei der Konstruktion der Aufgaben berücksichtigt wurden. Die Variation des Aufgabenmaterials der Subskala *analytisches Problemlösen* gestaltet sich dabei Bezug nehmend auf das BIS-Modell (Jäger, 1982) folgendermaßen:

- numerisches Aufgabenmaterial → Aufgabengruppe *Tabellenverständnis*
- verbales Aufgabenmaterial → Aufgabengruppe *Textverständnis*
- figural-bildhaftes Aufgabenmaterial → Aufgabengruppe *Diagrammverständnis*

**Fachraum-Plan**  
Mo | Di | Mi | Do | Fr

**Postfach**  
email verfassen

1. Schülerzahlen

2. Abi - Was nun?

3. Konferenz

**Telefon**

**Fachraum-Plan: Dienstag (Di)**

Lieber Thomas!

Für die Klasse 10a brauche ich dringend **am Dienstag** noch einen zusätzlichen Raum im Fach Physik am frühen Nachmittag. Bitte möglichst einen Raum ohne Experimentiermöglichkeit; wenn nicht möglich dann mit Experimentieren.

Viele Grüße, Linus

**Klasse 10a (Phy)** möglichst ohne Experimentiermöglichkeit (Exp.)

► 1 Stunde; früh nachmittags

10a (Phy) nachmittags

**Anmerkung:**  
Bitte Blöcke einmal anklicken und langsam per *drag&drop* bewegen (kein Doppelklick).

	Zeit	Fachraum 1 Bio/ Che	Fachraum 2 Bio/ Che/ Phy Exp.	Fachraum 3 Che/ Phy Exp.	Fachraum 4 Che/ Phy
Vormittag	8:00 - 8:45	8a (Bio) 8:00-8:45	8d (Bio) Exp.		9a (Chemie)
	8:50 - 9:35		2 Std./ vormittags		2 Std./ vormittags
	10:00 - 10:45	7a (Bio)	10a (Bio) Exp.	7c (Chemie) Exp.	10b (Physik)
	10:50 - 11:35	2 Std./ vormittags	2 Std./ vormittags	3 Std./ 10:00-12:30	2 Std./ 10:00-11:35
	11:45 - 12:30				
<b>Mittag</b>					
Nachmittag	13:30 - 14:15	7b (Bio) 13:30-14:15	9b(Che)13:30-14:15 Exp	8b (Chemie) Exp.	10c (Chemie)
	14:20 - 15:05			2 Std./ nachmittags	2 Std./ nachmittags
	15:15 - 16:00		8c (Bio) nachm., Exp.		

W Fertlg

Instruktionsfenster öffnen

03.58

**Fachraum-Plan**  
Mo | Di | Mi | Do | Fr

**Postfach**  
email verfassen

1. Schülerzahlen

2. Abi - Was nun?

3. Konferenz

**Telefon**

**Postfach: 3. Konferenz**

Lieber Thomas, hier kommen die angefragten Kennzahlen der letzten Jahre für alle Schulen des Landkreises.

Gruß, Jutta  
(Konrektorin)

**Krankenquote Lehrer/innen**

Jahr	1998	1999	2000	2001	2002
Prozent	4	5	2	6	5

**Diagramm Kennzahlen**

Jahr	Ausgaben (Veränderung gegenüber Vorjahr)	Zahl Lehrerstellen (Veränderung gegenüber Vorjahr)	Zahl Schüler/innen (Veränderung gegenüber Vorjahr)
1998	2.2	0.5	2.8
1999	-0.2	2.0	2.2
2000	0.8	1.0	2.0
2001	1.2	0.5	4.0
2002	2.0	-0.2	2.5

**Sie sollen die Zahlen der letzten Jahre für die Rektorenkonferenz vorbereiten. Welche Aussagen sind richtig oder falsch?**

1. 1999 war ein(e) Lehrer(in) durchschnittlich ca. 5 Tage krank.  richtig  falsch
2. Im Jahr 2000 gab es ein Prozent mehr Lehrerstellen.  richtig  falsch
3. Insgesamt sind zwischen 1998 und 2002 die Lehrerstellen stärker gestiegen als die Ausgaben.  richtig  falsch
4. Der Richtwert der Schullehrkräfte von mind. 90% Übernahme teilt sich im Jahr 2002 erreicht worden.  richtig  falsch
5. Zwischen 2000 und 2001 ist die Zahl der Schüler/innen um ca. 2% gestiegen.  richtig  falsch
6. Die Zahl der Schüler im Jahr 2000 ist etwa doppelt so hoch wie die der Lehrer.  richtig  falsch
7. 2002 wurde mehr Geld pro Schüler/inn ausgegeben als 2001.  richtig  falsch
8. 2000 waren die Ausgaben höher als 1999.  richtig  falsch

**schließen**

W Fertlg

Instruktionsfenster öffnen

01.14

**Abbildung 20: Operationalisierungsbeispiele der Aufgabengruppen *Wochenplan* und *Diagrammverständnis***

Fachraum-Plan

Mo | Di | Mi | Do | Fr

Postfach

email verfassen

1. Schülerzahlen

2. Abi - Was nun?

3. Konferenz

Telefon

### Postfach: 2. Abi - Was nun?

Lieber Herr Thomasen,

Hier kommt der angefragte Text, den Sie für die Vorbereitung des Vortrags "Abitur - Was nun" haben wollten.  
Gruß, Beate (Schulsekretärin)

Für die für Schulabsolventen anstehende persönliche Laufbahnentscheidung bezüglich der Studien- und Ausbildungswahl gehören neben der Information über die Rahmenbedingungen der Ausbildung und späteren Berufsausübung Bücher, Informationsbroschüren, Datenbanken, Webseiten und andere Medienbeiträge sowie persönliche Gespräche und die Auseinandersetzungen mit nahestehenden Menschen zu den entscheidenden Orientierungshilfen. Bei diesem für Schulabsolventen zumeist schwierigen Entscheidungsprozeß sind Angebote der Arbeitsämter als auch der Besuch einschlägiger Beratungsstellen in Schulen und Universitäten, bei Frauen-Beauftragten und teilweise sogar die Nutzung von Angeboten privater Anbieter überlegenswert.

**Sie wollen Knackpunkte aus dem Text ableiten. Welche der folgenden Aussagen gibt Teilaspekte des Textinhalts korrekt wieder?**

1. Im Leben von jungen Menschen gehören Entscheidungsprozesse für oder gegen die Studien- und Berufswahl zu den wichtigsten Abschnitten.  korrekt  nicht korrekt
2. Durch Nutzung von Angeboten privater Anbieter können teilweise einschlägige Beratungsstellen für den schwierigen Entscheidungsprozeß gefunden werden.  korrekt  nicht korrekt
3. Die Angebote privater Anbieter werden von Schulabsolventen teilweise genauso häufig genutzt wie die Angebote der Arbeitsämter.  korrekt  nicht korrekt
4. Schriftliche Informationsmittel sowie das persönliche Gespräch gehören neben Medienbeiträgen zu den wichtigsten Orientierungshilfen beim Besuch von Studien- und Beratungseinrichtungen.  korrekt  nicht korrekt
5. Gespräche mit nahestehenden Menschen sind Orientierungshilfen im zumeist schwierigen Entscheidungsprozeß der persönlichen Laufbahnentscheidung von Schulabsolventen.  korrekt  nicht korrekt
6. Neben den persönlichen Gesprächen sind Informationen aus dem Internet aus Büchern, Zeitungen und Zeitschriften einschlägige Orientierungshilfen.  korrekt  nicht korrekt

[schließen](#)

W

Instruktionenfenster öffnen

04.47

Fertig
Arbeitsplatz

Fachraum-Plan

Mo | Di | Mi | Do | Fr

Postfach

email verfassen

1. Schülerzahlen

2. Abi - Was nun?

3. Konferenz

Telefon

### Postfach: 1. Schülerzahlen

Hallo Thomas,

Du wolltest die Tabelle mit den Schülerzahlen der letzten Jahre haben, weil die Schulaufsichtsbehörde ja plant, bei Schulen mit starkem Ausländeranteil vermehrt Deutsch-Förder-Kurse anzubieten.  
Viele Grüße,  
Beate

	1999	2000	2001	2002
Anzahl ausländischer SchülerInnen	9	16	25	36
Anzahl deutscher SchülerInnen	56	65	75	86
Gesamtzahl	65	81	100	122

Tabelle 1: Zahl neuer Schülerinnen der letzten Jahre  
(unterschieden nach Staatsangehörigkeit)

**Fragen:**

1. Sie sollen der Behörde melden, mit welcher Zahl deutscher und ausländischer SchülerInnen Sie 2003 rechnen, wenn die Entwicklung genauso weitergeht. Bitte tragen Sie die Zahlen in die Tabelle ein.

	2003
Anzahl ausländischer SchülerInnen	
Anzahl deutscher SchülerInnen	
Gesamtzahl	
2. Ihre Schulbehörde unterstützt die Schulen mit hohem Ausländeranteil mit zusätzlichen Deutsch-Förderkursen für Ausländer. In welchem Jahr wurde / wird die Schwelle von 22,5% Ausländeranteil neuer Schüler/innen überschritten?  
 1999     2000     2001     2002     später als 2002
3. Angenommen, die Entwicklung der Schülerzahlen geht über die nächsten Jahre genauso weiter, wie werden sich dann die Zahlen neuer Schüler entwickeln? Wird es in 10 Jahren mehr Deutsche oder ausländische Schüler geben?  
 mehr deutsche Schüler     mehr ausländische Schüler     in etwa gleich viele     kann aus den Daten nicht abgeleitet

[schließen](#)

W

Instruktionenfenster öffnen

00.13

Fertig
Arbeitsplatz

**Abbildung 21: Operationalisierungsbeispiele der Aufgabengruppen *Textverständnis* und *Tabellenverständnis***

The image displays two screenshots of a simulation interface, likely for a training or assessment environment. Both screenshots show a top navigation bar with 'Fachraum-Plan' (Mo | Di | Mi | Do | Fr), 'Postfach' (email verfassen), and 'Telefon' (phone icon). The 'Postfach' section contains a list of tasks: 1. Schülerzahlen, 2. Abi - Was nun?, 3. Konferenz, 4. Email Malern?, and 5. Email Eltern.

**Top Screenshot (15.03):** The main content area shows a text-based task. A speech bubble on the left contains the text: 'Hallo, hier Schneider, Fachleiter Deutsch! Ich werde gleich aufbrechen, um rechtzeitig zur Lehrprobe bei Ihnen zu sein. In welchem Gebäudeteil soll die nun stattfinden?'. The main text area contains a question: 'Meine Antwort: Die Lehrprobe findet im folgenden Teil statt:'. Below this are radio buttons for 'im Nordteil', 'im Ostteil', 'im Südteil', 'im Westteil', and 'Weiß ich nicht'. A red 'Auflegen' button is visible. Below the question is a list of six statements with 'korrekt' and 'nicht korrekt' radio buttons for each. The bottom status bar shows 'Instruktionenfenster öffnen' and the time '15.03'.

**Bottom Screenshot (07.22):** The main content area shows an email composition screen with fields for 'An:', 'Betreff:', and 'Text:'. A blue popup window titled 'Benachrichtigung!' is overlaid on the screen, containing the text: 'Sie haben eine neue Email erhalten! Bitte klicken Sie auf OK und schauen in Ihr Postfach.' Below the text is an envelope icon and an 'O.K.' button. The bottom status bar shows 'Instruktionenfenster öffnen' and the time '07.22'.

Abbildung 22: Operationalisierungsbeispiele der Aufgabengruppe *Umgang mit Störungen*

#### **4.4.5 Benutzerfreundlichkeit der Mensch-Maschine-Interaktion**

Besonderes Augenmerk wurde bei der Programmierung und Gestaltung sowohl des *webPostkorbs* als auch der *webPlanungsaufgabe* auf einen leichten und benutzerfreundlichen Mensch-Maschine-Dialog gelegt, der möglichst noch ein wenig unterhaltend und spielerisch anmutet. Neben der oben beschriebenen Einweisung in die Navigation und dem Grad der Interaktivität (vgl. Schulmeister, 2002) ist dabei die Gestaltung der Oberfläche und der Bedienungselemente entscheidend. So wurde versucht, bekannte Abläufe aus dem Alltagsleben auf die virtuelle Situation zu übertragen, um sie für die Testperson möglichst intuitiv bedienbar zu machen. Störeinflüsse wie eingehende E-Mails und Anrufe sind beispielsweise so programmiert, dass die Bedienung möglichst einfach und alltagsnah erscheint.

Telefonanrufe sind als „Pull-Medium“ zu verstehen. Das bedeutet, dass die angerufene Testperson aktiv (durch Klicken des blinkenden Telefonsymbols) den Anruf annehmen muss. Der Telefonsound wird dabei durch einen visuellen Eindruck verstärkt. Wie in der Realität muss sich die Testperson aktiv entscheiden, das Telefongespräch anzunehmen oder es zu unterlassen. Der Anrufversuch ist nach mehrmaligem Klingeln beendet.

Bei eingehenden E-Mails handelt es sich typischerweise um ein „Push-Medium“. Das bedeutet, dass eingegangene E-Mails jederzeit einsehbar und bearbeitbar sind. Allerdings ist, sobald eine selbstverfasste E-Mail abgeschickt wurde, die E-Mail nicht mehr revidierbar. Diese im Alltag gelernten Abläufe sind in der gleichen Logik innerhalb des *webPostkorbs* gestaltet.

Die Aufgaben des Wochenplans sind mit Hilfe der vorher geübten Methode des „Drag&Drop“ zu lösen. Dabei muss ein Objekt angeklickt und mit der Computermaus positioniert werden. Das bedeutet, dass die Testpersonen die Zeitblöcke in den verschiedenen Plänen frei bewegen und platzieren können.



Die Bedienung des *webPostkorbs* bietet insgesamt einen hohen Grad an Interaktivität verglichen mit herkömmlichen E-Assessment-Verfahren oder anderen computergestützten Simulationsverfahren. In der Taxonomie der Interaktivität von Schulmeister (2002, s. o.) entspräche dies mindestens der Stufe III (die Gesamtbedienung kann auch als Interaktivität der Stufe IV verstanden werden). Die Stufe III ist nach Schulmeister (2002) bedeutend für die Motivation von Benutzern.

#### **4.5 Planen in der *webPlanungsaufgabe***

Planen in der *webPlanungsaufgabe* wird gemäß J. Funke und Fritz (1995) verstanden als verhaltensbasierte Planerstellung und Planausführung. Zur Planerstellung gehören demnach die Teilleistungen Abfolgen erkennen, Randbedingungen erfassen, Zwischenzielbildung, Verfügbarkeit von Alternativen und Angemessenheit der Lösung, während sich die Planausführung in die Bereiche Planüberwachung, Fehlerdiagnostik, Planrevision und Planverwerfung unterteilen lässt. Die *webPlanungsaufgabe* wurde daher so gestaltet, dass die Teilleistungen des Planens zur Bearbeitung der Aufgabe typischerweise nötig sind. Die Besonderheit dabei ist, dass die Aspekte der Planausführung - wie beispielsweise die Revision eines Plans - durch Eingreifen der Testperson möglich sind.

##### **4.5.1 Szenario und Einweisung**

Die *webPlanungsaufgabe* ist in eine berufstypische Rahmengeschichte eingebettet. In der Einweisung erfährt die Testperson Genaueres über das Szenario. Es wird erklärt, dass der Proband für die nächsten Minuten bitte die Aufgaben des fiktiven Schulleiters Thomas Thomasen übernehmen solle. Herr Thomasen sei Leiter eines Ganztagsgymnasiums in einer mittelgroßen Stadt und leider plötzlich erkrankt. Da in der kommenden Woche die Projektwoche anstehe, müsse umgehend die Planung für die Medienräume durchgeführt werden.

Gruppe X	Gruppe Y
10 SchülerInnen	20 SchülerInnen
▶ 3 Stunden	▶ 2 Stunden
X (3 Std.)	Y (2 Std.)

Zeit	Raum 1 10 Plätze	Raum 2 20 Plätze
1. Stunde		
2. Stunde		
3. Stunde		x

Keine Gruppe bewegt!

Sie sollen nun zwei Gruppen auf 2 Räume verteilen.

Um die Gruppen auf die Räume zuzuteilen, müssen sie per **"drag&drop"** die jeweiligen Blöcke in die Räume ziehen.

**Beispiel:**  
Sie haben erkannt, dass Gruppe x mit 10 Schüler/innen in den Raum 1 von der 1. bis zur 3. Stunde zugeteilt werden kann.

Vorgehen:

- Klicken Sie mit der linken Maustaste auf die **Mitte des Blocks** Gruppe X und lassen Sie die Maustaste gedrückt.
- Ziehen Sie die Gruppe X möglichst genau auf die freien Felder im Raum 1.
- Maustaste loslassen ▶ der Raum ist belegt.
- Sie können die Kästen frei in der Tabelle bewegen oder auch zurück zum Ausgangspunkt.

▶ Unterhalb der Tabelle wird Ihnen angezeigt, ob die Gruppen richtig platziert sind.

Haben Sie die Bedienung verstanden?  
Wenn ja, dann klicken Sie bitte auf "weiter".

Gruppe A	Gruppe B	Gruppe C	Gruppe D	Gruppe E	Gruppe F	Gruppe G	Gruppe H
20 SchülerInnen mit TV ▶ 4 Stunden; vormittags	30 SchülerInnen ohne TV ▶ 1 Stunde; vormittags	30 SchülerInnen ohne TV ▶ 3 Stunden; 10:00-12:30 Uhr	28 SchülerInnen ohne TV ▶ 2 Stunden; früh vormittags	18 SchülerInnen ohne TV ▶ 1 Stunde; 13:30-14.15 Uhr	10 SchülerInnen ohne TV ▶ 3 Stunden; irgendwann	20 SchülerInnen mit TV ▶ 2 Stunden; vormittags	20 SchülerInnen mit TV ▶ 1 Stunde; 8:00-8:45 Uhr
A (4 Std.)	B (1 Std.) vorm.	C (3 Std.)	D (2 Std.)	E (1 Std.) 13:30-14:15	F (3 Std.)	G (2 Std.)	H (1 Std.) 8:00-8:45
vormittags	▶ 2 Stunden; nachmittags	10:00 - 12:30	früh vormittags	▶ 1 Stunde; letzte Stunde, um einen Film anzusehen	irgendwann	vormittags	▶ 1 Stunde; nachmittags; es kommen noch 5 Schüler dazu
	B (2 Std.) nachmittags		▶ 1 Stunde;				H (1 Std.) nachm.

Bitte versuchen Sie, die Zeitblöcke der acht Gruppen per **drag & drop** auf die fünf Räume zu verteilen. Dabei sind die Vorgaben wie Gruppengröße, Zeit, TV-Bedarf zu berücksichtigen. Erst wenn Sie alle Gruppe verteilt haben, sollen Sie bitte den Test selbstständig beenden, in dem Sie "Fertig" anklicken!

Außerdem sind folgende Dinge zu beachten:

- Die *kleinen* Räume (Raum 4 und Raum 5 mit 20 Plätzen) sollen möglichst ausgenutzt werden.
- Die Gruppen sollen möglichst in *einem* Raum bleiben können, wenn sie vormittags und nachmittags Kurse haben.
- Raum 1* soll so wenig wie möglich genutzt werden, weil dort später noch weitere Gruppen hineinsollen.

Instruktionsfenster schliessen

	Raum 4 10 Plätze	Raum 5 20 Plätze	Raum 6 10 Plätze	Raum 7 20 Plätze	Raum 8 10 Plätze	Raum 9 20 Plätze
10:00 - 11:30						
11:45 - 12:30						
Mittag	x	x	x	x	x	x
Nachmittag						
13:30 - 14:15						
14:20 - 15:05						
15:15 - 16:00						

Abbildung 23: Übungsaufgabe und geöffnetes Instruktionsfenster während der Verfahrensbearbeitung der *webPlanungsaufgabe*

Vor der eigentlichen Bearbeitung erhalten die Testpersonen eine umfangreiche Einweisung in die Aufgabenstellung und in die Bedienung. Dabei sind einige Übungsaufgaben (s. Abbildung 23) zu durchlaufen.

Die Testperson erfährt, dass sowohl die Qualität des erarbeiteten Plans als auch die dafür benötigte Zeit bewertet wird. Die genaue Aufgabeninstruktion erscheint direkt vor der Verfahrensbearbeitung, sie ist allerdings auch während des Tests immer aufrufbar, um einen möglichen Einfluss der Merkfähigkeit zu minimieren.

#### **4.5.2 Inhaltliche Merkmale**

Die Semantik ist analog zum *webPostkorb* ebenfalls in den Rahmen des Schulalltags von pädagogischen Führungskräften eingebettet. Um die inhaltliche Validität zu gewährleisten, wurde eine typische Planungsaufgabe aus dem Schulalltag weiterentwickelt: der Stunden- oder Raumplan. Oberflächlich handelt es sich somit um die Erstellung eines Raumplans. Auf der tieferen Ebene ist zur Lösung der *webPlanungsaufgabe* typischerweise Planungsfähigkeit notwendig. Aus den Interviews mit den Schulleitungen zeigt sich, dass der Schulstundenplan zumeist in den Schulferien - mit nur wenig Zeitdruck durch das Tagesgeschäft - erstellt wird. Die Konzeption der *webPlanungsaufgabe* beinhaltet daher keine Zeitbeschränkung und keine dynamischen Störeinflüsse.

#### **4.5.3 Operationalisierungen**

Um eine möglichst breite Facette von Planungsanforderungen innerhalb des Verfahrens abzudecken, wurde die Schwierigkeit der Aufgabenmerkmale a priori systematisch variiert hinsichtlich räumlicher, zeitlicher, materieller und logischer Randbedingungen. Wichtige Eigenschaften des Umfelds von Planungskonstellationen gemäß von der Weth und Strohschneider (2002) sind umgesetzt worden, wie die unterschiedliche Anzahl von transparenten

und weniger transparenten Vorgaben und Zielen. Abgeleitet aus den empirischen Forschungsergebnissen wurden folgende formale Einflussgrößen, die die Schwierigkeit einer Planungsaufgabe bestimmen, variiert (vgl. J. Funke & Fritz, 1995; J. Funke & Krüger, 1995; Pascha et al., 2001):

- Vorhandensein zeitlicher Vorgaben
- Intransparenz / Zielvorgabe
- Notwendigkeit des Findens von Alternativlösungen
- unterschiedliche Randbedingungen
- Zwang zur Unvollständigkeit

In der *webPlanungsaufgabe* müssen per „Drag&Drop“ Zeitblöcke unter Berücksichtigung unterschiedlicher Randbedingungen in den Zeitplan gebracht werden. Dabei sind zum Teil konfligierende Ziele zu vereinbaren oder Alternativlösungen zu finden. Nicht alle Blöcke können perfekt untergebracht werden (Zwang zur Unvollständigkeit, vgl. J. Funke & Krüger, 1995). Die Reihenfolge der Bearbeitung innerhalb der *webPlanungsaufgabe* ist dem Benutzer - wie typischerweise in der Realität auch - frei gestellt. Alle Blöcke können während der Simulation zurückgesetzt oder anders platziert werden. Die Auswertung basiert auf der Bewertung der Güte pro platzierten Zeitblock. Dabei ist entscheidend, gegen wie viele Randbedingungen die gefundene Lösung jeweils verstößt. Dem liegt der Gedanke zu Grunde, dass in der beruflichen Realität schlecht zugeordnete Stunden besondere Härten bedeuten. Beispielsweise wird ein Lehrer, der eine Stunde mit Mediennutzung geplant hat, in einem Raum ohne mediale Einrichtungen zwar unterrichten können, es bedeutet jedoch eine gewisse Umstellung. Eine Stunde, die allerdings zeitlich und räumlich völlig falsch platziert wird, ist kaum zielgerecht abzuhalten. Aufgrund der Abhängigkeit der einzelnen Blöcke wird ein Gesamtwert gebildet. Insgesamt wird dabei die Qualität der Aufgabenbearbeitung als Indikator der Planungsleistung in der *webPlanungsaufgabe* verstanden.

Analog zum Vorgehen beim *webPostkorb* wurde auch bei der *webPlanungsaufgabe* versucht, die Mensch-Maschine-Interaktion möglichst benutzerfreundlich, interaktiv und unterhaltend zu gestalten. So wohnt der Bearbeitung der *webPlanungsaufgabe* insgesamt ein hoher Grad an Interaktivität inne, der in der Taxonomie von Schulmeister (2002) die Stufe III erreichen würde.

**Tabelle 14: Überblick über die Konzeption der *webPlanungsaufgabe* zum Planen**

### Konzeption der *webPlanungsaufgabe* zum Planen

- auswertungsobjektives, internetbasiertes Simulationsverfahren mit Musterlösung
- Schreibtisch-Simulation mit semantischer Einbettung in den Schulalltag
- Planen als verhaltensbasierte Planerstellung und Planausführung (mit den dazugehörigen Teilleistungen)
- Aktivierung der typischen kognitiven Prozesse des Planens
- systematische Variation schwierigkeitsvariierender formaler Aufgabenmerkmale, wie räumlicher, zeitlicher, materieller und logischer Randbedingungen



## 5 Fragestellung und Hypothesen, Messinstrumente, Evaluationsvorgehen

Im Mittelpunkt dieses Kapitels stehen die Darstellung der Fragestellung und Hypothesen, die Vorstellung der eingesetzten Messinstrumente sowie der Überblick über das Evaluationsvorgehen und die durchgeführten Studien.

### 5.1 Fragestellung und Hypothesen

Die im Rahmen dieser Arbeit durchgeführten Untersuchungen verfolgen die Überprüfung der Fragestellung, inwieweit die hier entwickelten E-Assessment-Verfahren diagnostischen Qualitätsstandards im Sinne der Gütekriterien gerecht werden. Dabei interessieren aufgrund des gewählten Konstruktionsansatzes insbesondere die Inhalts- und Konstruktvalidierung sowie Befunde zur Objektivität, Reliabilität, Benutzerfreundlichkeit, Fairness und Akzeptanz.

Hinsichtlich der Prüfung der Konstruktvalidität liegt das Hauptaugenmerk auf dem Zusammenhang zu Intelligenztestleistungen. Daher kommen verschiedene Intelligenztestverfahren zum Einsatz, und die Zusammenhänge sollen an verschiedenen Stichproben überprüft werden. Der Forschungsstand zum Zusammenhang zwischen Intelligenztestleistungen und Problemlösen und Planen sowie die daraus abgeleiteten Implikationen für diese Arbeit wurden bereits im Abschnitt 2.2 genauer dargestellt, daher wird im Folgenden nur auf die wesentlichsten Aspekte fokussiert.

Es wird angenommen, dass Intelligenz einen Einfluss auf die Fähigkeiten im Planen und Problemlösen hat, so dass für diese Arbeit ein signifikanter, zumindest mittlerer Zusammenhang zwischen den Leistungen in Intelligenztests und den Leistungen in der *webPlanungsaufgabe* und im *webPostkorb* erwartet wird. Dabei wird angenommen, dass die Subskalen des *webPostkorbs* unterschiedlich hoch mit der Intelligenztestleistung zusammenhängen. Abgeleitet aus der Forschung wird erwartet, dass die Skala *analytisches Problemlösen* des *webPostkorbs* substantiell und höher

mit Intelligenztestleistung korreliert als die Skala *dynamisches Problemlösen*, für die ebenfalls ein überzufälliger, wenngleich kleinerer Zusammenhang angenommen wird. Aufgrund der definitorischen Nähe der Begriffe „Planen“ und „Problemlösen“ wird erwartet, dass die Leistungen in den Online-Verfahren *webPostkorb* und *webPlanungsaufgabe* ebenfalls signifikant korrelieren.

Hinsichtlich des Zusammenhangs der Handlungs- und Lageorientierung (Kuhl, 1994) mit den Leistungen in der *webPlanungsaufgabe* und im *webPostkorb* wird Folgendes erwartet: Es ist anzunehmen, dass die Bearbeitung der beiden internetbasierten Verfahren im Allgemeinen von den Testpersonen als Belastungssituation - wenngleich nicht als bedrohlich - wahrgenommen wird. In einer solchen Situation gehen nach Kuhl (1994) handlungsorientierte Personen effizienter mit der Affektregulation um. Diese Fähigkeit wird insbesondere durch die Skala „Grad der Entscheidungs- und Handlungsplanung“ (HOP) erfasst.

Es ist somit für die Untersuchungen zu erwarten, dass der Grad der Entscheidungs- und Handlungsplanung zumindest einen indirekten Einfluss auf Leistungen im *webPostkorb* und in der *webPlanungsaufgabe* hat.

Hinsichtlich der Zusammenhänge mit der HOM-Skala lassen sich im Vorhinein keine eindeutigen Richtungen formulieren. Kuhl (1981) kann zwar zeigen, dass Leistungsdefizite nach Misserfolgsinduktion nur bei Lageorientierten auftreten, doch eine solche Induktion liegt hier nicht vor. Somit ist der systematische Zusammenhang der Leistungsmaße mit der HOM-Skala im Sinne einer „gelernten Hilflosigkeit“ (vgl. Kuhl, 1994) nicht anzunehmen. Für die Korrelation zwischen der Handlungsorientierung bei der Tätigkeitsausführung (HOT) und den Leistungsmaßen in den internetbasierten Verfahren lässt sich ebenfalls keine eindeutige Richtung im Vorwege bestimmen. Kuhl und Wassiljew (1985) geben an, dass HOT als ein Maß für (intrinsische) Tätigkeitszentrierung versus Zielorientierung aufgefasst werden kann. Ob eine Tätigkeitszentrierung gegenüber einer Zielorientierung innerhalb des *webPostkorbs* oder der *webPlanungsaufgabe* einen systematischen Vorteil bedeuten würde, lässt sich allerdings nicht genau



postulieren. Für die Skalen HOM und HOT sind somit a priori keine Richtungen der Korrelationen vorherzusagen.

Im Theorieteil wurde dargelegt, dass Planungs- und Problemlösefähigkeit als generelle, eignungsrelevante Personmerkmale von Führungskräften eingeschätzt werden können. Daher wird hinsichtlich der Kriteriumsvalidität angenommen, dass bewährte Führungsnachwuchskräfte in den hier entwickelten E-Assessment-Verfahren signifikant besser abschneiden als Novizen.

Hinsichtlich der Akzeptanz wird aufgrund des umfangreichen Entwicklungsprozesses erwartet, dass die beiden Simulationsverfahren von den Teilnehmern überwiegend als gelungen wahrgenommen werden. Einige Autoren weisen auf die überlegene Akzeptanz simulativer Verfahren hin (u. a. U. Funke, 1995a; Höft & U. Funke, 2001). Daher wird hier angenommen, dass die Simulationsverfahren im Vergleich mit herkömmlichen Testverfahren höhere Akzeptanz bei den Testpersonen finden.

## 5.2 Messinstrumente

Neben den im vorangegangenen Kapitel vorgestellten E-Assessment-Verfahren *webPostkorb* und *webPlanungsaufgabe* kommen im Rahmen der Vor- und Hauptuntersuchung weitere Verfahren zur Messung der Intelligenz, des Personmerkmals „Handlungs- und Lageorientierung“ und der Computererfahrung und -einstellung zum Einsatz. Aufgrund der besonderen Bedeutung des Intelligenzeinflusses für die vorliegende Arbeit werden diesbezüglich unterschiedliche Verfahren berücksichtigt. Der folgende Abschnitt stellt die Messinstrumente kurz vor. Erweiterte Kennwerte der Stichproben der Vor- und Hauptuntersuchung, die für die Auswertung entscheidend sind, werden im Rahmen der Ergebnisdarstellung diskutiert.

### **5.2.1 Intelligenztestverfahren**

#### *Berliner Intelligenzstruktur-Test (BIS-K-S) von Jäger, Süß und Beauducel (1997)*

Der BIS-4 Test ist eine Neuentwicklung von Jäger et al. (1997), die nach dem Ansatz der strukturellen Testtheorie von Cattell (1971) vorgenommen wurde. Die fähigkeitstheoretische Grundlage des Tests bildet das integrativ konzipierte, hierarchische und bimodale Berliner Intelligenzstrukturmodell von Jäger (1982, 1984), das im Abschnitt 2.2.2 vorgestellt wurde. Der Berliner Intelligenzstruktur-Test (Jäger et al., 1997) gilt allgemein als etabliertes, valides Intelligenztestverfahren und kam im Rahmen der Problemlöseforschung vermehrt zum Einsatz. Dabei zeigen sich insbesondere zur operativen Fähigkeit Verarbeitungskapazität (K) des Berliner Intelligenzstrukturmodells (Jäger, 1982) substantielle Zusammenhänge (s. Kapitel 2). Süß (1996) zieht aus seinen umfangreichen Untersuchungen zum Einfluss der Intelligenz auf das Problemlösen in computergestützten Szenarios das Fazit, dass „die spezifischere Fähigkeit *Verarbeitungskapazität*, wie erwartet, höher und konsistenter mit den Steuerungsleistungen korrelierte als die *Allgemeine Intelligenz*.“ (S. 142, Hervorhebungen im Original). Daher ist es angezeigt, diese operative Fähigkeit im Rahmen der Hauptuntersuchung zu messen. Es wird aus Gründen der Ökonomie die Kurzversion BIS-K-S zur Erfassung der Verarbeitungskapazität verwendet. Dabei kommen folgende sechs Aufgaben zum Einsatz:

- Zahlenreihe (ZN)
- Schätzen (SC)
- Wortanalogien (WA)
- Tatsache-Meinung (TM)
- Analogien (AN)
- Figurenauswahl (FA)

Jeweils zwei Aufgaben sind den Inhaltsbereichen „numerisch“, „verbal“ und „figural-bildhaft“ zugeordnet. Die figurale Aufgabe „Charkow“ wird aus

inhaltlichen und ökonomischen Gründen durch die Aufgabe Figurenauswahl (FA) ersetzt.

*Intelligenz-Struktur-Test (I-S-T 70) von Amthauer (1970)*

Mit dem Intelligenz-Struktur-Test von 1970 liefert Amthauer einen mehrdimensionalen Intelligenztest. Angedachte Einsatzfelder sind Berufsberatung und Eignungsdiagnostik. Der Test besteht aus insgesamt neun Aufgabengruppen, von denen im Rahmen dieser Arbeit aus ökonomischen Gründen folgende vier Aufgaben eingesetzt werden.

- Analogien (AN) zur Messung der Kombinationsfähigkeit
- Gemeinsamkeiten (GE) zur Messung sprachlicher Abstraktionsfähigkeit
- Rechenaufgaben (RA) zur Messung praktisch-rechnerischen Denkens
- Figurenauswahl (FA) zur Messung der Vorstellungsfähigkeit

Laut Amthauer (1970) werden dabei mit den einzelnen Aufgabengruppen unterschiedliche Fähigkeitsaspekte erfasst. Vorteile des I-S-T 70 sind neben befriedigenden Gütekriterien die Ökonomie, Überschaubarkeit und Anwenderfreundlichkeit (Brocke, Beauducel & Tasche, 1998). Gerade aufgrund der beachtlichen Kriteriumsvaliditäten (vgl. Schmidt-Atzert & Deter, 1993) eignet sich der I-S-T 70 zum Einsatz im Rahmen der Eignungsdiagnostik und zur hier anvisierten Aufdeckung eines möglichen Zusammenhangs zu den Verfahren *webPostkorb* und *webPlanungsaufgabe*. Kritik wird dagegen geäußert hinsichtlich des Umstands, dass durch die unterschiedliche Zahl von Skalen die Fähigkeitsbereiche im I-S-T-Maß für Allgemeine Intelligenz unterschiedlich gewichtet werden. Die verkürzte Version des I-S-T 70 wird hier im Rahmen der Voruntersuchung eingesetzt.

*Intelligenz-Struktur-Test (I-S-T 2000 R) von Amthauer, Brocke, Liepmann und Beauducel (2001)*

Mit der Überarbeitung des I-S-T 70 zum I-S-T 2000 R wurde versucht, die Kritik am alten Verfahren zu berücksichtigen und den Test hinsichtlich des aktuellen Stands der Intelligenzstrukturforschung anzupassen. Dabei basiert

der I-S-T 2000 R auf einem neuen, theoretisch begründeten und empirisch bestätigten Strukturkonzept (Amthauer et al., 2001). So sind sechs der ursprünglich neun Aufgabengruppen in überarbeiteter Form erhalten geblieben (Analogien, Gemeinsamkeiten, Zahlenreihen, Figurenauswahl, Satzergänzung und Würfelaufgaben). Es wurden zusätzlich drei Aufgabengruppen (Vorzeichenaufgaben, Rechenaufgaben ohne verbalen Anteil und Matrizenaufgaben) sowie ein Test zum Allgemeinwissen neu konstruiert. Jede Fähigkeit wird im I-S-T 2000 R mit mehreren Aufgabentypen erfasst. Durch dieses Vorgehen soll die kritisierte Gleichsetzung einzelner Aufgabentypen mit interessierenden Fähigkeiten vermieden werden. Das Grundmodul des I-S-T 2000 R kommt im Rahmen der Hauptuntersuchung dieser Arbeit zum Einsatz.

### **5.2.2 Handlungs- und Lageorientierung**

Es ist anzunehmen, dass der Erfolg im Planen und Problemlösen nicht ausschließlich von den kognitiven Fähigkeiten eines Probanden abhängt, sondern auch von nicht-kognitiven Personmerkmalen. Für die Leistungen in computergestützten Szenarios und Postkorbverfahren konnte diese Annahme in empirischen Studien bisher nicht eindeutig bestätigt werden. Inwieweit die Handlungs- und Lageorientierung nach Kuhl (1994) Einfluss auf die Leistungen im *webPostkorb* und in der *webPlanungsaufgabe* hat, soll im Rahmen der Hauptuntersuchung überprüft werden. Nach Kuhl (1994) betrifft das Konstrukt „Handlungs- und Lageorientierung“ die Fähigkeit, „in Belastungssituationen reduzierten handlungsbahnenden Affekt herauf zu regulieren (HOP = Selbstmotivierung) und in bedrohlichen Situationen negativen Affekt herab zu regulieren (HOM = Selbstberuhigung)“ (S. 3). Demnach beeinflussen die Komponenten der Handlungsorientierung die Effizienz der Affektregulation, d. h. die Fähigkeit, vorhandene affektregulatorische Kompetenzen unter Belastung bzw. Stress einzusetzen.

Nach Kuhl (1995) gilt das Verfahren als geeignet, um eignungs- und leistungsdiagnostisch entscheidende Personmerkmale zu messen. Im HAKEMP-90 werden folgende drei Subskalen betrachtet:

- Handlungsorientierung nach Misserfolgserfahrung (HOM)
- Grad der Entscheidungs- und Handlungsplanung (HOP)
- Handlungsorientierung bei der Tätigkeitsausführung (HOT)

Das Verfahren kommt hier im Rahmen der Hauptuntersuchung zum Einsatz.

### **5.2.3 Computererfahrung und Computereinstellung**

Einige Studien zeigen, dass die Computererfahrung und Computereinstellung die Leistungen in computergestützten Verfahren beeinflussen können (u. a. M. Kersting, 1999; Klieme et al., 2001; Süß, 1996). Diese Variablen wurden bisher jedoch nur recht selten und zumeist mit nur einem einzigen Item erfasst. Die Annahme hinsichtlich des Einflusses der Computererfahrung ist, dass Vorerfahrungen im Computerumgang zur Entlastung der kapazitätsbegrenzten kognitiven Ressourcen führen und dadurch die Leistungen in computergestützten Verfahren begünstigen (M. Kersting, 1999). Es wird bei der Verfahrenskonstruktion zwar versucht, durch die Darbietung einer umfangreichen Einführung mit Übungsbeispielen Unterschiede in der Vorerfahrung auszugleichen, inwieweit aber die Computererfahrung trotzdem einen Einfluss auf die Leistungen hat, soll im Rahmen der Vor- und Hauptuntersuchung überprüft werden. Zur Messung der Computererfahrung werden den Testteilnehmern dabei in Anlehnung an die Untersuchungen von M. Kersting (1999) elf neu konstruierte Fragen zur Computererfahrung vorgelegt, die die Vorerfahrung und die Vertrautheit mit verschiedenen Computerprogrammen zum Inhalt haben.

Zur Erfassung der Computereinstellung kommen Ausschnitte des Inventars zur Computerbildung (INCOBI) von Richter, Naumann und Groeben (2001) zum Einsatz.

Im INCOBI sollen die Testpersonen einschätzen, inwieweit sie den Computer als nützliches Werkzeug bzw. unbeeinflussbare Maschine beim Lernen und Arbeiten erleben. Auch die positiven bzw. negativen gesellschaftlichen und kulturellen Auswirkungen der Computertechnologie werden dabei erhoben. Zur Validierung legten Richter, Naumann und Groeben (2001) das Instrument einer Stichprobe von 101 Personen mit hoher Computer-Anwendungsexpertise sowie 51 Personen mit niedriger Expertise vor. Es lassen sich deutliche Differenzen zwischen den Gruppen der Anwendungsexperten und -novizen sowie in der Gesamtstichprobe signifikante Korrelationen mit Maßen der tatsächlichen Computernutzung finden. Nach Einschätzung der Autoren hat sich das Inventar dabei als reliables, konstrukt- und kriteriumsvalides Instrument erwiesen.

Für die Stichprobe der Hauptuntersuchung konnte hier für die elf eingesetzten Items zur Computererfahrung ein Cronbachs-Alpha-Wert in Höhe von .72 und für die sechs Items zur Erfassung der Computereinstellung in Höhe von .67 ermittelt werden. Die Computereinstellung ist dabei zu  $r = .33$  mit der Computererfahrung korreliert.

