

UNIVERSITÄT DORTMUND

FAKULTÄT FÜR HUMANWISSENSCHAFTEN UND THEOLOGIE

Die Qualität von Gruppenaufgaben in der Personalauswahl

Gutachter:

Prof. Dr. phil. Dr. med. Michael Kastner

Priv. Doz. Dr. phil. Joachim Vogt

Dissertation zur Erlangung
des akademischen Grades
eines Doktors der Philosophie

Vorgelegt von

Barbara Kolzarek

Dipl. Volkswirtin, Orga.-Psych. (Univ.-Zert.)

Feldstraße 28

D - 58119 Hagen

im Wintersemester 2005/2006

Danksagung

Es haben viele Personen auf unterschiedliche Weise dazu beigetragen, dass die vorliegende Arbeit entstehen konnte. Ich möchte mich daher ganz herzlich bedanken:

Bei meinen Eltern, die mich immer unterstützt haben.

Bei Herrn Prof. Dr. Dr. Michael Kastner und Herrn PD Dr. Joachim Vogt. Sie haben meine Forschungen von Anfang an begleitet.

Bei Dr. Lars Adolph, mit dem ich in zwei Jahren gemeinsamer Projektarbeit viel lernen konnte.

Bei meiner Kollegin Tanja Nolle, mit der ich die Höhen und Tiefen, die diese Promotion mit sich brachte, teilen konnte.

Bei Maren Maziul, mit der ich alles diskutieren konnte, die alles gelesen und mir wesentliche Korrekturhinweise gegeben hat.

Bei Heike Knieriem, die oft und immer wieder den Verwaltungsdschungel gelichtet hat.

Bei Tim Hagemann, der mir durch einen zwar kurzen, aber dafür nicht betriebsblinden Blick auf meine Arbeit noch einmal wertvolle Hinweise geben konnte.

Bei den Mitarbeiterinnen und Mitarbeitern der OPZ und des BMVg, die mir die Datenerhebung in ihrem Hause ermöglicht haben. Insbesondere bei Herrn Rühle, der sehr engagiert die wissenschaftlichen Anforderungen mit den operativen Anforderungen verknüpft hat, sowie bei Nicole Arndt, die mich zu erschreckend frühen Morgenstunden aus reinem Interesse bei der Datenerhebung unterstützt hat.

Inhalt

Tabellenverzeichnis	7
Abbildungsverzeichnis	9
Abkürzungsverzeichnis	10
Einleitung	11
1 Gruppenaufgaben in der Personaldiagnostik	13
1.1 Assessment Center – eine Definition?	13
1.2 Gütekriterien für Assessment Center – methodologische Einordnung	16
1.2.1 Objektivität	16
1.2.2 Reliabilität	17
1.2.3 Validität	18
1.3 Stand der Forschung – Empirische Ergebnisse zu den Gütekriterien	22
1.3.1 Reliabilität	22
1.3.2 Prognostische Validität	22
1.3.3 Konstruktvalidität	25
1.3.4 Kontentvalidität	26
1.3.5 Soziale Validität	27
1.4 Einflussfaktoren auf die Gütekriterien von Gruppenaufgaben	28
1.5 Formen und Anwendungsfelder von Gruppenaufgaben	31
1.5.1 Gruppendiskussion	33
1.5.2 Rollenspiel	35
1.5.3 Planspiel	36
1.5.4 Konstruktionsübung	37
1.5.5 Präsentation	38
2 Analyse des bestehenden und Entwicklung des neuen Verfahrens	39
2.1 Rahmenbedingungen und Besonderheiten des Verfahrens	39
2.2 Das bestehende Verfahren	40
2.3 Bewertung des bestehenden Verfahrens	42
2.4 Theoretische und organisationsspezifische Fundierung der Konstrukte	44
2.4.1 Auswahl der Zieldimensionen für die Gruppenaufgaben	45
2.4.2 Führungskompetenz	53
2.4.3 Sozialkompetenz	58
2.4.4 Teamkompetenz	64
2.4.5 Planungsverhalten	66
2.4.6 Unabhängigkeit bzw. Abhängigkeit der Konstrukte	66

2.5	Empfehlungen zur Verbesserung des Verfahrens	68
2.5.1	Allgemeine Empfehlungen	68
2.5.2	Empfehlungen für die Planspiele	69
2.6	Die neuen Gruppenaufgaben und die Verhaltenscheckliste	72
2.6.1	Entwicklung der Verhaltenscheckliste	72
2.6.2	Die Planspielaufgaben zur Situation „Krisenmanagement“	73
2.6.3	Die Planspielaufgaben zur Situation „Ressourcenengpass“	73
3	Methodik der empirischen Studie	74
3.1	Datenerhebung	74
3.1.1	Design	74
3.1.2	Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung	77
3.1.3	Stichprobe	78
3.2	Hypothesen der Arbeit	79
3.2.1	Auswirkung der Interventionen auf den Bewertungsprozess	80
3.2.2	Objektivität / Reliabilität	82
3.2.3	Differenzierung der Bewerber	82
3.2.4	Interne Konstruktvalidität	83
3.2.5	Externe Konstruktvalidität	84
3.3	Statistische Verfahren	85
3.3.1	Die Intraklassenkorrelation als Zusammenhangsmaß und als Varianzaufklärungsmaß	85
3.3.2	Die Multitrait-Multimethod-Matrix (MTMM-Matrix)	87
4	Ergebnisdarstellung I: Datenmaterial, Reliabilität und Differenzierung der Bewerber	89
4.1	Deskriptive Darstellung des Datenmaterials	89
4.2	Ergebnisse zur Hypothese 1 – Beobachterreliabilität und Verhaltenscheckliste	92
4.2.1	Beobachterreliabilität und Verhaltenscheckliste bei den alten Aufgaben	93
4.2.2	Beobachterreliabilität und Verhaltenscheckliste bei den neuen Aufgaben	94
4.2.3	Zusammenfassung zur Hypothese 1: Verhaltenscheckliste und Beobachterreliabilität	96
4.3	Ergebnisse zur Hypothese 2 – Beobachterreliabilität und neue Aufgaben	97
4.3.1	Beobachterreliabilität, alte versus neue Aufgaben ohne VCL	97
4.3.2	Beobachterreliabilität, alte versus neue Aufgaben mit Verhaltenscheckliste	98
4.3.3	Zusammenfassung zur Hypothese 2: neue Aufgaben und Beobachterreliabilität	99
4.4	Ergebnisse zur Hypothese 3: Differenzierung der OB und VCL	100
4.4.1	Die Differenzierung der Bewerber gelingt bei den alten Aufgaben mit VCL besser als ohne	100

4.4.2	Die Differenzierung der Bewerber gelingt bei den neuen Aufgaben mit VCL besser als ohne	100
4.4.3	Zusammenfassung zur Hypothese 3: Die Differenzierung der OB gelingt mit VCL besser als ohne	101
4.5	Ergebnisse zur Hypothese 4: Differenzierung der OB und neue Aufgaben	101
4.5.1	Die Differenzierung der OB gelingt mit den neuen Aufgaben besser als mit den alten (jeweils ohne VCL)	101
4.5.2	Die Differenzierung der OB gelingt mit den neuen Aufgaben besser als mit den alten (jeweils mit VCL)	101
4.5.3	Zusammenfassung zur Hypothese 4: Die Differenzierung der OB gelingt mit den neuen Aufgaben besser als mit den alten	102
5	Ergebnisdarstellung II: Interne Konstruktvalidierung	103
5.1	Ergebnisse zur Hypothese 5: Konstruktvalidität und VCL	103
5.1.1	Die konvergente Validität ist mit VCL größer als ohne	103
5.1.2	Die diskriminante Validität ist mit VCL größer als ohne	104
5.1.3	Zusammenfassung zur Hypothese 5: Die Konstruktvalidität ist mit VCL größer als ohne	106
5.2	Ergebnisse zur Hypothese 6: Konstruktvalidität und neue Aufgaben	107
5.2.1	Die konvergente Validität ist bei den neuen Aufgaben größer als bei den alten Aufgaben	107
5.2.2	Die diskriminante Validität ist bei den neuen Aufgaben größer als bei den alten Aufgaben	107
5.2.3	Zusammenfassung zur Hypothese 6: Die Konstruktvalidität ist bei den neuen Aufgaben höher als bei den alten	108
5.3	Anmerkung zur diskriminanten Validität	109
6	Ergebnisdarstellung III: Externe Konstruktvalidierung	110
6.1	Ergebnisse zur Hypothese 7: Externe Konstruktvalidität und VCL	110
6.1.1	Die externe konvergente Validität verbessert sich mit Einsatz der VCL beim alten Planspiel	110
6.1.2	Die externe diskriminante Validität verbessert sich mit Einsatz der VCL	111
6.1.3	Zusammenfassung externe Konstruktvalidität und VCL	117
6.1.4	Nomologisches Netzwerk	118
6.2	Ergebnisse zur Hypothese 8: Externe Konstruktvalidität und neue Aufgaben	118
6.2.1	Die externe konvergente Validität verbessert sich mit Einsatz des neuen Planspiels zum Krisenmanagement unter Einsatz der VCL	119
6.2.2	Die externe diskriminante Validität verbessert sich mit Einsatz des neuen Planspiels zum Krisenmanagement unter Einsatz der VCL	120
6.2.3	Zusammenfassung: Externe Konstruktvalidität und neue Aufgabe	125
6.2.4	Nomologisches Netzwerk	126

7 Zusammenfassende Darstellung der Ergebnisse und Diskussion	127
7.1 Tabellarische Übersicht der Ergebnisse	127
7.2 Objektivität	129
7.3 Reliabilität	130
7.4 Differenzierung	133
7.5 Konstruktvalidität	134
7.5.1 Kritische Diskussion der MTMM-Analyse	135
7.5.2 Wirkungsweise der VCL auf die Konstruktvalidität	137
7.6 Externe Konstruktvalidität	138
7.6.1 Nomologisches Netzwerk	139
7.6.2 Eignung des BIP™ zur externen Validierung	141
7.7 Kontentvalidität	141
7.8 Ökonomie und Nützlichkeit	142
8 Ausblick	144
Literaturverzeichnis	146
Anhang	156
Beschreibung der Gesamtheit, aus der die Stichprobe stammt	176