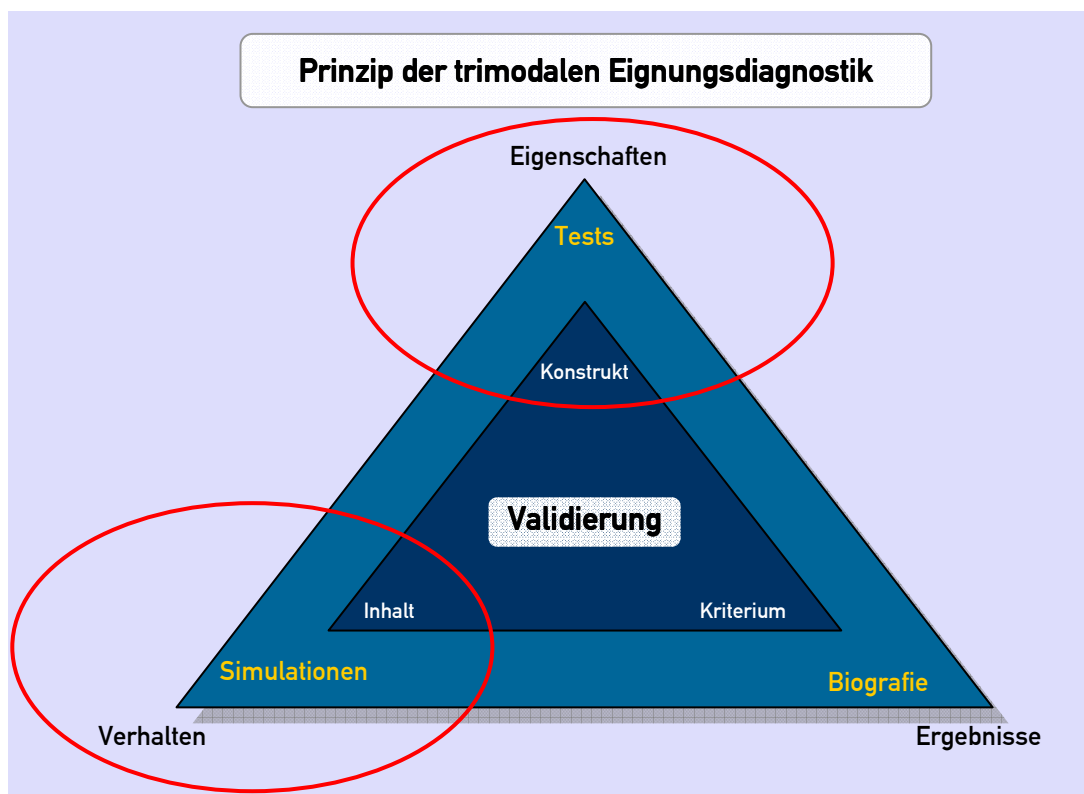
#### **5.2.4 Eigener Fragebogen**

Mit einem eigenen Fragebogen sollen Aspekte der Benutzerfreundlichkeit, Akzeptanz, Inhalts- oder Augenscheinvalidität und Fairness erfasst werden. Dabei interessiert insbesondere, wie die entwickelten Verfahren aus Benutzersicht bewertet werden. Der Fragebogen enthält eine Reihe von Aussagen, die die Testpersonen auf einer fünfstufigen Skala von „stimmt gar nicht“ bis „stimmt völlig“ einschätzen.

Die interne Konsistenz (Cronbachs-Alpha) für das Gesamturteil zum *webPostkorb* beträgt .73 und zur *webPlanungsaufgabe* .80 (Ergebnisse der Hauptuntersuchung).

### 5.3 Evaluationsvorgehen

Aus der Validierungslogik<sup>17</sup> der gewählten eignungsdiagnostischen Ansätze leitet sich das Evaluationsvorgehen ab (s. rot markierte Bereiche in Abbildung 24). Wie in der Abbildung dargestellt, bilden die Inhaltsvalidierung und die Konstruktvalidierung die Hauptziele.



**Abbildung 24: Validierungslogik der verschiedenen eignungsdiagnostischen Ansätze (aus Schuler & Höft, 2001, S. 95, erweitert um die roten Markierungen)**

Die Inhaltsvalidierung ist im Wesentlichen über eine umfassende, anforderungsgeleitete Verfahrenskonstruktion zu gewährleisten (vgl. von Rosenstiel, 2000; Schuler, 2001; Wottawa, 1995). Der hier anvisierte inhaltsvalide, anforderungsbezogene Entwicklungsprozess der Verfahren beinhaltet verschiedene, regelgeleitete Schritte, die im Mittelpunkt der Darstellung des vierten Kapitels standen (s. o.).

<sup>17</sup> Das Prinzip der trimodalen Eignungsdiagnostik von Schuler und Höft (2001) wurde in Kapitel 3 dargestellt und soll daher hier nicht weiter erläutert werden.

Zur Evaluation der Testgüte der Verfahren und insbesondere zur Konstruktvalidierung wurde eine ganze Reihe von Studien durchgeführt. Tabelle 15 gibt einen Überblick über die Evaluationsstudien zur Konstruktion der *webPlanungsaufgabe* und des *webPostkorbs*. Dabei werden die Stichprobengröße, die eingesetzten Verfahren bzw. die erhobenen Daten sowie die jeweiligen Ziele der Studien vorgestellt.

Die Voruntersuchung dient insbesondere dazu, möglichst erste Ergebnisse zur Aufgabengüte und zur Benutzerfreundlichkeit zu liefern, um die Verfahren weiter zu optimieren. Ziel der Hauptuntersuchung ist die Gewinnung von Befunden zur Testgüte, insbesondere zur Konstruktvalidität.

Es wird versucht, die Untersuchungsbedingungen zu variieren. So sollen die universitären Stichproben die Verfahren im Rahmen eines Self-Assessments jeweils einmal online von zu Hause aus und einmal im kontrollierten Setting in der Universität durchlaufen. Die Teilnehmer erhalten dabei per E-Mail ein individuelles Passwort. Innerhalb eines Zeitraums von einer Woche können sie sich in das Programm einloggen und die Verfahren bearbeiten.

Die *webPlanungsaufgabe* (Version Wirtschaft) wird zudem unter Ernstfallbedingungen im Sinne eines E-Assessment-Einsatzes durchgeführt. Dies ist von besonderer Bedeutung, da die meisten Untersuchungen lediglich den Ernstfall simulieren, jedoch keine tatsächliche Bewerbungssituation betrachten.



**Tabelle 15: Überblick über die Evaluationsstudien zur Entwicklung der *webPlanungsaufgabe* und des *webPostkorbs***

Entwicklungsschritt / Studie	N	eingesetzte Verfahren / erhobene Daten	Ziele
Voruntersuchung - Stichprobe <i>webPlanungsaufgabe</i>	35	<ul style="list-style-type: none"> <li>• <i>webPlanungsaufgabe</i></li> <li>• Logfiles, Tracking</li> <li>• I-S-T 70</li> <li>• Abiturnote, Zwischenprüfungsnote</li> <li>• Fragen zur Akzeptanz, Zufriedenheit</li> <li>• Fragen zu Computereinstellung und Computererfahrung</li> </ul>	<ul style="list-style-type: none"> <li>• Aufgabenkonstruktion</li> <li>• Benutzerfreundlichkeit, Akzeptanz und Fairness</li> <li>• Schwierigkeitsanalyse</li> <li>• erste Überprüfung der psychometrischen Qualität</li> <li>• Zusammenhang mit Intelligenz</li> </ul>
Voruntersuchung - Stichprobe <i>webPostkorb</i>	38	<ul style="list-style-type: none"> <li>• <i>webPostkorb</i> (Erstversion)</li> <li>• Logfiles, Tracking</li> <li>• I-S-T 70</li> <li>• Abiturnote, Zwischenprüfungsnote</li> <li>• Fragen zur Akzeptanz, Zufriedenheit</li> <li>• Fragen zu Computereinstellung und Computererfahrung</li> </ul>	<ul style="list-style-type: none"> <li>• Aufgabenkonstruktion</li> <li>• Benutzerfreundlichkeit, Akzeptanz und Fairness</li> <li>• Schwierigkeitsanalyse</li> <li>• erste Überprüfung der psychometrischen Qualität</li> <li>• Zusammenhang mit Intelligenz</li> </ul>
Expertenbefragung	7	<ul style="list-style-type: none"> <li>• <i>webPlanungsaufgabe</i></li> <li>• <i>webPostkorb</i> (Erstversion)</li> <li>• offene Rückmeldungen von Schulleitungsmitgliedern</li> <li>• offene Rückmeldungen von E-Assessment-Experten</li> </ul>	<ul style="list-style-type: none"> <li>• Aufgabenkonstruktion</li> <li>• Benutzerfreundlichkeit</li> <li>• Inhaltsvalidität</li> </ul>
Programmierung der internetbasierten Endversionen der <i>webPlanungsaufgabe</i> (Version Wirtschaft) und des <i>webPostkorbs</i> sowie Beta-Testungen			
Hauptuntersuchung - Stichprobe <i>webPostkorb</i>	66	<ul style="list-style-type: none"> <li>• <i>webPostkorb</i> (Endversion)</li> <li>• Logfiles, Tracking</li> <li>• BIS-K-S</li> <li>• HAKEMP-90</li> <li>• Abiturnote, Zwischenprüfungsnote</li> <li>• Fragen zur Akzeptanz, Zufriedenheit</li> <li>• Fragen zu Computereinstellung und Computererfahrung</li> </ul>	<ul style="list-style-type: none"> <li>• Konstruktvalidierung (Handlungs- und Lageorientierung, Wiederholung der Prüfung des Zusammenhangs mit Intelligenz)</li> <li>• Benutzerfreundlichkeit, Akzeptanz und Fairness</li> <li>• Kriteriumsvalidierung (Schul- und Studienerfolg)</li> </ul>
Hauptuntersuchung - Stichprobe <i>webPlanungsaufgabe</i>	45	<ul style="list-style-type: none"> <li>• <i>webPlanungsaufgabe</i></li> <li>• <i>webPostkorb</i> (Endversion)</li> <li>• Logfiles, Tracking</li> <li>• Abiturnote, Zwischenprüfungsnote</li> <li>• BIS-K-S</li> <li>• HAKEMP-90</li> <li>• Fragen zur Akzeptanz, Zufriedenheit</li> <li>• Fragen zu Computereinstellung und Computererfahrung</li> </ul>	<ul style="list-style-type: none"> <li>• Konstruktvalidierung (Handlungs- und Lageorientierung, Wiederholung der Prüfung des Zusammenhangs mit Intelligenz)</li> <li>• Kriteriumsvalidierung (Schul- und Studienerfolg)</li> </ul>
Hauptuntersuchung - Stichprobe <i>webPlanungsaufgabe</i> (Version Wirtschaft)	369	<ul style="list-style-type: none"> <li>• <i>webPlanungsaufgabe</i> (Version Wirtschaft)</li> <li>• E-Assessment zu kognitiver Leistungsfähigkeit</li> <li>• Telefoninterview (N = 178)</li> <li>• I-S-T 2000 R (N = 41)</li> <li>• Assessment-Center (N = 41)</li> <li>• Abiturnote, Diplomnote</li> </ul>	<ul style="list-style-type: none"> <li>• Konstruktvalidierung (kognitive Leistungsfähigkeit, soziale &amp; motivationale Personmerkmale, Wiederholung der Prüfung des Zusammenhangs mit Intelligenz)</li> <li>• Ermittlung der inkrementellen Validität gegenüber kogn. Leistungsfähigkeit</li> <li>• Benutzerfreundlichkeit, Akzeptanz und Fairness</li> <li>• Kriteriumsvalidierung (Schul-, Studienerfolg, Erfolg in Personalauswahlverfahren)</li> </ul>
Hauptuntersuchung - Stichprobe <i>webPlanungsaufgabe</i> (Wirtschaft)	19	<ul style="list-style-type: none"> <li>• <i>webPlanungsaufgabe</i> (Version Wirtschaft)</li> </ul>	<ul style="list-style-type: none"> <li>• Kriteriumsvalidierung (Extremgruppenvergleich bewährte Führungsnachwuchskräfte vs. Bewerber)</li> </ul>

Zur Evaluation der Testgüte des *webPostkorbs* und der *webPlanungsaufgabe* kommen die Verfahren der klassischen Testtheorie zum Einsatz. Von der Anwendung eines Testmodells im Rahmen der probabilistischen Testtheorie kann nach Rost (1996) abgesehen werden, wenn sich auf Grund der Vorgabe eines festen Zeitintervalls auch die Anzahl der gelösten Aufgaben als eine metrische Variable verstehen lassen. Die Durchführung der Aufgaben des *webPostkorbs* kann als ein solcher Anwendungsfall verstanden werden, daher wird hier von der Modellprüfung abgesehen (vgl. Schuler & Klingner, 2005). Hinsichtlich der Aufgaben der *webPlanungsaufgabe* ist eine Anwendung eines Testmodells ebenfalls nicht angezeigt, da die einzelnen Aufgaben nicht unabhängig voneinander sind. Dies ist allerdings eine Voraussetzung für die Modellprüfung.

Ein Nebeneffekt der Anwendung der Verfahren der klassischen Testtheorie zeigt sich darin, dass die Evaluationsergebnisse ggf. einfacher mit Ergebnissen zur Testgüte anderer Testverfahren verglichen werden können. Die Verfahren, die nach der klassischen Testtheorie evaluiert werden, stellen weiterhin die absolute Mehrheit aller Verfahren dar. Rost (1999, S. 140) schätzt, dass etwa 95 Prozent aller Verfahren auf der klassischen Testtheorie aufbauen. Näheres zur Kritik an der klassischen Testtheorie findet sich u. a. bei Rost (1996).

Die Ergebnisse der in Tabelle 15 dargestellten Studien sind Mittelpunkt der Ausführungen der nächsten beiden Kapitel.

## 6 Ergebnisse der Voruntersuchung

Für die Voruntersuchung kamen die *webPlanungsaufgabe* und die Testvorform des *webPostkorbs* zum Einsatz. Ziel der Studie ist eine erste Überprüfung der anvisierten inneren Struktur des *webPostkorbs* sowie der Bewertung der beiden Verfahren aus Benutzersicht (s. Tabelle 15). Auch der erwartete Zusammenhang mit der Intelligenztestleistung wird in diesem Rahmen untersucht. Die Befunde sollen helfen, die Testvorformen für den Einsatz im Rahmen der Hauptuntersuchung weiter zu optimieren. Um die psychometrische Qualität des eingesetzten Intelligenztestverfahrens für die vorliegende Stichprobe sicherzustellen, werden dafür ebenfalls Skalenkennwerte berechnet. Die Ergebnisse der Voruntersuchung zum *webPostkorb* und zur *webPlanungsaufgabe* werden im Folgenden jeweils im gleichen Abschnitt dargestellt, da sich die Befunde zumeist auf beide Verfahren beziehen.

### 6.1 Stichproben

#### *webPlanungsaufgabe*

An der Voruntersuchung zur *webPlanungsaufgabe* nahmen 35 Studierende der Studiengänge Lehramt, Diplom-Pädagogik und Sozialpädagogik der Universität Lüneburg teil. Das mittlere Alter der 31 weiblichen und vier männlichen Testpersonen beträgt 26.6 Jahre (SD = 7.2). Die mittlere Testleistung in den vier Aufgaben des eingesetzten Intelligenztests I-S-T 70 (Amthauer, 1970) liegt bei einem standardisierten Wert von 103. Dies entspricht einem Prozentrang von 60 in Bezug auf die Altersnorm für Abiturienten (Amthauer, 1970). Die Erhebung fand im Juni 2002 in den Computerräumen der Universität statt und erfolgte freiwillig und unentgeltlich. Die Probanden erhielten per E-Mail eine individuelle Rückmeldung über ihre Ergebnisse.

#### *webPostkorb*

Die Besonderheit der Durchführung der Voruntersuchung zum *webPostkorb* ist, dass die Studierenden sich von zu Hause oder von den Computerräumen

der Universität aus zu einer von ihnen gewünschten Zeit einloggen und die Aufgaben bearbeiten konnten. Dieser Teil der Untersuchung fand somit - dem Ziel der Variation der Untersuchungsbedingungen folgend - online statt. An der Voruntersuchung zum *webPostkorb* nahmen dabei 23 der 35 Studierenden aus der Untersuchung zur *webPlanungsaufgabe* teil. Dies entspricht einem Anteil von 66 Prozent. Signifikante Unterschiede zwischen den zwölf Abbrechern und den 23 Teilnehmern des zweiten Teils der Voruntersuchung in Bezug auf Alter, I-S-T 70 Testleistungen und Leistungen in der *webPlanungsaufgabe* konnten nicht festgestellt werden (T-Test sowie Levene-Test alle  $p$ -Werte  $> .05$ ). Darüber hinaus bearbeiteten 15 weitere Studierende der Studiengänge Lehramt, Diplom-Pädagogik und Sozialpädagogik der Universität Lüneburg ausschließlich den *webPostkorb* und einen anschließenden Fragebogen. Das den jeweiligen Analysen zu Grunde liegende  $N$  wird an den entsprechenden Stellen berichtet. Das mittlere Alter der 38 Studierenden der Stichprobe *webPostkorb* (25 weiblich und 13 männlich) beträgt 27.7 Jahre ( $SD = 7.0$ ). Die Erhebung fand im Juni 2002 statt und erfolgte freiwillig und unentgeltlich. Die Probanden erhielten per E-Mail eine individuelle Rückmeldung über ihre Ergebnisse.

### *Expertenbefragung*

Es konnten zu den beiden Verfahren *webPostkorb* und *webPlanungsaufgabe* offene Rückmeldungen von Schulleitungsmitgliedern und ausgewählten E-Assessment-Experten gewonnen werden. Insgesamt sieben Personen gaben Einschätzungen hinsichtlich der Aufgabenkonstruktion, Benutzerfreundlichkeit und Inhaltsvalidität.

## **6.2 Itemkennwerte**

Aufgrund der unterschiedlichen Konstruktionen der zwei Verfahren kommt nur der *webPostkorb* mit seiner Struktur für eine Betrachtung auf Itemniveau in Frage. Die *webPlanungsaufgabe* wurde als simulative Gesamtaufgabe konzipiert, so dass die verschiedenen zu bewältigenden Teilaufgaben nicht

unabhängig voneinander sind. Die Beantwortung eines Aufgabenteils hat gewisse Auswirkungen auf die Beantwortung anderer Aufgabenteile. Lokale stochastische Unabhängigkeit ist jedoch die Voraussetzung für eine Berechnung von itembezogenen Kennwerten (vgl. Kubinger, 1993). Die Bestimmung der Itemschwierigkeit und Itemtrennschärfe sowie der internen Konsistenz mit Hilfe von Cronbachs-Alpha ist somit nicht sinnvoll durchzuführen. Daher erfolgt die Datenanalyse der *webPlanungsaufgabe* bezogen auf den Summenscore. Die Betrachtung der Ergebnisse auf Skalenniveau ist Gegenstand der Darstellung des Abschnitts 6.3.

#### *Itemkennwerte webPostkorb*

Die Qualität der Items des *webPostkorbs* soll hinsichtlich der Kriterien Schwierigkeit, Trennschärfe, interne Konsistenz und Dimensionalität überprüft werden (vgl. Lienert & Raatz, 1998). Wie unter Abschnitt 4.4 ausgeführt, ist der *webPostkorb* der Voruntersuchung inhaltlich in fünf Aufgabengruppen eingeteilt, die wiederum zwei Subskalen zugeordnet sind. Die Mittelwerte, Standardabweichungen und Trennschärfen der Items des *webPostkorbs* der Voruntersuchung gibt Tabelle 16 wieder. In der letzten Spalte sind die Cronbachs-Alpha-Werte der Gesamtskala abgebildet, die sich ergäben, wenn das jeweilige Item nicht mit einbezogen würde. Die Werte für die Itemschwierigkeiten mit Inangriffnahmekorrektur bewegen sich in der Spanne zwischen .16 und .76 bei einer mittleren Schwierigkeit von .51. Dabei liegen über 70 Prozent der Werte zwischen .34 und .68. Somit wird im mittleren Bereich der Schwierigkeitsskala die größte Differenziertheit erlangt. Dies kann als positiver Befund gewertet werden.

Die Werte der Standardabweichungen streuen zwischen .37 bis .51 bei einem mittleren Wert von .47. Für die part-whole korrigierten Trennschärfen werden Größen zwischen .15 und .82 ermittelt. Der Mittelwert aller Trennschärfen liegt dabei bei .50 bezogen auf die Aufgabengruppe.

Tabelle 16: Itemkennwerte des *webPostkorbs* (Voruntersuchung)

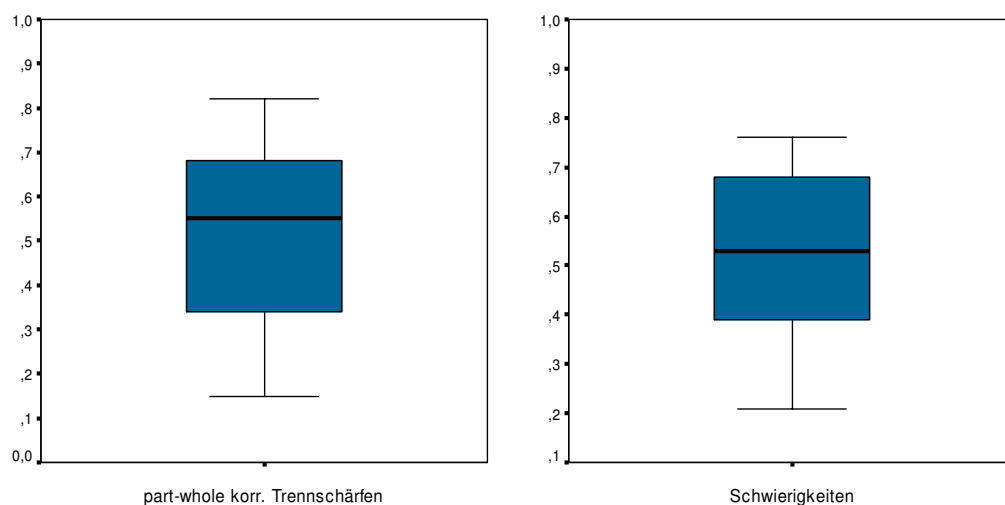
Subskala	Aufgaben- gruppe	Item	Mittel- wert	SD	Trenn- schärfe	Alpha ohne	
ana- lytisches Problem- lösen	Tabellen- verständnis	Item_1	.61	.50	.54	.84	
		Item_2	.61	.50	.45	.84	
		Item_3	.53	.50	.67	.84	
		Item_4	.39	.50	.33	.85	
		Item_5	.37	.49	.72	.83	
	Text- verständnis	Item_1	.76	.43	.68	.84	
		Item_2	.68	.47	.78	.83	
		Item_3	.68	.47	.56	.84	
		Item_4	.61	.50	.68	.84	
		Item_5	.68	.47	.82	.84	
		Item_6	.50	.51	.60	.84	
		Item_7	.39	.50	.52	.85	
	Diagramm- verständnis	Item_1	.76	.43	.69	.84	
		Item_2	.68	.47	.71	.84	
		Item_3	.24	.43	.61	.84	
		Item_4	.34	.48	.35	.84	
		Item_5	.50	.51	.62	.84	
		Item_6	.21	.41	.60	.84	
		Item_7	.16	.37	.40	.85	
	dyna- misches Problem- lösen	Wochen- plan	Item_1	.71	.46	.25	.84
			Item_2	.68	.47	.17	.84
Item_3			.53	.51	.33	.84	
Item_4			.29	.46	.49	.84	
Umgang mit Störungen		Item_1	.42	.50	.17	.85	
		Item_2	.76	.43	.15	.86	
		Item_3	.68	.47	.34	.85	

Anmerkung: N = 38.

Es wird deutlich, dass die Aufgabengruppe *Umgang mit Störungen* die niedrigsten Itemtrennschärfen aufweist mit einer mittleren korrigierten Trennschärfe von .22, während in den Aufgabengruppen der Subskala *analytisches Problemlösen* hohe Itemtrennschärfen mit einer mittleren korrigierten Trennschärfe von .60 für die Subskala zu finden sind.

Mit Ausnahme der Aufgabengruppe *Umgang mit Störungen* ergeben sich somit zumeist sehr hohe Trennschärfen (vgl. Beurteilungsrichtlinien für Testkennwerte nach Fisseni, 1997).

Die Verteilung der Mittelwerte (Schwierigkeiten) der Items wird im Boxplot (Abbildung 25, rechts) grafisch verdeutlicht. Die farbige Fläche wird durch das 25 und 75 Prozent Perzentil begrenzt, während die durchgezogene Linie in der Box den Median beschreibt. Die oberen und unteren horizontalen Linien außerhalb der farbigen Fläche kennzeichnen die Spannweite der Verteilung.



**Abbildung 25: Boxplots der korrigierten Trennschärfen und Schwierigkeiten des *webPostkorbs* (Voruntersuchung, N = 38)**

Die Trennschärfen des *webPostkorbs* werden in Abbildung 25 ebenfalls als Boxplot dargestellt. Hinsichtlich der part-whole korrigierten Trennschärfen und Schwierigkeiten ergeben sich somit insgesamt befriedigende Werte (vgl. Fissini, 1997). Lediglich die Itemreihenfolge nach aufsteigender Schwierigkeit muss in zwei Fällen angepasst werden.

### 6.3 Skalenkennwerte

In der Tabelle 17 sind die deskriptiven Skalenkennwerte für den *webPostkorb* und für die *webPlanungsaufgabe* wiedergegeben.

**Tabelle 17: Deskriptive Statistiken des *webPostkorbs* und der *webPlanungsaufgabe* (Voruntersuchung)**

Subskala	Aufgaben- gruppe	N	Mittel- wert	SD	Schiefe	Mittlere Iteminter- korrelation		Cronbachs Alpha	
<i>ana- lytisches Problem- lösen</i>	<i>Tabellen- verständnis</i> (5 Items)	38	2.50	1.78	.06	.39	.26	.77	.87
	<i>Text- verständnis</i> (7 Items)		3.63	2.25	-.69	.54		.87	
	<i>Diagramm- verständnis</i> (7 Items)		4.53	1.78	.14	.39		.84	
<i>dyna- misches Problem- lösen</i>	<i>Wochenplan</i> (4 Items)	38	2.84	1.48	.02	.21	.17	.51	.58
	<i>Umgang mit Störungen</i> (3 Items)		2.84	1.24	-.58	.16		.37	
Gesamt- score <i>webPK</i>		38	15.39	5.58	-.31	.20		.85	
Summen- score <i>webPLA</i>		35	16.80	3.15	-.93				
benötigte Zeit <i>webPLA</i>			408.03	170.41	1.16				

Anmerkung: *webPK*: *webPostkorb*; *webPLA*: *webPlanungsaufgabe*.



*Rohwerteverteilung, Homogenität und interne Konsistenz des webPostkorbs*

Der Mittelwert des Gesamtscores des *webPostkorbs* liegt bei 15.39 Punkten (Median 15 Punkte) mit einer Standardabweichung von 5.58. Die Schiefe der Verteilung weicht für die Stichproben der Voruntersuchung nicht signifikant von einer symmetrischen Verteilung ab.

Hinsichtlich der Überprüfung der Homogenität schlagen Bortz und Döring (1995) sowie Lienert und Raatz (1998) die Berechnung der mittleren Iteminterkorrelationen vor. Wenngleich der *webPostkorb* nicht als homogener Test konzipiert wurde, sondern als Kombination aus Konstrukt- und Simulationsverfahren, so spräche eine möglichst hohe mittlere Iteminterkorrelation innerhalb der Aufgabengruppen bzw. Subskalen für die anvisierte innere Struktur des Verfahrens. Die für die Aufgabengruppen des *analytischen Problemlösens* ermittelten Werte zwischen .39 und .54 bewegen sich im guten Bereich - selbst im Verhältnis zu rein konstruktbezogenen Instrumenten (s. Tabelle 17, vgl. Briggs & Cheek, 1986). Für die Aufgabengruppen *Wochenplan* und *Umgang mit Störungen* ergeben sich niedrigere Werte von .21 bzw. .16. Da Problemlösen hier hinsichtlich der zwei Teilkonstrukte *dynamisches Problemlösen* und *analytisches Problemlösen* operationalisiert wird, ist es von Interesse zu überprüfen, inwieweit diese Teilkonstrukte hohe Homogenitäten auf Subskalen-Ebene aufweisen. Briggs und Cheek (1986) schlagen für eindimensionale Instrumente einen Akzeptanzbereich für Gesamttest-Homogenitäten von .20 bis .40 vor. Nach den Autoren gilt für diesen Bereich, dass ein Konstrukt hinreichend homogen gemessen wird, ohne dass die inhaltliche Breite durch übermäßige Redundanz eine zu große Einschränkung erfährt. Für die Subskala *dynamisches Problemlösen* kann eine mittlere Iteminterkorrelation von .17 ermittelt werden. Dieser Wert liegt somit etwas unterhalb der vorgeschlagenen Grenze von Briggs und Cheek (1986), während sich die mittlere Iteminterkorrelation der Subskala *analytisches Problemlösen* hingegen mit .26 innerhalb des Akzeptanzbereichs bewegt. Auch die mittlere Iteminterkorrelation liegt für den Gesamtscore mit .20 noch innerhalb der von Briggs und Cheek (1986) vorgeschlagenen Spanne für Gesamttest-Homogenitäten für eindimensionale Instrumente.

Die interne Konsistenz des *webPostkorbs* - berechnet mit Hilfe von Cronbachs-Alpha - liegt für das Gesamtverfahren mit  $\alpha = .85$  im guten Bereich für ein simulatives Testverfahren. Die Reliabilitätsschätzungen mittels Cronbachs-Alpha bewegen sich für die einzelnen Aufgabengruppen zwischen .37 und .87. Die internen Konsistenzen erweisen sich als gut für die drei Aufgabengruppen zum *analytischen Problemlösen* (Werte zwischen .77 und .87, Gesamt-Subskalenwert von .87), während die Cronbachs-Alpha-Werte für die Skalen *Umgang mit Störungen* und *Wochenplan* mit .37 und .51 nicht zufrieden stellend sind. Allerdings gilt es dabei auch die geringe Zahl von Items pro Aufgabengruppe zu berücksichtigen. Bei einer Verlängerung der Aufgabengruppe *Wochenplan* auf 16 Items würde nach der Spearman-Brown-Formel eine interne Konsistenz von  $\alpha_{16} = .81$  resultieren.

#### *Dimensionsanalyse webPostkorb*

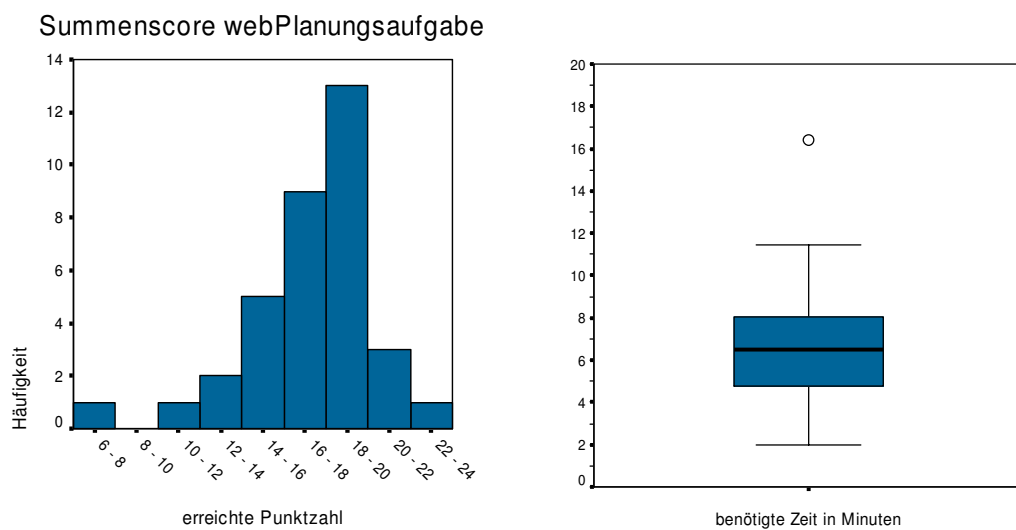
Zur Gewinnung weiterer Hinweise, inwieweit die anvisierte innere Struktur des *webPostkorbs* mit den Daten der Voruntersuchung vereinbar ist, soll eine faktorenanalytische Dimensionsanalyse (Hauptkomponentenanalyse mit Varimax-Rotation) berechnet werden. Die Voraussetzungen der Berechnung einer Faktorenanalyse sind Gegenstand des Abschnitts 7.6, so dass hier auf eine Darstellung verzichtet werden kann. Die Berechnung ergibt eine Zwei-Faktorenlösung bei einem Eigenwert größer als eins mit einer Trennung in *dynamisches Problemlösen* und *analytisches Problemlösen*. Durch die zwei Faktoren werden dabei 63 Prozent der Varianz erklärt. Die Aufgabengruppen laden jeweils auf nur einem Faktor hoch. Die Ausnahme stellt die Aufgabengruppe *Umgang mit Störungen* dar, die auf beiden Faktoren hohe Ladungen vorweist.

Dieser Befund unterstützt das Ergebnis der Skalenanalyse, wonach die Gruppe *Umgang mit Störungen* die Teilleistung im *dynamischen Problemlösen* nicht zuverlässig erfasst. Für die anderen Aufgabengruppen - insbesondere zur Subskala *analytisches Problemlösen* - können demgegenüber in der Voruntersuchung positive Befunde ermittelt werden, die auf eine zuverlässige Abbildung des angedachten Aufbaus des *webPostkorbs* hinweisen.

Insgesamt erweisen sich die Befunde zu Skalenkennwerten und innerer Struktur des *webPostkorbs* somit als gut. Lediglich die Kennwerte für die Aufgabengruppe *Umgang mit Störungen* deuten hier auf einen Revisionsbedarf hin. Diese Aufgabengruppe soll daher für die Testendform überarbeitet und in die Aufgabengruppe *Wochenplan* integriert werden. Dadurch sind für die Hauptuntersuchung die sieben Items der Subskala *dynamisches Problemlösen* zu bearbeiten. Fünf Items der Subskala müssen lediglich leicht überarbeitet werden, um sie inhaltlich auf die neue Aufgabengruppe *Wochenplan Nachmittag* anzupassen. Zwei Items der Aufgabengruppe *Umgang mit Störungen* lassen sich nicht durch Anpassung integrieren und werden daher neu entwickelt.

#### *Rohwerteverteilung der webPlanungsaufgabe*

Die Verteilung der Rohwerte der *webPlanungsaufgabe* hat ein Mittel von 16.8 Punkten (Median = 17 Punkte) bei einer Standardabweichung von 3.15 (s. Tabelle 17). Die Verteilung erweist sich als leicht rechtssteil bei einem Schiefe-Wert von -.93. Die Teilnehmer benötigen im Mittel sechs Minuten und 48 Sekunden für die Aufgabe (SD = 170.41 Sek.). Die Abbildung 26 veranschaulicht die Ergebnisse grafisch.



**Abbildung 26: Histogramm der Punkterohwerte und Boxplot der benötigten Zeit *webPlanungsaufgabe* (Voruntersuchung, N = 35)**

Auffällig sind die deutlich linkssteile Verteilung (Schiefe-Werte von 1.16) und die relativ große Spannbreite der benötigten Zeit von zwei bis über sechzehn Minuten (im Boxplot als Extremwert mit einem Kreis dargestellt). Dies liegt an der Konzeption des Verfahrens. Die Testpersonen können die Aufgabe selbständig beenden. Allerdings zeigt sich auch, dass der größte Teil der Teilnehmer (über drei Viertel) die Aufgabe innerhalb einer Zeit von vier bis acht Minuten absolviert.

#### *Skalenkennwerte und Rohwerteverteilungen des I-S-T 70*

Für die vier eingesetzten Aufgaben des Intelligenz-Struktur-Tests (I-S-T 70) von Amthauer (1970) wurden ebenfalls die Verteilungen der Rohwerte sowie Skalenkennwerte berechnet, um zu überprüfen, inwieweit die Daten für die hier angestrebten Zusammenhangsanalysen geeignet erscheinen (s. Abbildung 27).

Die Mittelwerte der Untertests rangieren in der Stichprobe der Voruntersuchung zwischen 9.3 (Rechenaufgaben) und 12.3 Punkten (Gemeinsamkeiten), während die Standardabweichungen zwischen 3.22 (Gemeinsamkeiten) und 3.87 (Figurenauswahl) liegen. Die Streuungen bewegen sich dabei im Bereich der Eichstichprobe (Amthauer, 1970).

Ein Verlust an Varianz, der sich auf die Korrelationen zu anderen Leistungsmaßen auswirken könnte, ist somit - trotz der Zusammensetzung der Stichprobe mit überwiegend weiblichen Studierenden des Lehramts, der Sozialpädagogik und der Diplom-Pädagogik - nicht zu erwarten.

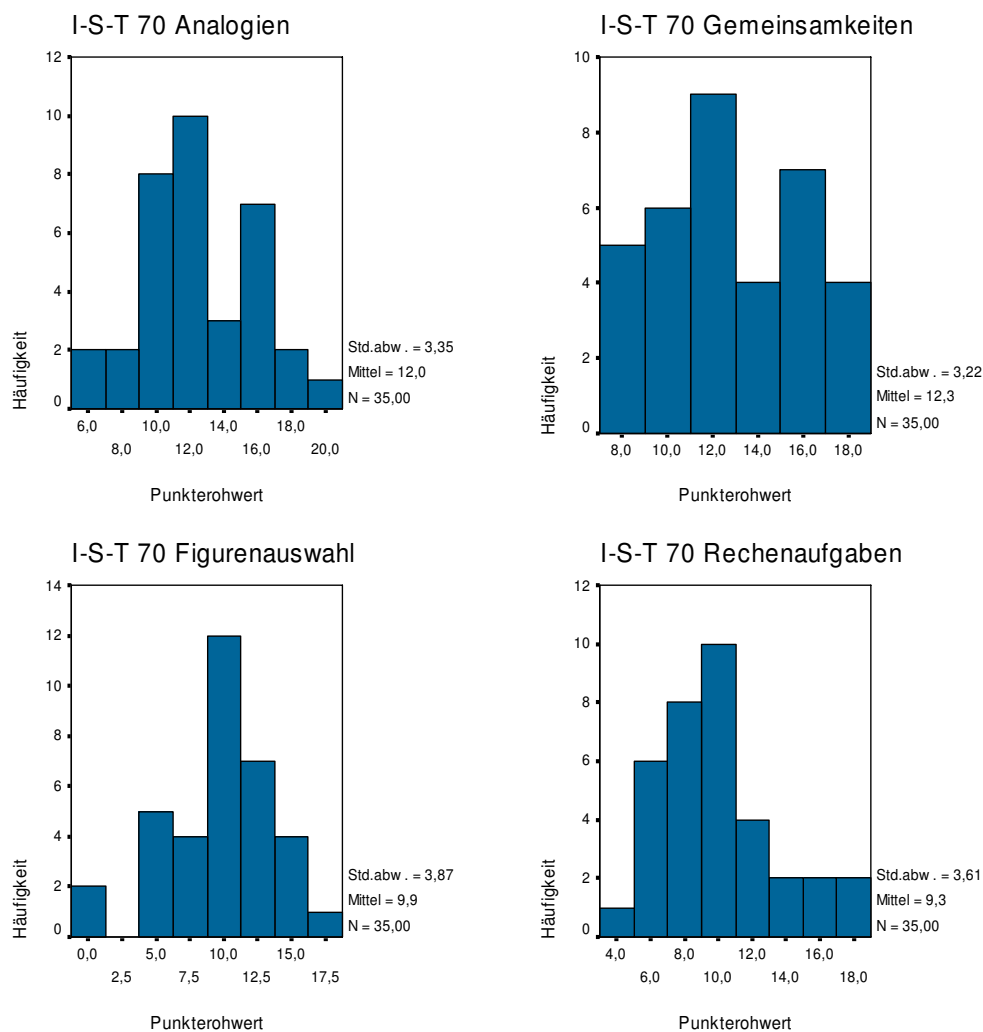


Abbildung 27: Rohwertverteilungen der vier I-S-T 70 Aufgaben (Voruntersuchung)

## 6.4 Benutzerfreundlichkeit, Akzeptanz und Fairness

### *Ergebnisse des Trackingverfahrens und der Logfile-Analyse*

Neben der oben vorgestellten benutzerorientierten Verfahrensentwicklung soll die anvisierte Benutzerfreundlichkeit der Testvorformen im Rahmen der Voruntersuchung auch empirisch untersucht werden. Dies erfolgt mit Hilfe eines zählpixelbasierten Trackingverfahrens sowie der Analyse der Protokolldateien (Logfiles). Auf diese Weise werden die Datensätze aller Teilnehmer der Voruntersuchung protokolliert. Sowohl beim *webPostkorb* als auch bei der *webPlanungsaufgabe* wurden keine Abbrecher während der

Bearbeitung registriert. Dies kann als erster Hinweis auf eine zumutbare Verfahrensdurchführung gewertet werden.

Das Verfahren *webPostkorb* ist mit einigen Freiheitsgraden konstruiert, so dass es zu überprüfen gilt, inwieweit die Teilnehmer die verschiedenen Aufgaben erkannt haben. Anhand der individuellen Verläufe ist ersichtlich, dass alle Aufgaben des *webPostkorbs* von den Probanden selbständig geöffnet wurden. Dieses Ergebnis spricht für die anvisierte intuitive und leicht erlernbare Bedienung.

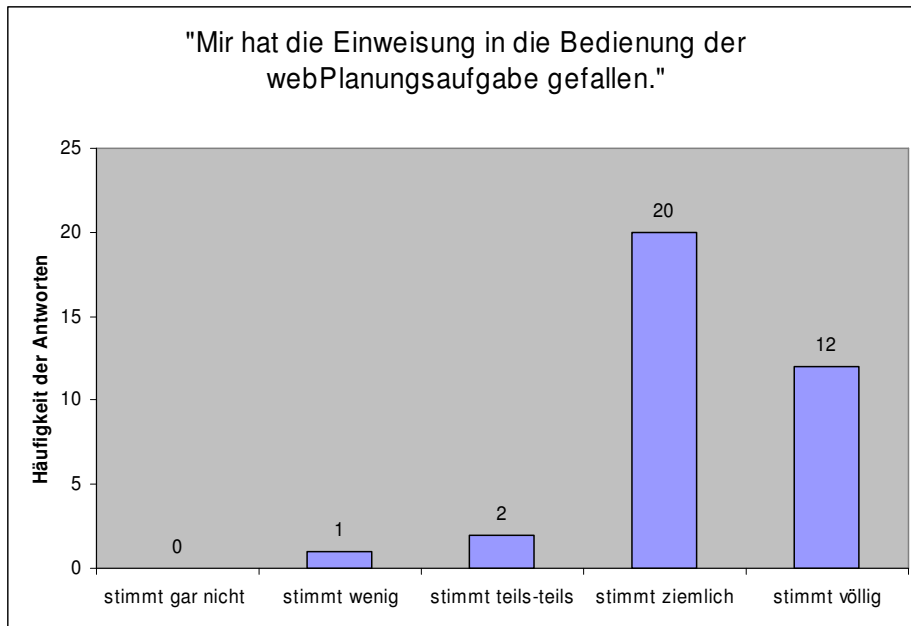
#### *Expertenbefragung*

Eine Reihe von ausgewählten E-Assessment-Experten hat die Testvorformen ebenfalls bewertet. Dadurch konnten offene Rückmeldungen von Spezialisten der internetgestützten Eignungsdiagnostik gewonnen werden. Darüber hinaus wurden einige Schulleitungsmitglieder um offene Rückmeldungen gebeten.