Tabellenverzeichnis

Tabelle 1-1: Überblick über veröffentlichte Validitätsstudien zum AC-Verfahren nach Jeserich (1981)	23
Tabelle 1-2: Übersicht über weitere Studien zur prognostischen Validität.....	24
Tabelle 1-3: Misst ein AC Qualitäten, die für das Management notwendig sind? (zit. nach Jeserich, 1991, S.315)	27
Tabelle 2-1: Begriffe im Zusammenhang mit Sozialer Intelligenz (Wegner, 1989).....	60
Tabelle 2-2: Drei Gruppen sozialer Kompetenzen nach Kanning (2003).....	61
Tabelle 3-1: Übersicht zum Design der Datenerhebung	76
Tabelle 3-2: Untersuchungskonzept	77
Tabelle 4-1: Datenmaterial deskriptiv, Häufigkeiten	89
Tabelle 4-2: Beobachterurteile deskriptiv 1, Mittelwert und Standardabweichung SD	91
Tabelle 4-3: Beobachterurteile deskriptiv 2, Mittelwert und Standardabweichung SD	92
Tabelle 4-4: ICC-Werte der OPZ-Beobachter zum alten Planspiel ohne und mit VCL.....	93
Tabelle 4-5: ICC-Werte der externen Beobachter zum alten Planspiel ohne und mit VCL.....	93
Tabelle 4-6: ICC-Werte der OPZ-Beobachter zum neuen Planspiel RE ohne und mit VCL.....	94
Tabelle 4-7: ICC-Werte der OPZ-Beobachter zum neuen Planspiel KM ohne und mit VCL.....	95
Tabelle 4-8: ICC-Werte der externen Beobachter zum neuen Planspiel RE ohne und mit VCL.....	95
Tabelle 4-9: ICC-Werte der externen Beobachter zum neuen Planspiel KM ohne und mit VCL.....	96
Tabelle 4-10: ICC-Werte der OPZ-Beobachter, Rundgespräch versus PS RE ohne VCL.....	97
Tabelle 4-11: ICC-Werte der OPZ-Beobachter, PS versus PS KM ohne VCL	97
Tabelle 4-12: ICC-Werte der externen Beobachter, Rundgespräch versus PS RE, ohne VCL.....	98
Tabelle 4-13: ICC-Werte der externen Beobachter, Planspiel versus PS KM, ohne VCL	98
Tabelle 4-14: ICC-Werte der OPZ-Beobachter, altes Planspiel versus PS KM, mit VCL	99
Tabelle 4-15: ICC-Werte der externen Beobachter, altes Planspiel versus PS KM, mit VCL.....	99
Tabelle 5-1: Konvergente Validität ohne und mit VCL	103
Tabelle 5-2: Diskriminante Validitätskoeffizienten, Planspiel alt, ohne und mit VCL	104
Tabelle 5-3: Diskriminante Validitätskoeffizienten, Planspiel KM ohne und mit VCL.....	105
Tabelle 5-4: Diskriminante Validitätskoeffizienten, Planspiel RE ohne und mit VCL	106
Tabelle 5-5: Konvergente Validität alte versus neue Aufgaben (jeweils ohne VCL)	107
Tabelle 5-6: Diskriminante Validitätskoeffizienten, Rundgespräch und PS RE	108
Tabelle 6-1: Externe konvergente Koeffizienten für Prüfgruppenmittelwerte (ohne und mit VCL) der OPZ-Beobachter und BIP™-Ergebnisse	110

Tabelle 6-2: Externe konvergente Koeffizienten für Prüfgruppenmittelwerte (ohne und mit VCL) der externen Beobachter und BIP TM -Ergebnisse.....	111
Tabelle 6-3: Koeffizienten für Prüfgruppenmittelwerte (ohne und mit VCL) der OPZ-Beobachter zur Dimension Führung und BIP TM -Ergebnisse	112
Tabelle 6-4: Koeffizienten für Urteilsmittelwerte (ohne und mit VCL) der externen Beobachter zur Dimension Führung und BIP TM -Ergebnisse	113
Tabelle 6-5: Koeffizienten für Prüfgruppenmittelwerte (ohne und mit VCL) der OPZ-Beobachter zur Dimension Soziale Kompetenz und BIP TM -Ergebnisse	114
Tabelle 6-6: Koeffizienten für Urteilsmittelwerte (ohne und mit VCL) der externen Beobachter zur Dimension Soziale Kompetenz und BIP TM -Ergebnisse	115
Tabelle 6-7: Koeffizienten für Prüfgruppenmittelwerte (ohne und mit VCL) der OPZ-Beobachter zur Dimension Planungsverhalten und BIP TM -Ergebnisse	116
Tabelle 6-8: Koeffizienten für Urteilsmittelwerte (ohne und mit VCL) der externen Beobachter zur Dimension Planungsverhalten und BIP TM -Ergebnisse	117
Tabelle 6-9: Externe konvergente Koeffizienten für Prüfgruppenmittelwerte (altes PS und PS KM, mit VCL) der OPZ-Beobachter und BIP TM -Ergebnisse	119
Tabelle 6-10: Externe konvergente Koeffizienten für Prüfgruppenmittelwerte (altes PS und PS KM, mit VCL) der externen Beobachter und BIP TM -Ergebnisse.....	119
Tabelle 6-11: Externe Koeffizienten für Prüfgruppenmittelwerte (altes PS und PS KM, mit VCL) der OPZ-Beobachter und BIP TM -Ergebnisse	120
Tabelle 6-12: Externe Koeffizienten für Urteilsmittelwerte (altes PS und PS KM, mit VCL) der externen Beobachter zur Dimension Führung und BIP TM -Ergebnisse.....	121
Tabelle 6-13: Externe Koeffizienten für Prüfgruppenmittelwerte (altes PS und PS KM, mit VCL) der OPZ-Beobachter zur Dimension Soziale Kompetenz und BIP TM -Ergebnisse	122
Tabelle 6-14: Externe Koeffizienten für Urteilsmittelwerte (altes PS und PS KM, mit VCL) der externen Beobachter zur Dimension Soziale Kompetenz und BIP TM -Ergebnisse	123
Tabelle 6-15: Externe Koeffizienten für Prüfgruppenmittelwerte (altes PS und PS KM, mit VCL) der OPZ-Beobachter zur Dimension Planungsverhalten und BIP TM -Ergebnisse	124
Tabelle 6-16: Externe Koeffizienten für Urteilsmittelwerte (altes PS und PS KM, mit VCL) der externen Beobachter zur Dimension Planungsverhalten und BIP TM -Ergebnisse	125
Tabelle 7-1: Zusammenfassende Darstellung der Hypothesenüberprüfung.....	127
Tabelle A A: Darstellung der Gruppenaufgaben und Urteilsdimensionen der Untersuchung	Fehler! Textmarke nicht definiert.
Tabelle A B: Beschreibung der Gesamtheit, aus der die Stichprobe stammt.....	176

Abbildungsverzeichnis

Abbildung 1-1: Reliabilitätsarten nach Wirtz & Caspar (2002).....	18
Abbildung 2-1: Das 2003 bestehende Verfahren im Überblick – Methoden und Dimensionen	41
Abbildung 2-2: Systematik der Personalauswahl nach Schuler (1996).	45
Abbildung 2-3: Liste möglicher Anforderungsdimensionen für die Offizierbewerbersauswahl:	47
Abbildung 2-4: Facetten der Sozialkompetenz nach Kastner (1999).....	59
Abbildung 2-5: Das neue Verfahren im Überblick – Methoden und Dimensionen.....	74
Abbildung 3-1: Wesentliche Einflussfaktoren auf den Bewertungsprozess.....	80
Abbildung 3-2: MTMM-Matrix für drei Traits (x,y,z) und drei Methoden (I,II,III)	87
Abbildung 6-1: Dimensionen des BIP TM zur Überprüfung der konvergenten Validität	110

Abkürzungsverzeichnis

AC	Assessment Center
BIP™	Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung
ex. Beo.	Externe Beobachter
GSV	Gruppensituationsverfahren
OB	Offizierbewerber/in
OPZ	Offizierbewerberprüfzentrale
PS	Planspiel
PS KM	Planspiel Krisenmanagement
PS RE	Planspiel Ressourcenengpass
RG	Rundgespräch
VCL	Verhaltenscheckliste

Einleitung

Die vorliegende Arbeit beschäftigt sich mit dem Einsatz von Gruppenaufgaben in der Personalauswahl am Beispiel der Offizierbewerberauswahl der Bundeswehr.

Seit Bestehen der Bundeswehr gibt es Aktivitäten, um das komplexe Gefüge der Personalauswahl und -entwicklung an die vielfältigen politischen, gesellschaftlichen und wissenschaftlichen Entwicklungen anzupassen. Insbesondere die Auswahl und Ausbildung der Führungskräfte unterliegt daher einem, durch die Größe der Organisation bedingt, allmählichen, aber ständigen Veränderungsprozess. So ist im Laufe der Jahrzehnte ein umfangreiches Personalauswahlsystem entwickelt worden, das sowohl Verfahren der Leistungsdiagnostik als auch der Persönlichkeitsdiagnostik umfasst. In der Offizierbewerberprüfzentrale (OPZ) in der Mudra-Kaserne in Köln werden zurzeit Offizierbewerber (OB) im Rahmen eines zweieinhalbtägigen Verfahrens unter anderem auf ihre allgemeine Offizierseignung geprüft.

Das Auswahlverfahren für Offizierbewerber beinhaltet ein breites Methodenspektrum. Die Bewerberinnen und Bewerber absolvieren computergestützte Testverfahren, schreiben einen Aufsatz, werden auf Studieneignung geprüft, durchlaufen ein Interview und ein Gruppenverfahren, werden auf ihre medizinische Eignung untersucht und bestreiten einen Fitness-Test.

Die vorliegende Untersuchung entstand im Rahmen eines Auftrags des Bundesverteidigungsministeriums, bestimmte Module des Auswahlverfahrens für Offizierbewerber (OB) zu analysieren und weiter zu entwickeln. Die Arbeit beschränkt sich auf das dort eingesetzte Gruppenverfahren. Zur Verbesserung der Lesbarkeit wird auf eine durchgängige gleichzeitige Verwendung der weiblichen und männlichen Form, wie z.B. Offizierbewerberin und Offizierbewerber, verzichtet.

Zielsetzung der Arbeit ist es, die Wirkung der entwickelten methodischen Veränderungen des Gruppenverfahrens auf die Beurteilungsqualität zu untersuchen.

Im ersten Teil der Arbeit wird die Rolle von Gruppenaufgaben in der Personaldiagnostik, insbesondere in ihrem Haupteinsatzbereich, den Assessment Centern (AC), dargestellt. Der Begriff des AC wird konkretisiert und für seine Verwendung in dieser Arbeit auf die Zielsetzung der Personalauswahl eingegrenzt. Im Anschluss werden die testtheoretischen Gütekriterien vorgestellt und ihre Anwendbarkeit im AC diskutiert. Das Kapitel 1.3 gibt einen Überblick über den Stand der Forschung zur Bestimmung von Gütekriterien in Assessment Centern und Kapitel 1.4 widmet sich speziell den empirischen Befunden zu Gruppenaufgaben. Im abschließenden Kapitel 1.5 wird die häufig fehlende Konkretisierung des Begriffs „Gruppenaufgabe“ im Zusammenhang mit Personalauswahl thematisiert und eine Abgrenzung der verschiedenen Aufgabenformen und ihrer Anwendungsfelder angeboten.

Der zweite Teil der Arbeit beschreibt das bestehende Gruppenverfahren in der OPZ sowie die Entwicklung und Begründung der Veränderungen im Rahmen des Projektes. Die Offizierbewerberauswahl der Bundeswehr zeichnet sich durch einige Besonderheiten im Vergleich zu den üblichen Verfahren zur Führungskräfteauswahl aus, die in der Beurteilung und

Weiterentwicklung des Verfahrens Berücksichtigung finden müssen. Nach einer Beschreibung dieser Rahmenbedingungen und Besonderheiten folgen die Beschreibung des bestehenden Verfahrens sowie seine qualitative Bewertung. In Kapitel 2.4 werden die Zieldimensionen des zukünftigen Gruppenverfahrens aus theoretischen Grundlagen sowie organisationsspezifischen Anforderungen hergeleitet. Auf dieser Basis werden die Empfehlungen zur Verbesserung des Verfahrens abgeleitet und ihre konkrete Umsetzung, die Struktur der neuen Gruppenaufgaben sowie eine Verhaltenscheckliste zur Unterstützung der Beobachtung und Beurteilung, beschrieben.

Mit dem dritten Teil beginnt der empirische Teil der Arbeit. Die Methodik der Studie, das Design und die Instrumente der Datenerhebung, sowie die Stichprobe werden vorgestellt. Die Gestaltung der Datenerhebung wurde zum Teil durch organisatorische Rahmenbedingungen beeinflusst. Aus diesem Grund werden die Hypothesen im Anschluss an dieses Kapitel dargestellt. Kapitel 3.3 beschließt diesen Teil mit einer Begründung der Auswahl der statistischen Verfahren sowie einer Beschreibung ihrer Anwendungsmöglichkeiten.

Die Teile 4 bis 6 enthalten die Ergebnisdarstellung. In Teil 4 wird eine deskriptive Darstellung des Datenmaterials gegeben sowie die Ergebnisse zur Reliabilität der Prüf-Urteile und zur Differenzierung der Bewerber dargestellt. Teil 5 enthält die Ergebnisse zur internen Konstruktvalidierung während in Teil 6 die Ergebnisse zur externen Konstruktvalidierung dargestellt werden.

Teil 7 schließlich fasst die Ergebnisse noch einmal zusammen und stellt sie in den inhaltlichen Zusammenhang des Veränderungsprozesses in der OPZ. Die Ergebnisse werden sowohl vor dem Hintergrund dieses Praxisbezuges als auch bezüglich methodischer Aspekte der Studie kritisch diskutiert und bewertet.

Die Arbeit schließt mit einem kurzen Ausblick, in dem Anknüpfungspunkte und weitere Forschungsfragen zu den wesentlichen Ergebnissen genannt werden.

1 Gruppenaufgaben in der Personaldiagnostik

Gruppenverfahren werden im Rahmen der Personaldiagnostik im Wesentlichen als Module im Rahmen eines Assessment Centers (AC) eingesetzt. Als theoretische Grundlagen im Umgang mit Gruppenaufgaben stehen also vor allem die Ergebnisse aus den folgenden Forschungsfeldern zur Verfügung. Es sind Erkenntnisse aus dem Bereich der Sozialpsychologie und der Gruppendynamik, weiter aus der differentiellen Psychologie und der Methodenlehre sowie aus der Diagnostik und dem spezifischen Bereich der Assessment Center-Forschung. Während das Forschungsfeld der Sozialpsychologie und der Gruppendynamik Hinweise für die Gestaltung von Gruppenaufgaben und den Prozess der Beobachtung liefert, bieten die Methoden der Assessment Center-Forschung die Grundlagen zur Überprüfung der erstellten Gruppenaufgaben. Auch wenn im Focus dieser Arbeit insbesondere die Gruppenaufgaben stehen und nicht das gesamte AC, sollen dennoch zur Einführung einige Grundlagen zum Thema Assessment Center dargestellt werden. Der Schwerpunkt wird hierbei auf die Aspekte des AC gelegt, die vor dem Hintergrund der Zielsetzung der Arbeit besondere Bedeutung haben. Eine kurze generelle Darstellung des Entwicklungs- und Durchführungsprozesses von Assessment Centern ist bei Fisseni und Fennekels (1995) zu finden.

1.1 Assessment Center – eine Definition?

Die Idee des AC geht historisch zurück auf die Entwicklung des Offizierauswahlverfahrens, das von der deutschen Wehrmacht in den 20' er Jahren entwickelt wurde. Methodisch sind seine Wurzeln in der Entwicklung der Psychotechnik zu sehen, als deren Begründer Hugo Münsterberg angesehen wird, und die auf der Annahme basiert, dass menschliches Verhalten in Segmente zerlegbar und damit beobachtbar und messbar ist. Einen ausführlichen geschichtlichen Abriss geben Domsch und Jochum (1989).

Grundsätzlich versteht man heute unter einem Assessment Center die

„... Kombination mehrerer unterschiedlicher Einzelmethoden zu einem Verfahrenssystem, das zur Einschätzung aktueller Kompetenzen wie zur Prognose künftiger beruflicher Entwicklung und Bewährung Verwendung findet“ (Schuler & Funke, 1989, S. 303).

Eine andere Definition stellt folgende Merkmale in den Vordergrund (vgl. Lattmann, 1989):

- Ein AC dient der Eignungsbeurteilung der Teilnehmenden durch Erfassung der Leistungsfähigkeit, der Arbeitseinstellung und der Entwicklungsfähigkeit.
- Beurteilt wird eine Mehrzahl von Personen durch eine Mehrzahl von Beurteilern.
- Beurteilt werden Merkmale der Person, die einen Anforderungsbezug aufweisen.
- Die Beurteilung erfolgt auf Grund einer Mehrzahl von Verfahren, insbesondere verhaltensbezogener Verfahren.

Laut Kleinmann (1997) hat sich im deutschsprachigen Raum die Definition von Jeserich durchgesetzt: Danach

... verstehen wir unter der Assessment-Center-Methode ein systematisches Verfahren zur qualifizierten Feststellung von Verhaltensleistungen bzw. Verhaltensdefiziten, das von mehreren Beobachtern gleichzeitig für mehrere Teilnehmer in bezug auf vorher definierte Anforderungen angewandt wird.“ (Jeserich, 1981, S. 33f)

In den letzten beiden Definitionen wird noch einmal deutlich die Verhaltensbezogenheit des Verfahrens herausgestellt.

Der inhaltliche Schwerpunkt von Assessment Centern liegt meist in der Beurteilung von Führungs- und Sozialverhalten im weiteren Sinne. Das AC kann mit zwei Zielsetzungen eingesetzt werden: der Personalauswahl und der Diagnostik individuenbezogener Fördermaßnahmen. Die weiteren Ausführungen dieser Arbeit beziehen sich auf die Zielsetzung der Personalauswahl.

Soweit einige Hinweise zu allgemeingültigen Merkmalen des Diagnoseverfahrens Assessment Center. Darüber hinaus gilt, dass Assessment Center zwar im Wesentlichen auf der gleichen Methodik beruhen, darüber hinaus jedoch jedes einzelne AC eine individuelle unternehmensspezifische Entwicklung ist. In diesem Sinne handelt es sich um ein nicht standardisiertes Verfahren (vgl. Fennekels, 1987). Die Entwicklung eines AC durchläuft idealtypisch folgende Schritte:

Der erste Schritt besteht aus einer organisationsspezifischen Anforderungsanalyse der Zielposition. Hierfür stehen unterschiedliche Methoden zur Verfügung, die sich auf die Analyse von Aufgaben, von Handlungen, von Handlungserfordernissen und von Fähigkeitserfordernissen stützen. Als einzelne Instrumente seien beispielhaft Interviewverfahren, die Methode kritischer Ereignisse oder auch die Repertory Grid Technik genannt (vgl. hierzu Donat & Moser, 1989 und Kratzer, 1991). Im Ergebnis liegt eine Reihe von Anforderungen vor, die in messbare Personenmerkmale überführt werden müssen. Es gibt unterschiedliche Ansätze, welches psychodiagnostische Beschreibungskonzept von Personenmerkmalen den zu messenden Beurteilungskriterien zu Grunde gelegt wird. Einerseits liegt ein Schwerpunkt des AC in der Beurteilung konkreten Verhaltens. Dies kann nur situationsspezifisch erfolgen. Andererseits wird immer wieder auf Eigenschaften, sog. Traits, also definitionsgemäß transsituativ stabile Persönlichkeitsmerkmale, als Beurteilungskriterium zurückgegriffen.

Im zweiten Schritt wird versucht, die Anforderungen der Zieldimension in entsprechenden Aufgaben abzubilden. Dabei besteht das so genannte

„... Missing-link-Problem, das darin besteht, einen eindeutigen Weg vom Phänomenbereich „Arbeit“ in den Phänomenbereich „Personenmerkmale“ zu finden. Bei allen bisher bekannten Methoden ist man auf die diesbezüglichen Einschätzungen durch Experten angewiesen, die, wie man inzwischen weiß, leider alles andere als perfekt sind.“ (Schuler, 1989, S.240)

Lattmann gibt folgende Übersicht über die am häufigsten eingesetzten Aufgabenarten zur Verhaltensbeurteilung (Lattmann, 1989, S.31):

- Schriftliche Einzelübungen
- Mündliche Einweg-Kommunikationen in einem sozialen Umfeld
- Dyadische interaktive Übungen
- Übungen in Kleingruppen

Diese können ergänzt werden durch anamnestische und eigenschaftserschließende Verfahren, sowie Leistungstests.