Die Ergebnisse zeigen, dass die Verfahren *webPostkorb* und *webPlanungsaufgabe* insgesamt als gelungen bewertet werden. Einzelne Anregungen zur Gestaltung der Einweisungsseiten des *webPostkorbs* fanden in der Testendform Berücksichtigung.

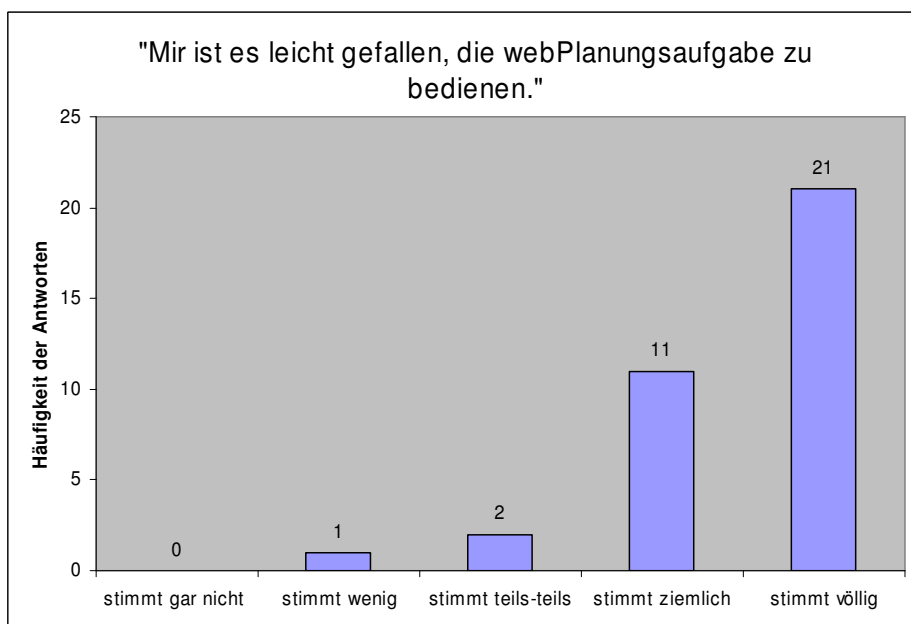
#### *Ergebnisse der Befragung der Testpersonen der Stichprobe webPlanungsaufgabe*

Es gilt hinsichtlich der angestrebten benutzerfreundlichen Verfahrensentwicklung zu überprüfen, inwieweit sich die hier gewählte Vorgehensweise in der Einschätzung der Benutzer widerspiegelt. Daher sollten die Testteilnehmer im Anschluss an die Bearbeitung die Verfahren aus Benutzersicht bewerten. Sowohl Einschätzungen zur Akzeptanz, Verständlichkeit und Zufriedenheit als auch die weiter unten dargestellten Ergebnisse zur Testfairness und der subjektiv wahrgenommenen Schwierigkeit wurden erhoben. Wie oben aufgeführt, erhalten die Testteilnehmer vor der eigentlichen Testbearbeitung eine ausführliche Erläuterung der Bedienung und Navigation. Insofern ist es interessant zu erfahren, wie die Benutzer diesen Aspekt einschätzen.



**Abbildung 28: Bewertung der Bedienungseinweisung der webPlanungsaufgabe (Voruntersuchung)**

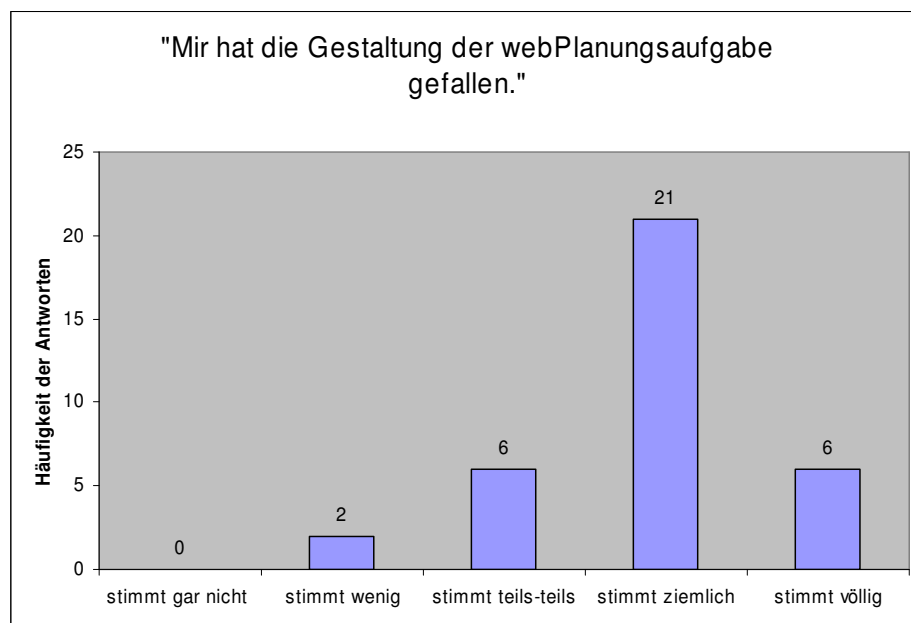
Wie Abbildung 28 zeigt, beantworten die Aussage „Mir hat die Einweisung in die Bedienung der webPlanungsaufgabe gefallen“ 32 der 35 Personen mit „stimmt ziemlich“ oder „stimmt völlig“.



**Abbildung 29: Bewertung der Bedienungsfreundlichkeit der webPlanungsaufgabe (Voruntersuchung)**

Ähnlich positiv wird auch die Bedienung des Programms bewertet. Die Aussage „*Mir ist es leicht gefallen, die webPlanungsaufgabe zu bedienen*“ beantworten ebenfalls 32 der 35 Personen mit „*stimmt ziemlich*“ oder „*stimmt völlig*“ (s. Abb. 29).

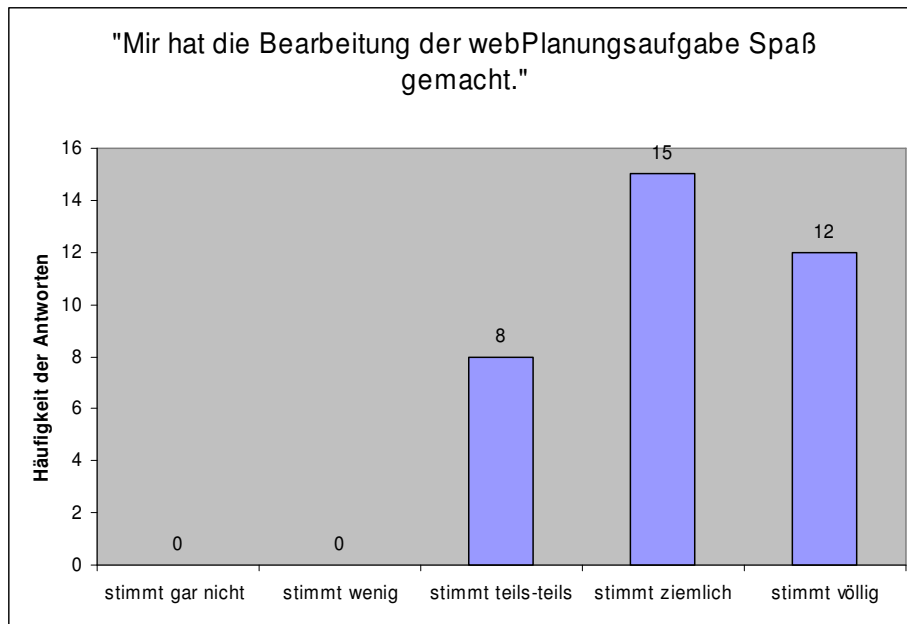
Dem Statement „*Mir hat die Gestaltung der webPlanungsaufgabe gefallen*“ stimmen über drei Viertel der Befragten ziemlich oder völlig zu (s. Abbildung 30).



**Abbildung 30: Bewertung der Gestaltung der webPlanungsaufgabe (Voruntersuchung)**

Darüber hinaus sollte das subjektive Erleben der Benutzer während der Aufgabenbearbeitung untersucht werden. Daher wurde die Frage gestellt, inwieweit die *webPlanungsaufgabe* den Probanden Spaß gemacht hat. Trotz der kognitiven Anforderungen, die die Bearbeitung der *webPlanungsaufgabe* an die Testpersonen stellt, stimmen wiederum über drei Viertel (27 von 35) der Teilnehmer der Aussage: „*Mir hat die Bearbeitung der Aufgaben Spaß gemacht*“ ziemlich oder völlig zu. Die restlichen acht Personen bewerten die Aussage zumindest noch mit „*stimmt teils-teils*“ (s. Abb. 31).





**Abbildung 31: Bewertung der Zufriedenheit bei der Bearbeitung der webPlanungsaufgabe (Voruntersuchung)**

Insgesamt zeigt sich somit eine sehr positive Bewertung des Verfahrens durch die Benutzer. Die Ergebnisse können als Beleg für die weitestgehende Akzeptanz der *webPlanungsaufgabe* sowie für die Zufriedenheit der Teilnehmer verstanden werden. Die Teilnehmer der Voruntersuchung bewerteten auch den *webPostkorb*. Dabei ergeben sich ähnlich positive Ergebnisse wie für die *webPlanungsaufgabe*. Auf die genaue Darstellung wird hier verzichtet, da diese Aspekte innerhalb der Hauptuntersuchung an einer größeren Stichprobe ebenfalls überprüft wurden (s. Abschnitt 7.7).

#### *Verfahrensbewertung webPlanungsaufgabe in Abhängigkeit von der erzielten Leistung und der wahrgenommenen Schwierigkeit*

Unter den Gesichtspunkten der Testfairness sowie im Hinblick auf einen möglichen Einsatz im Rahmen von Personalmarketing-Maßnahmen (s. o.) ist es von Bedeutung, wie Testpersonen die Verfahren bewerten. Neben den oben dargestellten persönlichen Einschätzungen ist es darüber hinaus interessant, ob Personen, die im Verfahren schlechter abschneiden, das Verfahren insgesamt auch schlechter bewerten (vgl. Gilliland, 1994). Dazu wurden die erzielten Punkterohwerte mit dem aufsummierten Wert der Benutzerzufriedenheit korreliert. Das Ergebnis ist, dass sich zwischen diesen

Variablen keine substanziellen Zusammenhänge für die Stichprobe der Voruntersuchung ergeben. Auch bei einer Extremgruppenbildung nach besonders guten und niedrigen Leistungen in der *webPlanungsaufgabe* zeigen sich keinerlei signifikante Unterschiede hinsichtlich der Bewertung des Verfahrens. Neben dem direkten Abschneiden wird in dieser Untersuchung auch nach der subjektiv wahrgenommenen Schwierigkeit gefragt. Auch hier können nur Nullkorrelationen zwischen der Bewertung der Verfahren und der subjektiv wahrgenommenen Schwierigkeit ermittelt werden.

#### *Korrelative Befunde zum konstruktremden Einfluss*

Von besonderer Bedeutung für die Bewertung der Fairness und der Benutzerfreundlichkeit eines eignungsdiagnostischen Verfahrens ist die Analyse möglicher konstruktremder Einflüsse. Wie oben geschildert, sind die Überprüfung des Einflusses der Computererfahrung und der Einstellung zum Medium Computer bzw. Internet für alle internet- und computergestützten Verfahren entscheidend. Neben dem Zusammenhang zwischen den erzielten Werten in den Leistungsverfahren und der Computererfahrung und Computereinstellung wurde im Rahmen der Voruntersuchung auch der Einfluss des Alters untersucht.<sup>18</sup> Wie in Tabelle 18 abgebildet, bewegen sich die meisten Korrelationen unter der zumeist als kritisch bezeichneten Schwelle von  $r = .30$  (vgl. Klieme et al., 2001). Im Hinblick auf den *webPostkorb* liegen nur die Korrelationen mit der Computererfahrung mit  $r = .32$  knapp über der Schwelle (in der Tabelle markiert).

Insgesamt zeigen die Ergebnisse, dass konstruktremde Einflüsse auf das Ergebnis in der *webPlanungsaufgabe* und dem *webPostkorb* bis auf eine Ausnahme durchgehend niedrig und nicht signifikant sind.

---

<sup>18</sup> Auf die Berechnung des Einflusses des Geschlechts wurde aufgrund von lediglich vier männlichen Testpersonen in der Stichprobe verzichtet.

**Tabelle 18: Korrelationen zwischen der *webPlanungsaufgabe* sowie dem *webPostkorb* und konstrukt fremden Variablen | Ziel:  $r < .3$  (Voruntersuchung)**

Verfahren / Skala	N	Alter	Computer- erfahrung	Computer- einstellung
komb. Gesamtscore <i>webPlanungsaufgabe</i>	35	.05	.21	.11
Summenscore Planen <i>webPlanungsaufgabe</i>		-.05	-.08	.05
Gesamtscore Problemlösen <i>webPostkorb</i>	38	-.22	<b>.32</b>	.00

\* Kodierung: 1 = weiblich; 2 = männlich.

## 6.5 Erste Befunde zur Validität

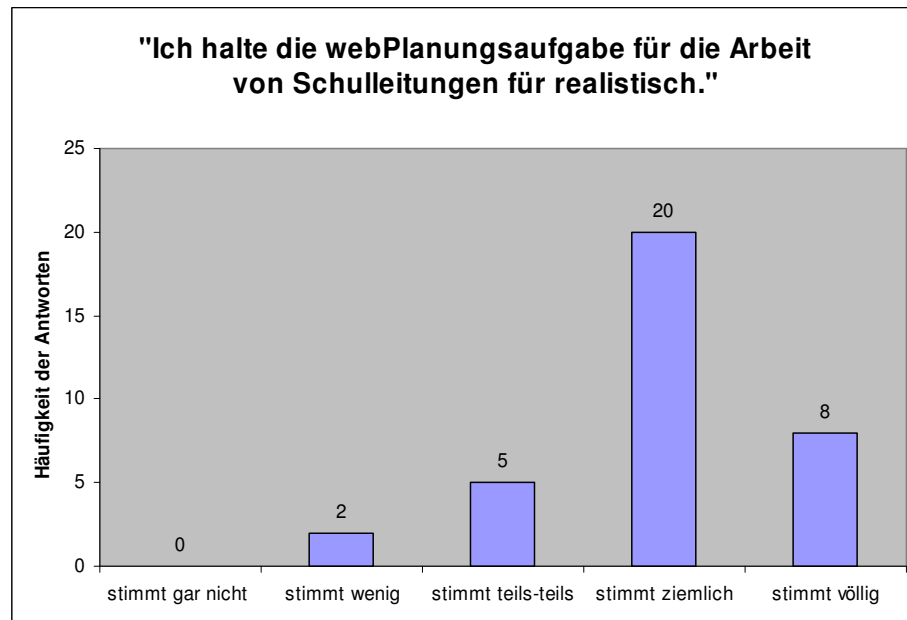
In der Voruntersuchung konnten bereits erste Ergebnisse zu Aspekten der Validität gesammelt werden. Vor dem Hintergrund der Verfahrenskonstruktion sind dabei insbesondere die Konstruktvalidität und die Inhaltsvalidität von Bedeutung.

### *Zur Inhalts- bzw. Augenscheinvalidität*

Eine Reihe von Schulleitungsmitgliedern hat die Testvorformen begutachtet. Dabei konnten offene Rückmeldungen bezüglich der Aufgabenkonstruktion und Realitätsnähe gewonnen werden. Die Aussagen zeigen, dass die Verfahren *webPostkorb* und *webPlanungsaufgabe* insgesamt als realitätsnah bewertet werden.

Neben der inhaltsvaliden Entwicklung mit der Einbeziehung von Experten und Studierenden in allen Phasen der Konstruktion (s. o.) sollten darüber hinaus auch die Einschätzungen der Testpersonen gesammelt werden. Dabei interessiert, wie die Benutzer den Zusammenhang zwischen den Verfahrensinhalten und dem zu messendem Verhalten bewerten. Dieser Zusammenhang wird zumeist als Augenscheinvalidität bezeichnet und ist insofern von Interesse, als dass die postulierte Überlegenheit der Akzeptanz

von Simulationsverfahren gegenüber herkömmlichen Tests vermutlich auch auf der zumeist höheren Augenscheinvalidität basiert (vgl. Höft & U. Funke, 2001).



**Abbildung 32: Bewertung der Realitätsnähe der webPlanungsaufgabe (Voruntersuchung)**

Die Studierenden wurden um eine Bewertung hinsichtlich der Aussage „Ich halte die webPlanungsaufgabe für die Arbeit von Schulleitungen für realistisch“ gebeten. 28 der 35 Studierenden beantworteten diese Aussage mit „stimmt ziemlich“ oder „stimmt völlig“ (s. Abbildung 32). Es zeigt sich somit, dass die große Mehrheit der befragten Testpersonen die webPlanungsaufgabe insgesamt als realistisch für den Beruf von Schulleitungen einschätzt. Dies ist als Hinweis für die anvisierte Augenscheinvalidität der webPlanungsaufgabe zu werten.

Die Teilnehmer der Voruntersuchung bewerteten auch den webPostkorb hinsichtlich der Augenscheinvalidität. Dabei ergeben sich ähnlich positive Ergebnisse wie für die webPlanungsaufgabe. Auf die genaue Darstellung wird hier verzichtet, da dieser Aspekt innerhalb der Hauptuntersuchung an einer größeren Stichprobe ebenfalls überprüft wurde (s. Abschnitt 7.4).

### Zur Konstruktvalidität

Die 35 Teilnehmer der Stichprobe *webPlanungsaufgabe* bearbeiteten vier Aufgaben des Intelligenz-Struktur-Tests von Amthauer (1970). Davon durchliefen 23 Personen auch den *webPostkorb*. Einen Überblick über die Zusammenhänge zwischen den Ergebnissen in den beiden internetbasierten Verfahren und den Testleistungen im I-S-T 70 gibt Tabelle 19. Aufgrund der recht kleinen Stichprobe sind nur Korrelationen über  $r > .32$  (Stichprobe *webPlanungsaufgabe*) und  $r > .40$  (Stichprobe *webPostkorb*) signifikant. Die statistische Power für das Aufdecken mittlerer Effekte (Effektstärke  $r = .30$ ) beträgt für die Stichprobe der Voruntersuchung ( $N = 35$ ) somit  $1 - \beta = .57$ .<sup>19</sup>

**Tabelle 19: Korrelationen zwischen dem *webPostkorb* sowie der *webPlanungsaufgabe* und Aufgaben des I-S-T 70 (Voruntersuchung)**

Verfahren	Skala	N	I-S-T 70 Rohwert (ausgewählte Aufgaben)
<i>webPostkorb</i>	<i>dynamisches Problemlösen</i>	23	.33
	<i>analytisches Problemlösen</i>		.38
	Gesamtscore Problemlösen		.49*
<i>webPlanungs- aufgabe</i>	Summenscore	35	.27
	benötigte Zeit		-.10
	komb. Gesamtscore		.31

\* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

Der Punkterohwert der *webPlanungsaufgabe* korreliert mit den aufsummierten Rohwerten der Subtests Analogien, Gemeinsamkeiten, Figurenauswahl und Rechenaufgaben des I-S-T 70 in Höhe von  $r = .27$  ( $N = 35$ ). Die höchste Korrelation ergibt sich zwischen der Aufgabe 5 (Rechenaufgabe) und dem Punkterohwert in Höhe von  $r = .35$ . Wird der

<sup>19</sup> Die Berechnung erfolgte nach Faul und Erfelder (1992).

Punkterohwert mit der von den Teilnehmern benötigten Zeit kombiniert (Summe der z-transformierten Werte), zeigen sich geringfügig höhere Korrelationen zu den Werten im Intelligenz-Struktur-Test. So korreliert der kombinierte, z-transformierte Punkt- und Zeitwert mit dem Gesamtrohwert der I-S-T 70 Aufgaben in Höhe von  $r = .31$ , wobei sich wiederum der höchste Einzelzusammenhang mit der Rechenaufgabe ergibt ( $r = .43$ ).

Stellt man eine Rangreihe der Teilnehmer aufgrund ihrer Ergebnisse im I-S-T 70 und in der *webPlanungsaufgabe* auf, so hängen diese beiden Reihen substantiell in Höhe von  $r$  (Spearman-Rho) =  $.38$  für den Punktrohwert und  $r = .42$  für den kombinierten Punkte- und Zeitwert zusammen. Die höchsten Werte ergeben sich für die Rangkorrelationsberechnungen zwischen der Rechenaufgabe und dem Punktwert in Höhe von  $r = .49$  bzw. dem kombinierten Punkt- und Zeitwert in Höhe von  $r = .56$ .

Es zeigen sich somit insgesamt erwartungsgemäße Zusammenhänge zwischen der Leistung in der *webPlanungsaufgabe* und der Intelligenztestleistung in mittelstarker Höhe. Die Werte liegen im Bereich der Korrelationen zwischen den Subtests des I-S-T 70 (Amthauer, 1970). Dieses Ergebnis soll durch eine Replikationsstudie im Rahmen der Hauptuntersuchung kontrolliert werden.

Die Korrelation zwischen dem Gesamtergebnis im *webPostkorb* (Rohwert) und der über vier I-S-T 70 Aufgaben aufsummierten Intelligenztestleistung bewegt sich mit  $r = .49$  im signifikanten Bereich, selbst bei der Stichprobengröße von  $N = 23$ . Auch zwischen den Subskalen und der Leistung im I-S-T 70 ergeben sich mittelstarke Zusammenhänge. Die Werte liegen bei  $r = .38$  zwischen I-S-T 70 und *analytischem Problemlösen* und  $r = .33$  zwischen I-S-T 70 und *dynamischem Problemlösen*. Diese Ergebnisse können als erster Hinweis für einen Zusammenhang der Intelligenztestleistung mit den Leistungen im *webPostkorb* gewertet werden. Aufgrund des Untersuchungsdesigns mit der Online-Durchführung des *webPostkorbs* von zu Hause aus und der recht kleinen Stichprobengröße sollen diese Ergebnisse in der Hauptuntersuchung nochmals überprüft werden.

## 6.6 Zusammenfassung und Implikationen für die Hauptuntersuchung

### *Benutzerfreundlichkeit, Akzeptanz und Fairness*

Die Voruntersuchung konnte erste positive Befunde für die Bewährung der Vorgehensweise bei der Konstruktion sowie für die Testgüte der Verfahren liefern. Hinsichtlich der Benutzerfreundlichkeit, Akzeptanz und Fairness zeigen sich insgesamt sehr zufrieden stellende Ergebnisse. Die Testpersonen bewerten die *webPlanungsaufgabe* als realistisch und benutzerfreundlich. Auch die Experten schätzen die Verfahren als gelungen ein. Lediglich der Einfluss der Computererfahrung auf die Leistung im *webPostkorb* liegt mit  $r = .32$  knapp über der angelegten Schwelle von  $.30$ . Wenngleich dieser Befund bei einer kleinen Stichprobe von  $N = 38$  ermittelt wurde, sollen die Einweisungsseiten des *webPostkorbs* für die Hauptuntersuchung erweitert werden, um computerunerfahrenen Personen die Navigation und Benutzerführung noch näher zu bringen und einen möglichen Einfluss der Computererfahrung zu minimieren.

### *Aufgabenanalyse*

Die Aufgaben- und Skalenkennwerte zeichnen insgesamt ein befriedigendes Bild. Die Befunde zu Skalenkennwerten und innerer Struktur sind gut, lediglich die Aufgabengruppe *Umgang mit Störungen* weist Revisionsbedarf auf. Daher soll diese Gruppe für die Testendform überarbeitet und in die Aufgabengruppe *Wochenplan* integriert werden. Dadurch sind für die Hauptuntersuchung fünf Items des *webPostkorbs* zu überarbeiten, da sie inhaltlich auf die neue Aufgabengruppe *Wochenplan Nachmittag* angepasst werden müssen. Zwei Items der Aufgabengruppe *Umgang mit Störungen* lassen sich nicht integrieren und werden daher neu entwickelt. Die Reihenfolge der Items ist in zwei Fällen zu ändern. Der größte Teil der Items kann allerdings ohne Änderungen in die Hauptuntersuchung übernommen werden.

**Tabelle 20: Überblick über die wichtigsten Implikationen für die weitere Verfahrensentwicklung des *webPostkorbs***

### Implikationen für die weitere Verfahrensentwicklung des *webPostkorbs*

- Überführung der Skala *Umgang mit Störungen* in die Aufgabengruppe *Wochenplan Nachmittag* - dadurch Überarbeitung von fünf Items und zwei Neuentwicklungen
- Erweiterung der Einweisungsseiten

#### *Untersuchungsbedingungen*

Als mögliche Einsatzformen der Verfahren kommen Self-Assessment- und E-Assessment-Anwendungen in Betracht (s. Kapitel 3). Die Untersuchungsbedingungen sind für die Hauptuntersuchung zu variieren. Im Rahmen der Voruntersuchung erfolgte der Einsatz der *webPlanungsaufgabe* als Selbsteinschätzungsverfahren (Self-Assessment). Daher soll die *webPlanungsaufgabe* in der Hauptuntersuchung unter Ernstfallbedingungen überprüft werden. Da der *webPostkorb* den Testpersonen der Voruntersuchung online dargeboten wurde, soll das Verfahren in der Hauptuntersuchung im kontrollierten Setting der universitären Computerräume zum Einsatz kommen.

#### *Optimaler Stichprobenumfang*

Die erwarteten Zusammenhänge zwischen der Problemlöseleistung im *webPostkorb*, der Planungsleistung in der *webPlanungsaufgabe* und der Intelligenztestleistung konnten in der Voruntersuchung bestätigt werden. Diese ersten Hinweise auf Konstruktvalidität der Verfahren gilt es in der Hauptuntersuchung zu replizieren. Daher wurde a priori der optimale Stichprobenumfang nach Faul und Erdfelder (1992) ermittelt. Als Entscheidungsniveau für die Annahme einer Hypothese wird für die Hauptuntersuchung eine Irrtumswahrscheinlichkeit von  $p < .05$  bei einer Teststärke von  $1 - \beta = .80$  festgelegt (vgl. Bortz & Döring, 1995). Zur Aufdeckung eines mittleren Effekts (vgl. J. Cohen, 1992) liegt der optimale



Stichprobenumfang demnach für die Hauptuntersuchung bei 64 Probanden (kritisches  $t(62) = 1.67$ ). Dieser Stichprobenumfang gewährleistet, dass ein Signifikanztest mit einer Wahrscheinlichkeit von 80 Prozent zu einem signifikanten Ergebnis führt, wenn die spezifische Alternativhypothese den Populationsverhältnissen entspricht. Das Risiko einer Fehlentscheidung bei Annahme der Alternativhypothese aufgrund eines signifikanten Ergebnisses entspricht hierbei dem Signifikanzniveau.

**Tabelle 21: Überblick über die wichtigsten Implikationen für die Hauptuntersuchung**

Implikationen für die Hauptuntersuchung
<ul style="list-style-type: none"><li>• Replikationsstudie zum Zusammenhang mit Intelligenztestleistung</li><li>• Einsatz weiterer Intelligenztestverfahren<ul style="list-style-type: none"><li>→ Berliner Intelligenzstruktur-Test (Jäger et al., 1997)</li><li>→ I-S-T 2000 R (Amthauer et al., 2001)</li></ul></li><li>• optimaler Stichprobenumfang bei angenommener mittlerer Effektgröße, Teststärke von .8 und 5 % Signifikanzniveau: <math>N = 64</math></li><li>• Variation der Untersuchungsbedingungen</li><li>• weitere Überprüfung der Testgüte der Verfahren</li></ul>

#### *Einsatz weiterer Intelligenztest-Verfahren*

In der Voruntersuchung kamen vier Aufgaben des I-S-T 70 von Amthauer (1970) zur Anwendung. Für die Hauptuntersuchung sollen modernere Verfahren der Intelligenzmessung Berücksichtigung finden. So lassen sich durch den Einsatz verschiedener Verfahren eventuelle Methodeneffekte des I-S-T 70 kontrollieren. Zum einen wird die Weiterentwicklung des I-S-T 70, der I-S-T 2000 R von Amthauer et al. (2001), verwendet und zum anderen kommt die Kurzversion des Berliner Intelligenzstruktur-Tests (BIS-K-S) von Jäger et al. (1997) zum Einsatz. Dieser Test fand im Rahmen der Problemlöseforschung bereits des Öfteren Anwendung und hat sich dabei als geeignetes Verfahren erwiesen.



## 7 Ergebnisse der Hauptuntersuchung

Innerhalb der Hauptuntersuchung kamen die *webPlanungsaufgabe* und die Weiterentwicklung des *webPostkorbs* zum Einsatz. Ziel der Hauptuntersuchung ist die Gewinnung von Hinweisen hinsichtlich der Gütekriterien. Dabei liegt der hier gewählten Kombination von Simulations- und Konstruktansatz die Validierungslogik der Inhaltsvaliditäts- und Konstruktvaliditätsprüfung zu Grunde. Die angestrebte inhaltsvalide Entwicklung der Verfahren wurde bereits weiter oben dargestellt, daher steht im Mittelpunkt dieses Kapitels die Vorstellung der Ergebnisse zur Konstruktvalidität. Die Befunde der Hauptuntersuchung zum *webPostkorb* und zur *webPlanungsaufgabe* werden im Folgenden jeweils im gleichen Abschnitt dargestellt, da sich die Ergebnisse zumeist auf beide Verfahren beziehen.

### 7.1 Stichproben

#### *webPostkorb*

An der Hauptuntersuchung zum *webPostkorb* nahmen 66 Studierenden der Studiengänge Lehramt, Diplom-Pädagogik und Sozialpädagogik der Universität Lüneburg teil. Davon hatte genau die Hälfte das Grundstudium bereits absolviert und die Zwischenprüfung bestanden. Im Sinne einer Variation der Untersuchungsbedingungen fand die Durchführung des *webPostkorbs* unter Aufsicht in den Computerräumen der Universität statt, während in der Voruntersuchung das Verfahren online von zu Hause aus bearbeitet wurde. Das mittlere Alter der 66 Studierenden der Stichprobe *webPostkorb* (55 weiblich und 11 männlich) beträgt 24.6 Jahre (Median = 24 Jahre, SD = 3.9). Die mittlere Testleistung in den sechs Aufgaben der Kurzversion des Berliner Intelligenzstruktur-Tests BIS-K-S (Jäger et al., 1997) liegt bei einem Punktwert von 582 (SD = 37.6). Dies würde einem Standardwert von 93 (Prozentrangwert von 24) bezogen auf die Altersnorm

von 18- bis 19-Jährigen<sup>20</sup> entsprechen. Die Erhebung wurde im Oktober 2002 durchgeführt und erfolgte freiwillig und unentgeltlich. Die Probanden erhielten per E-Mail eine individuelle Rückmeldung über ihre Ergebnisse.

#### *webPlanungsaufgabe*

An der Hauptuntersuchung zur *webPlanungsaufgabe* nahmen 45 der 66 Studierenden aus der Untersuchung zum *webPostkorb* teil. Dies entspricht einem Anteil von 68 Prozent. Das den jeweiligen Analysen zu Grunde liegende N wird an den entsprechenden Stellen berichtet. Die Besonderheit der Durchführung ist, dass die Studierenden sich von zu Hause oder den Computerräumen der Universität aus zu einer von ihnen gewünschten Zeit einloggen und die Aufgabe bearbeiten konnten. Dieser Teil der Untersuchung fand somit online statt. Signifikante Varianzunterschiede zwischen den sechzehn Abbrechern und den 45 Teilnehmern des zweiten Teils der Hauptuntersuchung in Bezug auf Alter, BIS-Testleistungen und Leistungen im *webPostkorb* konnten nicht festgestellt werden (Levene-Test, alle p-Werte > .05). Allerdings zeigen sich bei den Mittelwertvergleichen signifikant höhere BIS- und *webPostkorb*-Werte für die 45 Teilnehmer des zweiten Teils der Hauptuntersuchung ( $p < .05$ ). Weitere bedeutsame Unterschiede gibt es nicht. Das mittlere Alter der 45 Studierenden der Stichprobe *webPlanungsaufgabe* (39 weiblich und sechs männlich) beträgt 24.5 Jahre (SD = 3.9). Die Erhebung fand im Oktober 2002 statt und erfolgte freiwillig und unentgeltlich. Im Anschluss an die Bearbeitung der *webPlanungsaufgabe* bearbeiteten die Probanden den HAKEMP-90 (Kuhl, 1994) sowie einen selbst entwickelten Fragebogen. Die Untersuchungsteilnehmer erhielten per E-Mail eine individuelle Rückmeldung über ihre Ergebnisse.

#### *webPlanungsaufgabe bei Führungsnachwuchskräften*

Neben den universitären Stichproben konnte eine spezielle Version der *webPlanungsaufgabe* bei 369 Bewerbern (179 weiblich und 190 männlich)

---

<sup>20</sup> Für den BIS (Jäger et al., 1997) werden nur Normwerte für 16- bis 17-Jährige und 18- bis 19-Jährige geliefert. Zum Zwecke des Vergleichs wurde hier die Altersnorm der 18- bis 19-Jährigen verwendet.

auf Führungsnachwuchsstellen und 19 bewährten Führungsnachwuchskräften durchgeführt werden. Das mittlere Alter beträgt 27.5 Jahre (SD = 2.9). Die *webPlanungsaufgabe* wurde dabei semantisch dem Berufsumfeld eines Managers in der freien Wirtschaft angepasst. Besonderheit dieser Untersuchung ist, dass die Kandidaten die Aufgabe unter Ernstfallbedingungen im Rahmen eines E-Assessments zur Personalauswahl von „High-Potentials“ eines weltweit operierenden Konzerns durchführten, während die universitären Stichproben die Aufgaben jeweils im Rahmen eines Self-Assessments (zur Unterscheidung s. Abschnitt 3.1) bearbeiteten. Die Durchführung erfolgte dabei gemäß Lienert und Raatz (1998), wobei das neue, zu analysierende Verfahren (hier die *webPlanungsaufgabe*) zusammen mit anderen Tests unter Ernstfallbedingungen dargeboten, jedoch nicht mit in die Bewertung einbezogen wurde. Zu berücksichtigen ist, dass die 369 Bewerber bereits aus einer Zahl von etwa 3.500 Bewerber/innen anhand von Kriterien wie Lebenslauf, Alter und absolvierte Praktika vorselektiert wurden. Neben den Bewerbern durchliefen 19 bewährte Führungsnachwuchskräfte, die bereits erfolgreich im Unternehmen arbeiten, die *webPlanungsaufgabe*. Erhebungszeitraum war April bis September 2004.

## 7.2 Itemkennwerte

Wie bereits unter Abschnitt 6.1 erläutert, kommt nur der *webPostkorb* aufgrund seiner Struktur für eine Itemanalyse in Frage. Die Items des überarbeiteten *webPostkorbs* werden hier wiederum hinsichtlich der Kriterien Schwierigkeit, Trennschärfe, interne Konsistenz und Dimensionalität überprüft (vgl. Lienert & Raatz, 1998). Die Ergebnisse der Voruntersuchung führten dazu, dass der *webPostkorb* inhaltlich leicht überarbeitet wurde, so dass in der Endform die Aufgabengruppe *Umgang mit Störungen* komplett revidiert und in die Aufgabengruppe *Wochenplan* integriert wurde.

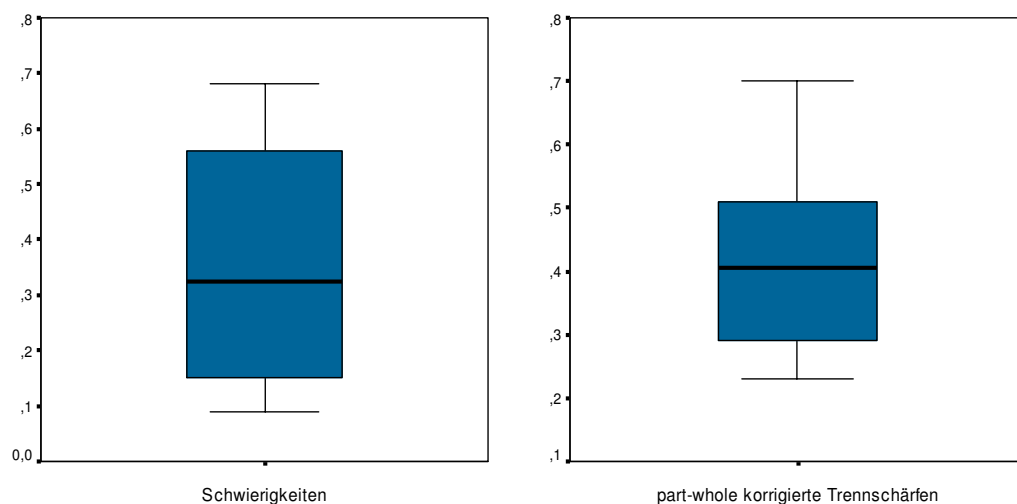
Tabelle 22: Itemkennwerte des *webPostkorbs* (Hauptuntersuchung)

Subskala	Aufgaben- gruppe	Item	Mittel- wert	SD	Trenn- schärfe	Alpha ohne
<i>ana- lytisches Problem- lösen</i>	<i>Tabellen- verständnis</i>	Item_1	.58	.48	.56	.78
		Item_2	.58	.50	.70	.77
		Item_3	.45	.50	.66	.77
		Item_4	.45	.50	.31	.77
		Item_5	.33	.48	.38	.77
		Item_6	.17	.38	.63	.77
	<i>Text- verständnis</i>	Item_1	.68	.47	.59	.78
		Item_2	.62	.49	.48	.77
		Item_3	.33	.48	.43	.77
		Item_4	.30	.46	.37	.77
		Item_5	.14	.39	.33	.77
		Item_6	.11	.31	.24	.78
		Item_7	.11	.31	.24	.78
	<i>Diagramm- verständnis</i>	Item_1	.62	.49	.44	.77
		Item_2	.41	.50	.29	.77
		Item_3	.32	.47	.29	.78
		Item_4	.32	.47	.51	.77
		Item_5	.30	.46	.41	.79
		Item_6	.20	.40	.35	.78
		Item_7	.15	.36	.66	.77
	<i>dyna- misches Problem- lösen</i>	<i>Wochenplan Vormittag</i>	Item_1	.56	.50	.40
Item_2			.41	.50	.28	.79
Item_3			.09	.29	.46	.78
<i>Wochenplan Nachmittag</i>		Item_1	.62	.49	.23	.79
		Item_2	.24	.43	.44	.78
		Item_3	.11	.31	.27	.78

Anmerkung: N = 66.

Die Mittelwerte, Standardabweichungen und Trennschärfen der Items des *webPostkorbs* gibt Tabelle 22 wieder. Es ergeben sich Schwierigkeiten in der Spannweite von .09 bis .68 bei einer mittleren Schwierigkeit von .35. Dabei liegen annähernd drei Viertel der Schwierigkeiten zwischen .20 und .68. Innerhalb der Aufgabengruppen sind jeweils Items verschiedener Schwierigkeitsgrade vertreten. Es wird deutlich, dass die Itemschwierigkeiten für diese Stichprobe mit sieben Items unter .20 relativ niedrig ausfielen. Dies kann mit der etwas unterdurchschnittlichen mittleren Testintelligenz der Stichprobe zusammenhängen (s. genauer Abschnitt 7.3).

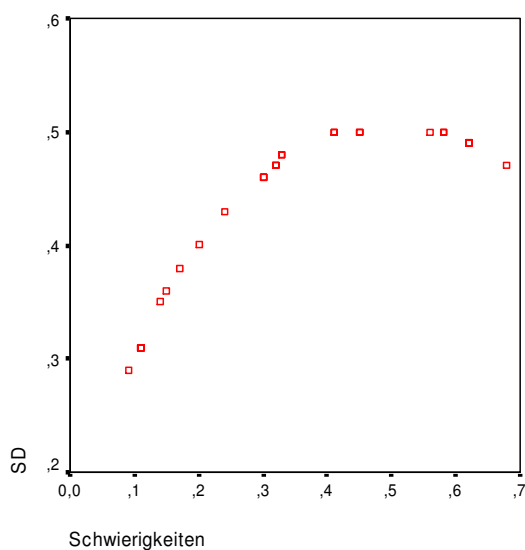
Die Werte der Standardabweichungen streuen zwischen .29 bis .50 bei einem mittleren Wert von .44. Die Verteilung der Mittelwerte (Schwierigkeiten mit Inangriffnahmekorrektur) der Items wird im Boxplot (Abbildung 33) grafisch verdeutlicht. Die farbige Fläche wird durch das 25 und 75 Prozent Perzentil begrenzt, während die durchgezogene Linie in der Box den Median beschreibt.



**Abbildung 33: Boxplots der Schwierigkeiten mit Inangriffnahmekorrektur und der part-whole korrigierten Trennschärfen *webPostkorb* (Hauptuntersuchung, N = 66)**

Die Itemtrennschärfen des *webPostkorbs* werden in Abbildung 33 ebenfalls als Boxplot wiedergegeben. Hinsichtlich der part-whole korrigierten Trennschärfen ergeben sich insgesamt zufrieden stellende Werte. Die mittlere korrigierte Trennschärfe für die 26 Items der Endversion liegt bei .42 (Median .40). Die Trennschärfen streuen dabei zwischen .23 und .70.