Im dritten Schritt gilt es, den Beobachtungs- und Beurteilungsprozess zu organisieren. Ziel ist es dabei, eine möglichst analytische Beurteilung des gezeigten Verhaltens der Teilnehmer zu ermöglichen. Um dies zu erreichen, wird auf eine Trennung der Erhebung der Auswertungsdaten und ihrer deutenden Beurteilung besonderer Wert gelegt. Diese Trennung sollte daher auch instrumentell und strukturell in jedem AC verankert sein. Weitere Einflussfaktoren auf die Urteile stellen die Systematik der Zuordnung von Beobachtern und Teilnehmern dar, sowie die Art und Weise, wie die Einzelbewertungen der Beobachter zu einem Gesamturteil und schließlich zu einem Endurteil zusammengeführt werden. Diese die Organisation des Gesamt-AC betreffenden Aspekte sollen an dieser Stelle nicht weiter vertieft werden. Weitere Faktoren, die konkret die Gestaltung und Qualität der Beobachtung und Beurteilung der Gruppenaufgaben betreffen, werden in den folgenden Kapiteln behandelt.

Einzelne Assessment Center können sich also bezüglich vieler Kriterien deutlich unterscheiden. Kleinmann (1997) gibt folgende Übersicht möglicher Differenzierungskriterien:

- Zielsetzung,
- Anzahl der Beobachter und Teilnehmer,
- Art der Bewertungsfindung,
- verwendete Verfahrenselemente,
- Dauer des Verfahrens,
- verwendete Rotationstechniken,
- Art der Anforderungen.

Dies gilt es bei einem Vergleich, z.B. von statistischen Gütekriterien, zu berücksichtigen. Entweder es liegen ausreichend Hintergrundinformationen über die Ausgestaltung der jeweiligen Assessment Center vor, oder es bleibt nur der Rückgriff auf einige Richtwerte, die auf der Basis von Meta-Analysen gewonnen wurden. Zusammenfassend lässt sich daher festhalten:

„Das Assessment-Center besteht in der Anwendungspraxis in einer Vielzahl unterschiedlichster Varianten, weshalb nur mit Vorsicht von der Praxis des Assessment-Centers gesprochen werden kann“ (Kompa, 1989, S. 25).

1.2 Gütekriterien für Assessment Center – methodologische Einordnung

Üblicherweise wird die Güte von Assessment Centern mit Hilfe der Kriterien der klassischen Testtheorie bewertet, also Objektivität, Reliabilität und Validität, sowie den Kriterien der Nützlichkeit und Ökonomie.

Fisseni und Fennekels (1995) geben eine kurze Zusammenfassung, in wiefern die Anwendung dieser Verfahren auf das AC problematisch zu sehen ist. Das AC entspricht von seiner Konzeption her eher der kriteriumsorientierten Testtheorie. Die in den einzelnen Aufgaben gemessenen Merkmale werden nicht an einer Normstichprobe relativiert, sondern ins Verhältnis gesetzt zu einem zuvor mit Hilfe der Anforderungsanalyse definierten Ideal. Da jedoch zur Beurteilung der Messqualität in der kriteriumsorientierten Testtheorie keine eigene Messtheorie zur Verfügung steht, muss auf die Konzepte der klassischen Testtheorie zurückgegriffen werden.

1.2.1 Objektivität

Unter Objektivität wird die Unabhängigkeit eines Verfahrens von der Person des Untersuchers verstanden (Lienert & Raatz, 1998). Genauer werden Durchführungs-, Auswertungs- und Interpretationsobjektivität unterschieden.

Die Durchführungsobjektivität lässt sich in einem AC über entsprechende Standardisierung der Abläufe und einzelnen Instruktionen recht gut verfolgen. Insbesondere bei den Gruppenaufgaben jedoch, deren Verlauf durch die einzelnen Teilnehmer selbst und die entstehende Gruppendynamik bestimmt wird, ist eine Grenze erreicht. Auch der Einsatz von Rollenspieler kann aus Gründen der Praktikabilität und der Effektivität kaum sinnvoll erscheinen. Da es im AC vorrangig um soziale Interaktion und Verhaltensbeobachtung geht, sind hier dem Ziel der Durchführungsobjektivität implizite Grenzen gesetzt.

Die Auswertungsobjektivität und Interpretationsobjektivität beziehen sich auf die Vorgehensweisen, mit denen gleiche Items gleichartig kategorisiert werden. Für die Gruppenaufgaben bedeutet dies, dass gleichen Verhaltensweisen auch gleiche Anforderungsdimensionen zugeordnet werden. Dieses Ziel kann unterstützt werden durch eine klare Definition der Anforderungsdimensionen mit ihren Kategorien sowie durch eine sorgfältige Beobachterschulung. Zudem muss eine möglichst einheitliche Vorgehensweise der Beobachter bei der Aggregation der Einzelwerte zum Gesamturteil angestrebt werden. Maukisch (1989) problematisiert, dass in der Regel der Bezug zwischen einzelnen Verhaltensweisen und einer definierten Aufgabensituation nicht explizit hergestellt wird.

„Es bleibt allenfalls der Improvisation des Beobachters vor Ort überlassen, jeweils intuitiv und ad hoc festzustellen, auf welche Aufgabenklasse sich ein jeweils beobachtetes Verhalten bezogen hat. Das Vorbringen neuer Vorschläge in einer Gruppendiskussion ist aber beispielsweise eine sehr unterschiedlich zu bewertende Verhaltensweise, je nachdem, ob sie zu Beginn oder am Ende eines „brainstorming“ erfolgt. ... Die Ausblendung bzw. fehlende Festlegung der Situationsklassen dürfte fast zwangsläufig zu einer traitorientierten Kodierung der Verhal-

tensweisen als „signs“ für generalisierte Persönlichkeitseigenschaften führen.“ (Maukisch, 1989, S. 263f.)

Die Objektivität von Assessment Centern wird in der Regel über die Messung der Beobachterübereinstimmung bestimmt.

1.2.2 Reliabilität

Mit der Reliabilität wird die Zuverlässigkeit eines Verfahrens beschrieben. Die Reliabilität bezeichnet die Abgrenzung von wahren und fehlerhaftem Anteil eines Messwerts und beantwortet die Frage, ob ein Instrument das, was es misst, genau misst. Es werden in der klassischen Testtheorie vier verschiedene Reliabilitätsarten unterschieden (Schelten, 1980):

- a) Retestreliabilität: hier wird mit denselben Teilnehmern ein Test zwei Mal durchgeführt und die Übereinstimmung der Ergebnisse überprüft. Abgesehen von dem relativ großen Aufwand, ein vollständiges AC mehrfach durchzuführen, ist die Bestimmung der Retestreliabilität grundsätzlich möglich. Zu bedenken sind dabei zum einen mögliche Lerneffekte auf Teilnehmerseite, mögliche Mängel bezüglich der Durchführungsobjektivität in den weniger standardisierten Modulen und die Veränderung der Gruppendynamik in den Gruppenaufgaben. Die Bestimmung der Retestreliabilität empfiehlt sich daher nur für hoch standardisierte AC-Module oder Gesamt-AC.
- b) Paralleltestreliabilität: diese Reliabilität wird als die Übereinstimmung zwischen den Ergebnissen eines Tests mit den Ergebnissen eines Paralleltests mit den gleichen Probanden bestimmt. Auch hier gilt, dass insbesondere im Rahmen der Gruppenaufgaben aufgrund der geringeren Standardisierbarkeit keine sehr hohen Koeffizienten zu erwarten sind, die Bestimmung für stark standardisierte Module oder Gesamt-AC aber durchaus sinnvoll sein kann.
- c) Halbierungsreliabilität: zur Bestimmung werden die Ergebnisse der zwei Hälften eines Tests auf ihre Übereinstimmung überprüft. Diese Reliabilitätsart eignet sich kaum für die Beurteilung eines Gesamt-AC, da sich kaum inhaltlich deckungsgleiche Teile abgrenzen lassen. Für einzelne Module, z.B. Einzelaufgaben wie Postkorb oder Leistungstests kann die Halbierungsreliabilität eingesetzt werden.
- d) Konsistenz: dieses Maß beschreibt, in wie weit alle Items eines Tests von den Probanden in gleicher Richtung beantwortet werden. Hier gilt das gleiche wie für die Halbierungsreliabilität.

In der Regel wird die Reliabilität im AC als Beobachterreliabilität angegeben, also mit dem gleichen Maß, das auch zur Bestimmung der Objektivität genutzt wird. Die Vermengung der Begriffe Objektivität und Reliabilität im Zusammenhang mit der Beobachtung von Verhalten entsteht dadurch, dass der Beobachter selbst auch als Messinstrument verstanden wird. Reliabilität wird dann verstanden als Maß für die Genauigkeit, mit der der Beobachter ein bestimmtes Verhaltens- oder Persönlichkeitsmerkmal erfasst.