Lienert und Raatz (1998, S. 31) weisen darauf hin, dass für mittelschwere Items im Allgemeinen die höchsten Streuungen vorliegen. Dieser Zusammenhang zeigt sich auch bei der hier untersuchten Stichprobe, wie in Abbildung 34 zu sehen. Die schwierigen Items weisen niedrigere Werte in der Standardabweichung auf, während im mittleren Bereich die höchsten Streuungen aufzufinden sind. An diesem Streudiagramm wird auch ersichtlich, dass nur wenige Items hohe Schwierigkeitswerte aufweisen. Für diese Stichprobe bewegen sich die meisten Items im mittelschweren bis schweren Bereich.



**Abbildung 34: Streudiagramm der Schwierigkeiten mit Inangriffnahmekorrektur und der Standardabweichungen für die Items des *webPostkorbs* (Hauptuntersuchung)**

### 7.3 Objektivität, Reliabilität und Skalenkennwerte

Die Durchführungs-, die Auswertungs- und die Interpretationsobjektivität der Verfahren *webPostkorb* und *webPlanungsaufgabe* können aufgrund der programmgesteuerten, computergestützten Vorgabe und Auswertung insgesamt als gegeben eingeschätzt werden. Die Darbietung erfolgt vollständig auf dem Bildschirm eines Rechners. Dabei sind die rechnergestützte Ablaufkontrolle und die kennwertebezogene Interpretation gewährleistet. Es wurde darauf geachtet, dass sämtliche Instruktionen zur Durchführung der Verfahren verständlich und mit kurzen obligatorischen



Übungsbeispielen illustriert sind. Analog zur Voruntersuchung ergeben sich auch bei den Stichproben der Hauptuntersuchung keinerlei Datenverluste.

Es ist allerdings zu berücksichtigen, dass bei einer eventuellen Online-Anwendung der Verfahren von zu Hause aus die Durchführungsobjektivität nicht letztgültig kontrollierbar ist. Dies gilt allerdings grundsätzlich für die rein internetbasierte Durchführung eignungsdiagnostischer Verfahren. Wenngleich hier darauf geachtet wurde, die Verfahren so zu programmieren, dass sie auf allen Rechnern annähernd identisch ablaufen (s. o.), so können ggf. Umgebungsfaktoren und Eingabemedien unterschiedlich sein. Diese denkbare Beeinträchtigung der Durchführungsobjektivität ist allerdings nicht den Verfahren an sich zuzuschreiben, da der objektive Ablauf möglich ist, wie beispielsweise im Rahmen der Vor- und Hauptuntersuchung beim kontrollierten Einsatz in Computerräumen.

#### *webPostkorb*

Deskriptive Statistiken der Aufgabengruppen und Skalen des *webPostkorbs* sind in Tabelle 23 ersichtlich. Die Mittelwerte für die fünf Aufgabengruppen liegen zwischen 1.01 und 2.42 bei Standardabweichungen zwischen .66 und 1.31. Zu berücksichtigen ist dabei die unterschiedliche Itemzahl pro Aufgabengruppe. Mit der Ausnahme der Aufgabengruppe *Tabellenverständnis* werden hinsichtlich der Schiefe der Aufgabengruppen positive Werte errechnet, was auf eher linkssteile Verteilungen deuten lässt. Die Werte weichen allerdings nicht signifikant von einer symmetrischen Verteilung ab.

**Tabelle 23: Deskriptive Statistiken und Cronbachs-Alpha-Werte des *webPostkorbs* (Hauptuntersuchung)**

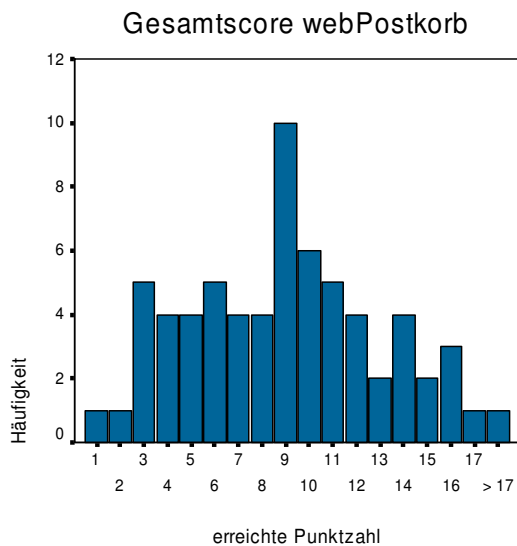
Skala	Aufgabengruppe	Mittelwert	SD	Schiefe	Mittlere Iteminterkorr.	Cronbachs-Alpha	
<i>analytisches Problemlösen</i>	<i>Tabellenverständnis</i> (6 Items)	2.39	1.05	-.05	.38	.85	.80
	<i>Textverständnis</i> (7 Items)	2.29	1.06	.33	.22	.68	
	<i>Diagrammverständnis</i> (7 Items)	2.42	1.31	.83	.28	.76	
<i>dynamisches Problemlösen</i>	<i>Wochenplan Vorm.</i> (3 Items)	1.02	.66	.48	.22	.62	
	<i>Wochenplan Nachm.</i> (3 Items)	1.01	.68	.58			
Gesamt <i>webPK</i>		9.12	4.58	.83	.19	.79	

Anmerkung: N = 66; *webPK*: *webPostkorb*.

Der Mittelwert des ungewichteten Summenscores für den gesamten *webPostkorb* beträgt 9.12 Punkte bei einer Standardabweichung von 4.58. Abbildung 35 visualisiert die Verteilung des Gesamtscores für den *webPostkorb* in Form eines Histogramms.

Hinsichtlich der Überprüfung der Homogenität schlagen Bortz und Döring (1995) und Lienert und Raatz (1998) die Berechnung der mittleren Iteminterkorrelationen vor. Wenngleich in der Hauptuntersuchung insgesamt etwas niedrigere Homogenitäts-Werte zu verzeichnen sind als in der Voruntersuchung, so können insgesamt befriedigenden Ergebnisse ermittelt werden (s. Tabelle 23). Für die Aufgabengruppen bzw. Subskalen streuen die Werte dabei zwischen .22 und .38. Aufgrund der niedrigen Itemzahl pro

Aufgabengruppe werden die Kennwerte für die Subskala *Wochenplan* und nicht für die Aufgabengruppen *Wochenplan Vormittag* und *Wochenplan Nachmittag* berechnet.



**Abbildung 35: Histogramm für den Gesamtscore des *webPostkorbs* (Hauptuntersuchung)**

Da Problemlösen hier hinsichtlich der zwei Teilkonstrukte *dynamisches Problemlösen* und *analytisches Problemlösen* operationalisiert wird, ist es von Interesse zu überprüfen, inwieweit diese Teilkonstrukte hohe Homogenitäten auf Subskalen-Ebene aufweisen (vgl. Abschnitt 6.3). Die Werte für die Subskalen bewegen sich mit *.22* (*dynamisches Problemlösen*) und *.26* (*analytisches Problemlösen*) innerhalb des vorgeschlagenen Akzeptanzbereichs (vgl. Briggs & Cheek, 1986). Lediglich die mittlere Iteminterkorrelation für den Gesamtscore liegt mit *.19* leicht unterhalb der von Briggs und Cheek (1986) vorgeschlagenen Spanne für Gesamttest-Homogenitäten für eindimensionale Instrumente.

Die interne Konsistenz des *webPostkorbs* - berechnet mit Hilfe von Cronbachs-Alpha - bewegt sich für die Gesamtskala *webPostkorb* mit  $\alpha = .79$  im guten Bereich für ein simulatives Testverfahren. Berechnet man das 95-Prozent-Konfidenzintervall, so ergeben sich bei  $N = 66$  Testpersonen als untere Grenze ein  $\alpha_u$  von *.70* und als obere Grenze ein  $\alpha_o$  von *.85*. Die

Reliabilitätsschätzungen mittels Cronbachs-Alpha liegen für die einzelnen Aufgabengruppen zwischen .62 und .85. Für die Aufgabengruppen *Textverständnis* und *Wochenplan* bleiben die Werte relativ niedrig. Für die Subskala *analytisches Problemlösen* ergibt sich ein Wert von .80. Dabei liegen fast alle Cronbachs-Alpha-Werte unter denen der Voruntersuchung. Dieser Effekt könnte mit den relativ niedrigen Mittelwerten im *webPostkorb* und BIS-K-S und einer möglichen Varianzeinschränkung für diese Stichprobe zusammenhängen. Diese Problematik wird weiter unten genauer beleuchtet. Wenngleich Kröner (2001) und Etzel und Küppers (2000) für computergestützte Simulationsverfahren teilweise noch etwas höhere interne Konsistenzen zwischen .78 und .91 ermitteln konnten, handelt es sich insgesamt bei den hier beschriebenen Kennzahlen - angesichts der nach den Übersichten von U. Funke (1995a) und Schippmann et al. (1990) zu erwartenden Reliabilität - um vergleichsweise hohe Werte. U. Funke (1995a) berichtet von Studien zu computergestützten Szenarios, die Retest-Reliabilitäten zwischen .40 und .80 aufweisen. Schippmann et al. (1990) zitieren Studien zu Postkorbverfahren, die Halbierungsreliabilitäten zwischen .50 und .60 sowie Paralleltestreliabilitäten zwischen .20 und .40 ermitteln.

Die Interkorrelationen der Aufgabengruppen und der part-whole korrigierte Zusammenhang zwischen Aufgabengruppenscore und Gesamtergebnis sind in Tabelle 24 aufgeführt. Es zeigt sich, dass die drei Aufgabengruppen zum *analytischen Problemlösen* alle substantiell miteinander zusammenhängen. Die Höhe der Korrelationen variiert dabei zwischen .27 und .34. Die mittlere Interkorrelation der drei Aufgabengruppen *Tabellenverständnis*, *Textverständnis* und *Diagrammverständnis* des *webPostkorbs* liegt bei  $r = .30$  ( $p < .01$ ). Die Aufgabengruppen des *analytischen Problemlösens* korrelieren dabei mit dem Gesamtscore noch ein wenig stärker in Höhe von .32 bis .40 (part-whole korrigiert). Lediglich für die Aufgaben zum *dynamischen Problemlösen* liegt die Korrelation zwischen dem *Wochenplan* und dem Gesamtscore mit einem Wert von .16 im nicht signifikanten Bereich.

**Tabelle 24: Interkorrelationen der Subskalen des *webPostkorbs* (Hauptuntersuchung)**

Skala	Subskala				
	Aufgaben- gruppe	Wochenplan (Gesamt)	Tabellen- verständnis	Text- verständnis	Diagramm- verständnis
<i>dynamisches Problemlösen</i>	<i>Wochenplan (Gesamt)</i>				
<i>analytisches Problemlösen</i>	<i>Tabellen- verständnis</i>	.22			
	<i>Text- verständnis</i>	.12	.29*		
	<i>Diagramm- verständnis</i>	.02	.27*	.34**	
Gesamtscore Problemlösen <i>webPostkorb</i>		.16	.39**	.40**	.32**

Anmerkung: N = 66.

\* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

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

### *Implikationen für die Bildung von Subskalen und einem Gesamtscore für den webPostkorb*

Die Befunde zur Interkorrelation der Aufgabengruppen, zur mittleren Iteminterkorrelation, zur inneren Konsistenz und vor allem auch die unter Abschnitt 7.6.1 hinsichtlich der Faktorenstruktur dargestellten Ergebnisse legen die Bildung der Summenscores *analytisches Problemlösen* und *dynamisches Problemlösen* nahe. Die Ergebnisse hinsichtlich der Bildung eines Gesamtscores für den *webPostkorb* sind nicht ganz so eindeutig, wengleich der Cronbachs-Alpha-Wert von .79 - berechnet über alle Items - und die mittlere Iteminterkorrelation (s. Tabelle 23) Hinweise auf die interne Konsistenz des gesamten Verfahrens geben. Da es sich beim *webPostkorb* um ein Simulationsverfahren handelt, das als Ganzes zu bearbeiten ist, bietet sich auch aufgrund inhaltlicher Gesichtspunkte eine Zusammenfassung an. Die Bildung eines Gesamtscores *webPostkorb*

erscheint somit als sinnvoll. Dieser Wert soll daher in die folgenden Analysen mit einbezogen werden.

### *webPlanungsaufgabe*

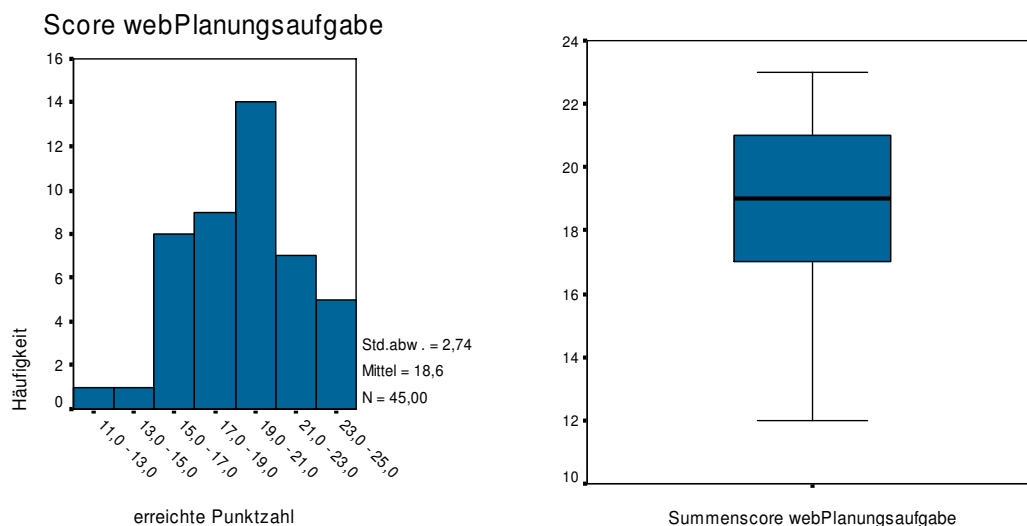
Aufgrund der Struktur der *webPlanungsaufgabe* ist die Bestimmung der Kennwerte auf Itemniveau, wie Itemschwierigkeit und Itemtrennschärfe sowie die Bestimmung der internen Konsistenz mit Hilfe von Cronbachs-Alpha, nicht sinnvoll durchzuführen. Daher folgen im Weiteren Analysen auf Skalenniveau. Die Skalenkennwerte für die studentische Stichprobe der Hauptuntersuchung (Schulleitungsversion der *webPlanungsaufgabe*) und für die Stichprobe der Führungsnachwuchskräfte (Wirtschaftsversion der *webPlanungsaufgabe*) zeigt Tabelle 25. Dabei ist zu berücksichtigen, dass für die Führungsnachwuchskräfte die *webPlanungsaufgabe* semantisch dem Berufsumfeld eines Managers angepasst wurde und somit die beiden Stichproben nicht direkt miteinander zu vergleichen sind.

**Tabelle 25: Deskriptive Statistiken der *webPlanungsaufgabe* (Schulleitungs- und Wirtschaftsversion, Hauptuntersuchung)**

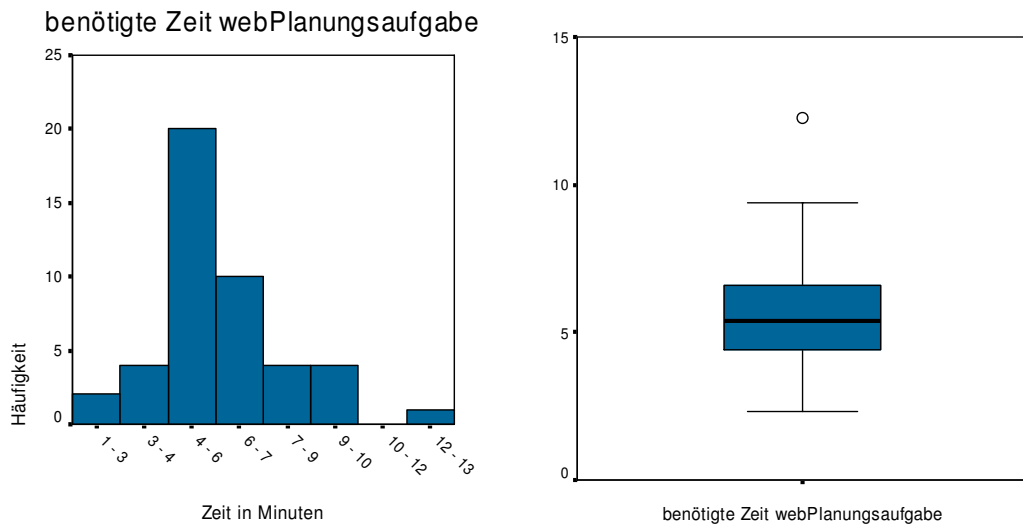
	Skala	N	Mittelwert	SD	Schiefe
<b>Studierende</b>	Summenscore	45	18.58	2.74	-.23
	benötigte Zeit (Sek.)		352.26	124.55	.65
	komb. Gesamtscore		.06	1.21	-.22
<b>Bewerber</b>	Summenscore	369	18.46	3.80	-.82
	benötigte Zeit (Sek.)		430.90	198.19	1.61
	komb. Gesamtscore		.03	1.25	-.79

Die Verteilung der Rohwerte der *webPlanungsaufgabe* (Schulleitungsversion) für die Stichprobe der Studierenden hat ein Mittel von

18.58 Punkten bei einer Standardabweichung von 2.74. Der Median liegt bei 19 Punkten. Die Teilnehmer benötigen im Mittel fünf Minuten und 52.26 Sekunden für die Aufgabe (SD = 124.55 Sek.). Die Bewerber auf Führungsnachwuchsstellen erreichen einen Mittelwert von 18.46 Punkten (SD = 3.80) bei einer mittleren Zeitdauer von 430.90 Sekunden (SD = 198.19). Die ermittelten Schiefewerte liegen für die Summen- und Gesamtscores jeweils im leicht negativen Bereich, was auf schwach rechtssteile Verteilungen schließen lässt. Die Schiefewerte hinsichtlich der benötigten Zeit sind deutlich positiv. Dies weist auf linkssteile Verteilungen hin und entspricht der Erwartung bei der Vorgabe freier Zeiteinteilung. Die Abbildungen 36 bis 38 veranschaulichen die Ergebnisse grafisch mit Hilfe von Histogrammen und Boxplots.



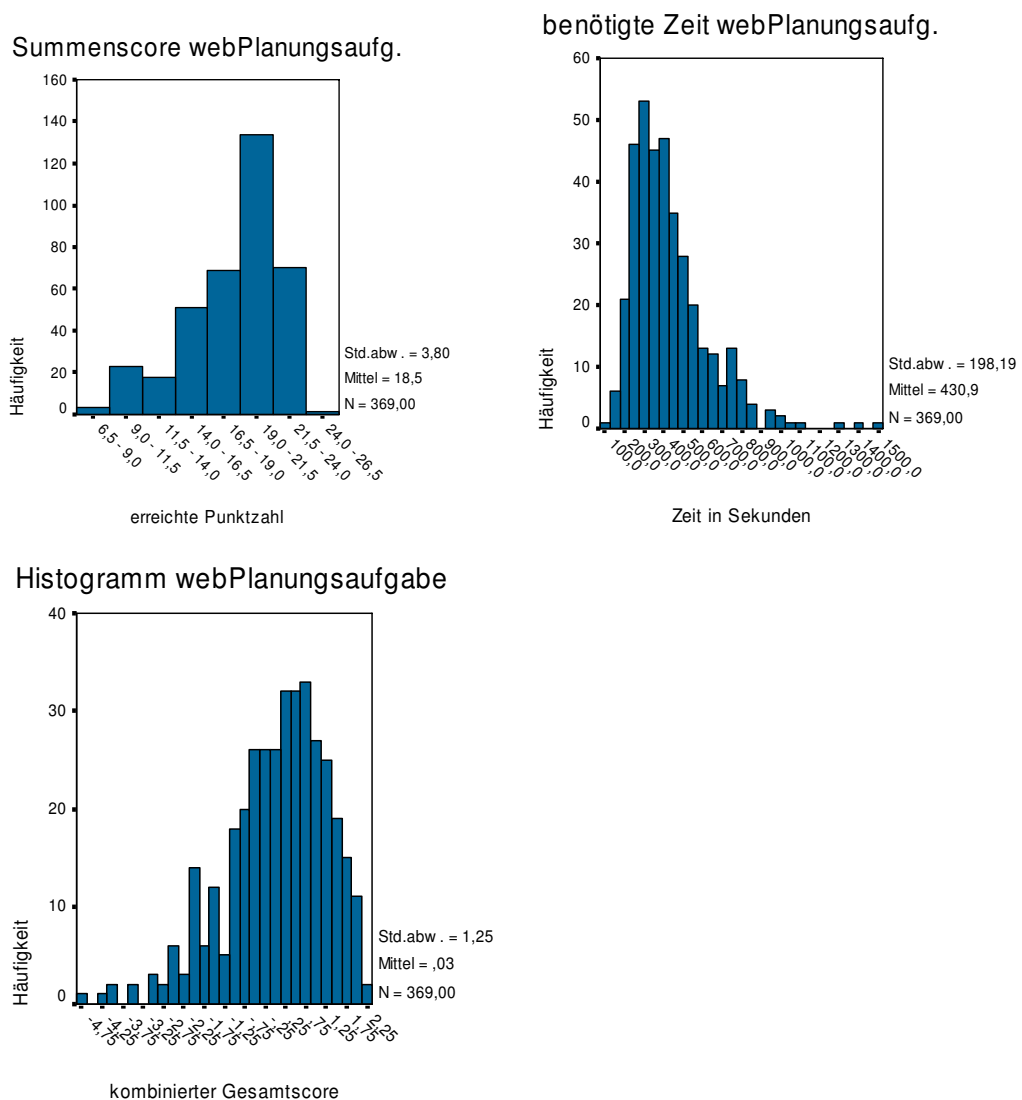
**Abbildung 36: Histogramm und Boxplot für die Punkterohwerte der webPlanungsaufgabe (Stichprobe Studierende, N = 45)**



**Abbildung 37: Histogramm und Boxplot für die benötigte Zeit zur Bearbeitung der *webPlanungsaufgabe* (Stichprobe Studierende, N = 45)**

Die *webPlanungsaufgabe* ist nicht zeitbegrenzt. Das bedeutet, dass die Testpersonen das Verfahren selbständig beenden, sobald die Aufgabe aus ihrer Sicht bewältigt ist. Trotzdem scheint sich eine Tendenz aus der Voruntersuchung auch in der Hauptuntersuchung zu bestätigen: Die meisten Personen benötigen zwischen vier und acht Minuten für die Aufgabe. Über drei Viertel der Testteilnehmer der Hauptuntersuchung absolvieren das Verfahren innerhalb dieser Zeit. Um sowohl die Zeit als auch die erreichte Punktzahl innerhalb der *webPlanungsaufgabe* in einem Wert zu betrachten, wird ein aufsummierter, z-transformierter Gesamtscore gebildet. Die kombinierten Gesamtscores liegen dabei um .06 für die Stichprobe der Studierenden und .03 für die Bewerber-Stichprobe (SD = 1.21 bzw. 1.25). Die Verteilung des kombinierten Gesamtscores der *webPlanungsaufgabe* gibt das unterste Histogramm in Abbildung 38 wieder.





**Abbildung 38: Histogramme für die webPlanungsaufgabe (Stichprobe Führungsnachwuchskräfte, N = 369)**

### *Kurzversion des Berliner Intelligenzstruktur-Tests (BIS-K-S)*

Für die sechs eingesetzten Aufgaben des Berliner Intelligenzstruktur-Tests von Jäger et al. (1997) wurden ebenfalls die Verteilungen der Rohwerte sowie Skalenkennwerte berechnet, um zu überprüfen, inwieweit die Daten für die hier angestrebten Zusammenhangsanalysen geeignet erscheinen. Im Rahmen der Hauptuntersuchung wurde die Kurzform (BIS-K-S) eingesetzt. Es sind dabei insgesamt sechs Aufgaben zum Einsatz gekommen. Jeweils zwei Aufgabentypen sind den drei Inhaltsbereichen „numerisch“, „verbal“ und „figural-bildhaft“ zugeordnet. Im Folgenden werden die Rohwerte der

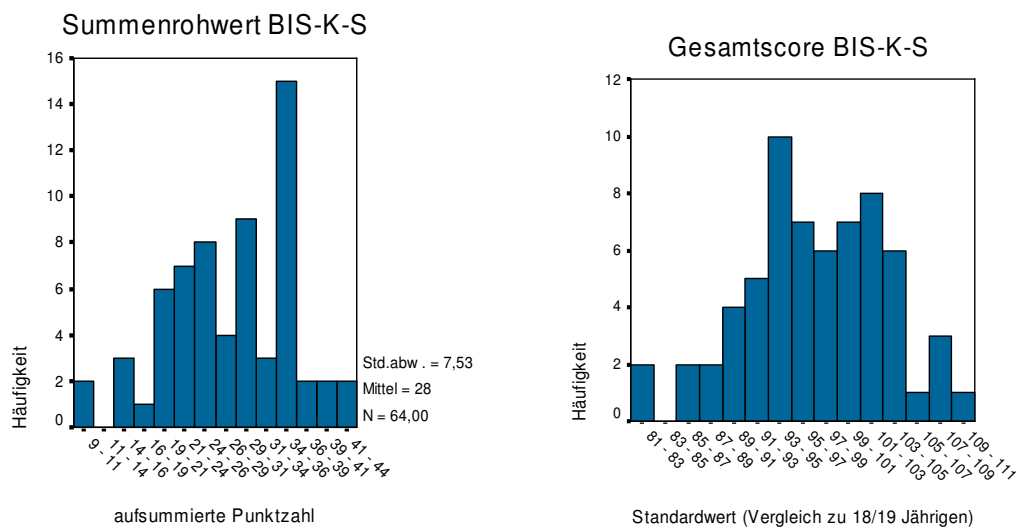
Aufgaben betrachtet. Eine Interpretation im Sinne einer Fähigkeit ist laut Jäger et al. (1997) erst durch Aggregation aller sechs Aufgaben-Rohwerte möglich. Mit Hilfe der Kurzsкала BIS-K-S kann nach Jäger et al. (1997) ein valides Fähigkeitsmaß zur Verarbeitungskapazität gewonnen werden.

**Tabelle 26: Deskriptive Statistiken der Rohwerte der sechs BIS-K-S Aufgaben und des zusammengefassten Summenscores (Hauptuntersuchung)**

Aufgabe	Mittelwert	Standardabweichung	Schiefe
Rohwert ZN (Zahlenreihe)	4.06	2.01	.31
Rohwert SC (Schätzen)	3.47	1.80	.04
Rohwert WA (Wortanalogien)	3.55	1.70	-.06
Rohwert TM (Tatsache-Meinung)	11.26	3.48	-.73
Rohwert AN (Analogien)	3.02	1.44	.52
Rohwert FA (Figurenauswahl)	2.67	1.34	.21
Rohwert BIS-K-S	28.00	7.53	-.38

Anmerkung: N = 64.

Die deskriptiven Statistiken für die Rohwerte der sechs BIS-K-S Aufgaben sind in Tabelle 26 aufgelistet. Die Aufgabe „Tatsache-Meinung“ ist aufgrund eines anderen Aufbaus nicht direkt mit den anderen fünf zu vergleichen. Bei diesem Untertest geht es um sechzehn Aussagen, die jeweils als Tatsache oder Meinung einzuschätzen sind. Daher ergibt sich ein höherer Mittelwert (11.26) und eine größere Standardabweichung (3.48). Die Mittelwerte der fünf Aufgaben mit ähnlichem Aufbau rangieren zwischen 2.67 und 4.06 bei Standardabweichungen zwischen 1.34 und 2.01. Die ermittelten Schiefewerte weichen nicht signifikant von einer symmetrischen Verteilung ab.



**Abbildung 39: Histogramme des Summenrohwerths und des Standardwerths des BIS-K-S (Stichprobe Studierende)<sup>21</sup>**

Die Verteilung der Summenrohwerthe der Kurzform zur Erfassung der Verarbeitungskapazität (BIS-K-S) ist in Abbildung 39 dargestellt. Bei einem Mittelwert von 28 liegt die Standardabweichung bei 7.53 (s. Tabelle 26). Außerdem werden die Standardwerte im Vergleich zur Normierungsstichprobe der 18- bis 19-Jährigen berechnet und als Histogramm wiedergegeben (s. Abbildung 39, rechtes Diagramm). Dies dient lediglich der Veranschaulichung, da von den Testautoren keine Standardwerte für ältere Populationen publiziert wurden (vgl. Jäger et al., 1997). Der Vergleich verdeutlicht jedoch, dass die Leistungen der studentischen Stichprobe der Hauptuntersuchung verhältnismäßig niedrig ausfielen. Der Mittelwert der aufsummierten Punktwerte des BIS-K-S entspricht lediglich einem Standardwert von 93 oder anders ausgedrückt einem Prozentrangwert von 24 bei einem Vergleich mit 18- bis 19-Jährigen. Die mittlere Verarbeitungskapazität der Testpersonen der Stichprobe ist somit als recht niedrig einzuschätzen. Dies kann gegebenenfalls zu Verzerrungen bei anderen Analysen führen. Allerdings liegt die Standardabweichung der Standardwerte mit 9.97 im normalen Bereich. Die Standardwerte der studentischen Hauptuntersuchungsstichprobe scheinen

<sup>21</sup> Zur Vergleichsrechnung wurde hier die Altersnorm der 18- bis 19-Jährigen angewendet.

somit lediglich „nach links verschoben“ im Vergleich zur Eichstichprobe der 18- bis 19-Jährigen. Dies sollte sich nicht bedeutend auf die hier relevanten Zusammenhangsanalysen zur Prüfung der Konstruktvalidität auswirken. Die niedrigen Mittelwerte könnten allerdings auch daraufhin weisen, dass einige Teilnehmer den Test möglicherweise nicht ernsthaft bearbeitet haben. Dies könnte sich dann negativ auf die Qualität der Daten auswirken. Diese Problematik ist allerdings nachträglich nicht letztgültig zu klären.

Zur weiteren Aufklärung der Datenqualität wurden die Aufgabeninterkorrelationen der Hauptuntersuchungstichprobe mit den Angaben aus dem Testheft von Jäger et al. (1997) verglichen. Hinsichtlich der Interkorrelationen der Aufgaben ergeben sich die in Tabelle 27 dargestellten Werte.

**Tabelle 27: Interkorrelationen der Aufgaben mit dem Gesamtscore BIS-K-S / Vergleich mit Werten des BIS-K der Normstichprobe (Jäger et al., 1997)**

<b>Aufgabe</b>	<b>Gesamtscore BIS-K-S</b> (Verarbeitungs- kapazität)	<b>Gesamtscore BIS-K*</b> Angaben aus dem Testheft (Jäger et al., 1997)
Rohwert ZN (Zahlenreihe)	.39	.50
Rohwert SC (Schätzen)	.46	.46
Rohwert WA (Wortanalogien)	.57	.42
Rohwert TM (Tatsache-Meinung)	.32	.40
Rohwert AN (Analogien)	.55	.52
Rohwert FA (Figurenauswahl)	.33	.35

Anmerkung: N = 64. Alle Korrelationen sind auf dem Niveau von .01 (2-seitig) signifikant. Die Korrelationen sind part-whole korrigiert.

\* Bei den Werten von Jäger et al. (1997) handelt es sich um Korrelationen zwischen den Aufgaben und der Verarbeitungskapazität des Gesamttests BIS-K.

Die mittlere Korrelation zwischen den Rohwerten für die einzelnen Aufgaben und dem Gesamtscore Verarbeitungskapazität der Kurzversion liegt bei  $r = .44$ . Die part-whole korrigierten Werte streuen dabei zwischen .32 und .57 (s. Tabelle 27). Dies entspricht in etwa den Angaben aus dem Testheft, die zum Vergleich in der rechten Spalte wiedergegeben sind (Jäger et al., 1997). Es ist zu beachten, dass die Testheftwerte auf Basis der Analyse der Eichstichprobe von  $N = 478$  Schülern entstanden sind. Dabei beziehen sich die Korrelationen - anders als im Rahmen dieser Arbeit - auf den Zusammenhang zwischen der Langversion zur Erfassung der Verarbeitungskapazität des Gesamttests (BIS-K) und den einzelnen Testaufgaben.

Die dargestellten Befunde legen insgesamt den Schluss nahe, dass die Verarbeitungskapazität der Teilnehmer mit Hilfe der Testaufgaben trotz der niedrigen Mittelwerte valide erfasst wird und somit die BIS-K-S-Daten der Hauptuntersuchung für Zusammenhangsanalysen zum *webPostkorb* und zur *webPlanungsaufgabe* hinreichend geeignet sind.

#### *HAKEMP-90*

Neben den Leistungsverfahren ist zur Testung möglicher Zusammenhänge mit relevanten Persönlichkeitsmerkmalen der HAKEMP-90 von Kuhl (1994) zum Einsatz gekommen. Es wurden für dieses Verfahren ebenfalls einige Analysen vorgenommen, um zu untersuchen, inwieweit die Daten für die hier angestrebte Zusammenhangsanalyse mit den Leistungen im *webPostkorb* und in der *webPlanungsaufgabe* geeignet erscheinen.

44 der 66 Personen der Hauptuntersuchung haben freiwillig dieses Messinstrument bearbeitet. Innerhalb des HAKEMP-90 werden folgende drei Merkmale untersucht (zur näheren Erläuterung s. o.):

- Handlungsorientierung nach Misserfolgserfahrung (HOM)
- Grad der Entscheidungs- und Handlungsplanung (HOP)
- Handlungsorientierung bei der Tätigkeitsausführung (HOT)

Tabelle 28: Häufigkeiten HAKEMP-90

Skala	Handlungsorientierung nach Misserfolgs-erfahrung (HOM)		Grad der Entscheidungs- und Handlungsplanung (HOP)		Handlungsorientierung bei der Tätigkeits-ausführung (HOT)	
	erreichte Punktzahl	Anzahl der Personen	Bewertung	Anzahl der Personen	Bewertung	Anzahl der Personen
0	4	Lageorientierung LOM (entspricht 34.1% der Befragten)	0	Lageorientierung LOP (43.2%)	0	Lageorientierung LOT (47.7%)
1	2		0		0	
2	1		2		0	
3	4		7		0	
4	4		6		0	
5	5	Handlungsorientierung HOM (65.9%)	4	Handlungsorientierung HOP (56.8%)	0	Handlungsorientierung HOT (52.3%)
6	7		5		2	
7	6		7		6	
8	4		7		5	
9	3		3		8	
10	2		1		4	
11	2		2		9	
12	0		0		10	
Gesamt	44		44		44	

Da laut Kuhl (1994) die drei Skalen des HAKEMP-90 mit unterschiedlichen Verhaltenskennwerten korrelieren, empfiehlt der Autor die Kennwerte für die Skalen getrennt zu berechnen. Tabelle 28 gibt Aufschluss über die Häufigkeitsverteilungen der Punktwerte für die drei Skalen. Der „Cut-off-Wert“ zwischen Lage- und Handlungsorientierung ist jeweils abhängig von der Skala. Zur Illustration sind die Bereiche farblich abgesetzt. Von den 44 Testpersonen, die den HAKEMP-90 bearbeitet haben, sind 15 Personen (entspricht 34.1 Prozent) auf der Skala „Handlungsorientierung nach

Misserfolgserfahrung“ (HOM) als lageorientiert und 29 Personen (65.9 Prozent) als handlungsorientiert einzustufen. Hinsichtlich des Grads der Entscheidungs- und Handlungsplanung (HOP) gelten 43.2 Prozent der Befragten als lageorientiert gegenüber 56.8 Prozent Handlungsorientierten. Hinsichtlich der Handlungsorientierung bei der Tätigkeitsausführung (HOT) schätzen sich jeweils fast die Hälfte der Testpersonen als lageorientiert (47.7 Prozent) bzw. handlungsorientiert (52.3 Prozent) ein. Inwieweit die prozentuale Verteilung der Handlungs- und Lageorientierten innerhalb der Stichprobe der Hauptuntersuchung im Rahmen der erwarteten Größenordnung liegt, konnte den Handanweisungen des HAKEMP-90 (Kuhl, 1994) nicht entnommen werden.

Tabelle 29 stellt Skalenkennwerte zur inneren Konsistenz sowie die mittleren Iteminterkorrelationen dar. Die Werte für Cronbachs-Alpha liegen zwischen .56 und .76 und somit etwas niedriger als die Vergleichswerte aus der Eichstichprobe (Kuhl, 1994).

**Tabelle 29: Iteminterkorrelationen und interne Konsistenzen HAKEMP-90**

Statistiken	Stichprobe	HOM	HOP	HOT
mittlere Iteminterkorrelation	Hauptuntersuchung	.21	.11	.10
Cronbachs-Alpha	Hauptuntersuchung	.76	.61	.56
	Eichstichprobe (Kuhl, 1994)	.70	.78	.74

Anmerkung: N = 44.

In Tabelle 30 sind die Interkorrelationen der Skalen des HAKEMP-90 wiedergegeben. Die Werte in Klammern sind den Angaben Kuhls (1994) entnommen und dienen als Vergleich. Die Skalen HOM und HOP hängen signifikant in Höhe von  $r = .37$  zusammen, während die Korrelationen zur HOT-Skala negativ in Höhe  $r = -.23$  (HOM) und  $r = -.32$  (HOP) ausfallen. Die

Ergebnisse sind mit denen der Eichstichprobe vergleichbar, lediglich die Korrelation zwischen HOP und HOT im Rahmen dieser Untersuchung erweist sich mit  $-.32$  im Unterschied zu den Ergebnissen von Kuhl (1994) als signifikant. Dieser Befund ist jedoch mit den Angaben von Kuhl (1994) zu vereinbaren, wonach die Handlungsorientierung nach Misserfolgserfahrung und der Grad der Entscheidungs- und Handlungsplanung auch zu einem Wert zusammengefasst werden können. Handlungsorientierung bei der Tätigkeitsausführung (HOT) hingegen erfasst eine relativ eigenständige Komponente, die demnach nicht mit HOM und HOP zu kombinieren ist. Bei der Vorläuferversion des HAKEMP-90, dem HAKEMP-88, wurde von negativen Zusammenhängen zwischen HOT-Skala und HOM- bzw. HOP-Skala berichtet, wie sie hier im Rahmen der Hauptuntersuchung gefunden wurden (vgl. Kuhl, 1994).

**Tabelle 30: Interkorrelationen der HAKEMP-90-Skalen**

Skala	HOM	HOP
Handlungsorientierung nach Misserfolgserfahrung (HOM)		
Grad der Entscheidungs- und Handlungsplanung (HOP)	.37*   (.43)	
Handlungsorientierung bei der Tätigkeitsausführung (HOT)	-.23   (-.08)	-.32*   (.06)

Anmerkung: N = 64. Die Vergleichswerte der Eichstichprobe von Kuhl (1994) stehen in Klammern.

\* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

Trotz der Unsicherheiten in Bezug auf die prozentuale Verteilung der Handlungs- und Lageorientierten innerhalb der Stichprobe deuten die Befunde insgesamt darauf hin, dass es sich hier um valide Messungen handelt und daher die HAKEMP-90-Daten der Hauptuntersuchung zur Überprüfung der Konstruktvalidität des *webPostkorbs* und der *webPlanungsaufgabe* geeignet sind.



### *Prüfung der Leistungsskalen auf Normalverteilung*

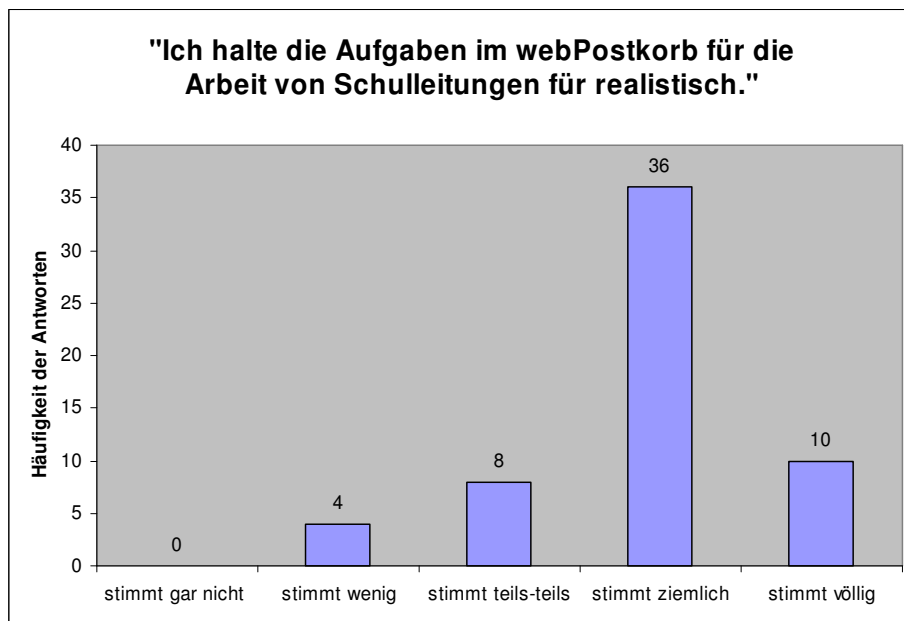
Für weitergehende Analysen ist die Annahme der Normalverteilungshypothese für die Leistungsskalen zu überprüfen. Dies erfolgt mit Hilfe des Kolmogorov-Smirnov-Anpassungstests. Für die erhobenen Leistungswerte des BIS-K-S, des *webPostkorbs* (Gesamtscore sowie die Subskalen *dynamisches Problemlösen* und *analytisches Problemlösen*) und der *webPlanungsaufgabe* (Summenscore sowie kombinierter Gesamtscore) ergeben sich p-Werte  $>.05$ . Es ist somit von hinreichend normalverteilten Werten auszugehen.

## **7.4 Inhaltsvalidität**

Die anforderungsgeleitete Verfahrenskonstruktion, die die wesentliche Grundlage für die Inhaltsvalidität von eignungsdiagnostischen Verfahren darstellt (vgl. von Rosenstiel, 2000; Schuler, 2001; Wottawa, 1995), wurde bereits weiter oben erörtert. Neben der inhaltvaliden Entwicklung mit der Einbeziehung von Experten und Studierenden in allen Phasen der Konstruktion (s. o.), sollten darüber hinaus analog zur Vorgehensweise im Rahmen der Voruntersuchung auch hier die Einschätzungen der Testpersonen gesammelt werden. Dabei interessiert, wie die Benutzer den Zusammenhang zwischen den Verfahrensinhalten und dem zu messendem Verhalten bewerten. Dieser als Augenscheinvalidität bezeichnete Zusammenhang ist insofern von Interesse, als dass die postulierte Überlegenheit der Akzeptanz von Simulationsverfahren gegenüber herkömmlichen Tests vermutlich auch auf der zumeist höheren Augenscheinvalidität basiert (vgl. Höft & U. Funke, 2001).

Die positiven Ergebnisse zur Augenscheinvalidität der *webPlanungsaufgabe* wurden bereits im Rahmen der Voruntersuchungsergebnisse dargelegt, so dass hier auf die Benutzerbewertungen des *webPostkorbs* fokussiert werden kann.

Abbildung 40 veranschaulicht, wie die befragten Studierenden die Aussage „Ich halte die Aufgaben im webPostkorb für die Arbeit von Schulleitungen für realistisch“ bewerten. Über drei Viertel der Studierenden stimmen dem Statement völlig oder ziemlich zu. Es zeigt sich somit, dass die große Mehrheit der Testpersonen den *webPostkorb* als realistisch für den Beruf von Schulleitungen einschätzt. Dies ist als Hinweis für die anvisierte Augenscheinvalidität des *webPostkorbs* zu werten.



**Abbildung 40: Bewertung der Realitätsnähe des *webPostkorbs* (Hauptuntersuchung)**

## 7.5 Kriteriumsvalidität

### *Zusammenhänge zur Abiturnote und Diplomnote*

Die Teilnehmer der Hauptuntersuchung wurden nach ihrer Abitur- und Zwischenprüfungsnote befragt. Tabelle 31 stellt die Zusammenhänge zwischen den Noten und den eingesetzten Leistungsverfahren dar. Die unterschiedliche Stichprobengröße ist dem Umstand geschuldet, dass elf Personen keine Angaben zur Abiturnote machten. Drei der 33 Studierenden, die die Zwischenprüfung bereits absolviert hatten, gaben die Note nicht an, so dass die Korrelationen zur Zwischenprüfung auf Basis von  $N = 30$  Studierenden ermittelt wurden. Alle Korrelationen haben erwartungsgemäß

negative Vorzeichen. Die Zusammenhänge zur Abiturnote liegen zwischen  $r = -.08$  und  $r = -.35$ . Die Zusammenhänge zur Zwischenprüfungsnote sind noch stärker ausgeprägt, und zwar in Höhe von  $r = -.24$  bis  $r = -.47$ . Aufgrund der kleinen Stichprobengröße weisen lediglich die Korrelationen des Summenscores bzw. des Kombinationsscores der *webPlanungsaufgabe* mit der Abiturnote bzw. mit der Zwischenprüfungsnote Signifikanz auf ( $r = -.35$  sowie  $r = -.47$ ). Der Zusammenhang zwischen der Abitur- und der Zwischenprüfungsnote ist substantiell in Höhe von  $r = .58$ .

**Tabelle 31: Korrelationen der Abitur- und der Zwischenprüfungsnote mit den Ergebnissen in den Leistungsverfahren (Hauptuntersuchung, Stichprobe Studierende)**

Skala	N	Abiturnote	N	Note der Zwischenprüfung
BIS-K-S	55	-.08	30	-.28
Gesamtscore <i>webPostkorb</i>		-.12		-.24
Summenscore <i>webPlanungsaufgabe</i>	41	-.26	24	-.47*
komb. Gesamtscore <i>webPlanungsaufgabe</i>		-.35*		-.26
Abiturnote			30	.58**

\* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

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

Wird die Note der Zwischenprüfung als Operationalisierung des Kriteriums „Studienerfolg“ verstanden, so ist gerade die Korrelation zum Summenscore der *webPlanungsaufgabe* aufschlussreich. Der Anteil gemeinsamer Varianz ist mit 22 Prozent beachtlich, berücksichtigt man die zeitökonomische Messung<sup>22</sup> innerhalb der *webPlanungsaufgabe*. Schmidt und Hunter (1998, 2000) weisen auf den Anteil inkrementeller Validität verschiedener

<sup>22</sup> Im Vergleich zu den meisten Leistungstestverfahren (vgl. Sarges & Wottawa, 2005) ist die *webPlanungsaufgabe* mit einer mittleren Bearbeitungsdauer von etwa sechs Minuten ein sehr kurzes Verfahren.

Personalauswahlverfahren gegenüber kognitiven Leistungsverfahren hin (vgl. Abschnitt 3.1). Selbst bei Kontrolle der Verarbeitungskapazität (BIS-K-S) bleibt die partielle Korrelation zwischen Zwischenprüfungsnote und Summenscore mit  $r = -.42$  noch substantiell.

**Tabelle 32: Hierarchische Regression von Verarbeitungskapazität, Planungsleistung und Problemlöseleistung auf das Kriterium Studienerfolg (Note der Zwischenprüfung)**

Prädiktoren	Kriterium Studienerfolg (Note der Zwischenprüfung)					
	r	R	Beta	R <sub>diff</sub>	R <sup>2</sup>	R <sup>2</sup> <sub>korr</sub>
Verarbeitungskapazität (BIS-K-S)	-.23	-.23	-.01			
Planungsleistung ( <i>webPlanungsaufgabe</i> )	-.47*	-.47*	-.41	.24*		
Problemlöseleistung ( <i>webPostkorb</i> )	-.30*	-.48*	-.13	.01	.23	.12

Anmerkung: N = 24; \*  $p < .05$ ; Beta, R, R<sup>2</sup> und das adjustierte R<sup>2</sup><sub>korr</sub> für das Gesamtmodell mit drei Prädiktoren.

Trotz der kleinen Stichprobengröße ist es interessant zu betrachten, inwieweit die *webPlanungsaufgabe* und der *webPostkorb* weitere Varianzanteile des Kriteriums Studienerfolg erklären können. Die Ergebnisse der hierarchischen Regression stellt Tabelle 32 dar mit den ermittelten Werten der Einzelkorrelationen (r), multiplen Korrelationen (R) sowie dem gemeinsamen Varianzanteil zwischen Kriterium und Prädiktor (R<sup>2</sup> bzw. R<sup>2</sup><sub>korr</sub>) nach Berücksichtigung aller unabhängigen Variablen. Im ersten Schritt ergibt sich ein R von  $-.23$  bei Einschluss der Variable Verarbeitungskapazität des BIS-K-S. Im zweiten Schritt wurde die Leistung in der *webPlanungsaufgabe* als Prädiktor einbezogen. Dadurch steigt das R signifikant von  $-.23$  auf  $-.47$  an ( $F_{\text{change}} = 4.51$ ,  $p < .05$ ). Dies bedeutet eine zusätzliche Varianzaufklärung von sechzehn Prozent. Nach Einschluss der Problemlöseleistung im *webPostkorb* kann die multiple Korrelation kaum mehr erhöht werden (R =  $-.48$ ). Für die hier vorliegende Stichprobe erweist sich die

Planungsleistung in der *webPlanungsaufgabe* als bedeutendste Variable mit einem standardisierten Regressionskoeffizienten (Beta) von  $-.41$ .

In der Stichprobe der potenziellen Führungsnachwuchskräfte wurden die Abiturnote und die Diplomnote erhoben. Die unterschiedlichen Stichprobengrößen sind auf die jeweiligen Auswahlstufen zurückzuführen. Dabei zeigen sich nur wenige signifikante Korrelationen zu den eingesetzten Auswahlverfahren (s. Tabelle 33). Das Abschneiden im Telefoninterview hängt demnach zu  $r = -.25$  bzw.  $-.27$  mit der Abiturnote zusammen. Der Zusammenhang zwischen Abiturnote und Diplomnote ist mit  $r = .13$  zwar gering, bei einem N von 369 allerdings ebenfalls überzufällig. Alle anderen Korrelationen sind nicht signifikant.

**Tabelle 33: Korrelationen der Abitur- und der Zwischenprüfungsnote mit den Leistungsverfahren der Führungsnachwuchskräfte-Stichprobe**

Verfahren	N	Abitur- note	Diplom- note
Abiturnote	369	1.00	.13*
Summenscore <i>webPlanungsaufgabe</i>		-.03	-.07
komb. Gesamtscore <i>webPlanungsaufgabe</i>		.05	-.12
unternehmensspezifisches E-Assessment kognitive Leistungsfähigkeit		-.02	.02
Telefoninterview bestanden	178	-.25**	-.07
Summenscore Telefoninterview		-.27**	-.02
I-S-T 2000 R Gesamtscore	41	-.16	-.08
Assessment-Center bestanden		-.02	.00
Summenscore Assessment-Center		.12	.10

\* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

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

Die Ergebnisse des substanziellen Zusammenhangs zwischen *webPlanungsaufgabe* und „Studienerfolg“ aus der universitären Hauptuntersuchung können für die Stichprobe der Führungsnachwuchskräfte nicht wiederholt werden. Die Korrelationen der *webPlanungsaufgabe* mit der

Diplomnote liegen bei  $r = -.07$  bzw.  $r = -.12$ . Es handelt sich bei dieser Stichprobe allerdings um eine vorselektierte Gruppe an Bewerbern. Durch diese Vorauswahl wird die Varianz vermutlich eingeschränkt, was die niedrigen Korrelationen bedingt haben könnte. Entgegen den hier ermittelten Werten zeigt die Metaanalyse von Hell, Trapmann, Weigand, Hirn & Schuler (2005) Zusammenhänge von .46 zwischen Abitur- und Studiennoten und Korrelationen in Höhe von .36 zwischen der Allgemeinen Intelligenz und Studiennoten.

**Tabelle 34: Hierarchische Regression von Abiturnote, kognitiver Leistungsfähigkeit und Planungsleistung auf das Kriterium Studienerfolg (Note der Diplomprüfung)**

Prädiktoren	Kriterium Studienerfolg (Diplomnote)					
	r	R	Beta	R <sub>diff</sub>	R <sup>2</sup>	R <sup>2</sup> <sub>korrr</sub>
Abiturnote	-.13*	-.13*	-.14			
kognitive Leistungsfähigkeit (unternehmensspezifisches E-Assessment)	.02	-.13*	.07	.00		
Planen (komb. Score der <i>webPlanungsaufgabe</i> , Wirtschaftsversion)	-.12	-.19*	-.16	.06*	.04	.03

Anmerkung: N = 362; \*  $p < .05$ ; Beta, R, R<sup>2</sup> und das adjustierte R<sup>2</sup><sub>korrr</sub> für das Gesamtmodell mit drei Prädiktoren.

Neben der bivariaten Analyse ist es in diesem Zusammenhang interessant zu untersuchen, inwieweit die Leistung in der *webPlanungsaufgabe* (Wirtschaftsversion) weitere Varianzanteile des Kriteriums Studienerfolg über die Prädiktoren Abiturnote und kognitive Leistungsfähigkeit hinaus erklären kann. Die Ergebnisse der hierarchischen Regression stellt Tabelle 34 dar mit den ermittelten Werten der Einzelkorrelationen (r), multiplen Korrelationen (R) sowie dem gemeinsamen Varianzanteil zwischen Kriterium und Prädiktor (R<sup>2</sup> bzw. R<sup>2</sup><sub>korrr</sub>) nach Berücksichtigung aller unabhängigen Variablen. Im ersten Schritt ergibt sich ein R von -.13 bei Einschluss des Prädiktors Abiturnote. Im zweiten Schritt wurde die kognitive Leistungsfähigkeit im

unternehmensspezifischen E-Assessment als Variable einbezogen. Die multiple Korrelation (R) erweist sich dabei als unverändert. Nach Einschluss des Regressors Planungsleistung in der *webPlanungsaufgabe* (Wirtschaftsversion) kann die multiple Korrelation signifikant um  $R_{\text{diff}} = .06$  auf  $R = -.19$  gesteigert werden ( $F_{\text{change}} = 8.08$ ,  $p < .05$ ). Dies entspricht einer inkrementellen Varianzaufklärung von etwa zwei Prozent. Die gesamte aufgeklärte Varianz des Kriteriums beim Modell mit drei Prädiktoren liegt allerdings mit vier Prozent sehr niedrig, dies kann wiederum durch die Varianzeinschränkung aufgrund der starken Vorselektion bedingt sein.

#### *Einsatz der webPlanungsaufgabe im Rahmen von Personalauswahlverfahren*

Die Wirtschaftsversion der *webPlanungsaufgabe* konnte unter Ernstfallbedingungen im Rahmen der Personalauswahl durchgeführt werden. Die Zusammenhänge zwischen der Planungsleistung in der *webPlanungsaufgabe* und den anderen eingesetzten Auswahlverfahren sind in Tabelle 35 beschrieben. Es ergeben sich signifikante, niedrige Korrelationen in Höhe von  $r = .12$  bis  $.16$  zwischen den Leistungen in der *webPlanungsaufgabe* und dem Telefoninterview. Ein substanzieller Zusammenhang zum Abschneiden im eingesetzten Assessment-Center liegt nicht vor - die Werte liegen um Null. Dies ist nicht weiter verwunderlich, da die Konzeption dieses Assessment-Centers in erster Linie auf die Erfassung sozialer und kommunikativer Fähigkeiten ausgerichtet ist und analytische Fähigkeiten, wie sie typischerweise in Postkorbverfahren ermittelt werden, nicht erhoben werden. Die Leistung im E-Assessment korreliert signifikant in Höhe von  $r = .28$  bzw.  $r = .37$  mit dem Abschneiden in der *webPlanungsaufgabe*. Das E-Assessment besteht aus drei kognitiven Leistungsverfahren. Dabei sind in Anlehnung an das Berliner Intelligenzstrukturmodell numerische, verbale und figural-bildhafte Aufgabeninhalte zu bearbeiten. Die Leistung in den drei Verfahren wird zu einem „Gesamtscore E-Assessment“ aufsummiert. Daher ist dieses Ergebnis auch als Beleg für die Konstruktvalidität der *webPlanungsaufgabe*

(Wirtschaftsversion) zu verstehen (vgl. genauer Abschnitt 7.6.4 Zusammenhang mit kognitiven Leistungsverfahren).

**Tabelle 35: Zusammenhang zwischen der *webPlanungsaufgabe* und den Bausteinen des Personalauswahlverfahrens**

Personalauswahlverfahren	N	<i>webPlanungsaufgabe</i> - Wirtschaftsversion		
		Summenscore	benötigte Zeit	kombinierter Gesamtscore
Summenscore E-Assessment	369	.28**	-.19**	.37**
Bestanden Telefoninterview	178	.14**	-.07	.16**
Summenscore Telefoninterview		.12*	-.07	.14**
Bestanden AC	41	n. s.	n. s.	n. s.
Summenscore AC		n. s.	n. s.	n. s.

\* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

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

### *Extremgruppenvergleich webPlanungsaufgabe*

Im Rahmen der Stichprobe der Führungsnachwuchskräfte konnten auch die Werte bei 19 bewährten Führungsnachwuchskräften erhoben werden. Diese Personen sind bereits seit geraumer Zeit erfolgreich im Unternehmen tätig. Hinsichtlich der Werte der *webPlanungsaufgabe* ergibt sich ein signifikanter Mittelwertsunterschied zwischen den bewährten Führungsnachwuchskräften und den Bewerbern (s. Tabelle 36). Für den Summenscore berechnen sich Unterschiede hinsichtlich der Varianz und des Mittelwerts. Der erzielte Mittelwert der 19 Führungsnachwuchskräfte liegt mit 21.00 höher als bei der Vergleichsstichprobe der Bewerber. Die Standardabweichung wiederum weist mit 2.29 für die Führungsnachwuchskräfte-Stichprobe einen niedrigeren Wert aus. Die Signifikanztests ergeben überzufällige Kennzahlen ( $T = -2.81$ ,  $df = 386$ ,  $p < .05$  bzw.  $F = 6.94$ ,  $p < .05$ ). Für den kombinierten Score der *webPlanungsaufgabe* kann ebenfalls ein signifikanter



Mittelwertsunterschied ermittelt werden ( $T = -2.44$ ,  $df = 386$ ,  $p < .05$ ). Diese Befunde können als Hinweise auf die Kriteriumsvalidität der Wirtschaftsversion der *webPlanungsaufgabe* interpretiert werden.

**Tabelle 36: Extremgruppenvergleich zwischen Bewerbern und bewährten Führungsnachwuchskräften in der *webPlanungsaufgabe* (Wirtschaftsversion)**

Skala	Status	N	Mittelwert	SD
Summenscore	Bewerber	369	18.46	3.84
	Führungsnachwuchskräfte	19	21.00	2.29
benötigte Zeit (Sek.)	Bewerber	369	430.90	198.19
	Führungsnachwuchskräfte	19	413.21	159.46
kombinierter Gesamtscore	Bewerber	369	.03	1.25
	Führungsnachwuchskräfte	19	.70	.90

## 7.6 Konstruktvalidität

Ein wichtiges Anliegen dieser Arbeit liegt in der Überprüfung der Konstruktvalidität der Verfahren. Es wurden daher „konstruktnahe“ Messinstrumente - wie die oben beschriebenen Intelligenztests - eingesetzt, um Zusammenhangsanalysen vornehmen zu können. Darüber hinaus gilt es den Einfluss des Persönlichkeitsmerkmals „Handlungs- und Lageorientierung“ und die Verfahrensanwendung im Rahmen von realen Personalauswahlverfahren unter Ernstfallbedingungen zu untersuchen. Bevor diese Aspekte im Mittelpunkt der Darstellung stehen, soll allerdings einleitend zunächst die anvisierte theoretische Struktur des *webPostkorbs* überprüft werden.

### **7.6.1 Dimensionsanalysen des *webPostkorbs* („innere“ Konstruktvalidität)**

Wie oben erläutert, wird im *webPostkorb* aufgrund der Heterogenität des Konstrukts „Problemlösefähigkeit“ zwischen den Facetten *dynamisches Problemlösen* und *analytisches Problemlösen* unterschieden. Es gilt daher die Fragestellung zu erörtern, inwieweit sich der aus der Theorie abgeleitete inhaltliche Aufbau des *webPostkorbs* mit den entsprechenden Operationalisierungen in der Empirie bestätigt und das Verfahren die inhaltlichen Facetten differenzieren kann.

Dabei sind verschiedene Analysen zum Einsatz gekommen: zum einen die explorative Hauptkomponentenanalyse und zum anderen zur zusätzlichen Absicherung die konfirmatorische Faktorenanalyse.

An die Durchführung von Faktorenanalysen sind eine Reihe von Voraussetzungen zu knüpfen. Eine Stichprobengröße von  $N = 60$  ist laut der Studie von MacCallum, Widaman, Zhang und Hong (1999) ausreichend, wenn - neben der Annahme der multivariaten Normalverteilung - die Kommunalitäten mindestens .60 betragen. Diese Aspekte treffen für den vorliegenden Fall zu (Ergebnisse s. u.). Einen weiteren Anhaltspunkt dafür, ob die Variablenauswahl für eine Faktorenanalyse angebracht ist, gibt der Kaiser-Meyer-Okin-Koeffizient. Demnach sind die hier vorliegenden Daten geeignet, da sich der KMO-Wert als größer .60 erweist (vgl. Bühner, 2004). Trotz der kleinen Stichprobengröße können die Voraussetzungen für die Berechnung von Faktorenanalysen somit als gegeben angenommen werden.

Die Ergebnisse der explorativen Faktorenanalysen sind in Tabelle 37 wiedergegeben. Als Ausgangspunkt dient die Korrelationsmatrix der fünf Aufgabengruppen. Neben der orthogonalen Rotation erscheint aufgrund der zu erwartenden Interdependenzen zwischen den dynamischen und analytischen Problemlösefähigkeiten auch eine schiefwinklige Rotation angemessen. Die Werte der schiefwinkligen Rotation nach der Hauptachsenmethode sind zuerst angegeben, die beiden weiteren Werte sind nach der schiefwinkligen und orthogonalen Rotation der

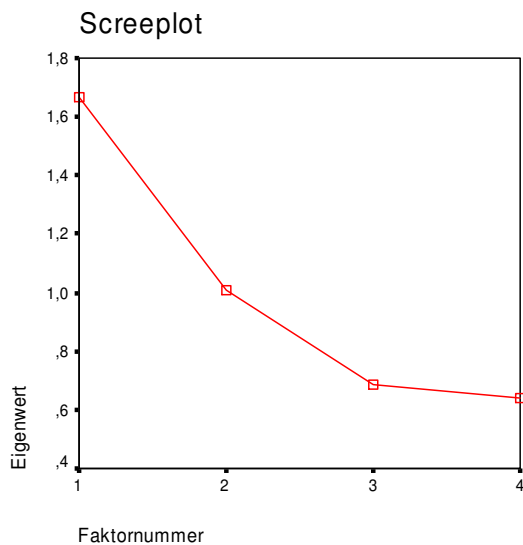
Hauptkomponentenanalyse berechnet. Alle Verfahren ergeben eine zweifaktorielle Struktur bei einem Eigenwert größer als eins mit einem Faktor „analytische Problemlösefähigkeit“ und einem Faktor „dynamische Problemlösefähigkeit“. Die Aufgabengruppen *Wochenplan Vormittag* und *Wochenplan Nachmittag* laden dabei jeweils hoch auf dem zweiten Faktor, während die anderen drei Aufgabengruppen stark mit dem ersten Faktor zusammenhängen (s. Tabelle 37).

**Tabelle 37: Dimensionsanalysen des *webPostkorbs* (Hauptuntersuchung)**

Aufgabengruppe	Komponenten (Faktoren)	
	analytische Problemlösefähigkeit (F1)	dynamische Problemlösefähigkeit (F2)
<i>Wochenplan Vormittag</i>	-0.04 / -0.06 (.01)	.46 / .81 (.80)
<i>Wochenplan Nachmittag</i>	.02 / -0.07 (.12)	.76 / .81 (.80)
<i>Tabellenverständnis</i>	.45 / .63 (.64)	.18 / .47 (.29)
<i>Textverständnis</i>	.58 / .76 (.76)	-.03 / .13 (.07)
<i>Diagrammverständnis</i>	.63 / .79 (.77)	-.14 / -.12 (-.13)
Varianzaufklärung (in %)	42.1	25.2

Anmerkung: N = 66. In Klammern stehen die Ergebnisse der orthogonalen Faktorenstruktur.

Dieses Ergebnis bestätigt die anvisierte Konstruktion des *webPostkorbs* mit zwei Subskalen. Insgesamt werden 67.3 Prozent der Gesamtvarianz durch die beiden Faktoren erklärt. Dabei fallen 42.1 Prozent der Varianzaufklärung auf den ersten Faktor „analytische Problemlösefähigkeit“ und 25.2 Prozent auf den Faktor „dynamische Problemlösefähigkeit“. Für die schiefwinklige Faktorenrotation (Hauptachsenanalyse) ist in Abbildung 41 der Screeplot dargestellt.



**Abbildung 41: Screeplot der Faktorenanalyse *webPostkorb* (Hauptuntersuchung)**

Als Voraussetzungen zur Berechnung einer konfirmatorischen Faktorenanalyse werden neben der theoretischen Fundierung im Allgemeinen das Vorliegen normalverteilter manifester Variablen sowie eine ausreichende Stichprobengröße genannt (vgl. Backhaus, Erichson, Plinke & Weiber, 2003). Die Daten genügen der ersten statistischen Voraussetzung sowohl auf Ebene der Aufgabengruppen als auch hinsichtlich der Multinormalverteilung (Exzess = .21; „c.r.“-Wert = .10). Allerdings liegt der Stichprobenumfang deutlich unter der zumeist angegebenen Grenze von  $N = 100$ . Eine andere Empfehlung (vgl. Bagozzi, 1981 zitiert nach Backhaus et al., 2003) wiederum lautet, dass die Voraussetzung zur Berechnung konfirmatorischer Faktorenanalysen erfüllt sei, wenn die Differenz aus der Stichprobenanzahl und der Anzahl der zu schätzenden Parameter 50 überschreite. Wenngleich diese Bedingung im Rahmen dieser Untersuchung erfüllt wird, sollen die folgenden Analysen nicht als hypothesenprüfend verstanden werden, da es nicht auszuschließen ist, dass die in Tabelle 38 genannten Indizes aufgrund der relativ kleinen Stichprobe überschätzt werden. Die Berechnung der konfirmatorischen Faktorenanalyse soll vielmehr die Ergebnisse der oben beschriebenen explorativen Faktorenanalyse zusätzlich überprüfen. Eine tiefgehende Auseinandersetzung mit den Voraussetzungen der Strukturgleichungs-

modelle findet sich bei Backhaus et al. (2003), Hu und Bentler (1999), Jöreskog und Sörbom (1989, 1993) sowie Pfeifer und Schmidt (1987).

**Tabelle 38: Kennwerte für die Anpassungsgüte verschiedener faktorenanalytischer Modelle (Hauptuntersuchung)**

Modell	$\chi^2$	df	p	GFI	AGFI	CFI	RMSEA	LO90	HI90
Generalfaktor-Modell	14.4	7	.05	.91	.81	.63	.13	.02	.22
Zwei-Faktoren-Modell (korreliert)	3.3	5	.66	.98	.94	1.00	.00	.00	.14

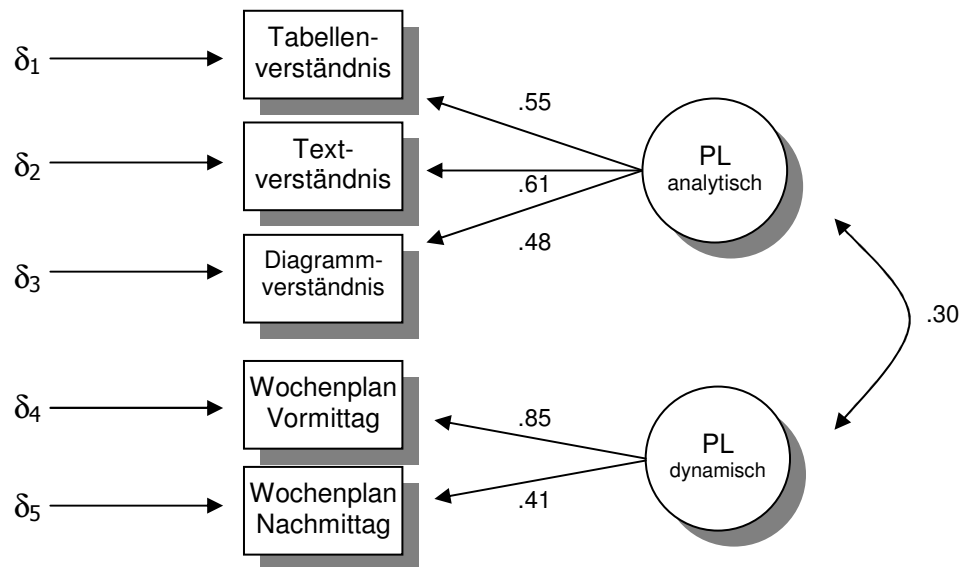
Anmerkung: N = 66; df: Freiheitsgrade; p: Irrtumswahrscheinlichkeit; GFI: Goodness-of-Fit-Index; AGFI: Adjusted Goodness-of-Fit-Index; CFI: Comparative-Fit-Index; RMSEA: Root-Mean-Square-Error-of-Approximation; LO90, HI90: unteres und oberes 90-prozentiges Konfidenzintervall des RMSEA.

Die Ergebnisse der konfirmatorischen Faktorenanalyse gibt Tabelle 38 wieder. Es werden zwei Modelle getestet, die sich aus der Theorie ableiten lassen - zum einen ein Modell mit einem Generalfaktor „Problemlösefähigkeit“ und zum anderen das Zwei-Faktoren-Modell mit der Unterscheidung zwischen „dynamischer Problemlösefähigkeit“ und „analytischer Problemlösefähigkeit“. Die beste Anpassung an die empirischen Daten zeigt analog zu den Ergebnissen der explorativen Faktorenanalysen (s. o.) das Modell mit zwei korrelierten Faktoren. Der  $\chi^2$ -Hypothesentest ergibt ein nicht signifikantes p von .66 ( $\chi^2 = 3.27$  bei fünf Freiheitsgraden), so dass eine hinreichende Modellpassung vorliegt.<sup>23</sup> Backhaus et al. (2003) nennen als Grenzwert  $p > .10$ . Der Goodness-of-Fit-Index (GFI) und der Adjusted Goodnes-of-Fit-Index (AGFI) liegen für das Zwei-Faktoren-Modell bei .98 und .94 und damit im Akzeptanzbereich größer als .90 (vgl. Backhaus et al., 2003). Für das Generalfaktor-Modell werden

<sup>23</sup> Im Gegensatz zum üblichen Vorgehen beim Hypothesentesten wird hier ein Modell angenommen, sobald die  $\chi^2$ -Prüfgröße insignifikant wird, das heißt, möglichst groß ist.

Werte von .91 (GFI) und .81 (AGFI) ermittelt, die sich auch noch im annähernd akzeptablen Bereich bewegen, jedoch im Vergleich deutlich niedriger liegen.

Gerade bei kleinen Stichproben ist laut Bühner (2004) der Root-Mean-Square-Error-of-Approximation (RMSEA) anzuwenden. Verzerrungen aufgrund eines kleinen Stichprobenumfangs führen demnach eher zur Ablehnung des Modells als zu einer Annahme. Laut Hu und Bentler (1999) sollte der RMSEA kleiner .08 sein. Für das Zwei-Faktoren-Modell beträgt der RMSEA-Wert .0 bei einer unteren Grenze des 90-prozentigem Konfidenzintervalls von .0 (LO90) und einer oberen von .14 (HI90). Der ermittelte Wert des RMSEA liegt somit innerhalb der von Hu und Bentler (1999) geforderten Grenzen. Allerdings ist mit einer gewissen Wahrscheinlichkeit damit zu rechnen, dass das Modell nicht passt, da die obere Grenze des Konfidenzintervalls mit .14 größer als die anvisierte von .08 ist. Der Comparative-Fit-Index (CFI) weist für das Zwei-Faktoren-Modell mit 1.0 einen optimalen Wert aus und liegt somit deutlich über den von Backhaus et al. (2003) und Hu und Bentler (1999) geforderten Grenzen von .90 beziehungsweise .95. Die im Rahmen der Maximum Likelihood Schätzung ermittelten standardisierten Regressionsgewichte der Aufgabengruppen sind in Abbildung 42 ersichtlich. Die Werte entsprechen geschätzten Korrelationen zwischen latenter und beobachteter Variable und liegen hier zwischen .41 und .85. Die quadrierten Werte kommen dem erklärten Anteil der Varianz in der beobachteten Variable durch die latente Variable gleich. Zumindest die Schätzwerte für *Wochenplan Nachmittag* und *Diagrammverständnis* sind mit 17 bzw. 23 Prozent erklärter Varianz durch die hypothetischen Konstrukte „dynamische Problemlösefähigkeit“ und „analytische Problemlösefähigkeit“ recht niedrig. Die anderen drei Aufgabengruppen werden zu 30, 37 und 72 Prozent durch die latente Variable erklärt. Die beiden Faktoren dynamische Problemlösefähigkeit und analytische Problemlösefähigkeit korrelieren im Zwei-Faktoren-Modell in Höhe von  $r = .30$ .



Anmerkung: N = 66; PL analytisch: analytische Problemlösefähigkeit; PL dynamisch: dynamische Problemlösefähigkeit.

**Abbildung 42: Pfaddiagramm des Zwei-Faktoren-Modells mit geschätzten standardisierten Regressionsgewichten (*webPostkorb*, Hauptuntersuchung)**

Insgesamt bildet das Zwei-Faktoren-Modell die empirischen Daten gut ab. Die Ergebnisse der Dimensionsanalysen bestätigen somit die anvisierte Konstruktion des *webPostkorbs* mit zwei Subskalen und der Unterscheidung zwischen den Facetten *dynamisches Problemlösen* und *analytisches Problemlösen*, wenngleich aufgrund der kleinen Stichprobe weitere Untersuchungen sinnvoll erscheinen.

### **7.6.2 Zusammenhang mit dem Personmerkmal Handlungs- und Lageorientierung**

Im Rahmen der Hauptuntersuchung soll der Zusammenhang der Leistungen in der *webPlanungsaufgabe* und im *webPostkorb* mit dem nicht-kognitiven Persönlichkeitsmerkmal „Handlungs- und Lageorientierung“ überprüft werden.

Wie in Abschnitt 5.3.2 ausgeführt, betrifft das Konstrukt „Handlungs- und Lageorientierung“ nach Kuhl (1994) die Fähigkeit, in Belastungssituationen reduzierten handlungsbahnenden Affekt herauf (HOP = Selbstmotivierung) und in bedrohlichen Situationen negativen Affekt herab zu regulieren (HOM = Selbstberuhigung). Demnach wirkt sich die Handlungs- und Lageorientierung auf die Effizienz der Affektregulation aus. Darunter ist laut Kuhl (1994) die Fähigkeit zu verstehen, vorhandene affektregulatorische Kompetenzen unter Belastung bzw. Stress einzusetzen.

Wie in Abschnitt 5.1 dargelegt, ist es anzunehmen, dass die Bearbeitung der beiden internetbasierten Verfahren von den Testpersonen als Belastungssituation wahrgenommen wird. In einer solchen Situation gehen nach Kuhl (1994) handlungsorientierte Personen effizienter mit der Affektregulation um. Es wird somit erwartet, dass der Grad der Entscheidungs- und Handlungsplanung (HOP-Skala) zumindest einen indirekten Einfluss auf Leistungen im *webPostkorb* und in der *webPlanungsaufgabe* hat.

Hinsichtlich der Zusammenhänge mit der HOM-Skala lassen sich im Vorhinein keine eindeutigen Richtungen formulieren. Für die Korrelation zwischen der Handlungsorientierung bei der Tätigkeitsausführung (HOT) und den Leistungsmaßen in den internetbasierten Verfahren lässt sich ebenfalls keine eindeutige Richtung im Vorwege bestimmen. Für die Skalen HOM und HOT sind somit a priori keine Richtungen der Korrelationen vorherzusagen.

Tabelle 39 gibt die ermittelten Zusammenhänge zwischen den HAKEMP-Skalen und den erhobenen Leistungswerten wieder. Es zeigen sich hinsichtlich der HOP-Skala und den Skalen der internetbasierten Verfahren erwartungskonforme positive Korrelationen zwischen  $r = .26$  und  $r = .42$ . Signifikante Mittelwertsunterschiede zwischen lage- und handlungsorientierten Teilnehmern konnten für den Gesamtscore *webPostkorb* und die Subskala *analytisches Problemlösen* ermittelt werden ( $T = -2.35$  bzw.  $-2.47$ ;  $df = 42$ ;  $p = .02$  bzw.  $.02$ ). Handlungsorientierte bearbeiten den *webPostkorb* insgesamt somit erwartungsgemäß erfolgreicher als Lageorientierte. Auch mit der *webPlanungsaufgabe* kommen handlungsorientierte Teilnehmer besser zurecht. Dieser Unterschied stellt



sich jedoch für die Stichprobe der Hauptuntersuchung als nicht signifikant heraus.

**Tabelle 39: Korrelationen zwischen HAKEMP-90-Skalen und erhobenen Leistungswerten**

Verfahren	Skalen	HOM	HOP	HOT
<i>web-Planungsaufgabe</i>	Summenscore Planen	.20	.26	.04
	komb. Gesamtscore Planen	.16	.30*	-.07
<i>webPostkorb</i>	<i>dynamisches Problemlösen</i>	.44**	.29*	.05
	<i>analytisches Problemlösen</i>	.02	.38*	.02
	Gesamtscore Problemlösen	.16	.42**	.05
Berliner Intelligenzstruktur-Test	Summenrohwert BIS-K-S	.13	.25	-.25

Anmerkung: N = 44.

\* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

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

Hinsichtlich der Zusammenhänge mit den anderen beiden HAKEMP-90-Skalen ergeben sich kaum signifikante Korrelationen. Lediglich der Wert zwischen der Aufgabengruppe *Wochenplan* des *webPostkorbs* und der HOM-Skala erweist sich mit  $r = .44$  als überzufällig. Zu dem gleichen Ergebnis gelangt auch der Mittelwertsvergleich zwischen lage- und handlungsorientierten Personen (HOM-Skala). So erzielen handlungsorientierte Teilnehmer ein signifikant besseres Ergebnis in dieser Aufgabengruppe ( $T = -2.50$ ;  $df = 42$ ;  $p = .02$ ). Dieser Befund lässt sich möglicherweise mit dem Sachverhalt erklären, dass sich für viele Testpersonen die Subskala *dynamisches Problemlösen* als recht anspruchsvoll gestaltet (verhältnismäßig niedrige Itemschwierigkeiten, s. o.) und somit als Misserfolgsinduktion im Sinne Kuhls (1994) verstanden werden

könnte. Demnach handeln Lageorientierte uneffektiver nach Misserfolg als Handlungsorientierte.

Laut Kuhl (1994) ergibt sich in Validierungsstudien, dass die HOP-Skala den Zusammenhang zwischen dem Grad der Intention zur Ausführung von (freiwilligen) Aktivitäten und dem Ausmaß der tatsächlichen Ausführung moderiert. Demnach sind Korrelationen zwischen Intention und Ausführung bei Handlungsorientierten erheblich größer als bei Lageorientierten. Interpretiert man die Leistungen im BIS-K-S, der *webPlanungsaufgabe* und dem *webPostkorb* als Indikator für die tatsächlichen Handlungen, so ist es interessant zu überprüfen, ob sich die Angaben von Kuhl (1994) auch bei der Stichprobe der Studierenden bestätigen lassen. Ein Vergleich der Korrelationen zwischen der HOP-Skala - unterteilt in die zwei Subgruppen der Handlungsorientierten und Lageorientierten - und den eingesetzten Leistungsverfahren ist in Tabelle 40 dargelegt. Es zeigen sich erwartungskonform durchweg höhere Korrelationen bei der Gruppe der Handlungsorientierten. Ob diese Korrelationen allerdings wie Kuhl (1994) in der Handanweisung formuliert „erheblich größer“ (S. 2) sind, lässt sich nicht abschließend beurteilen.

**Tabelle 40: Korrelationen zwischen der HOP-Skala und den eingesetzten Leistungsskalen**

Verfahren	Skala	HOP-Skala	
		Lageorientierte (N = 19)	Handlungsorientierte (N = 25)
<i>web-Planungsaufgabe</i>	Summenscore Planen	-.04	.44*
	komb. Gesamtscore Planen	.18	.50*
<i>webPostkorb</i>	<i>dynamisches Problemlösen</i>	.14	.25
	<i>analytisches Problemlösen</i>	.14	.20
	Gesamtscore Problemlösen	.17	.26
Berliner Intelligenzstruktur-Test	Summenrohwert BIS-K-S	-.09	.26

Es zeigt sich somit insgesamt, dass die Planungsleistung in der *webPlanungsaufgabe*, die Problemlöseleistung im *webPostkorb* und das Personmerkmal „Handlungs- und Lageorientierung“ größtenteils erwartungsgemäß zusammenhängen. Insbesondere der Grad der Entscheidungs- und Handlungsplanung (HOP-Skala) scheint die Planungs- und Problemlöseleistung zu beeinflussen. Aufgrund der kleinen Stichprobengröße sollten diese Ergebnisse jedoch in weiteren Studien überprüft werden.

### **7.6.3 Einsatz im Rahmen von Personalauswahlverfahren**

Die Leistungen innerhalb der Wirtschaftsversion der *webPlanungsaufgabe* wurden - wie schon erwähnt - im Rahmen des Personalauswahlprozesses eines Konzerns erhoben. Die Zusammenhänge zwischen den erzielten Ergebnissen in der *webPlanungsaufgabe* und den anderen eingesetzten Auswahlverfahren waren dabei Inhalt der Darstellung zur Kriteriumsvalidität (s. o.). Neben dem E-Assessment, das aus drei kognitiven Leistungsverfahren besteht, absolvieren die Bewerber auch ein strukturiertes Telefoninterview. Darin werden Persönlichkeitsmerkmale aus dem sozialen, motivationalen und biografischen Bereich erhoben. Die Ergebnisse der Faktorenanalyse, in die die Leistungen in den Personalauswahlverfahren E-Assessment (drei Leistungsverfahren), strukturiertes Telefoninterview (soziale und motivationale Merkmale) und *webPlanungsaufgabe* (kombinierter Gesamtscore) einbezogen wurden, sind in Tabelle 41 dargestellt.

**Tabelle 41: Faktorenanalyse der berufsrelevanten Personmerkmale eines Personalauswahlprozesses**

Skala	Komponenten (Faktoren)		
	soziale, motivationale Merkmale	kognitive Leistungsfähigkeit	Planungsfähigkeit
motivationale Merkmale Telefoninterview	.86	.11	
soziale Merkmale Telefoninterview	.87		
numerische Aufgaben E-Assessment		.61	
figural-bildhafte Aufgaben E-Assessment		.68	.32
verbale Aufgaben E-Assessment		.70	-.24
kombinierter Score <i>webPlanungsaufgabe</i>			.94
Varianzaufklärung (in %)	29.1	19.1	16.8

Anmerkung: N = 178; Extraktionsmethode: Hauptkomponentenanalyse; Rotationsmethode: Varimax mit Kaiser-Normalisierung; Ladungen unter .1 sind nicht dargestellt; Die Rotation ist in 4 Iterationen konvergiert.

Für die Stichprobe der 178 Personen, die sowohl E-Assessment, Telefoninterview und *webPlanungsaufgabe* durchliefen, ergibt sich eine Drei-Faktoren-Struktur (Hauptkomponentenanalyse mit Varimax-Rotation und Eigenwert > 1). Den ersten Faktor erklären die im Telefoninterview erhobenen sozialen und motivationalen Merkmale. Der zweite Faktor wird durch die drei Leistungsverfahren des E-Assessments beschrieben, während der kombinierte Score der *webPlanungsaufgabe* auf dem dritten Faktor hoch lädt. Nebenladungen ergeben sich auf dem dritten Faktor in Höhe von .32 für die figural-bildhaften und -.24 für die verbalen Aufgaben des E-Assessments.

Insgesamt werden 29.1 Prozent der Gesamtvarianz durch den ersten Faktor, 19.1 Prozent durch den zweiten und immerhin weitere 16.8 Prozent durch den dritten Faktor erklärt. Es zeigt sich somit, dass die *webPlanungsaufgabe* trotz der substantiellen Korrelationen mit den E-Assessment-Verfahren (s. Tabelle 35) einen relativ eigenständigen Bereich im Vergleich zu den eingesetzten Personalauswahlverfahren anspricht.