Wirtz und Caspar (2002, S. 18) stellen folgenden Überblick über die verschiedenen Arten von Reliabilitätsdefinitionen vor:

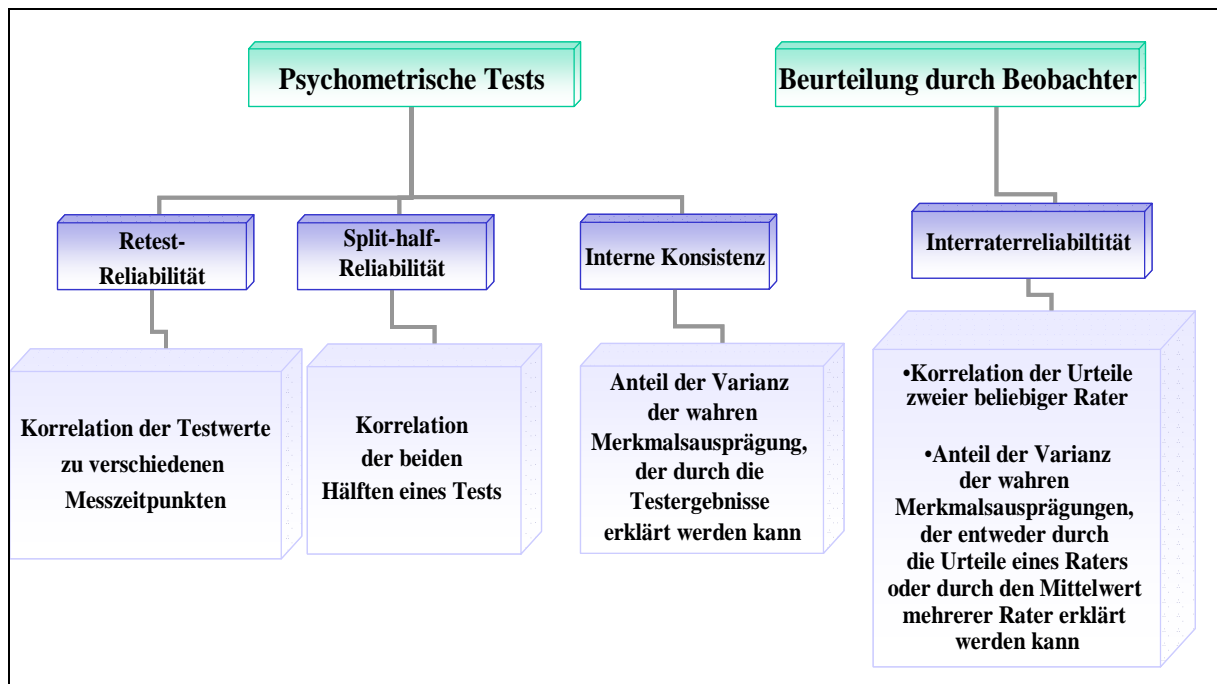


Abbildung 1-1: Reliabilitätsarten nach Wirtz & Caspar (2002)

Diese Abbildung verdeutlicht, dass Reliabilitätsmaße zum einen als Übereinstimmungsmaße zwischen Urteilen aufgefasst werden können, zum anderen aber auch als Varianzaufklärungsmaß im Sinne der internen Konsistenz. Im AC werden zur Bestimmung der internen Konsistenz die Urteile der Rater zu einem Teilnehmer als verschiedene Items eines Messinstrumentes aufgefasst. Urteilen mehrere Rater sehr ähnlich oder reliabel, so kann man davon ausgehen, dass von den Ratern dasselbe Merkmal erfasst wird.

Es wird jedoch mit der Reliabilität keine Aussage darüber getroffen, ob das gemessen wird, was gemessen werden soll. Dies ist die Frage der Validität.

1.2.3 Validität

Das Gütekriterium der Validität beschreibt die Angemessenheit von Schlussfolgerungen, die aus Diagnosewerten gezogen werden, wobei die soziale Validität die Perspektive der Bewerber berücksichtigt. Im Folgenden werden die verschiedenen Validitätsarten kurz erläutert, u. a. unter dem Aspekt ihrer Anwendbarkeit im AC:

- a) Kriteriumsbezogene Validität (konkurrente und prognostische Validität)
- b) Konstruktvalidität
- c) Kontentvalidität
- d) Soziale Validität

Ad a) Kriteriumsbezogene Validität

Die kriteriumsbezogene Validität wird auch als empirische Validität bezeichnet. Die geläufigste Form der Operationalisierung besteht im korrelativen Vergleich zweier Datenreihen – eines Prädiktors und eines Kriteriums. Werden Prädiktor und Kriterium zeitgleich erhoben, spricht man von konkurrenter Validität, wird das Kriterium zu einem späteren Zeitpunkt erhoben, von prognostischer Validität. Für Personalauswahlzwecke, zu denen das AC häufig eingesetzt wird, ist diese prognostische Validität besonders entscheidend. Zentrales Problem ist jedoch die Bestimmung der Kriterien, mit deren Hilfe Vorhersage oder Übereinstimmung überprüft werden. Gäbe es ein gesichertes Kriterium für die konkurrente Validierung, mit der die diagnostische Entscheidung einer Personalauswahl überprüft werden könnte, wäre es für die Praxis sinnlos, ein zusätzliches Verfahren einzusetzen. Stattdessen würde die Entscheidung auf der Basis des Kriteriums getroffen. Es ist in der Praxis im Grunde unmöglich ein prognostisches Kriterium zu finden, das keine Kontaminierung, also systematische Fehler enthält. Als mögliche Variablen für die prognostische Validierung werden Berufserfolg, Karriereschritte, Gehaltszuwachs oder Vorgesetztenurteil diskutiert (vgl. Fisseni, 1995). Da die Bestimmung der Kriteriumsvalidität nicht Bestandteil dieser Arbeit ist, soll auf eine tiefer gehende Diskussion an dieser Stelle verzichtet werden.

Ad b) Konstruktvalidität

Mit der Überprüfung der Konstruktvalidität wird die Frage danach gestellt, was ein Verfahren eigentlich misst. Konstrukte werden dabei als nicht beobachtbare Eigenschaften (latente Variablen) von Personen verstanden, die implizit durch die theoretisch zu erwartenden Beziehungen zu anderen latenten und beobachtbaren Variablen definiert werden (vgl. Cronbach & Meehl, 1955). Für das AC als eine Kombination verschiedener diagnostischer Verfahren, in denen zum Teil aber die gleichen Merkmale überprüft werden sollen, ergibt sich daraus die Anforderung, dass ein Konstrukt sich nicht durch die unterschiedlichen Messmethoden in seinem Gehalt verändern sollte. Es gibt in der Literatur ein sehr weit gefasstes Verständnis von Konstruktvalidität, in dem alle Schritte der Testentwicklung und –konstruktion ebenso berücksichtigt werden wie inhaltliche und kriteriumsbezogene Validität, mit dem Zweck, Hypothesen über theoretisch relevante Beziehungen zu prüfen und den Erkenntniswert einer diagnostischen Vorgehensweise zu erforschen (vgl. Schuler 1988). Während Guion (1987) ausführt, dass keine elaborierte Theorie beteiligt sein müsse, um ein Konstrukt abzugrenzen, ergeben sich aus einer solchen Sichtweise für eine theoriegeleitete Konstruktvalidierung notwendigerweise Probleme.

Ein Problem der Konstruktvalidität ist darin zu sehen, dass sie nicht durch einen exakten Kennwert bestimmbar ist. Fisseni beschreibt Konstruktvalidierung als einen Prozess, bei dem es

„... für die konkrete Durchführung keine Handlungsanweisung im Sinne eines Imperativs für Einzelschritte gibt.“ (Fisseni, 1990, S. 86)

In der Literatur findet sich eine Vielzahl von empirischen Möglichkeiten zur Konstruktvalidierung. Laubsch stellt folgende Auflistung zusammen (Laubsch, 2001):

- Korrelation mit Außenkriterien
- Korrelation mit anderen Test, die ähnliche Konstrukte erfassen
- Korrelation mit Tests, die andere Konstrukte erfassen
- Faktorenanalyse des zu validierenden Verfahrens gemeinsam mit Außenkriterien oder anderen Tests als Markiervariablen
- Experimentelle Variation von Variablen
- Testwiederholung
- Untersuchung von Gruppenunterschieden
- Testtheoretische Analyse der Items

Konstruktvalidierung ist also als eine Kombination theoretisch-psychologischer Vorüberlegungen bei der Entwicklung eines Verfahrens und der Explizierung seines nomologischen Netzwerkes im Nachhinein zu verstehen. Aus diesem Verständnis heraus ist zwangsläufig ein direkter Vergleich der Konstruktvaliditäten verschiedener AC oder AC-Aufgaben, der an der Oberfläche einzelner Koeffizienzwerte bleibt, nicht sinnvoll.

Ad c) Kontentvalidität

Der dritte Aspekt des Validitätsbegriffes bezieht sich zum einen auf die Frage, ob die Aufgabe repräsentativ für die Tätigkeit in der Zielposition ist. Lattmann (1989) spricht von gegebener inhaltlicher Validität, wenn die Aufgaben des AC ihren Gegenstand richtig erfassen. Geht man von Verhaltensstichproben aus, die den Schwerpunkt in AC-Verfahren darstellen, benötigt man für die Inhaltsvalidität einen möglichst sicheren Repräsentativitätsschluss, damit das Testverhalten als repräsentativ für das gewünschte Gesamtverhalten angesehen werden kann. Eine Möglichkeit, dies zu gewährleisten, besteht darin, direkt eine Stichprobe aus der Menge der realen Arbeitsaufgaben zu ziehen und den Bewerbern als Test vorzugeben (vgl. Schuler, 1989). Diese Methode hat sicherlich Vorteile, sofern es sich nicht um eine Auswahlentscheidung für eine Ausbildung handelt. Zu beachten ist jedoch, dass es Anforderungen gibt, die erst nach längerer zeitlicher Erstreckung zum Tragen kommen, bzw. erst in der Vernetzung verschiedener Anforderungskomponenten wirksam werden.

Zum anderen wird mit der Frage der Inhaltsvalidität die Affinität der verschiedenen Aufgaben zu den Dimensionen bewertet. Diese Bewertung findet in der Regel durch Expertenratings statt. Verlässt man bei der Konstruktion eines AC die konkrete Verhaltensebene und greift auf Konstrukte zurück, fehlen wissenschaftliche Nachweise über die Art der gewünschten Verbindung zwischen Aufgabenkonzeptionierung und Zieldimensionen. Um diese Verbindung möglichst optimal herzustellen, gibt es zwar verschiedene Methoden der Anforderungsanalyse, die hier nicht weiter beschrieben werden sollen. Letztlich bleibt es aber der Kreativität und Erfahrung des AC-Entwicklers überlassen, die Inhaltsvalidität zu gewährleisten. Bei der Inhaltsvalidität wirkt sich also das bereits genannte „missing-link-Problem“ aus.

Für das Verfahren in der OPZ ist der Rückgriff auf Konstrukte zur Bewertung der Bewerber aus zwei Gründen notwendig. Zum einen handelt es sich um eine Auswahlentscheidung für eine umfangreiche und zeitintensive Ausbildung. Zum anderen unterscheiden sich die späte-

ren Zielpositionen in ihren Tätigkeiten sehr stark. Im Auswahlverfahren werden also Fähigkeiten und Dimensionen geprüft, die als notwendig für eine erfolgreiche Absolvierung der Ausbildung erachtet werden.

Die Gestaltung der inhaltlichen Validität von AC-Aufgaben hat einen wesentlichen Einfluss auf die so genannte Augenscheingültigkeit und damit auf die Akzeptanz von Auswahlverfahren. Diese wiederum sind Facetten der sozialen Validität.

Ad d) Soziale Validität

Der Begriff der sozialen Validität bezieht sich auf die Wahrnehmung und Beurteilung des Verfahrens durch die Bewerber. Schuler und Stehle führten diesen Begriff erst 1983 ein. Das Konzept der sozialen Validität wird über vier Facetten beschrieben, die prinzipiell auch der empirischen Überprüfung zugänglich sein sollen (vgl. Schuler, 1989):

- Bewerber bekommen Informationen über die zur Selbstselektion relevanten Charakteristika von Arbeitsplatz und Organisation.
- Bewerber werden an der Entwicklung und Verwendung der eignungsdiagnostischen Instrumente beteiligt.
- Die Verfahrensdurchführung und der diagnostische Schluss sind für die Bewerber transparent.
- Die Urteilkommunikation erfolgt in rücksichtsvoller, verständlicher und nachvollziehbarer Form.

Gerade bei einem AC lassen sich diese Facetten gut berücksichtigen. Bisher ist die empirische Evidenz jedoch noch gering. Für das Verfahren der OPZ ist die Frage der sozialen Validität von besonderer Bedeutung. Mit der Gestaltung der o.g. Facetten wird der Organisationskultur Ausdruck gegeben. Aspekte der Akzeptabilität und Fairness des Verfahrens sollten explizit Beachtung finden, da es sich um einen öffentlichen Arbeitgeber handelt und zudem die Zahl der Bewerber sehr hoch ist. Es kommt hinzu, dass sich die Bewerber für einen nennenswerten Zeitraum verpflichten. Daher sollte diese Entscheidung auf einem fairen und auch für die Bewerber transparenten Prozess beruhen. Entsprechend wurden die Aspekte der sozialen Validität bei der Weiterentwicklung des Verfahrens in der OPZ explizit berücksichtigt.

Die Bedeutung der einzelnen Validitätskriterien ist unter pragmatischen Gesichtspunkten je nach Zweck eines AC unterschiedlich zu bewerten. Für ein Auswahl-AC ist sicherlich die Prognostische Validität das entscheidende Kriterium. Sie sagt aus, inwieweit die Organisation durch den Einsatz des AC in der Lage ist, ihre Personalauswahlentscheidungen erfolgreich zu treffen. Aus diesem Grunde konzentrierten sich zunächst die Forschungsarbeiten auf diesen Bereich. Mit der Zeit rückte jedoch auch die Frage nach der Konstruktvalidität stärker in den Mittelpunkt. In dem Maße, in dem die im AC gewonnenen Erkenntnisse auch für die Personalentwicklung nutzbar gemacht werden sollten, gewann die Frage danach, was im AC eigentlich bewertet wird, zunehmend an Bedeutung.

1.3 Stand der Forschung – Empirische Ergebnisse zu den Gütekriterien

1.3.1 Reliabilität

Zufriedenstellende Werte für die Reliabilität bilden die Voraussetzung für eine sinnvolle Auseinandersetzung mit Fragen der Validität. Scholz (1994) gibt im Anhang seiner Arbeit eine ausführliche Auflistung der Ergebnisse von Reliabilitätsuntersuchungen. In zwei Studien wurden ganze AC wiederholt durchgeführt, und es wurde eine Korrelation des Gesamtergebnisses von .73 und .74 gefunden. Die Retest-Reliabilität einzelner Aufgaben schwankt sehr stark zwischen .06 und .90, wobei jedoch die Mehrheit der aufgeführten Studien zufrieden stellende Werte erbringt. In der Untersuchung von Scholz (1994) ergeben sich Interraterreliabilitäten für einzelne Aufgaben und Dimensionen zwischen .43 und .95. Wird die Interraterreliabilität nach einer Aussprache der Beobachter gemessen, ergeben sich leicht höhere Werte zwischen .67 und .99. Seine Ergebnisse liegen damit in den gleichen Bereichen wie die einer Vielzahl weiterer Studien, die im Laufe von fast 40 Jahren Forschung zu diesem Thema publiziert wurden. Bass (1954) hat für Gruppendiskussionen eine Interraterreliabilität zwischen .61 und .84 vor einer Aussprache der Beobachter ermittelt. Gatewood et al. (1990) haben für Gruppendiskussionen Werte zwischen .69 und .99 nach Aussprache der Beobachter ermittelt.

1.3.2 Prognostische Validität

Der größte Teil der Forschungsarbeiten zur prognostischen Validität des AC wurde in den 60´er und 70´er Jahren verfasst. Die wohl bekannteste Studie, die auch einen Grundstein für die Verbreitung der AC-Methode darstellte, ist die so genannte Management Progress-Studie (Bray & Grant, 1966). 422 Führungsnachwuchskräfte der Firma AT&T wurden von Psychologen im Rahmen eines umfangreichen AC bezogen auf 25 Anforderungen eingeschätzt. Zusätzlich gaben die Beobachter eine Prognose ab, ob die Teilnehmer in das mittlere Management aufrücken würden. Diese Prognose wurde nach acht Jahren überprüft. Die Besonderheit dieser Studie bestand darin, dass die Ergebnisse des AC weder den Vorgesetzten noch den Teilnehmern mitgeteilt wurden, so dass es keine Kontamination im Sinne einer *self fulfilling prophecy* geben konnte. Nach 8 Jahren ließ sich ein Validitätskoeffizient von $r = .46$ errechnen. Zwar wurde u.a. kritisiert, dass das gewählte Kriterium „Aufstieg ins mittlere Management“ nicht sehr geeignet sei, da sich bei einer Überprüfung nach 16 Jahren zeigte, dass die meisten der Beurteilten entsprechend aufgestiegen waren, das Kriterium also nicht mehr ausreichend Varianz aufweist. Dennoch wurde diese Studie als eindrucksvoller Nachweis für die Validität des AC gewertet (vgl. Kleinmann, 1997).

Der Schwerpunkt der Forschung zur prognostischen Validität lag in den 70´er Jahren, danach verlagerte sich der Schwerpunkt stärker zu Fragen der Konstruktvalidität. Einen guten Überblick über die Vielzahl der Validitätsstudien geben Cohen, Moses & Byham (1974) und Thornton & Byham (1982) für den amerikanischen Raum, Jeserich (1981) für den deutsch-

sprachigen Raum. Die dort zitierten Validitätskoeffizienten schwanken zwischen -.25 und .78. Jeserich (1981, S. 323) hat folgenden tabellarischen Überblick zusammengestellt:

Tabelle 1-1: Überblick über veröffentlichte Validitätsstudien zum AC-Verfahren nach Jeserich (1981)

Quelle	Unternehmen	Anzahl d. Kandidaten	Kriterien	Validität
Bray & Grant (1966)	AT&T	422	Hierarchieebene, Gehaltszuwachs	.44 - .71
Bentz / 1966)	Sears & Roebuck	51	Kombination aus Vorgesetztenbeurt., Selbstbeurt., Gehaltszuwachs	.29 - .51
Bray & Campbell (1968)	AT&T	78	Spezielle Leistungsbeurt. am Arbeitsplatz	.51
Dodd & McNamara (1968)	IBM	64	Hierarchieebene	.38
Wollowick & McNamara (1969)	IBM	94	Zuwachs an Verantwortung	.37
Hinrichs (1969)	IBM	47	Hierarchieebene	.26
Carleton (1970)	SOHIO	122	Vorgesetztenbeurt., 30 - 60 Monate später	.65
Moses (1971)	AT&T	5943	Beförderung	.46
Moses & Boehm (1975)	AT&T	4848	Hierarchieebene	.37
Huck & Bray (1976)	AT&T	126	Vorgesetztenbeurteilung Aufstiegspotential	Weiß: .41, Schwarz: .35 .59 - .54
Hinrichs (1978)	IBM	30	Hierarchieebene	.65

Einige aktuellere Ergebnisse werden in der folgenden Tabelle dargestellt:

Tabelle 1-2: Übersicht über weitere Studien zur prognostischen Validität

Quelle	Studie	Ergebnis
Laubsch (2001)	Studie zu Assessment Center und Situativem Interview für Fluglotsenbewerber	Keine prognostische Validität für das AC
Thornton et al. (1992)	Metaanalyse auf der Basis von 50 Validitätsstudien	korrigierte mittlere Validität: $r = .37$
Wolf et al. (1995)	Validitätsüberprüfung am Beispiel von AC der Schweizerischen Kreditanstalt	$r = .50$ bis $r = .68$ je nach Bestimmung des Kriteriums

Diese Tabellen veranschaulichen, ohne Anspruch auf Vollständigkeit zu erheben, die Verschiedenartigkeit der einzelnen Studien hinsichtlich Teilnehmerzahlen, Kriterien und Validitätskoeffizienten. Die entscheidende Frage, von welchen Faktoren die Variation des Validitätskoeffizienten abhängt, wird mit Hilfe von Metaanalysen zu beantworten versucht.

Kleinmann zitiert eine Studie von Schmidt & Hunter (1978), die im Ergebnis feststellt, dass häufig ein Großteil der Variation des Validitätskoeffizienten auf folgende vier Artefakte zurückgeführt werden kann (Kleinmann, 1997, S.22):

- Größe der Stichprobe
- Mangelnde Reliabilität des Prädiktors
- Mangelnde Reliabilität des Kriteriums
- Eingeschränkter Messbereich des Kriteriums

Thornton et al. (1992) veröffentlichten die bisher umfangreichste Metaanalyse, in die 50 Validitätsstudien mit insgesamt 107 Validitätskoeffizienten eingingen. Sie haben einen Schätzwert für die von Artefakten bereinigte prognostische Validität ermittelt:

„Ergebnis der Studie von Thornton et al. (1992) ist, dass der beste Schätzwert für die prognostische Validität (korrigierte mittlere Validität) von Assessment-Centern bei $r = .37$ liegt, mit einer Varianz von $.017$. Hieraus kann geschlossen werden, dass die Validität von Assessment-Center Verfahren generalisierbar ist. ... Damit kann als gesichert gelten, dass Assessment-Center Verfahren im Vergleich zu anderen eignungsdiagnostischen Verfahren eine vergleichsweise hohe prognostische Validität aufweisen.“ (Kleinmann, 1997, S. 22)

Bei der Suche nach Moderatorvariablen für die prognostische Validität des AC kamen sie zu dem Ergebnis, dass es sich positiv auf die Höhe der Validität auswirkt, wenn:

- die Beurteiler Psychologen statt Manager waren,
- Peer-Ratings der anderen AC-Kandidaten einbezogen wurden,
- der prozentuale Anteil weiblicher Kandidaten hoch war,
- die Studien methodologisch fundiert waren. (Thornton et al. (1992) zitiert nach Kleinmann (1997))

Einen weiteren Einflussfaktor auf den Validitätskoeffizienten stellt der Zeitraum zwischen AC und Überprüfung des Kriteriums dar. Jeserich (1981) fasst die Ergebnisse verschiedener

Studien mit der Aussage zusammen, dass die Vorhersagen umso besser bestätigt werden konnten, je später das Kriterium gemessen wurde.