#### **7.6.4 Zusammenhang mit kognitiven Leistungsverfahren**

Der Zusammenhang zwischen Intelligenztestleistung und Problemlöseleistungen wurde - wie unter Abschnitt 2.2.5 vorgestellt - in zahlreichen Studien bestätigt. In dieser Arbeit konnten im Rahmen der Voruntersuchung erste Korrelationen zwischen der *webPlanungsaufgabe*, der Testvorform des *webPostkorbs* und den Testleistungen in Aufgaben des I-S-T 70 (Amthauer, 1970) ermittelt werden. Es zeigten sich dabei Zusammenhänge zwischen dem kombinierten Gesamtscore der *webPlanungsaufgabe* und der Intelligenztestleistung in Höhe von  $r = .31$  sowie zwischen den Leistungen im *webPostkorb* und den eingesetzten I-S-T 70 Aufgaben in Höhe  $r = .47$  (s. o.). Diese Ergebnisse sollen im Rahmen der Hauptuntersuchung durch eine Replikationsstudie überprüft werden.

Die ermittelten Zusammenhänge zwischen den Rohwerten der Verfahren innerhalb der Hauptuntersuchung sind in Tabelle 42 aufgelistet. Die drei Aufgabengruppen innerhalb der Skala *analytisches Problemlösen* korrelieren jeweils substantiell zwischen  $r = .34$  und  $.36$  mit dem BIS-K-S Wert (nicht in der Tabelle dargestellt). Die Subskala *analytisches Problemlösen* hängt erwartungsgemäß mit  $r = .48$  mit der Verarbeitungskapazität zusammen. Für die Aufgabengruppe *dynamisches Problemlösen* können keine signifikanten Zusammenhänge mit dem BIS-K-S gefunden werden, die Korrelation liegt bei  $.20$ . Dieser Zusammenhang ist somit etwas geringer als erwartet, lässt sich allerdings wahrscheinlich durch die hohen Schwierigkeiten der Aufgaben innerhalb der Subskala erklären. Der Gesamtscore *webPostkorb* korreliert

substantiell und erwartungsgemäß mit der Verarbeitungskapazität und weist mit  $r = .50$  den höchsten Wert aus.

**Tabelle 42: Korrelationen der Simulationsverfahren mit Intelligenztestleistungen (Hauptuntersuchung)**

Verfahren	Skala	N	BIS-K-S	I-S-T	Gesamt- score <i>web- Postkorb</i>
				2000 R	
<i>webPostkorb</i>	<i>analytisches Problemlösen</i>	64	.48**		
	<i>dynamisches Problemlösen</i>		.20		
	Gesamtscore Problemlösen		.50**		
<i>webPlanungsaufgabe Studierende</i>	Summenscore	45	.32*		.45**
	benötigte Zeit		-.12		.02
	komb. Gesamtscore		.35*		.32**
<i>webPlanungsaufgabe Führungsnachwuchs</i>	Summenscore	41		.28	
	benötigte Zeit			-.22	
	komb. Gesamtscore			.33*	

\* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

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

Die Korrelationen zwischen der Verarbeitungskapazität - gemessen durch den BIS-K-S Rohwert - und der *webPlanungsaufgabe* sind ebenfalls in Tabelle 42 dargestellt. Der kombinierte, z-transformierte Punkt- und Zeitwert hängt mit dem Rohwert BIS-K-S in Höhe von .35 zusammen, während die Korrelationswerte für den Summenscore und für die benötigte Zeit bei  $r = .32$  bzw.  $r = -.12$  liegen.

Die Nähe der Planungs- und Problemlöseleistungen, die durch die *webPlanungsaufgabe* bzw. den *webPostkorb* erfasst werden, zeigt sich ebenfalls in der Tabelle 42. Der Summenscore für die *webPlanungsaufgabe* hängt substanziell und erwartungsgemäß in Höhe von  $r = .45$  mit der Gesamtleistung im *webPostkorb* zusammen. Diese Korrelation bleibt auch bei Kontrolle der Verarbeitungskapazität signifikant bei einem Wert von  $r = .36$ .

In einer weiteren Studie konnte der Zusammenhang zwischen den Leistungen in der Wirtschaftsversion der *webPlanungsaufgabe* und der Intelligenztestleistung überprüft werden. 41 Bewerber auf Führungsnachwuchsstellen absolvierten im Rahmen des Assessment-Center-Auswahltags auch den I-S-T 2000 R von Amthauer et al. (2001). Die Korrelationen liegen bei  $r = .28$  und  $r = -.22$  für den Summenscore und für die benötigte Zeit. Der kombinierte Gesamtscore der *webPlanungsaufgabe* hängt signifikant in Höhe von  $r = .33$  mit dem I-S-T 2000 R Grundmodul zusammen. Dieser Befund ist interessant, da die 41 Bewerber bereits erfolgreich durch das E-Assessment und durch das Telefoninterview gekommen waren, der Zusammenhang somit auch innerhalb der stark selektierten Stichprobe der „Erfolgreichen“ in mittelstarker Höhe auftritt.

Insgesamt können die Ergebnisse der Voruntersuchung hinsichtlich des Zusammenhangs zwischen den Leistungen im *webPostkorb*, in der *webPlanungsaufgabe* und der Intelligenztestleistung repliziert werden. Dieser Befund ist insofern bedeutend, als dass ein möglicher Minderungseffekt der Reliabilitäten auf die Koeffizienten hier gar nicht berücksichtigt wurde. Es kann somit davon ausgegangen werden, dass erwartungsgemäße, substanzielle Zusammenhänge in mittelstarker Höhe zwischen Leistungen in der *webPlanungsaufgabe* und Intelligenztests sowie Leistungen im *webPostkorb* und Intelligenztests vorliegen.

Es stellt sich dabei in Bezug auf die Konstruktvalidität der Verfahren die Frage, durch welche Faktoren die Leistungen in der *webPlanungsaufgabe*

und im *webPostkorb* bestimmt werden. Insbesondere gilt es zu überprüfen, inwieweit die allgemeine Intelligenz als genereller Faktor die Leistungen im Wesentlichen determiniert oder ob etwa verschiedene Teilfaktoren, die die Problemlösefähigkeit repräsentieren, die Leistungen entscheidend beeinflussen.

Eine systematische Antwort lässt sich mit Hilfe von Strukturgleichungsmodellen geben. Wenngleich die Voraussetzung der multivariaten Normalverteilung hier innerhalb der von Bühner (2003) geforderten Grenzen liegt (Exzess = -2.10; „c.r.“-Wert = -.50), so sind die folgenden Analysen aufgrund der kleinen Stichprobengröße wiederum nicht als hypothesenprüfend zu verstehen.<sup>24</sup> Vielmehr sollen die hier erhobenen Daten mit den Befunden der PISA-Studie zur fächerübergreifenden Kompetenz Problemlösen (Klieme et al., 2001) verglichen werden, um weitere Hinweise hinsichtlich der Konstruktvalidität zu erhalten. Klieme et al. (2001) konnten in ihrer Untersuchung Zusammenhangsmuster verschiedener kognitiver Kompetenzen ermitteln und einzelne Faktoren bestimmen.

Für die studentische Stichprobe der Hauptuntersuchung ließen sich im vorangegangenen Abschnitt der Dimensionsanalyse bereits zwei Teilfaktoren der Problemlösekompetenz extrahieren. Darüber hinaus wurde im Rahmen der Hauptuntersuchung auch die Verarbeitungskapazität mittels BIS-K-S (Jäger et al., 1997) erhoben, so dass neben fünf Problemlöseindikatoren in Form der Aufgabengruppen des *webPostkorbs* auch die Rohwerte der BIS-K-S Aufgaben in die Analyse eingehen können.

Tabelle 43 gibt die Ergebnisse der konfirmatorischen Faktorenanalysen wieder. Es werden - analog zum Vorgehen der PISA-Studie zum Problemlösen als fächerübergreifender Kompetenz (Klieme et al., 2001) - drei Modelle geprüft, die sich aus der Theorie ableiten lassen. Dazu zählen ein Generalfaktor-Modell, ein Zwei-Faktoren-Modell mit der Unterscheidung zwischen kognitiver Grundfähigkeit (in Form der Verarbeitungskapazität) und

---

<sup>24</sup> Die genauen Voraussetzungen zur Berechnung von konfirmatorischen Faktorenanalysen wurden bereits an anderer Stelle diskutiert (s. Abschnitt 7.6.1).



einem allgemeinen Problemlösefaktor und ein Drei-Faktoren-Modell mit den Komponenten kognitive Grundfähigkeit, dynamische Problemlösefähigkeit und analytische Problemlösefähigkeit. Die beste Passung an die empirischen Daten zeigt das korrelierte Drei-Faktoren-Modell. Der  $\chi^2$ -Hypothesentest ergibt ein nicht signifikantes p von .26 ( $\chi^2 = 46.34$  bei 41 Freiheitsgraden), so dass von einer hinreichenden Modellpassung ausgegangen werden kann. Der Goodness-of-Fit-Index (GFI) und der Comparative-Fit-Index (CFI) beträgt für das Drei-Faktoren-Modell .90 beziehungsweise .95 und erweist sich damit als gut (vgl. Backhaus et al., 2003). Lediglich der Adjusted Goodness-of-Fit-Index (AGFI) liegt mit .85 etwas unter der Grenze von .90.

**Tabelle 43: Kennwerte für die Anpassungsgüte verschiedener faktorenanalytischer Modelle (Hauptuntersuchung)**

Modell	$\chi^2$	df	P	GFI	AGFI	CFI	RMSEA	LO90	HI90
Generalfaktor-Modell	72.2	46	.01	.86	.79	.76	.09	.05	.13
Zwei-Faktoren-Modell*	54.2	43	.12	.88	.83	.90	.06	.00	.11
Drei-Faktoren-Modell**	46.3	41	.26	.90	.85	.95	.05	.00	.10

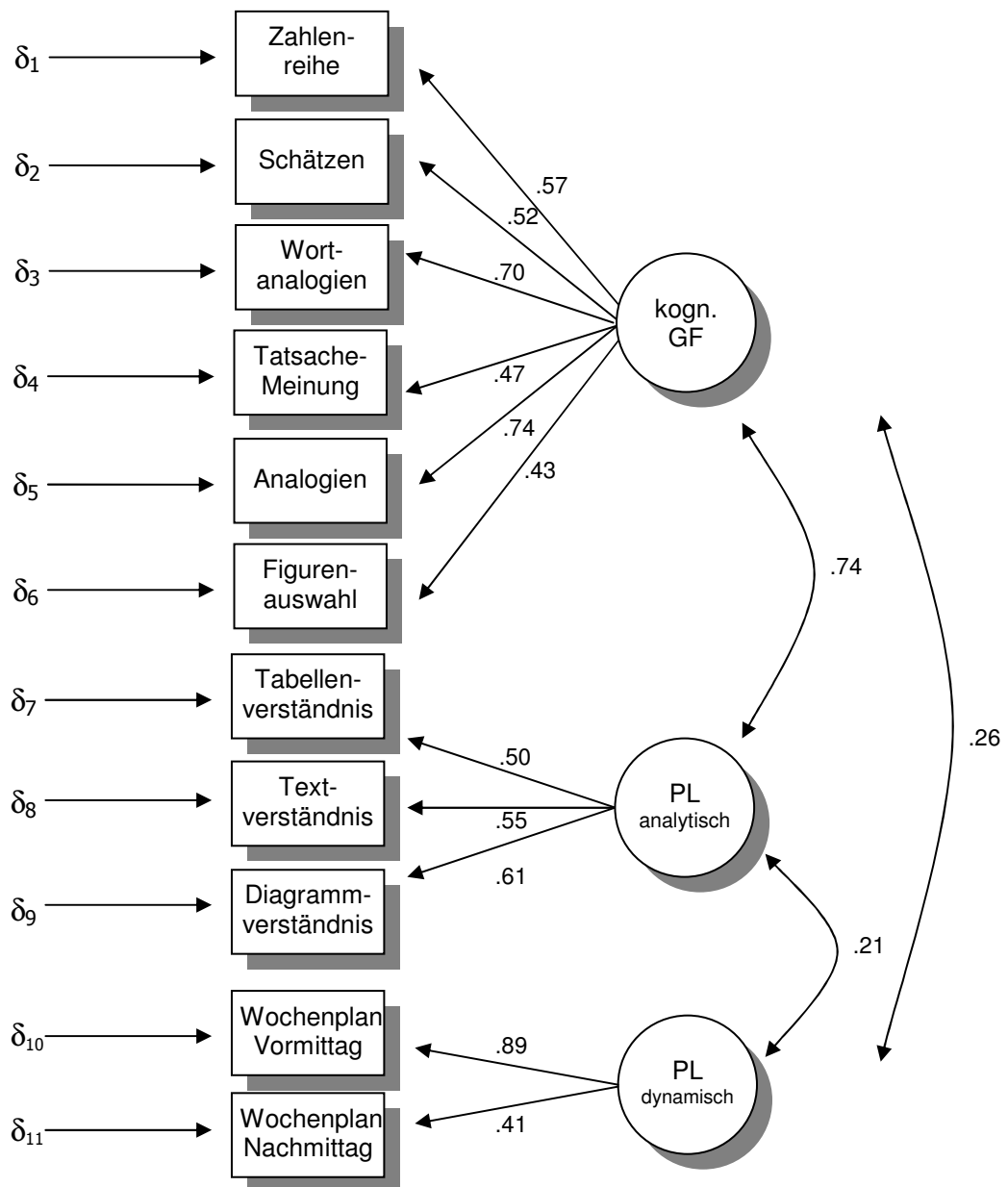
Anmerkung: N = 64; df: Freiheitsgrade; p: Irrtumswahrscheinlichkeit; GFI: Goodness-of-Fit-Index; AGFI: Adjusted Goodness-of-Fit-Index; CFI: Comparative-Fit-Index; RMSEA: Root-Mean-Square-Error-of-Approximation; LO90, HI90: unteres und oberes 90-prozentiges Konfidenzintervall des RMSEA.

\* Zwei-Faktoren-Modell mit kognitiver Grundfähigkeit und Problemlösefähigkeit (korreliert).

\*\*Drei-Faktoren-Modell mit kognitiver Grundfähigkeit, analytischer und dynamischer Problemlösefähigkeit (korreliert).

Laut Bühner (2004) ist gerade bei kleinen Stichproben der Root-Mean-Square-Error-of-Approximation (RMSEA) anzuwenden. Für das Drei-Faktoren-Modell beträgt der RMSEA-Wert .05 bei einer unteren Grenze des 90-prozentigem Konfidenzintervalls von .00 (LO90) und einer oberen von .10 (HI90). Der ermittelte RMSEA-Wert liegt somit mit hoher Wahrscheinlichkeit

innerhalb der von Hu und Bentler (1999) geforderten Grenzen von kleiner .08. Für das Zwei-Faktoren-Modell ist der Punktwert RMSEA mit .06 ebenfalls niedriger als die Schwelle von .08. Auch der Comparative-Fit-Index (CFI) weist für das Zwei-Faktoren-Modell mit .90 einen akzeptablen Wert aus.



Anmerkung: N = 64; kogn. GF: kognitive Grundfähigkeit; PL analytisch: analytische Problemlösefähigkeit; PL dynamisch: dynamische Problemlösefähigkeit.

**Abbildung 43: Pfaddiagramm des Drei-Faktoren-Modells mit geschätzten standardisierten Regressionsgewichten (Hauptuntersuchung)**

Die im Rahmen der Maximum Likelihood Schätzung ermittelten standardisierten Regressionsgewichte des Drei-Faktoren-Modells sind in Abbildung 43 ersichtlich. Die Werte rangieren zwischen .41 und .89. Die quadrierten Werte kommen dem erklärten Varianzanteil gleich und liegen hier zwischen 17 Prozent erklärter Varianz der Aufgabengruppe *Wochenplan Nachmittag* durch die hypothetische Variable dynamische Problemlösefähigkeit und 79 Prozent erklärter Varianz der Aufgabengruppe *Wochenplan Vormittag*.

Die Faktoreninterkorrelationen betragen in dem geschätzten Drei-Faktoren-Modell  $r = .74$  zwischen der kognitiven Grundfähigkeit und der analytischen Problemlösefähigkeit,  $r = .26$  zwischen der kognitiven Grundfähigkeit und der dynamischen Problemlösefähigkeit und  $r = .21$  zwischen den beiden Problemlösefähigkeitsfaktoren.

Insgesamt werden die empirischen Daten somit am besten durch das Drei-Faktoren-Modell mit zwei Problemlösefähigkeitsfaktoren abgebildet. Dieser Befund wird durch die Ergebnisse der PISA-Studie von Klieme et al. (2001) bestätigt. Auch bei Klieme et al. (2001) zeigt das Modell die beste Passung, das die Problemlösefähigkeit nicht als einen homogenen Faktor, sondern als ein aus verschiedenen Teilfaktoren zusammengesetztes Konstrukt abbildet. Innerhalb dieses siebenfaktoriellen PISA-Modells lassen sich ein Faktor der kognitiven Grundfähigkeit und drei getrennte Problemlösefähigkeitsfaktoren unterscheiden. Darüber hinaus können noch die PISA-Konstrukte „Lesekompetenz“, „mathematische Kompetenz“ und „naturwissenschaftliche Kompetenz“ identifiziert werden, die im Rahmen dieser Arbeit nicht erhoben wurden. Die drei Problemlösefähigkeitsfaktoren sind laut Klieme et al. (2001) zu verstehen als „Problemlösekompetenz in schriftlichen, analogen Verfahren“, als „Strategien beim Explorieren und Steuern eines komplexen Systems“ und als „Wissenserwerb in computergestützten Systemen“. Dabei ist der Faktor der „Problemlösekompetenz in schriftlichen, analogen Verfahren“ des PISA-Modells vergleichbar mit der hier bestimmten analytischen Problemlösefähigkeit, während dynamische Problemlösefähigkeit dem PISA-Konstrukt „Strategien beim Explorieren und Steuern eines komplexen Systems“ nahe kommt. Auch der Faktor der

kognitiven Grundfähigkeit kann sowohl in der PISA-Studie wie auch im Rahmen dieser Arbeit identifiziert werden. Lediglich der Problemlöseteilkfaktor des „Wissenserwerbs in computergestützten Systemen“ findet in der vorliegenden Untersuchung kein Pendant, da ein solches Konstrukt hier nicht angedacht war und keine Daten erhoben wurden, die dieses Konstrukt erfassen könnten.

Auch die Ergebnisse der konfirmatorischen Faktorenanalyse hinsichtlich der Zusammenhänge zwischen den Teilk Faktoren sind vergleichbar mit den Befunden der PISA-Studie. Während das Problemlösen als „strategisches Vorgehen in den Computerverfahren“ in der Untersuchung von Klieme et al. (2001) relativ niedrig mit den beiden anderen Teilkompetenzen und der kognitiven Grundfähigkeit korreliert, zeigt die Problemlösekompetenz in schriftlichen, analogen Verfahren eine große Nähe zu der kognitiven Grundfähigkeit. Dies entspricht dem hier ermittelten Befund, dass der Faktor kognitive Grundfähigkeit deutlich höher mit der analytischen Problemlösefähigkeit zusammenhängt als mit der dynamischen Problemlösefähigkeit und die beiden Problemlösefähigkeitsfaktoren untereinander ebenfalls niedrig korrelieren. Es wird somit deutlich, dass Problemlösefähigkeit als ein Konstrukt mit mehreren Teilk Faktoren, die jeweils eine unterschiedliche Nähe zur kognitiven Grundfähigkeit aufweisen, zu verstehen ist.

Zusammenfassend lässt sich feststellen, dass sich die hier ermittelten Zusammenhänge zwischen den Leistungen im *webPostkorb*, den Leistungen in der *webPlanungsaufgabe* und Intelligenztestleistungen in der theoretisch erwartbaren Größenordnung bewegen (vgl. M. Kersting, 1999; Klieme et al., 2001; Süß, 1996).

Dabei wird aus den Analysen deutlich, dass die erzielten Werte in den Verfahren nicht alleine durch die allgemeine Intelligenz bzw. die kognitive Grundfähigkeit erklärt werden können, sondern den Leistungen in der *webPlanungsaufgabe* und im *webPostkorb* jeweils eine eigenständige Fähigkeit zu Grunde liegt, die als „Planungsfähigkeit“ bzw.

„Problemlösefähigkeit“ mit den Facetten „analytische Problemlösefähigkeit“ und „dynamische Problemlösefähigkeit“ interpretierbar ist.

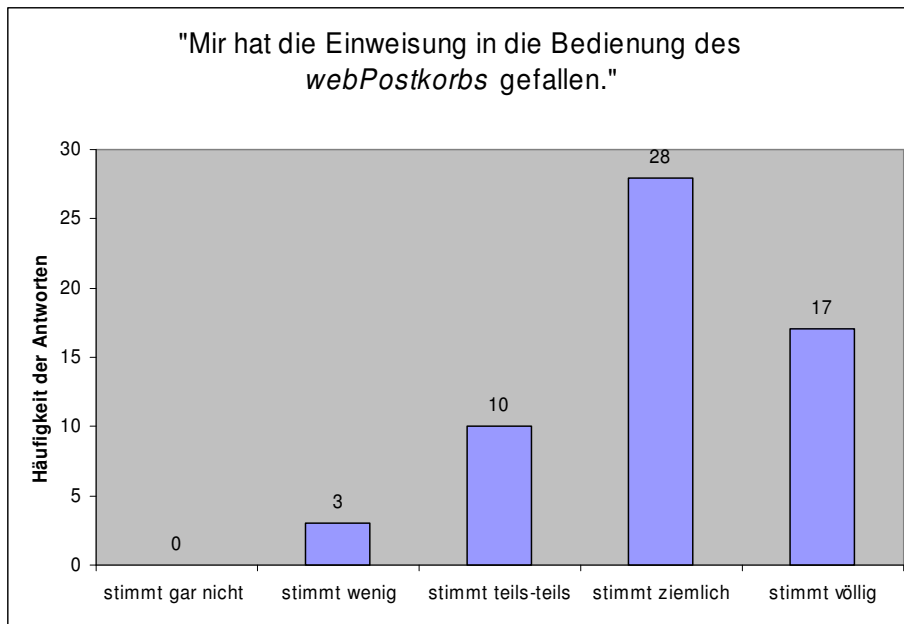
Insgesamt können die hier berichteten Befunde somit als Belege für die Konstruktvalidität der beiden internetbasierten Simulationsverfahren *webPostkorb* und *webPlanungsaufgabe* verstanden werden.

## 7.7 Akzeptanz und Fairness

Analog zur Vorgehensweise innerhalb der Voruntersuchung soll auch im Rahmen der Hauptuntersuchung überprüft werden, inwieweit sich die hier anvisierte Vorgehensweise der benutzerorientierten Verfahrensgestaltung und Verfahrensdurchführung in der Einschätzung der Probanden widerspiegelt. Daher wurden die Testpersonen im Anschluss an die Bearbeitung des *webPostkorbs* und der *webPlanungsaufgabe* aufgefordert, das jeweilige Verfahren aus Benutzersicht zu bewerten.

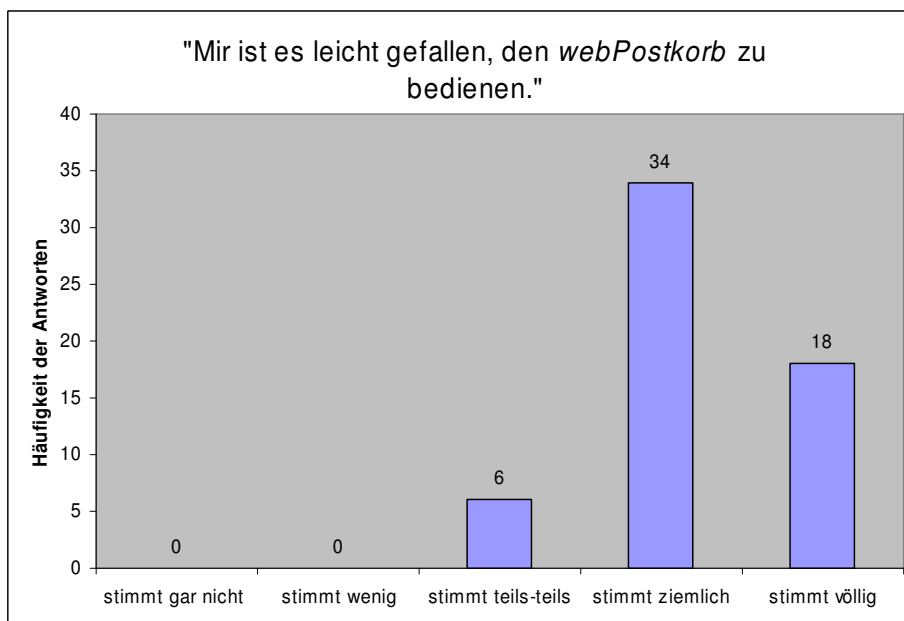
Im Mittelpunkt der Betrachtung stehen dabei Aspekte der Akzeptanz und der Fairness. Darüber hinaus sind auch Hinweise auf Verständlichkeit, Zumutbarkeit, Zufriedenheit und die subjektiv wahrgenommene Schwierigkeit sowie ein Vergleich der computergestützten Verfahren mit dem eingesetzten Intelligenztest hinsichtlich einiger relevanter Aspekte von Interesse. Nicht alle Teilnehmer haben sämtliche Fragen beantwortet. Daher kommt es zu Abweichungen von der Gesamtstichprobengröße der Hauptuntersuchung.

Insgesamt können die positiven Ergebnisse der Voruntersuchung bestätigt werden. Die Befunde sind als Beleg für die weitestgehende Fairness und Akzeptanz der *webPlanungsaufgabe* und des *webPostkorbs* sowie der Zufriedenheit der Teilnehmer mit diesen Verfahren zu verstehen. Da die Teilnehmerbewertungen der *webPlanungsaufgabe* bereits im Rahmen der Voruntersuchungsergebnisse im Mittelpunkt standen, sollen hier die Befunde zum *webPostkorb* näher betrachtet werden. Vor der eigentlichen Testbearbeitung erhalten die Testteilnehmer eine ausführliche Erläuterung der Bedienung und der Oberflächengestaltung.



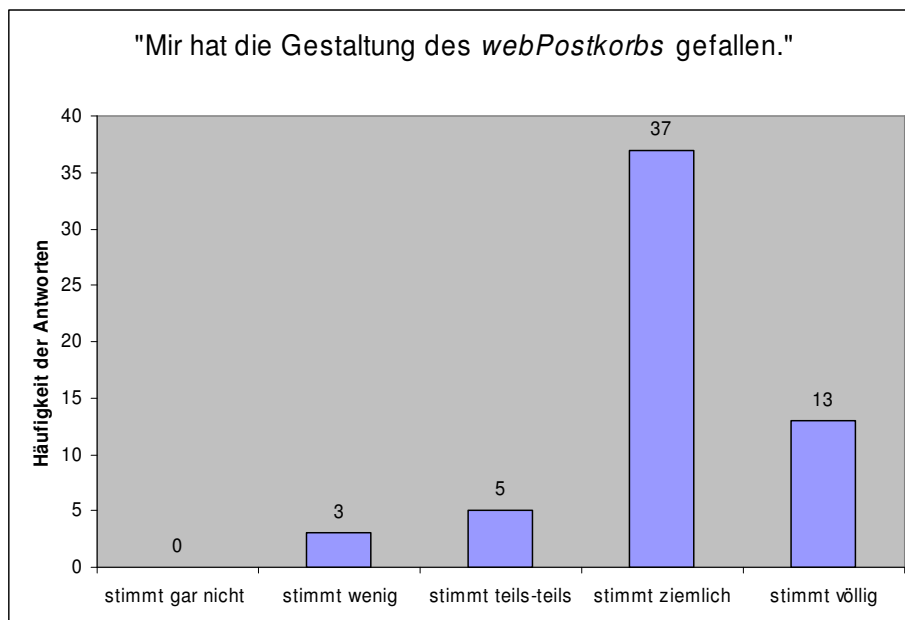
**Abbildung 44: Bewertung der Bedienungseinweisung des *webPostkorbs* (Hauptuntersuchung)**

Daher ist es interessant zu erfahren, wie die Teilnehmer die Einweisung bewerten. Die Aussage „*Mir hat die Einweisung in die Bedienung des webPostkorbs gefallen*“ wird von 45 der 58 Personen mit „*stimmt ziemlich*“ und „*stimmt völlig*“ beantwortet (Abb. 44).



**Abbildung 45: Bewertung der Bedienungsfreundlichkeit des *webPostkorbs* (Hauptuntersuchung)**

Ähnlich positiv bewerten die Teilnehmer auch die Bedienung des *webPostkorbs* (Abb. 45). Das Statement „*Mir ist es leicht gefallen, den webPostkorb zu bedienen*“ beantworten annähernd 90 Prozent der Befragten mit „*stimmt ziemlich*“ oder „*stimmt völlig*“ (52 von 58 Personen). Dieses Ergebnis ist insofern bemerkenswert, als dass der *webPostkorb* mit einem hohen Grad an Interaktivität ausgestattet ist. Es kann daher vermutet werden, dass die Bedienung höhere Anforderungen an die Probanden stellt als herkömmliche psychologische Testverfahren. Trotzdem bewerten die Testpersonen die Handhabung des *webPostkorbs* als leicht. Dies ist als Hinweis auf eine intuitive und leicht verständliche Bedienung zu werten.

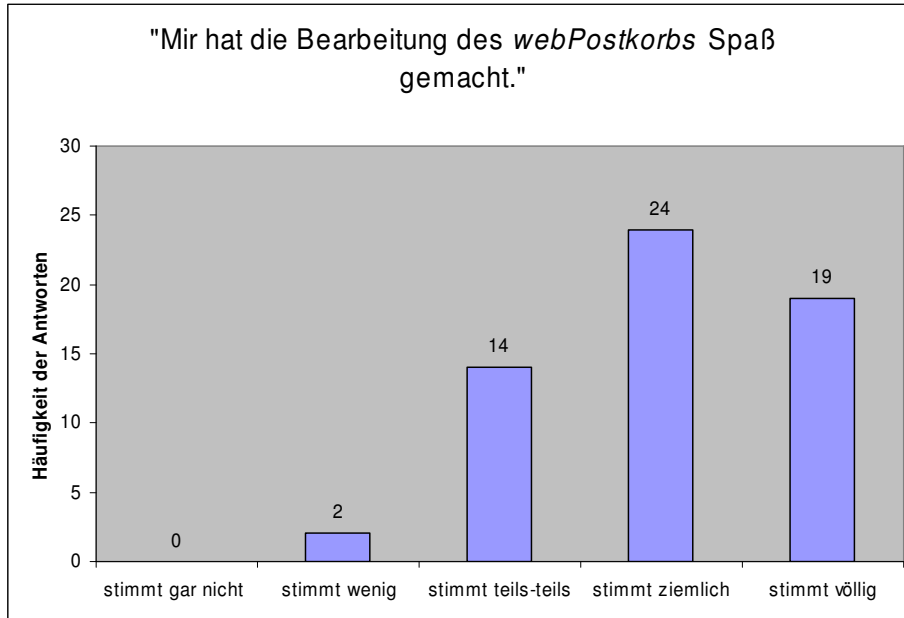


**Abbildung 46: Bewertung der Gestaltung des *webPostkorbs* (Hauptuntersuchung)**

Hinsichtlich der Aussage „*Mir hat die Gestaltung des webPostkorbs gefallen*“ stimmt ebenfalls die deutliche Mehrheit (50 von 58 Befragten) ziemlich oder völlig zu (Abb. 46).

Darüber hinaus sollte das subjektive Erleben der Benutzer während der Aufgabenbearbeitung befragt werden. Trotz der kognitiven Anforderungen, die die Bearbeitung des *webPostkorbs* an die Probanden stellt, beantworten ca. drei Viertel der Teilnehmer die Aussage „*Mir hat die Bearbeitung der Aufgaben Spaß gemacht*“ mit „*stimmt ziemlich*“ bzw. „*stimmt völlig*“.

Vierzehn Personen stimmen diesem Statement teils-teils zu und lediglich zwei können der Aussage wenig zustimmen (Abb. 47).

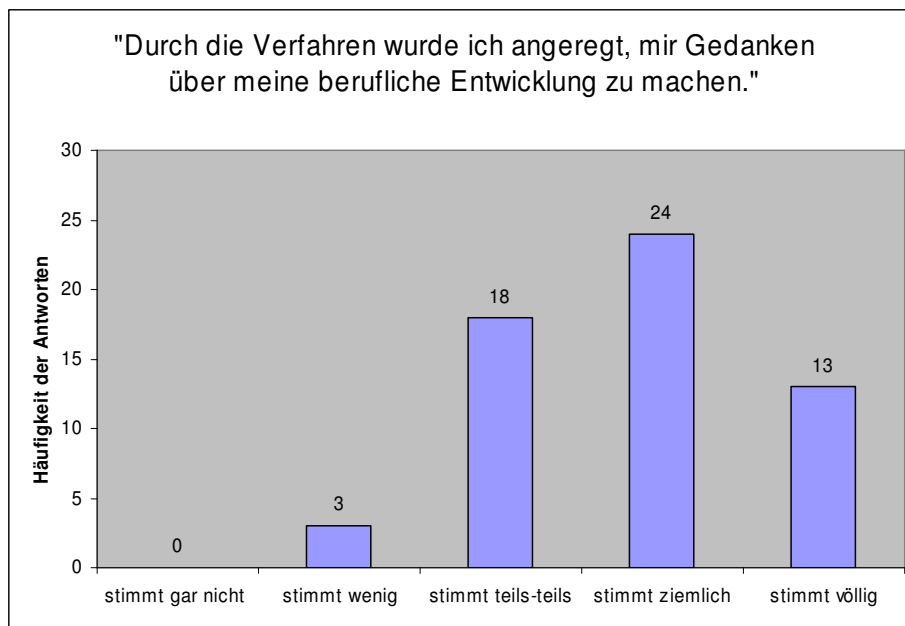


**Abbildung 47: Bewertung der Zufriedenheit bei der Bearbeitung des webPostkorbs (Hauptuntersuchung)**

Hinsichtlich des subjektiven Nutzens wurden die Teilnehmer gefragt, inwieweit die Simulationsverfahren dazu angeregt haben, sich Gedanken über die persönliche berufliche Entwicklung zu machen. Fast zwei Drittel der Befragten stimmen der Aussage „Durch die Verfahren wurde ich angeregt, mir Gedanken über meine berufliche Entwicklung zu machen“ ziemlich oder völlig zu (Abb. 48).

Nachdem die Teilnehmer Rückmeldungen über ihre Leistungen in den verschiedenen Verfahren (Intelligenztest, HAKEMP-90 und webPostkorb) erhalten haben, werden abschließend noch einige vergleichende Fragen gestellt. Dabei sollen die Teilnehmer einschätzen, welches der beiden Leistungsverfahren sie hinsichtlich bestimmter Aspekte für geeigneter halten oder ob die eingesetzte Kurzversion des Berliner Intelligenzstruktur-Tests von Jäger et al. (1997) mit dem webPostkorb vergleichbar sei.



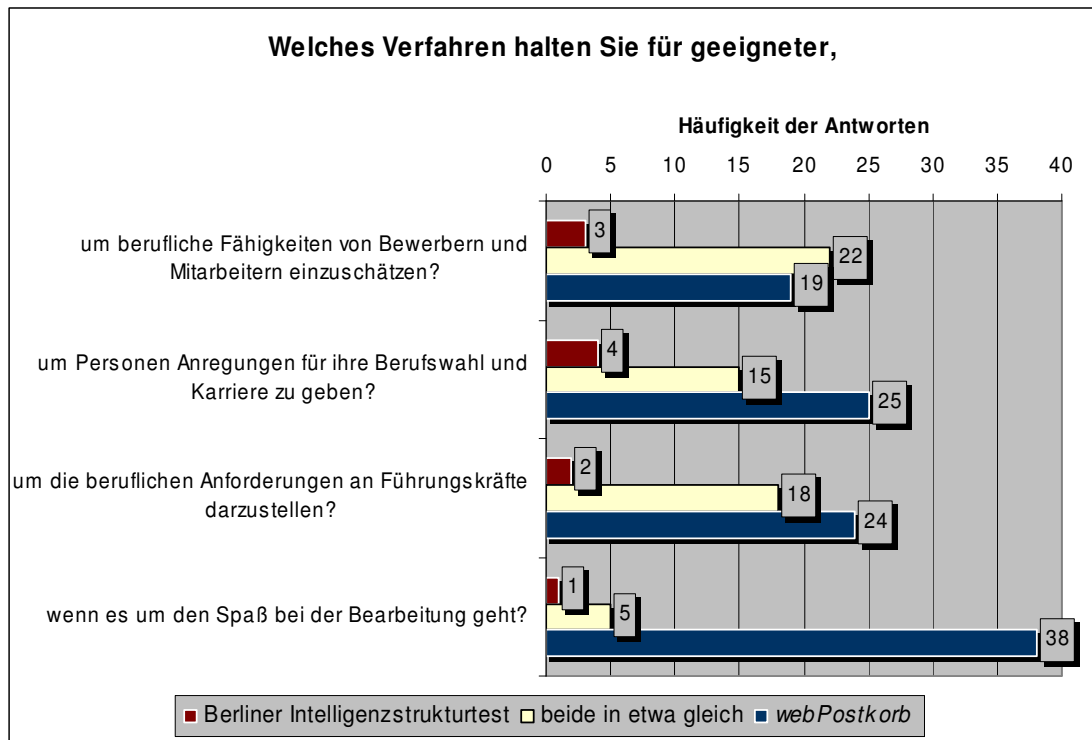


**Abbildung 48: Bewertung der Anregung zur Auseinandersetzung mit der beruflichen Entwicklung (Hauptuntersuchung)**

Es zeigt sich, dass der *webPostkorb* von den Personen bei allen erfassten Aspekten besser bewertet wird als der Intelligenztest BIS-K-S (s. Abbildung 49). So geben hinsichtlich der Einschätzung, welches Verfahren geeigneter wäre, um berufliche Fähigkeiten von Bewerbern und Mitarbeitern einzuschätzen, genau die Hälfte der 44 Personen an, dass beide Verfahrenstypen in etwa vergleichbar sind, während drei Personen den BIS-Test und 19 Personen den *webPostkorb* für geeigneter halten. Die Einschätzungen hinsichtlich der Anregung für die Berufswahl und Karriere und hinsichtlich der Verdeutlichung der beruflichen Anforderungen von Führungskräften offenbaren ein ähnliches Bild. Noch deutlicher zeigt sich die positive Bewertung des *webPostkorbs* hinsichtlich des Aspekts des Spaßes bei der Bearbeitung. So halten 85 Prozent der Befragten den *webPostkorb* für geeigneter, „wenn es um den Spaß bei der Bearbeitung geht“.

Insgesamt kann die Einschätzung einiger Autoren (u. a. U. Funke, 1995a; Höft & U. Funke, 2001), die die überlegene Akzeptanz simulativer Verfahren gegenüber herkömmlichen Intelligenztests hervorheben, bestätigt werden. Anders als in der Studie von M. Kersting (1998) wird das hier eingesetzte, simulative Verfahren *webPostkorb* durchgehend besser hinsichtlich

verschiedener Akzeptanzaspekte eingeschätzt als der Intelligenztest BIS-K-S. M. Kersting (1998) hat in einer Studie aufgezeigt, dass computergestützte Szenarios nicht grundsätzlich höhere Akzeptanzeinstufungen nach sich ziehen als Intelligenztestverfahren.



**Abbildung 49: Vergleich der beiden Leistungsverfahren BIS-K-S und *webPostkorb* hinsichtlich verschiedener Akzeptanzaspekte (Hauptuntersuchung)**

Wenngleich die Stichprobe hier recht klein ist, so besitzt dieser Befund möglicherweise Implikationen für die Praxis. Werden beim Einsatz von Self- oder E-Assessments auch Personalmarketing-Ziele verfolgt, so ist die Benutzereinschätzung der eingesetzten Instrumente von höchster Bedeutung (vgl. Abschnitt 3.2). Verfahren, die positiv von den Probanden bewertet werden - wie hier der *webPostkorb* - unterstützen angestrebte Personalmarketing-Botschaften. Die Benutzerfreundlichkeit der Verfahren ist im Rahmen der Hauptuntersuchung auch mit Hilfe eines zählpixelbasierten Trackingverfahrens sowie der Analyse der Protokolldateien (Logfiles) überprüft worden. Die Ergebnisse zeigen, dass es keine Abbrecher während der Verfahrensbearbeitung zu verzeichnen gibt. Dies kann als Hinweis auf

eine zumutbare Verfahrensdurchführung gewertet werden. Das Verfahren *webPostkorb* ist mit einigen Freiheitsgraden konstruiert, so dass es zu überprüfen gilt, inwieweit die Teilnehmer selbständig die verschiedenen Aufgaben bearbeitet haben. Die Analyse der Benutzerverläufe veranschaulicht, dass alle Aufgaben des *webPostkorbs* von den Probanden angesteuert wurden. Das positive Ergebnis der Voruntersuchung kann somit innerhalb der Hauptuntersuchung wiederholt werden. Diese Befunde sprechen für die anvisierte intuitive und leicht verständliche Benutzeroberfläche.

Auch die Testökonomie der Verfahren *webPostkorb* und *webPlanungsaufgabe* kann als positiv eingeschätzt werden. Mit einer durchschnittlichen Netto-Bearbeitungsdauer (reine Testzeit ohne Einweisung, Übungsbeispiele etc.) von unter einer halben Stunde und einer mittleren Brutto-Bearbeitungsdauer von ca. 40 bis 45 Minuten für beide Instrumente liegt der Gesamtzeitaufwand unter dem der meisten anderen simulationsbasierten Leistungsverfahren.

#### *Zusammenhang zwischen den erzielten Leistungen und der Bewertung der Verfahren*

Die Ergebnisse der Voruntersuchung hinsichtlich der Frage, inwieweit Personen, die in den Verfahren *webPlanungsaufgabe* und *webPostkorb* unterdurchschnittlich abschneiden, das Verfahren insgesamt auch schlechter bewerten, sollen in der Hauptuntersuchung noch einmal überprüft werden. Dazu werden wiederum die erzielten Punkterohwerte mit dem aufsummierten Wert der Items zur Benutzerzufriedenheit korreliert. Es zeigt sich analog zur Voruntersuchung, dass sich zwischen der Leistung in den Verfahren und der Bewertung Nullkorrelationen ergeben. Auch bei einer Extremgruppenbildung nach besonders guten und niedrigen Leistungen in den beiden Verfahren offenbaren sich keinerlei signifikante Unterschiede hinsichtlich der Bewertung. Neben dem direkten Abschneiden wird auch nach der subjektiv wahrgenommenen Schwierigkeit gefragt. Auch hier können nur

Nullkorrelationen zwischen der Verfahrensbewertung und der subjektiv wahrgenommenen Schwierigkeit ermittelt werden.

Entgegen den Ergebnissen von Gilliland (1994) lassen sich hier somit keine Hinweise finden, dass Personen, die in Verfahren schlechter abschneiden, diese auch insgesamt schlechter bewerten. Wenngleich nicht auszuschließen ist, dass die Ergebnisse anders ausgefallen wären bei einem Verfahrenseinsatz unter Ernstfallbedingungen, so sprechen die dargestellten Befunde insgesamt für die weitestgehende Teilnehmerakzeptanz der Verfahren.

#### *Korrelative Befunde zu konstruktremden Einflüssen*

Von besonderer Bedeutung für die Bewertung der Fairness eines Leistungsverfahrens ist die Analyse möglicher konstruktremder Einflüsse. In der Voruntersuchung lag der Zusammenhang zwischen der Computererfahrung und der Leistung im *webPostkorb* bei  $r = .32$ . Daher wurde die Einweisung für die Testform der Hauptuntersuchung überarbeitet und erweitert (s. Abschnitt 6.6).

Wie in Tabelle 44 ersichtlich, zeigen sich für die überarbeitete Version des *webPostkorbs* sowie für die *webPlanungsaufgabe* keinerlei substantielle Korrelationen über der Schwelle von  $r = .30$  zwischen den Gesamtergebnissen in den beiden internetbasierten Verfahren und den konstruktremden Einflüssen wie Geschlecht, Alter, Computererfahrung und Computereinstellung. Auch der Einfluss konstruktremder Variablen auf die Ergebnisse in den Subskalen des *webPostkorbs* bleibt deutlich unter der allgemein als kritisch bezeichneten Grenze von  $r = .30$  (vgl. Klieme et al., 2001). Aufgrund des Simulationsansatzes sind die Verfahren *webPlanungsaufgabe* und *webPostkorb* in spezifische, situative Kontexte eingebettet. Ziel der Diagnose ist die Erfassung der Konstrukte „Planungsfähigkeit“ und „Problemlösefähigkeit“. Es gilt daher auch zu überprüfen, inwieweit das spezielle Interesse am thematischen Kontext mit dem Leistungsergebnis kovariert. Wie in Tabelle 44 veranschaulicht, lassen sich keine substantiellen Korrelationen finden.

**Tabelle 44: Konstrukt fremde Einflüsse (Korrelationen mit den Testergebnissen) auf die Leistungen im *webPostkorb* und in der *webPlanungsaufgabe* (Hauptuntersuchung)**

Verfahren	Skalen	N	Geschlecht	Alter	Computerefahrung	Computereinstellung	thematisches Interesse
<i>web-Postkorb</i>	<i>dynamisches Problemlösen</i>	66	-.01	.06	.00	.16	.10
	<i>analytisches Problemlösen</i>		.21	-.11	.06	-.16	-.10
	Gesamtscore Problemlösen		.19	-.08	.04	-.08	.09
<i>webPLA</i>	Summenscore	45	.18	.11	.06	-.03	-.12
	komb. Gesamtscore		.16	.03	-.11	-.06	-.17

Anmerkung: Die Variable Geschlecht wurde mit 1 = weiblich und 2 = männlich kodiert; *webPLA*: *webPlanungsaufgabe*.

Die dargestellten Ergebnisse zur Akzeptanz und zur Fairness der *webPlanungsaufgabe* und des *webPostkorbs* sind insgesamt sehr zufriedenstellend und lassen den Schluss zu, dass es sich um faire und benutzerfreundliche Verfahren handelt.



## 8 Diskussion und Ausblick

### *Diskussion und Bewertung der Ergebnisse*

Hauptaugenmerk dieser Arbeit gilt der Darstellung der theoriegeleiteten, benutzerorientierten Entwicklung und der Güteprüfung der internetbasierten Simulationsverfahren *webPostkorb* und *webPlanungsaufgabe*. Darüber hinaus thematisiert die Arbeit die Vorteile und grundsätzliche Bedeutung internetgestützter Eignungsdiagnostik und stellt eine Klassifikation zum Blended-Assessment vor.

Die Optimierung und Evaluierung der Verfahren erfolgte durch empirische Studien an 104 Studierenden sowie 19 bewährten und 369 potenziellen Führungsnachwuchskräften. Die Verfahren wurden dabei sowohl im Rahmen eines Self-Assessments als auch unter echten Auswahlbedingungen im Rahmen eines E-Assessments eingesetzt. Gemäß der Validierungslogik der gewählten eignungsdiagnostischen Ansätze interessieren insbesondere die Inhalts- und Konstruktvalidität sowie die Benutzerfreundlichkeit, Akzeptanz und Fairness. Insgesamt zeichnet sich für beide Verfahren ein positives Bild hinsichtlich der Güteprüfung ab. Die wichtigsten Ergebnisse sollen im Folgenden diskutiert werden.

Aufgrund der programmgesteuerten, computergestützten Vorgabe und Auswertung können sowohl der *webPostkorb* als auch die *webPlanungsaufgabe* als objektiv eingeschätzt werden. Es ist allerdings zu berücksichtigen, dass bei einem eventuellen Online-Einsatz der Verfahren von zu Hause aus die Durchführungsobjektivität nicht letztgültig kontrollierbar ist. Dies gilt jedoch grundsätzlich für die internetbasierte Anwendung von Tests und ist nicht den Verfahren an sich zuzuschreiben. Ein objektiver Einsatz im kontrollierten Setting, wie er hier in der Vor- und Hauptuntersuchung erfolgte, ist sehr wohl möglich.

Die interne Konsistenz des *webPostkorbs* - berechnet mit Hilfe von Cronbachs-Alpha - bewegt sich für die Gesamtskala *webPostkorb* mit  $\alpha = .79$  im befriedigenden bis guten Bereich für ein simulatives Testverfahren. Hinsichtlich der Subskalen und Aufgabengruppen zeigen sich befriedigende Befunde zur Homogenität und internen Konsistenz. Angesichts der zu

erwartenden Reliabilitäten simulationsbasierter Verfahren (s. Übersichten von U. Funke, 1995a; Schippmann et al., 1990) handelt es sich um vergleichsweise hohe Werte.

Die Rohwerteverteilungen, Itemtrennschärfen und Itemschwierigkeiten des *webPostkorbs* können insgesamt als zufrieden stellend eingeschätzt werden. Allerdings zeigt sich, dass einige Items für die Stichprobe der Hauptuntersuchung zu schwierig waren. Dies hängt möglicherweise mit der etwas unterdurchschnittlichen mittleren Testintelligenz der Stichprobe zusammen.

Ein wichtiges Anliegen der Verfahrenskonstruktion war es, benutzerfreundliche und faire Verfahren zu entwickeln. Es wurde versucht, dies mit Hilfe einer Reihe von Strategien - wie der Einbeziehung von Experten und der Zielgruppe bei der Verfahrenskonstruktion, der Darbietung umfangreicher Einweisungs- und Instruktionsseiten sowie der wiederholten empirischen Überprüfung des Einflusses möglicher konstrukt fremder Variablen - zu gewährleisten.

Die Ergebnisse der Vor- und Hauptuntersuchung zur Benutzerfreundlichkeit, Akzeptanz und Fairness deuten auf den Erfolg der Vorgehensweise hin. Jeweils zwischen 75 und 90 Prozent der befragten Untersuchungsteilnehmer bewerten den *webPostkorb* und die *webPlanungsaufgabe* hinsichtlich der Bedienbarkeit, Einweisung, Gestaltung, Anregung zu Karriereentscheidungen und des Spaßes bei der Bearbeitung positiv oder sehr positiv.

Einige Studien haben gezeigt, dass die Computererfahrung und Computereinstellung die Leistungen in computergestützten Verfahren beeinflussen (u. a. M. Kersting, 1999; Klieme et al., 2001; Süß, 1996). Diese Variablen wurden in bisherigen Studien zur Güteprüfung internetbasierter Verfahren jedoch nur recht selten und zumeist nur mit einem einzigen Item erfasst. Daher war es Zielsetzung dieser Arbeit, die Computererfahrung und die Computereinstellung mit Hilfe eigener Skalen zu erheben und ihren Einfluss zu überprüfen. Die Analyse ergab, dass lediglich der Einfluss der Computererfahrung auf die Leistung im *webPostkorb* innerhalb der



Voruntersuchung mit  $r = .32$  knapp über der angelegten Schwelle von  $.30$  lag. Für die überarbeitete Endversion des *webPostkorbs* und für die *webPlanungsaufgabe* konnten keinerlei substantielle Zusammenhänge ermittelt werden. Ebenso zeigt sich kein entscheidender Einfluss weiterer konstruktremder Variablen wie Geschlecht, thematisches Interesse und Alter.

Mit steigender Interaktivität und steigenden Freiheitsgraden von internetbasierten Simulationsverfahren nehmen typischerweise auch die Anforderungen der Bearbeitung für die Probanden zu. Es galt daher zu überprüfen, inwieweit die Teilnehmer die verschiedenen Aufgaben erkannt haben. Die Analyse der Protokolldateien ergab, dass alle Aufgaben des *webPostkorbs* von den Probanden angesteuert wurden. Diese Befunde sprechen somit für die anvisierte intuitive und leicht verständliche Benutzeroberfläche.

Mit einer mittleren Gesamtbearbeitungsdauer von ca. 40 bis 45 Minuten für beide Instrumente inkl. der Einweisungen liegt der Gesamtzeitaufwand unter dem der meisten anderen simulationsbasierten Leistungsverfahren. Die Analysen des eingesetzten Trackingverfahrens bzw. der Logfiles zeigen keine Datenverluste und Abbrecher. Dies kann als Hinweis auf eine zumutbare Verfahrensdurchführung gewertet werden.

Bei einem Vergleich der eingesetzten Kurzversion des Berliner Intelligenzstruktur-Tests (BIS-K-S) mit dem *webPostkorb* schätzen die Befragten den *webPostkorb* in Bezug auf verschiedene Akzeptanzaspekte sowie dem Berufsbezug durchgehend als besser ein. Dies ist insbesondere im Zusammenhang mit eventuell angestrebten Personalmarketing-Zielen von Bedeutung. Hinsichtlich eines Einsatzes im Rahmen der Personalauswahl ist es nicht nur aufschlussreich, wie Verfahren insgesamt eingeschätzt werden, sondern auch, inwieweit Personen, die in Verfahren schlechter abschneiden, die Instrumente auch schlechter bewerten. Die Ergebnisse der Vor- und der Hauptuntersuchung belegen, dass die erzielten Leistungen und die selbst eingeschätzte subjektive Schwierigkeit der Bearbeitung keinerlei Einfluss auf die Bewertung der *webPlanungsaufgabe* und des *webPostkorbs* haben.

Zusammenfassend können die Verfahren somit als weitestgehend fair, akzeptiert und benutzerfreundlich eingeschätzt werden.

Wenngleich im Rahmen dieser Arbeit keine Analysen anhand von expliziten externen Kriterien des beruflichen Erfolgs durchgeführt wurden, so konnten dennoch erste interessante Ergebnisse hinsichtlich der Kriteriumsvalidität gefunden werden. Die *webPlanungsaufgabe* (Version Wirtschaft) korreliert signifikant in mittlerer Höhe mit dem Abschneiden innerhalb eines E-Assessments zur Personalauswahl von Führungsnachwuchskräften. Bei einem Mittelwertsvergleich der Leistungen in der Wirtschaftsversion der *webPlanungsaufgabe* erweisen sich bewährte Führungsnachwuchskräfte als signifikant besser als die Stichprobe der Bewerber. Dieser Befund kann als erster Hinweis dafür interpretiert werden, dass die mittels *webPlanungsaufgabe* (Version Wirtschaft) erhobenen Daten Prädiktoren für beruflichen Erfolg sein könnten.

Zukünftige Studien - insbesondere auch zum *webPostkorb* - sollten die kriterienbezogene Validität der Verfahren weitergehend beleuchten.

Besonderes Augenmerk dieser Arbeit liegt auf der Prüfung der Konstruktvalidität. Hintergrund der Verfahrensentwicklung ist u. a. die Analyse und die Berücksichtigung des empirischen Forschungsstands zu den Konstrukten „Planungsfähigkeit“ und „Problemlösefähigkeit“ (s. Kapitel 2). Im Theorieteil wurde aufgezeigt, dass Problemlösefähigkeit als ein heterogenes Konstrukt mit verschiedenen Facetten zu verstehen ist (Klieme et al., 2001). Dabei gilt es, komplexes, strategisches Problemlösen in dynamischen Situationen innerhalb computergestützter Szenarios von Problemlösen in weniger komplexen oder gar künstlichen Situationen zu unterscheiden. Letztere Facette hängt offensichtlich stark mit schlussfolgerndem Denken und daher Intelligenztestleistungen zusammen. Der Einfluss der Intelligenz - insbesondere der Verarbeitungskapazität - auf das komplexe Problemlösen in computergestützten Szenarios ist ebenfalls signifikant, wenngleich nur in mittlerer Größenordnung. Aufgrund der

begrifflichen Nähe zwischen Planen und schlussfolgerndem Denken (J. Funke & Fritz, 1995) wurde auch für das Planen ein Zusammenhang mit der Intelligenztestleistung erwartet. Es galt daher im Rahmen der Überprüfung der Konstruktvalidität zu untersuchen, inwieweit sich die hier vorgenommenen Operationalisierungen von Problemlöseleistung im *webPostkorb* und Planungsleistung in der *webPlanungsaufgabe* mit dem Forschungsstand vereinbaren lassen.

Dabei ist im Rahmen der Vor- und Hauptuntersuchung zum einen die innere Struktur und zum anderen der theoretisch zu erwartende Zusammenhang mit Intelligenztestleistungen näher betrachtet worden. Darüber hinaus wurde auch der Zusammenhang mit dem Personmerkmal „Handlungs- und Lageorientierung“ (Kuhl, 1994) überprüft.

Dimensionsanalysen zur Struktur des *webPostkorbs* bestätigen den anvisierten, aus der Theorie abgeleiteten, inhaltlichen Aufbau mit der Unterscheidung der zwei Subskalen *dynamisches Problemlösen* und *analytisches Problemlösen*.

Hinsichtlich des Zusammenhangs zu Intelligenztestverfahren konnten die positiven Ergebnisse der Voruntersuchung in mehreren Studien der Hauptuntersuchung repliziert werden. Es zeigen sich jeweils erwartungsgemäße, signifikante Zusammenhänge in mittelstarker Höhe zwischen der Planungsleistung in der *webPlanungsaufgabe* sowie der Problemlöseleistung im *webPostkorb* und den eingesetzten Intelligenztests.

Auch die relative Nähe der Subskalen des *webPostkorbs* zu der Intelligenztestleistung deckt sich mit den Ergebnissen der bisherigen Forschung - respektive der PISA-Studie zum Problemlösen von Klieme et al. (2001).

Lediglich der Zusammenhang der Subskala *dynamisches Problemlösen* mit der Verarbeitungskapazität weicht für die studentische Stichprobe der Hauptuntersuchung etwas von der erwarteten Höhe ab. Dies kann dem Umstand geschuldet sein, dass sich für diese Stichprobe die Aufgaben der Subskala *dynamisches Problemlösen* als ein wenig zu schwierig erwiesen.

Desweiteren konnte ein erwarteter, signifikanter Zusammenhang zwischen den Planungsleistungen in der *webPlanungsaufgabe* und den Problemlöseleistungen im *webPostkorb* aufgedeckt werden.

Aus den Analysen wird deutlich, dass die Leistungen in den Simulationsverfahren nicht alleine durch die allgemeine Intelligenz bzw. die kognitive Grundfähigkeit erklärt werden können, sondern den Leistungen in der *webPlanungsaufgabe* und im *webPostkorb* jeweils eine eigenständige Fähigkeit zu Grunde liegt, die als „Planungsfähigkeit“ bzw. „Problemlösefähigkeit“ mit den Facetten „analytische Problemlösefähigkeit“ und „dynamische Problemlösefähigkeit“ interpretierbar ist.

Die Ergebnisse zum Zusammenhang mit dem Personmerkmal „Handlungs- und Lageorientierung“ liegen ebenfalls im Rahmen der theoretisch zu erwartenden Größenordnungen.

Trotz der teilweise recht kleinen Stichproben können die ermittelten Befunde somit insgesamt als gute Hinweise auf die Konstruktvalidität der beiden internetbasierten Simulationsverfahren *webPostkorb* und *webPlanungsaufgabe* verstanden werden.

Die Sicherung der Inhaltsvalidität ist im Wesentlichen über eine umfassende, anforderungsgeleitete Verfahrenskonstruktion zu gewährleisten (vgl. von Rosenstiel, 2000; Schuler, 2001; Wottawa, 1995). Der hier verfolgte rationale, anforderungsbezogene Entwicklungsprozess der Verfahren beinhaltet verschiedene, regelgeleitete Schritte. Dazu gehörten die Berücksichtigung empirischer Tätigkeitsanalysen bei Schulleitungen zum Zwecke der Identifizierung berufstypischer Aufgaben und Anforderungen sowie die Bestimmung des Verhaltens bzw. der Fähigkeiten, die für die erfolgreiche Bewältigung dieser Aufgaben und Anforderungen entscheidend und notwendig sind.

Es wurde dargestellt, dass Planungs- und Problemlösefähigkeit erfolgskritische, eignungsrelevante Personmerkmale für die Arbeit von Schulleitungen bzw. Führungskräften im Allgemeinen sind. Darüber hinaus fanden die Erkenntnisse der psychologischen Forschung zu den Konstrukten und bisherigen methodischen Herangehensweisen Berücksichtigung, die für

die Verfahrenskonstruktion und das Evaluationsvorgehen entscheidend waren. Die Übersetzung der verschiedenen Aspekte in die Verfahrens- und Itemgestaltung erfolgte schließlich regelgeleitet und wurde in Vor- und Hauptuntersuchung hinsichtlich der verschiedenen Testgütekriterien überprüft und ggf. optimiert.

Die oben erwähnte Einbeziehung von Experten und der Zielgruppe bei den verschiedenen Schritten der Verfahrenskonstruktion galt dabei der fortlaufenden Inhaltsvalidierung der beiden Verfahren.

Auch die Ergebnisse zur Bewertung der Augenscheinvalidität durch Studierende weisen auf den Erfolg der anvisierten Verfahrenskonstruktion hin. Im Anschluss an die Bearbeitung des *webPostkorbs* und der *webPlanungsaufgabe* schätzten die Studierenden den Realitätsbezug der Verfahren ein. Es zeigt sich dabei, dass die große Mehrheit der Befragten den *webPostkorb* und die *webPlanungsaufgabe* als realistische Abbildungen beruflicher Anforderungen von Schulleitungen beurteilen.

Die Ergebnisse zur Inhalts- und Augenscheinvalidität deuten insgesamt darauf hin, dass der *webPostkorb* und die *webPlanungsaufgabe* als simulative Abbildungen eines Bereichs der erfolgsrelevanten Arbeit von Schulleitungen im Sinne einer Arbeitsprobe verstanden werden können.

Zusammenfassend konnten somit positive Hinweise hinsichtlich der Testgüte der Verfahren ermittelt werden. Ausblickend wären - neben der Bestätigung der Befunde an größeren Stichproben - insbesondere Studien zur populationsbezogenen Normierung, zur Kontrolle der Retest-Reliabilität und zur vertiefenden Überprüfung der Kriteriumsvalidität wünschenswert. Gerade im Hinblick auf die prädiktive Validität der Verfahren zur Vorhersage beruflicher Leistungen gilt es, zukünftig weitere Untersuchungen durchzuführen.

Es wäre sicherlich auch hilfreich, die Verfahren *webPostkorb* und *webPlanungsaufgabe* weiterzuentwickeln in Bezug auf verschieden schwere Versionen für unterschiedliche Zielgruppen oder hinsichtlich des Ausbaus zu adaptiven Testverfahren. Neben der bisher fokussierten resultatsorientierten Sichtweise könnte sich eine Analyse des Bearbeitungsprozesses für

differenziell-psychologische bzw. eignungsdiagnostische Fragestellungen als lohnend erweisen. Die zielgruppenorientierte Semantik, die mit der Wirtschaftsversion der *webPlanungsaufgabe* begonnen wurde, ist möglicherweise auch für andere Personengruppen und Berufswelten sinnvoll, wenngleich dadurch eine Güteprüfung jeweils neu durchzuführen wäre. Aus grundlagenorientierter Sicht könnte es ebenso von Interesse sein, die Rolle des Zeitdrucks und der verschiedenen Themenwechsel während der Bearbeitung zu beleuchten und mit dem Forschungsstand zu den Konstrukten in Einklang zu bringen.

#### *Rückblickende Bewertung des Vorgehens*

Neben der Diskussion der Ergebnisse der Güteprüfung ist auch die Betrachtung der hier gewählten Vorgehensweise von Interesse.

Dabei lässt sich rückblickend konstatieren, dass sich die realisierte Kombination von Konstrukt- und Simulationsansatz in der vorliegenden Arbeit bewährt hat. So konnten die Vorteile beider Ansätze verbunden werden.

Die Probleme hinsichtlich der Reliabilität und Validität von Problemlösemaßen bei komplexen Computerszenarios und bei computergestützten Postkorbverfahren ließen sich mit Hilfe der itembasierten und auswertungsobjektiven Struktur des *webPostkorbs* lösen. Durch diese Vorgehensweise war es möglich, die Analyseverfahren der klassischen Testtheorie zur weiteren Optimierung des *webPostkorbs* auch auf Itemebene sinnvoll einzusetzen. Inwieweit eine solche itembasierte Struktur auch für die kürzere *webPlanungsaufgabe* sinnvoll zu entwickeln ist, könnten weitere Arbeiten thematisieren.

Ein entscheidendes Ziel der Verfahrensgestaltung stellte die Realisierung einer möglichst weitgehenden Interaktivität dar (vgl. Schulmeister, 2002). Waren bisher im Internet zumeist nur einfache Übertragungen psychologischer Testverfahren mit Multiple-Choice-Logik zu finden, so hat diese Arbeit gezeigt, dass auch ein höherer Grad der Interaktivität zielgerecht realisierbar ist.

Allerdings ist mit zunehmender Interaktivität zumeist auch ein steigender inhaltlicher und technischer Aufwand verbunden, der nicht unerwähnt bleiben sollte und im Rahmen dieser Arbeit an den verschiedenen Stellen dokumentiert wurde.

Die Vorteile, die der höhere Grad an Interaktivität und die simulative Einbettung der Verfahren insbesondere im Hinblick auf die Akzeptanz und Zufriedenheit aufweisen (s. o.), scheint den höheren Aufwand jedoch durchaus zu rechtfertigen.

Die vorliegende Arbeit hat somit erste Erkenntnisse dazu geliefert, inwieweit sich auch verhaltensorientierte Simulationsverfahren internetgestützt abbilden und nach Testgütekriterien evaluieren lassen.

#### *Ausblick*

Bereits zum jetzigen Zeitpunkt besitzen Self- und E-Assessments Relevanz für das Personalmarketing und die Personalauswahl von Unternehmen und ebenfalls für die Auswahl und Beratung Studierender und Auszubildender (vgl. Kupka, 2005). Wagt man einen Blick auf die zukünftigen Entwicklungen, so ist insgesamt davon auszugehen, dass die praktische Bedeutung von Self- und E-Assessments und insbesondere der Einsatz multimedialer, interaktiver Simulationsverfahren aufgrund der genannten Vorteile weiter zunehmen wird (s. Kapitel 3).

Neben der Entwicklung neuartiger und interaktiver Verfahren ist für die Zukunft gerade die systematische Kombination verschiedener Verfahren und Ziele interessant. Wie umfassendes Blended-Assessment aussehen kann und welche Einflussgrößen für den Erfolg der Bewerberauswahl entscheidend sind, wurde bereits in Kapitel 3 vorgestellt. Inwieweit die systematische Kombination verschiedener Verfahren und Medien zu einer besseren Vorhersage der Berufs- oder Ausbildungsleistung bzw. Verbesserung der Personal- oder Studierendenauswahl beitragen kann, wäre dabei eine möglicherweise aufschlussreiche Forschungsfragestellung. Schmidt und Hunter (1998) haben mit ihrem Beitrag zur inkrementellen

Validität verschiedener Verfahren einen bedeutsamen Meilenstein gesetzt. Zukünftige Metaanalysen - insbesondere unter Berücksichtigung des Einsatzes interaktiver, multimedialer Verfahren und verschiedener Medien (wie beispielsweise online, offline, telefonbasiert) - könnten hier für weitere Aufklärung sorgen.

Interessant wäre dabei auch die Fragestellung, inwieweit sich die Strategien der Selbstauswahl (beispielsweise der Verbesserung der Grundquote mittels Self-Assessment) und Fremdauswahl (beispielsweise nach dem Ansatz des Blended-Assessments) optimal verbinden lassen, um insgesamt eine bestmögliche Auswahl zu erreichen.

Vor dem Hintergrund der demografischen Entwicklung<sup>25</sup> ist bei gleichzeitig steigendem Bewerbungsaufkommen<sup>26</sup> anzunehmen, dass sich die Suche geeigneter Nachwuchskräfte für die Unternehmen zukünftig zusehends schwieriger gestalten wird (vgl. Kopping, Diercks & Kupka, 2007). E- und Self-Assessments, die Personalmarketing- und Selbstauswahlaspekte betonen, werden in diesem Zusammenhang deutlich an Bedeutung gewinnen. Dabei ist zu hoffen, dass die Vertreter der Eignungsdiagnostik nicht nur die Erhöhung der prognostischen Validitätskoeffizienten im Sinne haben, sondern auch andere mögliche Wirkungshebel auf die Qualität einer Auswahlentscheidung betrachten. Effektive, umfassende Eignungsdiagnostik sollte nicht nur die Verbesserung des Validitätskoeffizienten beinhalten, sondern ebenso die Verbesserung der Grundquote bzw. der Selektionsquote sowie ggf. sogar die Verbesserung von Einflussgrößen innerhalb der Ausbildung bzw. Arbeit anvisieren, um eine insgesamt bestmögliche Auswahlentscheidung ermöglichen (s. Abschnitt 3.3).

---

<sup>25</sup> Die Zahl der 20- bis 29-Jährigen wird sich nach Einschätzung des Statistischen Bundesamt (2003) zukünftig von ihrem Stand 2006 von ca. 9.95 Millionen in den nächsten zwanzig Jahren um über eine Million auf 8.73 Millionen verkleinern. [s. a. [http://www.destatis.de/basis/d/bevoe/src/poppyrger\\_v1w1.svg](http://www.destatis.de/basis/d/bevoe/src/poppyrger_v1w1.svg)].

<sup>26</sup> Jüngerer Studien zufolge ist die Anzahl an Bewerbungen, die ein Hochschulabsolvent bis zur Einstellung verfasst, in den Jahren zwischen 2002 und 2005 sprunghaft um annähernd das Vierfache gestiegen (vgl. Berufsstart.de, 2005; StepStone, 2004). Inwiefern für die Erhöhung der derzeitige Arbeitsmarkt oder andere Faktoren wie etwa die einfache Vervielfältigungsmöglichkeit von E-Mail-Bewerbungen ursächlich ist, konnte nicht abschließend geklärt werden.



Neben der Wirtschaft zeichnet sich vermutlich auch der Bildungsbereich vermehrt als ein zukünftiger Anwendungsbereich aus, sowohl hinsichtlich der Studierendenauswahl als auch hinsichtlich der grundsätzlichen Ausbildungs- und Karriereberatung. Nicht zuletzt an der aktuellen Diskussion um die hochschulbezogene Vergabe von Studienplätzen zeigt sich die Bedeutung sinnvoller und umfassender Eignungsdiagnostik mit Studienberatung (vgl. Sieland, Schumacher & Neidhard, 2004). Dabei können anforderungsbezogene E- und Self-Assessments, die die spezifischen Rahmenbedingungen berücksichtigen, einen gewichtigen Beitrag leisten (Kupka, in Vorbereitung).

Inwieweit die psychologische Forschung diese Prozesse umfassend begleitet oder gar bestimmt, gilt es allerdings abzuwarten. Dabei wäre eine enge Verzahnung zwischen Wissenschaft und Praxis in diesem Feld wünschenswert und wichtig, da gerade die eignungsdiagnostische Forschung entscheidend zur Verbesserung von praktischen Auswahlentscheidungen beitragen kann und somit nicht nur Kosten- oder Zeitersparnisse bringen, sondern auch die teilweise gravierenden Folgen falscher Karriere- oder Auswahlentscheidungen für die beteiligten Personen reduzieren könnte.

Hinsichtlich der zukünftigen Entwicklung der internetbasierten Simulationsverfahren ist anzunehmen, dass Verfahren, die die Realität bzw. die erfolgskritischen Aspekte der Realität sinnvoll abbilden, weiter an Bedeutung gewinnen. Dabei werden der technische Fortschritt und die höhere Benutzerakzeptanz vieler Simulationsverfahren gegenüber herkömmlichen Tests ausschlaggebend sein. Womöglich können durch neuartige Verfahrensumgebungen, wie beispielsweise durch den Einsatz interaktiver, videobasierter Elemente, weitere, bisher nicht erklärte Varianzanteile des Berufs- oder Ausbildungserfolgs prognostiziert werden.

Allerdings bleibt zu fordern, dass bei zukünftigen, eignungsdiagnostischen Verfahrensentwicklungen der Evaluation neben der inhaltlichen Konzeption

und der sinnvollen Nutzung neuartiger (Internet-) Technologien zumindest ein ebenso großes Gewicht eingeräumt wird. Für Untersuchungen im Rahmen der angewandten Eignungsdiagnostik, die sich im Spannungsfeld zwischen Praxisrelevanz und methodischer Strenge bewegen, wird dabei die Herausforderung bestehen bleiben, jeweils eine optimale Balance zu finden.

## 9 Literatur

- Ach, N. (1905). *Über die Willenstätigkeit und das Denken*. Göttingen: Vandenhoeck & Ruprecht.
- Ackermann, H. & Wissinger, J. (Hrsg.). (1998). *Schulqualität managen: Von der Verwaltung der Schule zur Entwicklung von Schulqualität*. Neuwied: Luchterhand.
- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
- American Educational Research Association (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Amthauer, R. (1970). *Intelligenz-Struktur-Test 1970. (IST-70)*. Göttingen: Hogrefe.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)*. Göttingen: Hogrefe.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Anderson, J. R. (1996). *Kognitive Psychologie: eine Einführung* (2. Aufl.). Heidelberg: Spektrum.
- Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment*, 11, 121-136.
- Anderson, N., Lievens, F., Van Dam, K. & Ryan, A. M. (2004). Future perspectives on employee selection: Key directions for future research and practice. *Applied Psychology: An International Review*, 53, 487-501.

- Anzai, Y. & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review*, 86, 124-140.
- Arbeitsgemeinschaft der Schulleiterverbände Deutschlands (ASD) (1999). *Schulleitung in Deutschland. Ein Berufsbild in Entwicklung*. Stuttgart: Raabe.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2003). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. (10. Aufl.). Berlin: Springer.
- Barak, A. & Buchanan, T. (2004). Internet-based psychological testing and assessment. In R. Kraus, J. Zack & G. Stricker (Eds.), *Online counseling: A handbook for mental health professionals* (pp. 217-239). San Diego, CA: Elsevier Academic Press.
- Barak, A. & Cohen, L. (2002). Empirical examination of an online version of the self-directed search. *Journal of Career Assessment*, 10, 387-400.
- Barak, A. & English, N. (2002). Prospects and limitations of psychological testing on the Internet. *Journal of Technology in Human Services*, 19 (2/3), 65-89.
- Bartram, D. (2000). Internet recruitment and selection: Kissing frogs to find princes. *International Journal of Selection and Assessment*, 8, 261-274.
- Bartram, D. (2005). Computer-Based Testing and the Internet. In A. Evers, N. Anderson & O. Voskuil (Eds.), *The Blackwell Handbook of Personnel Selection* (pp. 399-418). Malden, MA: Blackwell Publishing Ltd.
- Bartram, D. & Brown, A. (2004). Online testing: Mode of administration and the stability of OPQ 32i scores. *International Journal of Selection and Assessment*, 12, 278-284.

- Baumert, J. & Leschinsky, A. (1986). Berufliches Selbstverständnis und Einflußmöglichkeiten von Schulleitern. *Zeitschrift für Pädagogik*, 2, 247-266.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.). (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske und Budrich.
- Beckmann, J. F. (1994). *Lernen und komplexes Problemlösen*. Bonn: Holos.
- Berufsstart.de (2005). *Studie zur Online-Bewerbung*. Großenkneten: Klaus Resch.
- Birri, T. & Sonderegger, J. (2000). *Unterlagen Lernberatung und Eignungsüberprüfung: Personale und Soziale Kompetenzen*. Rorschach: Seminar Eignungsabklärung.
- Bonsen, M. & Pfeiffer, H. (1998). Schulleitungsforschung in Deutschland. Forschung und Ergebnisse im Überblick. In H. Buchen, L. Horster & H. G. Rolf (Hrsg.), *Schulleitung und Schulentwicklung. Erfahrungen, Konzepte, Strategien* (S. 1-18). Stuttgart: Raabe.
- Bonsen, M. (2002). Schulleitungshandeln aus Lehrersicht - Einschätzungen zu schulentwicklungsbezogenen Handlungsdimensionen. In J. Wissinger & S. G. Huber (Hrsg.), *Schulleitung - Forschung und Qualifizierung* (S. 63-78). Opladen: Leske und Budrich.
- Bonsen, M. (2003). *Schule, Führung, Organisation. Eine empirische Studie zum Organisations- und Führungsverständnis von Schulleiterinnen und Schulleitern*. Münster: Waxmann.
- Bonsen, M. (2005, November). *Die Wirksamkeit von Schulleitung - Forschungsergebnisse und Konzepte*. Vortrag auf der

Schulleitungsfortbildung in Rheinland-Pfalz in Kooperation mit dem Ministère de l'Education Nationale et de la Formation Professionnelle, Luxemburg.

- Bonsen, M., Gathen, J. & Pfeiffer, H. (2002). Wie wirkt Schulleitung? Schulleitungshandeln als Faktor für Schulqualität. In H. G. Rolff, H. G. Holtappels, K. Klemm, H. Pfeiffer & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung, Band 12* (S. 287-322). Weinheim: Juventa.
- Bonsen, M., Gathen, J., Iglhaut, C. & Pfeiffer, H. (2002). *Die Wirksamkeit von Schulleitung. Empirische Annäherungen an ein Gesamtmodell schulischen Leitungshandelns*. Weinheim: Juventa.
- Booth, J. F. (1998). The user interface in computer-based selection and assessment: Applied and theoretical problematics of an evolving technology. *International Journal of Selection and Assessments*, 6, 1998, 61-82.
- Borman, W. (1982). Validity of Behavioral Assessment for Predicting Military Recruiter Performance. *Journal of Applied Psychology*, 67, 3-9.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. (2. Aufl.). Berlin: Springer.
- Bösel, R. M. (2001). *Denken. Ein Lehrbuch*. Göttingen: Hogrefe.
- Bray, D. W. (1964). The Management Progress Study. *American Psychologist*, 19, 419-420.
- Bray, D. W. & Grant, D. L. (1966). The assessment center in the measurements of potential for business management. *Psychological Monographs*, 80 (17), 1-27.

- 
- Briggs, S. R. & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54*, 106-148.
- Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control. *Quarterly Journal of Experimental Psychology, 29*, 181-201.
- Brocke, B., Beauducel, A. & Tasche, K. (1998). Der Intelligenz-Struktur-Test: Analysen zur theoretischen Grundlage und technischen Güte. *Diagnostica, 44*, 84-99.
- Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J. & Scholey, A. B. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods, 37*, 148-154.
- Buchner, A. (1999). Komplexes Problemlösen vor dem Hintergrund der Theorie finiter Automaten. *Psychologische Rundschau, 50*, 206-212.
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112*, 155-159.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct Validity in Psychological Tests, *Psychological Bulletin, 52*, 281-302.
- Diercks, J. & Kupka, K. (2004). eAssessment-Verfahren als Bestandteil des Personalauswahlprozesses [Online-Dokument]. *Crosswater-Systems, 15*, 1. URL: [http://www.crosswater-systems.com/ej5003\\_gastbeitrag\\_02\\_cyquest.htm](http://www.crosswater-systems.com/ej5003_gastbeitrag_02_cyquest.htm) [25.02.04].
- Diercks, J., Kupka, K. & Jägeler, T. (in Druck). Das SelfAssessment-Verfahren „Die Karrierejagd durchs Netz“. In J. Erpenbeck &

- L. von Rosenstiel (Hrsg.), *Handbuch Kompetenzmessung* (2. Aufl.). Stuttgart: Schäffer-Poeschel.
- Dommel, N. A. (1995). Postkörbe. In W. Sarges (Hrsg.), *Management-Diagnostik* (2. Aufl., S. 582-587). Göttingen: Hogrefe.
- Domsch, M. & Jochum, I. (1989). Zur Geschichte des Assessment Centers - Ursprünge und Werdegänge. In C. Lattmann (Hrsg.), *Das Assessment-Center-Verfahren der Eignungsbeurteilung. Sein Aufbau, seine Anwendung und sein Aussagegehalt* (S. 1-18). Heidelberg: Physica.
- Döring, N. (1997). Kommunikation im Internet: Neun theoretische Ansätze. In B. Batinic (Hrsg.), *Internet für Psychologen* (S. 267-298). Göttingen: Hogrefe.
- Döring, N. (1999). *Sozialpsychologie des Internet. Die Bedeutung des Internet für Kommunikationsprozesse, Identitäten, soziale Beziehungen und Gruppen*. Göttingen: Hogrefe.
- Dörner, D. (1979). *Problemlösen als Informationsverarbeitung* (2. Aufl.). Stuttgart: Kohlhammer.
- Dörner, D. (1986). Diagnostik der operativen Intelligenz. *Diagnostica*, 32, 290-308.
- Dörner, D. (1989). *Die Logik des Misslingens. Strategisches Denken in komplexen Situationen*. Hamburg: Rowohlt.
- Dörner, D., Kreuzig, H. W., Reither, F. & Stäudel, T. (Hrsg.). (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Dörner, D. & Reither, F. (1978). Über das Problemlösen in sehr komplexen Realitätsbereichen. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 25, 4, 527-551.



- Dörner, D., Schaub, H. & Strohschneider, S. (1999). Komplexes Problemlösen - Königsweg der Theoretischen Psychologie? *Psychologische Rundschau*, 50, 198-205.
- Drasgow, F. & Olson-Buchanan, J. B. (Eds.). (1999). *Innovations in computerized assessment*. Mahwah, NJ: Erlbaum.
- Dreher, M. & Oerter, R. (1985). Development of action planning during adolescence and early adulthood. In S. Friedmann, E. Scholnik & R. Cocking (Eds.), *Blueprints for thinking: The role of planning in cognitive development* (pp. 321-355). Cambridge, MA: University Press.
- Dubs, R. (1994). *Die Führung einer Schule: Leadership und Management*. Stuttgart: Franz Steiner.
- Dumas, J. S. & Redish, J. C. (1994). *A practical guide to usability testing*. Norwood, NJ: Ablex.
- Duncker, K. (1935). *Zur Psychologie des produktiven Denkens*. Berlin: Springer, Neudruck: 1974.
- Edelmann, W. (2000). *Lernpsychologie* (6. Aufl.). Weinheim: Beltz.
- Eichstaedt, J. (2005). Online-Assessment-Center. Ein neues Paradigma. In K.-H. Renner, A. Schütz & F. Machilek (Hrsg.), *Internet und Persönlichkeit* (S. 315-326). Göttingen: Hogrefe.
- Epstein, J., Klinkenberg, W. D., Wiley, D. & McKinley, L. (2001). Insuring sample equivalence across Internet and paper-and-pencil assessments. *Computers in Human Behavior*, 17, 339-346.
- Etzel, S. & Küppers, A. (2000). *Managementarbeitsprobe - MAP*. Göttingen: Hogrefe.
- Evers, L. (1995). *Planungskompetenz bei Führungskräften*. Unveröffentlichte Diplomarbeit, Universität Bonn.

- Fahrenberg, J., Hampel, R. & Selg, H. (1989). *Das Freiburger Persönlichkeits-Inventar FPI. Revidierte Fassung FPI-R und teilweise geänderte Fassung FPI-A1. Handanweisung* (5. Aufl.). Göttingen: Hogrefe.
- Faul, F. & Erdfelder, E. (1992). *GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS* [Computer-Programm]. Bonn: Psychologisches Institut der Universität Bonn.
- Fennekels, G. P. (1995). *PC-OFFICE 1.0. Postkorb zur Diagnose von Führungsverhalten (Handanweisung)*. Göttingen: Hogrefe.
- Fiedler, R. (2002). PC-Office. In U. P. Kanning & H. Holling (Hrsg.), *Handbuch personaldiagnostischer Instrumente* (S. 283-288). Göttingen: Hogrefe.
- Finke, A. & Eckl, M. (2001). Evolution E-Recruitment - Das Internet als Rekrutierungsmedium. In L. Hünninghausen (Hrsg.), *Die Besten gehen ins Netz - Report E-Recruitment: Innovative Wege bei der Personalauswahl* (S. 209-232). Düsseldorf: Symposium.
- Fischer, W. A. & Schratz, M. (1999). *Schule leiten und gestalten: Mit einer neuen Führungskultur in die Zukunft*. Innsbruck: Studien-Verlag.
- Fisseni, H. J. (1997). *Lehrbuch der psychologischen Diagnostik: mit Hinweisen zur Intervention* (2. Aufl.). Göttingen: Hogrefe.
- Fleishman, E. A. & Quaintance, M. K. (1984). *Taxonomies of Human Performance*. Orlando, FL: Academic Press.
- Frederiksen, N., Saunders, D. R. & Wand, B. (1957). The In-Basket test. *Psychological Monographs: General & Applied*, 71 (438).