Kritisch anzumerken ist, dass in allen Validitätsstudien nur die noch in der Organisation Verbliebenen berücksichtigt werden, nicht jedoch die inzwischen Ausgeschiedenen.

1.3.3 Konstruktvalidität

In der Logik des AC bedeutet Konstruktvalidität, dass zwischen den Urteilen, die über einen Teilnehmer für dieselbe Dimension in verschiedenen Aufgaben vergeben werden, ein hoher Zusammenhang zu erwarten ist, während für die Urteile, die im Rahmen einer Aufgabe für verschiedene Dimensionen vergeben werden, ein geringer Zusammenhang erwartet wird. Eine gängige Methode zur Bewertung der Konstruktvalidität ist die von Campbell und Fiske (1959) vorgeschlagene Multi-Trai-Multi-Method-Matrix (MTMM). Die Übereinstimmung der Messwerte eines bestimmten Merkmals in verschiedenen Methoden wird dabei als konvergente Validität bezeichnet, sie sollte möglichst hohe Werte annehmen. Die Übereinstimmung der Messwerte verschiedener Merkmale, die im Rahmen einer Methode erhoben werden, wird als diskriminante Validität bezeichnet. Dieser Koeffizient sollte möglichst gering ausfallen (vgl. Kap. 3.3.2)

Der erste Versuch einer Konstruktvalidierung auf Basis der Einzelbeurteilungen im Rahmen eines AC wurde von Sacket und Dreher (1982) unternommen. Sie haben die Ergebnisse aus drei Unternehmen mit insgesamt über 500 Teilnehmern an Assessment Centern untersucht und dabei festgestellt, dass die geschilderten Erwartungen bezüglich konvergenter und diskriminanter Validität nicht erfüllt wurden. Die erwarteten Urteilsdimensionen ließen sich in der Analyse der AC-Ergebnisse nicht abbilden. In einem Unternehmen betrug die durchschnittliche Korrelation zwischen Eigenschaften mit der gleichen Bezeichnung lediglich $r = .07$, während die Einstufungen verschiedener Merkmale innerhalb einer Methode mit $r = .64$ korreliert waren.

Dieses Ergebnis wurde sinngemäß in zahlreichen weiteren Studien repliziert (Archambeau, 1979; Bycio, Alvares & Hahn, 1987; Neidig, Martin & Yates, 1979; Neidig & Neidig, 1984; Neubauer, 1989; Robertson, Gratton & Sharpley, 1987; Russel, 1987). In der Studie von Russel (1987) beträgt die konvergente Validität durchschnittlich $.25$, die diskriminante Validität lag um $.50$.

„Weder konvergente noch diskriminante Validität ist in der Untersuchung von C.G. Russel (1987) gegeben. Die Anforderungsdimensionen werden also nicht transsituational erfasst (keine konvergente Validität) und sind auch nicht innerhalb der einzelnen Übungen hinreichend differenzierbar (keine diskriminante Validität)“ (Kleinmann, 1997, S. 32 f.).

Die Frage, in wie weit in den AC-Übungen tatsächlich unterscheidbare und abgrenzbare Konstrukte bewertet werden, konnte also nur sehr unbefriedigend beantwortet werden.

Ein weiterer Ansatz zur Überprüfung der Konstruktvalidität besteht in der Korrelation der einzelnen Urteile mit Außenkriterien, also mit anderen eignungsdiagnostischen Verfahren mit bekannter Konstruktvalidität. Scholz und Schuler (1993) führten eine Metaanalyse durch, in

die 55 derartige Studien eingingen. Die Persönlichkeitskonstrukte wurden nach kognitiven und nicht-kognitiven Persönlichkeitsvariablen differenziert. Die Metaanalyse ergab, dass allgemeine Intelligenz mit $r = .33$ als bester Prädiktor für das Abschneiden im AC dient. Soziale Kompetenz, Leistungsmotivation, Selbstvertrauen und Dominanz korrelieren mit Werten zwischen $r = .23$ und $r = .31$ mittelhoch mit dem AC-Ergebnis.

Kleinmann (1997) fasst den Forschungsstand zur Konstruktvalidität im AC in folgenden Punkten zusammen:

- Die intendierten Konstrukte werden mit Hilfe von AC-Verfahren in den meisten Studien zur Konstruktvalidität nur unzureichend erfasst.
- Es gibt empirisch belegt einzelne Variablen, die die konvergente Validität des AC beeinflussen.
- Bisher sind keine Variablen bekannt, die die diskriminante Validität des AC beeinflussen.
- Es sind Persönlichkeitskonstrukte bekannt, die mit dem AC-Ergebnis kovariieren.
- Es ist ungeklärt, warum die prädiktive Validität der AC-Verfahren vergleichsweise hoch ausfällt, obwohl die intendierten Konstrukte nicht erfasst werden.

Die empirischen Ergebnisse zum Nachweis der Konstruktvalidität sind also nach wie vor unbefriedigend. Insbesondere die Sicherstellung der diskriminanten Validität ist problematisch. Obwohl es also nicht gelingt, eine differenzierte Beurteilung verschiedener Konstrukte nachzuweisen, ist eine zufrieden stellende prognostische Validität des AC unbestritten.

1.3.4 Kontentvalidität

Die Zahl der empirischen Studien zur Kontentvalidität bleibt weit hinter der der Untersuchungen zur prädiktiven Validität zurück. Kernpunkt der Kontentvalidität ist die Frage nach der Angemessenheit und Beurteilbarkeit der Urteilsdimensionen. Hinter dem Konzept der Kontentvalidität verbirgt sich die Überprüfung des komplizierten Prozesses über eine Anforderungsanalyse die zu erhebenden Anforderungskategorien auszuwählen, konkrete Aufgaben zu schaffen, in denen diese Anforderungskategorien beobachtet und bewertet werden können und Merkmale festzulegen, mit denen die Anforderungskategorien operationalisiert werden können. Jeserich (1991) bietet ausführliche Beispiele, um diesen Prozess nachvollziehen zu können. Dabei treten in der Praxis Probleme auf durch die Vermischung von Konstrukten und Verhaltensweisen, von Verhaltensbedingungen und Verhaltenswirkungen. Soll die Qualität der in diesem Prozess getroffenen Entscheidungen überprüft werden, so ist dies nur durch Ratingverfahren und Indexbildungen verschiedener Expertengruppen für jeden Entwicklungsschritt möglich. Reilly und Henry (1990) zum Beispiel beschreiben einen solchen Prozess für die Entwicklung einer Verhaltenscheckliste. Sie lassen in mehreren Schritten 10 Beurteiler Items zu den Dimensionen Teamarbeit, Führung, Problemlösung, Arbeitsorientierung, Detailbewusstsein, Umgang mit Anweisungen, Planung und Organisation sowie Sicherheitsbewusstsein auswählen. Die endgültige Checkliste weist dann noch 273 Verhaltensweisen auf. An diesem Beispiel wird auch deutlich, dass Fragen der Inhaltsvalidität natürlich auch Bestandteil einer Konstruktvalidierung sind und sich nicht eindeutig voneinander trennen lassen.

Schuler (1989) zitiert das Ergebnis einer Studie von Ostroff und Schmitt (1987), wonach trotz aufwendig geprüfter befriedigender inhaltlicher Validität die Vorhersagegültigkeit eines AC zur Auswahl von Schulleitern sehr gering war. Schuler zieht daraus den Schluss, dass

„... die beiden Validierungsstrategien geeignet sind, unterschiedliche Fragen zu beantworten, und dadurch zu unterschiedlichen Zwecken einzusetzen sind.“
(Schuler, 1989, S. 238)

1.3.5 Soziale Validität

Die empirische Evidenz zur sozialen Validität ist recht gering. Jeserich (1991) stellt die Ergebnisse einer Studie von Dodd (1977) vor. 489 IBM-Mitarbeiter wurden zu einem AC befragt, an dem sie teilgenommen hatten. Es wird hier unterstellt, dass die Akzeptanz eines Verfahrens umso größer ist, je klarer ein Anforderungsbezug zu erkennen ist. Die u.a. gestellte Frage vermischt in gewisser Weise Kontentvalidität und Soziale Validität und ist den von Schuler vorgeschlagenen Dimensionen der sozialen Validität nur indirekt zuzuordnen. Die Frage lautete: „Misst ein AC Qualitäten, die für das Management notwendig sind?“

Tabelle 1-3: Misst ein AC Qualitäten, die für das Management notwendig sind? (zit. nach Jeserich, 1991, S.315)

	Überhaupt nicht	Nicht sehr viel	In mittelmäßigem Ausmaß	In großem Ausmaß	In sehr großem Ausmaß
Frühere Beobachter	0 %	5 %	57 %	35 %	3 %
Andere Manager	0 %	10 %	61 %	25 %	4 %

Diese Tabelle steht stellvertretend für die Ergebnisse verschiedener Studien (vgl. Jeserich, 1991) die durchgängig eine positive Bewertung des AC-Verfahrens ergeben haben. Interessanterweise ist diese Einschätzung sowohl bei Teilnehmern zu finden, die gut bewertet wurden, als auch mit nur einer geringen Abschwächung bei Teilnehmern, die nicht so positiv abgeschlossen haben.

Scholz (1994) fasst die Ergebnisse verschiedener Studien zur sozialen Validität des Assessment Center nach den von Schuler vorgeschlagenen Dimensionen folgendermaßen zusammen:

- Das Verfahren berücksichtigt die Information und Partizipation der Teilnehmer.
- Im Allgemeinen erfolgt eine ausführliche Urteilkommunikation.
- Die Aufgabenstellungen werden als transparent und realitätsnah erlebt, die Beurteilungsmerkmale dagegen eher nicht.
- Das Interview erweist sich als beliebter.
- Erfolgreiche Teilnehmer bewerten die Gruppendiskussion am positivsten, gefolgt von Vortrag, Postkorb und schriftlicher Aufgabe.
- Bei schlechter beurteilten Teilnehmern kehrt sich die Reihenfolge um.
- Das AC-Ergebnis wird überwiegend als zutreffend und nützlich beurteilt.