- 
- Frensch, P. A. & Funke, J. (Hrsg.). (1995). *Complex problem solving: The European perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fritz, A. & Funke, J. (1995). Übersicht über vorliegende Verfahren zur Planungsdiagnostik. In J. Funke & A. Fritz (Hrsg.), *Neue Konzepte und Instrumente zur Planungsdiagnostik* (S. 47-78). Bonn: Deutscher Psychologen Verlag.
- Fritz, A. & Hussy, W. (1995). „Zoo-Spiel“: Zur Analyse der Planungsfähigkeit bei Kindern. In J. Funke & A. Fritz (Hrsg.), *Neue Konzepte und Instrumente zur Planungsdiagnostik* (S. 229-258). Bonn: Deutscher Psychologen Verlag.
- Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? *Diagnostica*, 29, 283-302.
- Funke, J. (1985). Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kausalmodelle. *Zeitschrift für Psychologie*, 193, 435-457.
- Funke, J. (1986). *Komplexes Problemlösen. Bestandsaufnahme und Perspektiven*. Berlin: Springer.
- Funke, J. (1990). Systemmerkmale als Determinanten des Umgangs mit dynamischen Systemen. *Sprache & Kognition*, 9, 143-154.
- Funke, J. (1992). *Wissen über dynamische Systeme. Erwerb, Repräsentation und Anwendung*. Berlin: Springer.
- Funke, J. (1993). Computergestützte Arbeitsproben: Begriffsklärung, Beispiele sowie Entwicklungspotentiale. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 119-129.

- Funke, J. (1999). Komplexes Problemlösen: Ein Blick zurück und nach vorne. *Psychologische Rundschau*, 50, 194-197.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Funke, J. & Buchner, A. (1992). Finite Automaten als Instrumente für die Analyse von wissensgeleiteten Problemlöseprozessen: Vorstellung eines neuen Untersuchungsparadigmas. *Sprache & Kognition*, 11, 27-37.
- Funke, J., Buchner, A., Dörner, D., Süß, H. M. & Vollmeyer, R. (1999). Diskussionsrunde zum Themenheft „Komplexes Problemlösen. *Psychologische Rundschau*, 50, 229-233.
- Funke, J. & Fritz, A. (Hrsg.). (1995). *Neue Konzepte und Instrumente zur Planungsdiagnostik*. Bonn: Deutscher Psychologen Verlag.
- Funke, J. & Fritz, A. (1995). Über Planen, Problemlösen und Handeln. In J. Funke & A. Fritz (Hrsg.), *Neue Konzepte und Instrumente zur Planungsdiagnostik* (S. 1-45). Bonn: Deutscher Psychologen Verlag.
- Funke, J. & Glodowski, A.-S. (1990). Planen und Problemlösen: Überlegungen zur neuropsychologischen Diagnostik von Basiskompetenzen beim Planen. *Zeitschrift für Neuropsychologie*, 2, 139-148.
- Funke, J. & Krüger, T. (1993). „Plan-A-Day“ (PAD) [Computer-Programm]. Bonn: Psychologisches Institut der Universität Bonn.
- Funke, J. & Krüger, T. (1995). „Plan-A-Day“: Konzeption eines modifizierbaren Instruments zur Führungskräfte-Auswahl sowie erste empirische Befunde. In J. Funke & A. Fritz (Hrsg.), *Neue Konzepte und Instrumente zur Planungsdiagnostik* (S. 97-120). Bonn: Deutscher Psychologen Verlag.

- 
- Funke, J. & Rasche, B. (1988). *PC-Postkorb. Manual zur PC-Version der Postkorb-Übung*. Bonn: Psychologisches Institut der Universität Bonn.
- Funke, U. (1992). *Diagnostisches interaktives System zur Komplexitätssimulation (DISKo). Manual*. Filderstadt: Hofmann & Partner.
- Funke, U. (1993). Computergestützte Eignungsdiagnostik mit komplexen dynamischen Szenarios, *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 109-118.
- Funke, U. (1995a). Szenarien in der Eignungsdiagnostik und im Personaltraining. In B. Strauß & M. Kleinmann (Hrsg.), *Computersimulierte Szenarien in der Personalarbeit* (S. 145-216). Göttingen: Hogrefe.
- Funke, U. (1995b). Using complex problem solving tasks in personnel selection and training. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving. The European perspective* (pp. 219-240). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gediga, G., Schöttke, H. & Tücke, M. (1983). Problemlösen in einer komplexen Situation. *Archiv für Psychologie*, 135, 325-339.
- Geilhardt, T. & Mühlbradt, T. (Hrsg.). (1995). *Planspiele im Personal- und Organisationsmanagement*. Göttingen: Hogrefe.
- Gilliland, S. W. (1994). Effects of Procedural and Distributive Justice on Reactions to a Selection System. *Journal of Applied Psychology*, 79, 691-701.
- Glatzer, D. & Roth, S. (2002). *Assessment-Center als Instrument zur Potentialanalyse und als Basis zur Entwicklungsplanung für potentielle Führungskräfte. Abschlussbericht*. Stuttgart: Oberschulamt.

- Goldstein, H. W., Yusko, K. P., Bravermann, E. P., Smith, D. B. & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology*, 51, 357-374.
- Graf, J. (Hrsg.). (1992). *Planspiel - simulierte Realitäten für den Chef von morgen*. Speyer: Gabal.
- Grunwald, W. (1995). Aufgaben und Schlüsselqualifikationen von Managern. In W. Sarges (Hrsg.), *Management-Diagnostik* (2. Aufl., S. 194-206). Göttingen: Hogrefe.
- Hacker, W. (1986). *Arbeitspsychologie - Psychische Regulation von Arbeitstätigkeiten. Schriften zur Arbeitspsychologie, Band 41*. Bern: Huber.
- Hartung, S. & Schneider, I. (1995). In B. Strauß & M. Kleinmann (Hrsg.), *Computersimulierte Szenarien in der Personalarbeit* (S. 219-236). Göttingen: Hogrefe.
- Hasselmann, D. & Strauß, B. (1995). *Herausforderung Komplexität Baustein 2. Textilfabrik*. Hamburg: Windmühle.
- Hasselmann, D. (1993). *Computersimulierte komplexe Problemstellung in der Management-Diagnostik: die Theorie zum Konzept Herausforderung Komplexität*. Hamburg: Windmühle.
- Hayes-Roth, B. & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science*, 3, 275-310.
- Hell, B., Trapmann, S., Weigand, S., Hirn, J. O. & Schuler, H. (2005, Sept.). *Die Validität von Prädiktoren des Studienerfolgs - eine Metaanalyse*. Vortrag auf der 4. Tagung der Fachgruppe Arbeits- und Organisationspsychologie in der Deutschen Gesellschaft für Psychologie, Universität Bonn.

- 
- Hertel, G. & Konradt, U. (Hrsg.). (2004). *Human Resource Management im Inter- und Intranet*. Göttingen: Hogrefe.
- Hertel, G., Naumann, S., Konradt, U. & Batinic, B. (2002). Personality assessment via Internet: Comparing online and paper-and-pencil questionnaires. In B. Batinic, U. D. Reips & M. Bosnjak (Eds.), *Online Social Sciences* (pp. 115-133). Seattle, WA: Hogrefe & Huber.
- Herz, O. & Lohmann, A. (1999). *Leitbild für eine Schulleiterin, einen Schulleiter, ein Schulleitungsmitglied* [Online-Dokument]. URL: <http://www.schulleitung.de> [30.06.2001].
- Heuke, S. (2005). *Das Auswahlverfahren für das Traineeprogramm „UniTrain“ bei der Unilever Deutschland GmbH unter besonderer Berücksichtigung des eAssessments „unique.st“ - Eine Untersuchung zur sozialen Validität*. Unveröffentlichte Diplomarbeit, Helmut-Schmidt-Universität Hamburg.
- Höft, S. & Funke, U. (2001). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 135-173). Göttingen: Hogrefe.
- Hörmann, H. J. & Thomas, M. (1989). Zum Zusammenhang zwischen Intelligenz und komplexen Problemlösen. *Sprache & Kognition*, 1, 23-31.
- Hofmann, L. (2000). *Führungskräfte in Europa - Empirische Analyse zukünftiger Anforderungen*. Wiesbaden: Gabler.
- Holland, J. L. (1994). *The Self-directed Search (SDS)* (4th Ed.). Odessa: Psychological Assessment Resources.
- Hornke, L. F. & Kersting, M. (2003). DIN 33430 - Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen

Eignungsbeurteilungen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 47, 57-58.

Hornke, L. F. & Kersting, M. (2006). Optimizing Quality in the Use of Web / Computer based Testing for Personnel Selection. In D. Bartram & R. K. Hambleton (Eds.), *Computer-Based Testing and the Internet: Issues and Advances* (pp. 149-162). New York: Wiley.

Hu, L. & Bentler, P. M. (1999). Cut-off criteria for fit indexes in Covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling*, 6, 1-55.

Huber, M. (1997). Wie können Schulleitungen wirksam sein? In U. Moser, E. Ramseier, C. Keller & M. Huber (Hrsg.), *Schule auf dem Prüfstand. Eine Evaluation der Sekundarstufe I auf der Grundlage der „Third International Mathematics and Science Study“* (S. 215-245). Chur: Rüegger.

Huber, S. G. (1999a). School Effectiveness: Was macht Schule wirksam? Internationale Schulentwicklungsforschung (I). *schul-management*, 2, 10-17.

Huber, S. G. (1999b). School Improvement: Wie kann Schule verbessert werden? Internationale Schulentwicklungsforschung (II). *schul-management*, 3, 7-18.

Huber, S. G. (2002). Trends in der Qualifizierung von Schulleiterinnen und Schulleitern - Ausgewählte Ergebnisse einer internal-vergleichenden Studie. In J. Wissinger & S. G. Huber (Hrsg.), *Schulleitung - Forschung und Qualifizierung* (S. 215-234). Opladen: Leske und Budrich.

Huber, S. G. (2003). *Qualifizierung von Schulleiterinnen und Schulleitern im internationalen Vergleich: Eine Untersuchung in 15 Ländern*



*zur Professionalisierung von pädagogischen Führungskräften für Schulen.* Kronach: Wolters Kluwer.

Hultman, G. (1989). The State of the Art of School Administration: a review of facts and theory. *Scandinavian Journal of Educational Research*, 33, 2, 123-162.

Hünninghausen, L. (2001). Personalsuche als strategischer Erfolgsfaktor. In L. Hünninghausen (Hrsg.), *Die Besten gehen ins Netz - Report E-Recruitment: Innovative Wege bei der Personalauswahl* (S. 11-17). Düsseldorf: Symposion.

Hussy, W. (1984). *Denkpsychologie. Ein Lehrbuch. Band 1: Geschichte, Begriffs- und Problemlöseforschung, Intelligenz.* Stuttgart: Kohlhammer.

Hussy, W. (1985). Komplexes Problemlösen - Eine Sackgasse? *Zeitschrift für Experimentelle und Angewandte Psychologie*, 32, 55-77.

Hussy, W. (1989). Intelligenz und Problemlösen. *Diagnostica*, 35, 1-16.

Hussy, W. (1991a). Eine experimentelle Studie zum Intelligenzkonzept „Verarbeitungskapazität“. *Diagnostica*, 37, 314-333.

Hussy, W. (1991b). Komplexes Problemlösen und Verarbeitungskapazität. *Sprache & Kognition*, 10, 208-220.

Hussy, W. (1998). *Denken und Problemlösen* (2. Aufl.). Stuttgart: Kohlhammer.

International Test Commission (2006). International Guidelines on Computer-Based and Internet Delivered Testing. *International Journal of Testing*, 6, 2, 181-188.

ISO 9241-11 (1995). *Ergonomic requirements for office work with display terminals (VDTs): Guidance on usability.* Berlin: Beuth.

- Jäger, A. O. (1982). Mehrmodale Klassifikationen von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195-225.
- Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35, 21-35.
- Jäger, A. O. (1986). Validität von Intelligenztests. *Diagnostica*, 32, 272-289.
- Jäger, A. O. (1991). Beziehungen zwischen komplexem Problemlösen und Intelligenz - eine Einleitung. *Diagnostica*, 37, 287-290.
- Jäger, A. O., Süß, H. M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test, Form 4*. Göttingen: Hogrefe.
- Jeserich, W. (1981). *Mitarbeiter auswählen und fördern: Assessment-Center-Verfahren*. München: Hanser.
- Jeserich, W. (1995). Assessment-Center. In W. Sarges (Hrsg.), *Management-Diagnostik* (2. Aufl., S. 717-727). Göttingen: Hogrefe.
- Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior and Research Methods, Instruments & Computers*, 31, 433-438.
- Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7 - A guide to the program and applications* (2nd Ed.). Chicago: SPSS Publications.
- Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8 - Structural equation modeling with the SIMPLIS Command Language*. Chicago: Scientific Software.
- Karat, J. (1982). A model of problem solving with incomplete constraint knowledge. *Cognitive Psychology*, 14, 538-559.

- Keim, T., König, W., Westarp, F. von, Wendt, O. & Weitzel, T. (2005). *Recruiting Trends 2005 - Eine empirische Untersuchung der Top-1.000-Unternehmen in Deutschland und von 1.000 Unternehmen aus dem Mittelstand*. Frankfurt: Universität, Schwerpunkt Wirtschaftsinformatik und Informationswirtschaft.
- Kersting, K. (2004). How do you test on the Web? Responsibly. *APA Monitor on Psychology, 35*, 26.
- Kersting, M. (1998). Differentielle Aspekte der sozialen Akzeptanz von Intelligenztests und Problemlöseszenarien als Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie, 42*, 61-75.
- Kersting, M. (1999). *Diagnostik und Personalauswahl mit computergestützten Problemlöseszenarien. Zur Kriteriumsvalidität von Problemlöseszenarien und Intelligenztests*. Göttingen: Hogrefe.
- Kirbach, C., Montel, C., Oenning, S. & Wottawa, H. (2004). *Recruiting und Assessment im Internet. Werkzeuge für eine optimierte Personalauswahl und Potenzialerkennung*. Göttingen: Vandenhoeck & Ruprecht.
- Klauer, K. C. (1995). Grundlagen der Problemlöseforschung. In B. Strauß & M. Kleinmann (Hrsg.), *Computersimulierte Szenarien in der Personalarbeit* (S. 17-42). Göttingen: Hogrefe.
- Kleinmann, M. & Strauß, B. (1998). Validity and application of computer-simulated scenarios in personnel assessment. *International Journal of Selection and Assessment, 6*, 97-106.
- Kleiter, G. D. (1970). Trend-control in a dynamic decision-making task. *Acta Psychologica, 34*, 387-397.

- Klieme, E., Funke, J., Leutner, D., Reimann, P. & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik*, 47, 179-200.
- Klix, F. (1971). *Information und Verhalten*. Bern: Huber.
- Kluwe, R. H., Misiak, C. & Haider, H. (1991). The control of complex systems and performance in intelligence tests. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 227-242). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kluwe, R. H. (1995). Computergestützte Systemsimulationen. In W. Sarges (Hrsg.), *Management-Diagnostik* (2. Aufl., S. 572-577). Göttingen: Hogrefe.
- Kmetz, J. T. & Willower, D. J. (1982). Elementary School Principals' Work Behaviour. *Educational Administration Quarterly*, 18 (4), 62-78.
- Köhler, W. (1917). *Intelligenzprüfungen an Anthropoiden, I*. Berlin: Verlag der Königlichen Akademie der Wissenschaften.
- Konradt, U. & Hertel, G. (2004). Personalauswahl, Platzierung und Potenzialanalyse mit internetbasierten Verfahren. In G. Hertel & U. Konradt (Hrsg.), *Human Resource Management im Inter- und Intranet* (S. 55-71). Göttingen: Hogrefe.
- Konradt, U., Lehmann, K., Böhm-Rupprecht, J. & Hertel, G. (2003). Computer- und internetbasierte Verfahren der Berufseignungsdiagnostik: Ein empirischer Überblick. In U. Konradt & W. Sarges (Hrsg.), *E-Recruitment und E-Assessment* (S. 105-124). Göttingen: Hogrefe.
- Konradt, U. & Sarges, W. (2003). Suche, Auswahl und Förderung von Personal mit dem Intra- und Internet: Strategien,

- Zielrichtungen und Entwicklungspfade. In U. Konradt & W. Sarges (Hrsg.), *E-Recruitment und E-Assessment* (S. 3-17). Göttingen: Hogrefe.
- Kopping, N., Diercks, J. & Kupka, K. (2007). E-Assessments bei Unilever. In G. Maier & M. John (Hrsg.), *Eignungsdiagnostik in der Personalarbeit: Grundlagen, Methoden, Erfahrungen* (S. 405-422). Düsseldorf: Symposion.
- Kotovsky, K., Hayes, J. R. & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17, 248-294.
- Kreitler, S. & Kreitler, H. (1987). Conceptions and process of planning: the developmental perspective. In S. L. Friedman, E. K. Schlonick & R. R. Cooking (Eds.), *Blueprints for thinking* (pp. 205-272). Cambridge, MA: Cambridge University Press.
- Kröner, S. (2001). *Intelligenzdiagnostik per Computersimulation*. Münster: Waxmann.
- Krüger, R. (1983). Was tut der Rektor? Zum Berufsbild und Selbstverständnis des Schulleiters. *schul-management*, 4, 32-36.
- Kubinger, K. D. (1993). Testtheoretische Probleme der Computerdiagnostik. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 130-137.
- Kuhl, J. (1981). Motivational and functional helplessness: The moderating effect of action vs. state orientation. *Journal of Personality and Social Psychology*, 40, 155-170.
- Kuhl, J. (1994). *Kurzanweisung zum Fragebogen HAKEMP 90. Handlungskontrolle nach Erfolg, Misserfolg und prospektiv.*

Osnabrück: Universität, Fachbereich Psychologie, Abteilung für Differentielle Psychologie und Persönlichkeitsforschung.

- Kuhl, J. (1995). Handlungs- und Lageorientierung. In W. Sarges (Hrsg.), *Management-Diagnostik* (2. Aufl., S. 303-316). Göttingen: Hogrefe.
- Kuhl, J. & Wassiljew, I. (1985). An information-processing perspective on motivation: Intrinsic task-involvement, problem-solving, and the complexity of action plans. In G. D'Ydewalle (Ed.), *Cognition, information processing, and motivation* (pp. 505-522). Amsterdam: North-Holland.
- Kupka, K. (2005). Internetbasierte simulative Spiele und Self-Assessments als Hilfe für die Berufs-, Studien- und Ausbildungswahl. In U. Blötz (Hrsg.), *Planspiele in der beruflichen Bildung* (CD-Rom Publikation zum Fachbuch). Bielefeldt: Bertelsmann.
- Kupka, K. (in Vorbereitung). Eignungsdiagnose - zu kurz gegriffen? In B. Baumann, K. Kupka, M. Otto, A. Redlich, S. Schmah, J. Schmidt, R. Schulmeister & V. Turau (Hrsg.), *Studierendenauswahl im Umbruch*.
- Kupka, K. & Diercks, J. (in Druck). CyPRESS - die Kombination von Self- und eAssessment bei Gruner+Jahr. *Wirtschaftspsychologie aktuell*.
- Kupka, K., Diercks, J. & Kopping, N. (2004). Webbasierte Personalauswahl durch E-Assessment bei Unilever Deutschland. *Wirtschaftspsychologie aktuell*, 3, 24-28.
- Langenscheidt (2005). *Studienwörterbuch Englisch. Langenscheidt. Englisch-Deutsch / Deutsch-Englisch*. München: Langenscheidt.

- Lienert, G. A. (1969). *Testaufbau und Testanalyse* (3. Aufl.). Weinheim: Beltz.
- Lienert, G. A. & Ratz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Lievens, F. & Harris, M. M. (2003). Research on Internet recruiting and testing: Current status and future directions. In C. L. Cooper & I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology vol. 18* (pp. 131-165). Chichester, England: John Wiley.
- Lüer, G. & Spada, H. (1992). Denken und Problemlösen. In H. Spada (Hrsg.), *Allgemeine Psychologie* (S. 189-280). Bern: Huber.
- MacCallum, R. C., Widaman, K. F., Zhang, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.
- Maukisch, H. (1978). Einführung in die Eignungsdiagnostik. In A. Mayer (Hrsg.), *Organisationspsychologie* (S. 105-136). Stuttgart: Poeschel.
- Mayr, J. (2001). Career Counselling for Teachers. *Context - European Education Magazine*, 25, 11.
- Mayr, J. & Nieskens, B. (2004, März). *Umstiege - Aufstiege - Ausstiege: Selbsterkundungsverfahren als Hilfsmittel zur Gestaltung der Lehrerlaufbahn*. Vortrag auf dem Kongress "Bildung über die Lebenszeit", Universität Zürich.
- Michel, L. & Conrad, W. (1982). Theoretische Grundlagen psychometrischer Tests. In K. J. Groffmann & L. Michel (Hrsg.), *Enzyklopädie der Psychologie, Serie II: Psychologische Diagnostik, Bd. 1* (S. 1-129). Göttingen: Hogrefe.

- Miller, G. A., Galanter, E. & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston. Deutsche Übersetzung: (1973). *Strategien des Handelns. Pläne und Strukturen des Verhaltens*. Stuttgart: Ernst Klett.
- Münch, E. (1999). *Neue Führungsperspektiven in der Schulleitung. Kooperation zwischen Schulleiter und Stellvertreter*. Neuwied: Luchterhand.
- Musch, J. & Lieberei, W. (1997). *Eine auswertungsobjektive Postkorbübung für Assessment Center*. Berichte aus dem Psychologischen Institut der Universität Bonn. Band 23, Heft 1. Bonn: Psychologisches Institut der Universität.
- Musch, J. & Lieberei, W. (2001). *Bonner-Postkorb-Module (BPM). Die Postkörbe CaterTrans, Chronos, Minos und AeroWings*. Göttingen: Hogrefe.
- Naglieri, J., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A. & Velasquez, R. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist*, 59, 150-162.
- Neuberger, O. (1986). Mitarbeiterführung. In W. Sarges & R. Fricke (Hrsg.), *Psychologie für die Erwachsenen-Weiterbildung* (S. 411-416). Göttingen: Hogrefe.
- Newell, A. & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Newell, A., Shaw, J. C. & Simon, H. A. (1960). A variety of intelligent learning in a general problem solver. In M. C. Yovits & S. Cameron (Eds.), *Self-organizing systems* (pp. 153-189). Oxford: Pergamon Press.
- Nielsen, J. (1993). *Usability Engineering*. San Diego, CA: Academic Press.



- Nielsen, J. (2001). *Designing Web Usability*. München: Markt+Technik-Verlag.
- Normausschuss Gebrauchstauglichkeit und Dienstleitungen (2002). *Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. DIN 33430*. Berlin: Beuth.
- Oerter, R. & Dreher, M. (1998). Entwicklung des Problemlösens. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie. Ein Lehrbuch* (S. 561-621). Weinheim: Beltz.
- Österreich, R. (1981). *Handlungsregulation und Kontrolle*. München: Urban & Schwarzenberg.
- Pascha, A., Schöppe, B. & Hacker, W. (2001). Was macht Planen kompliziert? - Zum Einfluss von Aufgabenmerkmalen auf die Schwierigkeit von Abfolgeplanungen. *Zeitschrift für Psychologie*, 209, 245-276.
- Pfeifer, A. & Schmidt, P. (1987). *LISREL - Die Analyse komplexer Strukturgleichungsmodelle*. Stuttgart: Fischer.
- Pfeiffer, H. (2002). Forschung zur Schulleitung - Schwerpunkte und Perspektiven. In J. Wissinger & S. G. Huber (Hrsg.), *Schulleitung - Forschung und Qualifizierung* (S. 21-32). Opladen: Leske und Budrich.
- Polya, G. (1945). *How to solve it*. Princeton, NJ: Princeton University Press.
- Potosky, D. & Bobko, P. (2004). Selection testing via the Internet: Practical consideration and exploratory empirical findings. *Personnel Psychology*, 57, 1003-1034.
- Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg. *Zeitschrift für Psychologie*, 189, 79-100.

- Putz-Osterloh, W. (1985). Selbstreflexion, Testintelligenz und interindividuelle Unterschiede bei der Bewältigung komplexer Probleme. *Sprache & Kognition*, 4, 203-216.
- Putz-Osterloh, W. (1987). Gibt es Experten für komplexe Probleme? *Zeitschrift für Psychologie*, 193, 63-84.
- Putz-Osterloh, W. & Lemme, M. (1987). Knowledge and its intelligent application to problem solving. *The German Journal of Psychology*, 11, 286-303.
- Putz-Osterloh, W. & Lüer, G. (1981). Über die Vorhersagbarkeit komplexer Problemlöseleistungen durch Ergebnisse in einem Intelligenztest. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 28, 309-334.
- Raven, J. C. (1962). *Advanced Progressive Matrices, Set II*. London: H. K. Lewis.
- Reichert, U. & Dörner, D. (1988). Heuristiken beim Umgang mit einem einfachen dynamischen System. *Sprache & Kognition*, 1, 12-24.
- Richards, S. A. & Jaffee, C. L. (1972). Blacks supervising whites: A study of interracial difficulties in working together in a simulated organization. *Journal of Applied Psychology*, 56, 234-240.
- Richter, T., Naumann, J. & Groeben, N. (2001). Das Inventar zur Computerbildung (INCOBI): Ein Instrument zur Erfassung von Computer Literacy und computerbezogenen Einstellungen bei Studierenden der Geistes- und Sozialwissenschaften. *Psychologie in Erziehung und Unterricht*, 48, 1-13.
- Riediger, M. & Rolfs, H. (1998). Computergestützte Postkorbverfahren: Mailbox '90, PC-Office und PC-Postkorb "Seeblick". *Zeitschrift für Arbeits- und Organisationspsychologie*, 42, 43-50.

- Roest, F., Scherzer, A., Urban, E., Gangel, H. & Brandstätter, C. (1989). *Mailbox '90. Ein computergestütztes Test- und Trainingsverfahren zur Personalentwicklung*. Wien: SciCon und Beltz.
- Rolff, H. G. (1980). *Soziologie der Schulreform*. Weinheim: Beltz.
- Rosenbusch, H. S. & Huber, S. G. (2002). Organisation und Merkmale der Qualifizierung von Schulleiterinnen und Schulleitern in den deutschen Bundesländern. In J. Wissinger & S. G. Huber (Hrsg.), *Schulleitung - Forschung und Qualifizierung* (S. 111-128). Opladen: Leske und Budrich.
- Rosenstiel, L. von (2000). Potentialanalyse und Potentialentwicklung. In L. von Rosenstiel & T. Lang-von Wins (Hrsg.), *Perspektiven der Potentialbeurteilung* (S. 3-25). Göttingen: Hogrefe.
- Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Bern: Huber.
- Rost, J. (1999). Was ist aus dem Rasch-Modell geworden? *Psychologie Rundschau*, 50, (3), 140-156.
- Salgado, J. F. & Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assesses' perceptions and reactions. *International Journal of Selection and Assessment*, 11, 194-205.
- Sarges, W. (Hrsg.). (1995). *Management-Diagnostik* (2. Aufl.). Göttingen: Hogrefe.
- Sarges, W. (1995). Eignungsdiagnostische Überlegungen für den Managementbereich. In W. Sarges (Hrsg.), *Management-Diagnostik* (2. Aufl., S. 1-21). Göttingen: Hogrefe.
- Sarges, W. & Wottawa, H. (Hrsg.). (2004). *Handbuch wirtschaftspsychologischer Testverfahren, Band I:*

*Personalpsychologische Instrumente* (2. Aufl.). Legerich: Pabst.

Scharley & Partner (1991). *Seeblick / Lakeview computerized In-basket-test*. Konstanz: Autor.

Scharley & Partner (1998). *PC-Postkorb* [Demoversion des Computer-Programms]. Konstanz: Autor.

Schaub, H. (2001). *Persönlichkeit und Problemlösen*. Weinheim: Beltz.

Schaub, H. & Strohschneider, S. (1992). Die Auswirkungen unterschiedlicher Problemlöseerfahrung auf den Umgang mit einem unbekanntem komplexen Problem. *Zeitschrift für Arbeits- und Organisationspsychologie*, 36, 117-126.

Schippmann, J. S., Prien, E. P. & Katz, J. A. (1990). Reliability and validity of in-basket performance measures. *Personnel Psychology*, 43, 837-861.

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. (2000): Meßbare Personmerkmale: Stabilität, Variabilität und Validität zur Vorhersage zukünftiger Berufsleistung und berufsbezogenen Lernens. In M. Kleinmann & B. Strauß (Hrsg.), *Potentialfeststellung und Personalentwicklung* (2. Aufl., S. 15-43). Göttingen: Hogrefe.

Schmidt-Atzert, L. & Deter, B. (1993). Intelligenz und Ausbildungserfolg: Eine Untersuchung zur prognostischen Validität des IST-70. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 52-63.

- Schmitz, K. (1980). Gegenwärtige Schulprobleme - dargestellt am Wochenlauf eines Schulleiters. *Bildung und Erziehung*, 33, 6, 546.
- Scholz, G. & Schuler, H. (1993). Das nomologische Netzwerk des Assessment Centers: eine Metaanalyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 73-85.
- Schoppek, W. (1996). *Kompetenz, Kontrollmeinung und komplexe Probleme. Zur Vorhersage individueller Unterschiede bei der Systemsteuerung*. Bonn: Holos.
- Schuler, H. (1990). Personenauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 34, 184-191.
- Schuler, H. (2001). Arbeits- und Anforderungsanalyse. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 43-61). Göttingen: Hogrefe.
- Schuler, H. & Funke, U. (1995). Diagnose beruflicher Eignung und Leistung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (2. Aufl., S. 235-284). Bern: Huber.
- Schuler, H., Funke, U., Moser, K. & Donat, M. (1995). *Personalauswahl in F&E. Eignung und Leistung von Wissenschaftlern und Ingenieuren*. Göttingen: Hogrefe.
- Schuler, H. & Höft, S. (2001). Konstruktorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 93-133). Göttingen: Hogrefe.
- Schuler, H. & Klingner, Y. (2005). *Arbeitsprobe zur berufsbezogenen Intelligenz. Büro- und kaufmännische Tätigkeiten (AZUBI-BK)*. Göttingen: Hogrefe.

- 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.
- Schulmeister, R. (2002). Taxonomie der Interaktivität von Multimedia - Ein Beitrag zur aktuellen Metadaten-Diskussion. In it + ti (Hrsg.), *Informationstechnik und Technische Informatik. Ausgabe 4* (S. 193-199). München: Oldenburg.
- Schwarzer, R. & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). Windsor: Nfer-Nelson.
- Selz, O. (1913). *Über die Gesetze des geordneten Denkverlaufs*. Stuttgart: Spemann.
- Sieland, B. (2002). Verhaltensprävention - ein unverzichtbarer Schritt. Wie können Pädagogen ihre Gesundheit und Leistungsfähigkeit fördern? *Pädagogik*, 54, 7-8, 22-28.
- Sieland, B. (2003). Das Problem der Eignung in der Aus- und Fortbildung von Pädagogen. In T. Gabriel & M. Winkler (Hrsg.), *Heimerziehung. Kontexte und Perspektiven* (S. 84-93). München: Reinhardt.
- Sieland, B. & Nieskens, B. (2001). Diagnosegeleitete Laufbahnberatung und Lehrerbildung. In C. Hanckel, B. Jötten & K. Seifried (Hrsg.), *Schule zwischen Realität und Vision* (S. 197-205). Bonn: Deutscher Psychologen Verlag.
- Sieland, B., Schumacher, L. & Neidhard, E. (2004, September). *Potenzialanalysen im Lehramtsstudium*. Vortrag auf dem Kongress der Deutschen Gesellschaft für Psychologie. Göttingen.

- Simoneit, M. (1933). *Wehrpsychologie. Ein Abriß ihrer Probleme und praktischen Folgerungen*. Berlin: Verlag Bernard & Graefe.
- Smither, J., Millsap, R., Stoffey, R., Reilly, R. & Pearlmann, K. (1996). An experimental test of the influence of selection procedures on fairness perceptions, attitudes about the organization, and job pursuit intentions. *Journal of Business and Psychology*, 10, 297-318.
- Sonntag, K. (1999). Ermittlung tätigkeitsbezogener Merkmale: Qualifikationsanforderungen und Voraussetzungen menschlicher Aufgabenbewältigung. In K. Sonntag (Hrsg.), *Personalentwicklung in Organisationen* (2. Aufl., S. 157-179). Göttingen: Hogrefe.
- Soworka, B. H. & Sarges, W. (1995). Psychologische Konstrukte für Eignungsprädiktoren. In W. Sarges (Hrsg.), *Management-Diagnostik* (2. Aufl., S. 207-217). Göttingen: Hogrefe.
- Statistisches Bundesamt (2003). *Bevölkerung Deutschlands bis 2050. 10. koordinierte Bevölkerungsvorausberechnung*. Wiesbaden: Autor.
- StepStone Deutschland AG (2004). *Studie: Aktuelle Trends auf dem Bewerbermarkt*. Düsseldorf: Autor.
- Sternberg, R. J. (1982). Reasoning, problem solving, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 225-307). New York: Cambridge University Press.
- Stogdill, R. M. (1948). Personal factors associated with leadership. A survey of the literature. *Journal of Psychology*, 25, 35-71.
- Strauß, B. (1993). *Konfundierungen beim Komplexen Problemlösen*. Bonn: Holos.

- Strauß, B. (2000). Die Messung der praktischen Intelligenz von Managern mit Hilfe computersimulierter Szenarien. In L. von Rosenstiel & T. Lang-von Wins (Hrsg.), *Perspektiven der Potentialbeurteilung* (S. 3-25). Göttingen: Hogrefe.
- Strauß, B. & Kleinmann, M. (Hrsg.). (1995a). *Computersimierte Szenarien in der Personalarbeit*. Göttingen: Hogrefe.
- Strauß, B. & Kleinmann, M. (1995b). Computersimierte Szenarien in der Personalarbeit. Eine Einleitung. In B. Strauß & M. Kleinmann (Hrsg.), *Computersimierte Szenarien in der Personalarbeit* (S. 11-14). Göttingen: Hogrefe.
- Strauß, B. & Kleinmann, M. (1995c). Die formale Beschreibung computersimulierter Szenarien. In B. Strauß & M. Kleinmann (Hrsg.), *Computersimierte Szenarien in der Personalarbeit* (S. 105-125). Göttingen: Hogrefe.
- Streifert, S., Pogash, R. & Piasecki, M. (1988). Simulation-based assessment of managerial competence: Reliability and validity. *Personnel Psychology, 41*, 437-557.
- Strohschneider, S. (1991a). Problemlösen und Intelligenz: Über die Effekte der Konkretisierung komplexer Probleme. *Diagnostica, 4*, 353-371.
- Strohschneider, S. (1991b). System von Systemen! Kommentar zu dem Aufsatz Systemmerkmale als Determinanten des Umgangs mit dynamischen Systemen von Joachim Funke. *Sprache & Kognition, 10*, 109-113.
- Süß, H. M. (1996). *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.



- Süß, H. M. (1999). Intelligenz und komplexes Problemlösen: Perspektiven für eine Kooperation zwischen differentiell-psychometrischer und kognitionspsychologischer Forschung. *Psychologische Rundschau*, 50, 220-228.
- Süß, H. M., Kersting, M. & Oberauer, K. (1991). Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen. *Diagnostica*, 37, 334-352.
- Süß, H. M., Kersting, M. & Oberauer, K. (1993). Zur Vorhersage von Steuerungsleistungen an computersimulierten Systemen durch Wissen und Intelligenz. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 14, 189-203.
- Süß, H. M., Oberauer, K. & Kersting, M. (1993). Intellektuelle Fähigkeiten und die Steuerung komplexer Systeme. *Sprache & Kognition*, 12, 83-97.
- Sydow, H. (1990). Zur Entwicklung der Planungsfähigkeit im Kindesalter. *Zeitschrift für Psychologie*, 198, 431-441.
- Taylor, H. C. & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology*, 23, 565-578.
- Tent, L. (1984). Intelligenz und Problemlösefähigkeit. Kommentar. *Psychologische Rundschau*, 3, 152-155.
- Testkuratorium der Föderation deutscher Psychologenverbände (1986). Beschreibung der einzelnen Kriterien für die Testbewertung. *Diagnostica*, 32, 358-360.
- Tewes, U. (1981). *Hamburg-Wechsler Intelligenztest für Kinder (HAWIK-R)*. Bern: Huber.

- Thiemermann, F. J. (1998). *Schulleiter im Hauptberuf: Führung und Management in der Schule*. Neuwied: Luchterhand.
- Thornton, G. C. & Byham, W. C. (1982). *Assessment Centers and Managerial Performance*. New York, NY: Academic Press.
- Thornton, G. C., Hollenshead, J. D. & Larsh, S. (1997). Comparison of two measures of motivation to manage: Ethnic and gender differences. *Educational and Psychological Measurement*, 57, 241-253.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.
- TNS Emnid (2001). *Repräsentative Umfrage zur Nutzung des Internets im Auftrag des Jobportals Jobware* [Online-Dokument]. URL: <http://www.jobware.de/pz/pm/meldungen/111.html/> [30.01.2002].
- Ueckert, H. (1995). Planung, Organisation und Kontrolle. In W. Sarges (Hrsg.), *Management-Diagnostik* (2. Aufl., S. 254-262). Göttingen: Hogrefe.
- VanLehn, K. (1991). Rule acquisition events in the discovery of problem-solving strategies. *Cognitive Science*, 15, 1-47.
- Vernon, P. A. (1971). *The structure of human abilities*. London: Methuen.
- Vogt, H. M. (1998). *Persönlichkeitsmerkmale und komplexes Problemlösen: der Zusammenhang von handlungstheoretischen Persönlichkeitskonstrukten mit Verhaltensweisen und Steuerungsleistungen bei dem computersimulierten komplexen Szenario UTOPIA*. München: Hamp.

- Vollmeyer, R. & Funke, J. (1999). Personen- und Aufgabenmerkmale beim komplexen Problemlösen. *Psychologische Rundschau*, 50, 213-219.
- Vollmeyer, R. & Rheinberg, F. (1998). Motivationale Einflüsse auf Erwerb und Anwendung von Wissen in einem computersimulierten System. *Zeitschrift für Pädagogische Psychologie*, 12, 11-24.
- Wagener, D. (2001). *Psychologische Diagnostik mit komplexen Szenarios*. Lengerich: Pabst.
- Weber, D. & Busch, D. (2002). Recrutainment: Karriere- und Bewerbermanagement im Flow. *Wirtschaftspsychologie aktuell*, 1, 44-48.
- Wernimont, P. & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, 52, 372-376.
- Wertheimer, M. (1943). *Productive Thinking*. New York, NY: Harper & Row. Deutsche Übersetzung: (1956). *Produktives Denken*. Frankfurt am Main: Kramer.
- Westhoff, K., Hellfritsch, L. J., Hornke, L. F., Kubinger, K. D., Lang, F., Moosbrugger, H., Püschel, A. & Reimann, G. (Hrsg.). (2005). *Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen. Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (2. Aufl.). Lengerich: Pabst.
- Weth, R. von der & Strohschneider, S. (2002). Planungsprozesse aus psychologischer Sicht. In S. Strohschneider & R. von der Weth (Hrsg.), *Ja, mach nur einen Plan. Pannen und Fehlschläge - Ursachen, Beispiele, Lösungen* (2. Aufl., S. 12-34). Bern: Huber.

- Wiesner, B. (1995). *Diagnostik individueller Planungsprozesse. Zu Beziehungen zwischen ausgesagter Planungsneigung, realisiertem Planungsverhalten und Leistung*. Regensburg: Roderer.
- Wingert, O. (1999). Schulleiterauswahlverfahren in Oberösterreich. *schulmanagement*, 3, 19-21.
- Wissinger, J. (1996). *Perspektiven schulischen Führungshandelns: eine Untersuchung über das Selbstverständnis von Schulleitern*. Weinheim: Juventa.
- Wissinger, J. (2002). Schulleitung im internationalen Vergleich - Ergebnisse der TIMSS-Schulleiterbefragung. In J. Wissinger & S. G. Huber (Hrsg.), *Schulleitung - Forschung und Qualifizierung* (S. 45-61). Opladen: Leske und Budrich.
- Wissinger, J. & Huber, S. G. (2002). Schulleitung als Gegenstand von Forschung und Qualifizierung - Eine Einführung. In J. Wissinger & S. G. Huber (Hrsg.), *Schulleitung - Forschung und Qualifizierung* (S. 9-20). Opladen: Leske und Budrich.
- Wittmann, W. W. & Süß, H. M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem solving performances via Brunswik-symmetry. In P. L. Ackerman, P. C. Kyllonen & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait and content* (pp. 77-108). Washington, DC: American Psychological Association.
- Wittmann, W. W., Süß, H. M. & Oberauer, K. (1996). *Determinanten komplexen Problemlösens* (Bericht 9). Mannheim: Universität Mannheim, Lehrstuhl Psychologie II.
- Wolfmeyer, P. (1981). *Die schulinterne Verwaltungstätigkeit der Lehrer. Eine empirische Untersuchung in Grund-, Haupt, Real- und*

---

*beruflichen Schulen sowie den Gymnasien Nordrhein-Westfalens.* Kasellaun: Henn.

- Wottawa, H. (1995). Umsetzung von situationsdiagnostischen Erkenntnissen in personendiagnostische Überlegungen. In W. Sarges (Hrsg.), *Management-Diagnostik* (2. Aufl., S. 175-194). Göttingen: Hogrefe.
- Wottawa, H., Kirbach, C., Montel, C. & Oenning, C. (2005). Psychologiegestütztes Recruiting im Internet: Stand und Perspektiven. *Report Psychologie*, 30 (4), 160-166.
- Zimmerhofer, A. & Hornke, L. F. (2005). Konzeption einer webbasierten Studienberatung für Interessierte der Studienfächer Informatik, Elektrotechnik und Technische Informatik. In K.-H. Renner, A. Schütz & F. Machilek (Hrsg.), *Internet und Persönlichkeit* (S. 269-283). Göttingen: Hogrefe.