Im Rahmen der vorliegenden Studie steht die Konstruktvalidität im Mittelpunkt der Untersuchung. Auch wenn für ein Verfahren der Personalauswahl in erster Linie die kriteriumsbezogene, und hier insbesondere die prognostische Validität, von Interesse ist, ist es eine Frage der Konstruktvalidität, ob die erhobenen Daten als Ausgangsbasis für eine erste Karriereplanung bzw. für individuelle Personalentwicklungsmaßnahmen sinnvoll eingesetzt werden können.

1.4 Einflussfaktoren auf die Gütekriterien von Gruppenaufgaben

Entsprechend der Zielsetzung dieser Arbeit geht es um den Einsatz von Gruppenaufgaben zur Personalauswahl. Dies findet in der Regel im Rahmen eines AC statt. Lattmann (1989) stellt die Anforderungen an die Gestaltung von Gruppenaufgaben zusammen, die zur Erreichung des Prüfungszweckes erfüllt werden sollten.

Zunächst muss die mittels der Gruppenaufgabe zu prüfende Verhaltensform für die Lösung der Gruppenaufgabe erheblich sein. Eine rein kompetitiv gestaltete Gruppenaufgabe ist zur Prüfung empathischer Fähigkeiten z.B. wenig sinnvoll; denn um eine Wettbewerbssituation in einer Gruppe erfolgreich zu meistern, ist Empathie in der Regel nur bedingt erfolgsrelevant. Diese Anforderung ist dem Bereich der Inhaltsvalidität zuzurechnen.

Weiter sollte ein unterschiedlich ausgeprägtes Verhalten sich in unterschiedlich guten Ergebnissen niederschlagen. Diese Forderung lässt sich relativ gut auf leistungsorientierte Gruppenaufgaben anwenden. Auf die Beurteilung von Gruppenaufgaben, in denen es um die Bewertung von Persönlichkeitsmerkmalen geht, lässt sich diese Forderung nur schwer übertragen, da sich in der Regel keine direkte Kopplung zwischen dem Ergebnis einer Aufgabe und der Ausprägung von Persönlichkeitsmerkmalen herstellen lässt.

Drittens fordert Lattmann die Vergleichbarkeit der Ergebnisse der Teilnehmer, womit er zum einen auf die Wahrung der Durchführungsobjektivität abhebt, zum anderen auf die Beachtung der Bewerbergerechtigkeit. Für die OPZ ergibt sich daraus u.a. der Verzicht auf militärische Aufgabenstellungen, da diese die noch ungedienten Bewerber systematisch benachteiligen würden.

Schließlich formuliert er, wenn auch nur indirekt, die Anforderung transparenter Beurteilungsmerkmale. Die Teilnehmer sollen den Aufgabeninhalt in Übereinstimmung mit dem ihm zugrunde liegenden Beurteilungszweck wahrnehmen können. Nur so können Verzerrungen der Ergebnisse vermieden werden, die aus der unterschiedlich ausgeprägten Fähigkeit der Teilnehmer entstehen, die Beurteilungsdimensionen richtig zu erraten.

Die Problematik des „missing-link“ zwischen dem Phänomenbereich „Arbeit“ und dem Phänomenbereich „Personenmerkmal“ wurde bereits angesprochen. Lattmann (1989) konkretisiert diesen Aspekt mit zwei Anmerkungen. Zum einen ist der Erlebnisbezug der Teilnehmer eine nicht zu kontrollierende Variable. Aufgrund der simulativen Natur der Gruppenaufgaben weicht der Erlebnisbezug der einzelnen Teilnehmer unterschiedlich stark vom Erleben einer möglichen Realsituation ab. Es ist also offen, in wie weit ein Teilnehmer die in der Gruppenaufgabe erhobenen Personenmerkmale auch in der Realsituation zeigen wird. Zum anderen

lassen sich die Anforderungen der Zielposition in der Gruppenaufgabe nicht immer inhaltsgültig wiedergeben. Insbesondere der zeitliche Aspekt (Dauer einer Anforderung) und die Komplexität der Anforderungen weichen in der Gruppenaufgabe zwangsläufig mehr oder weniger stark von der Realsituation ab. Auch dies führt zu der ungeklärten Frage, ob und in wie weit der Teilnehmer das in der Simulation für kurze Zeit gezeigte Personenmerkmal, z.B. Durchsetzungsfähigkeit, auch in der Realsituation über lange Zeit zeigen kann. Hier kann nur der Schluss weiterhelfen, dass vermutlich ein Teilnehmer, der in der Gruppenaufgabe dem Ideal entspricht, die besten Voraussetzungen mitbringt, auch in der Realsituation den Anforderungen gerecht zu werden.

Neben der Konzeption der Gruppenaufgaben birgt der Beobachtungsprozess selbst eine Fülle möglicher Fehlerquellen, die die Qualität der Ergebnisse beeinflussen können. Der Vollständigkeit halber sollen daher die häufigsten Beobachterfehler hier kurz benannt werden. Jeserich (1991) zählt die gängigsten 4 Varianten von Beobachtungsfehlern auf:

- Der Primacy-Effekt: eine spontane unmittelbare Einschätzung in der Phase der ersten Kontaktaufnahme, die keinem Beurteilungsmerkmal zuzuordnen ist, die Urteile insgesamt aber dennoch positiv oder negativ beeinflusst.
- Der Halo-Effekt: durch einen einzigen negativen oder positiven Eindruck werden weitere Eigenschaften oder Verhaltensweisen überstrahlt.
- Der Kontrast-Effekt: die Bewertung eines Kandidaten wird durch das Auftreten des unmittelbar vorher erschienenen Kandidaten beeinflusst und die Orientierung an einer vorher festgelegten Idealnorm wird dadurch erschwert.
- Das Ähnlichkeitsphänomen: Menschen, die dem Beobachter im Verhalten oder anderen Bereichen ähneln, werden positiver beurteilt.

In der Literatur lassen sich über diese Darstellung hinausgehende unterschiedliche Unterteilungen der Beobachterfehler finden (z.B. Fisseni & Fennekels, 1995, oder Kleinmann, 1997). Kleinmann (1997) verweist auf die zu Grunde liegenden theoretischen Konzepte und Erklärungsansätze für die Vielzahl der möglichen Beobachterfehler. Er nennt attributionstheoretische Ansätze (Meyer & Försterling, 1993), Forschung zu Urteilsheuristiken (Strack, 1985), schemageleitete Informationsverarbeitungsprozesse (Schwarz, 1985), implizite Persönlichkeitstheorien (Krystofiak, Cardy & Newman, 1988), Vorurteilsforschung (Six, 1987) und theoretische Konzepte zur Erwartungsbildung (Bierhoff, 1986).

Vor diesem Hintergrund wird immer wieder die Notwendigkeit sorgfältiger Beobachtertrainings betont. Welchen Einfluss die Beobachterschulung genau hat, wurde von Scholz (1994) untersucht. Er fasst die Ergebnisse verschiedener Studien zum Thema Beobachtertraining für AC folgendermaßen zusammen:

„Die Vermittlung der Methodik scheint demnach unabdingbar zu sein, aber über einen bestimmten Zeitpunkt der Schulung hinaus ist anscheinend kaum mehr ein Zugewinn an Messqualität möglich.“ (Scholz, 1994, S. 15)

Dies deckt sich mit einer weiteren Studie von Latham, Wexley und Pursell (1975), deren Ergebnisse Jeserich (1991) vorstellt. Mit Hilfe eines neunstündigen Trainings konnte erreicht werden, dass Beobachter sechs Monate nach dem Training keinen der o.g. Fehler begingen,

während in einer Kontrollgruppe, die kein Training erhalten hatte, alle Fehler bis auf den des „ersten Eindrucks“ erneut auftraten. Wesentlich ist also einerseits die Feststellung, dass mit einem gezielten Training diese Beobachterfehler minimiert werden können und damit entscheidende Verbesserungen der Aussagegenauigkeit erreicht werden können. Andererseits ist die Ausweitung des Beobachtertrainings kein Allheilmittel zur Verbesserung der Urteilsqualität.

Neben einem Training der Beobachter ist der zweite wichtige Ansatzpunkt zur Erreichung einer ausreichenden Objektivität und Beurteilungsqualität die strikte Trennung von Beobachtung und Bewertung. Auf diese Weise soll der Schritt von einer individuellen, subjektiven und spontanen Beurteilung hin zu einer analytischen Beurteilung der dem Anforderungsbild entsprechenden Verhaltensformen ermöglicht und unterstützt werden.

Im Folgenden sollen einige Ergebnisse aus Studien vorgestellt werden, die gezielt verschiedene Moderatorvariablen der Gütekriterien untersucht haben.

- **Beobachterdiskussion:** Urteile, die nach einer Beobachterdiskussion abgegeben wurden, erwiesen sich in einer Studie von Jones (1981) als etwas reliabler als Urteile, die ohne eine Beobachterdiskussion abgegeben wurden (zitiert nach Scholz, 1994).
- **Qualifikation der Beobachter:** Es wurde die Frage untersucht, ob die Urteilsqualität durch die Hinzuziehung von Psychologen verbessert werden kann. Lattmann (1989) kommt zu dieser Frage zu keinem klaren Ergebnis. Während es Studien gibt (Byham, 1970, zitiert nach Lattmann, 1989), die keine Unterschiede feststellen, gelangten Thornton et al. (1992) zu der Aussage, dass durch die Einbeziehung von Psychologen die Genauigkeit der Beurteilung verbessert wurde.
- **Anzahl der Beobachtungsdimensionen:** Gaugler und Thornton (1989) (zitiert nach Kleinmann, 1997) ermittelten, dass eine Verringerung der zu beurteilenden Dimensionen sich positiv auf die konvergente Validität auswirkte, jedoch keine Verbesserung der diskriminanten Validität bewirkte. Sie vermuteten, dass Beobachter überlastet sind, wenn sie zu viele Dimensionen beachten müssen.
- **Einsatz einer Verhaltenscheckliste:** Reilly und Henry (1990) haben durch den Einsatz einer Verhaltenscheckliste eine deutliche Verbesserung der durchschnittlichen konvergenten Validität (von .24 auf .43) und eine leichte Verbesserung der durchschnittlichen diskriminanten Validität (von .47 auf .41) erreichen können. Auch sie argumentierten mit einer kognitiven Überlastung der Beobachter, die durch den Einsatz der Verhaltenscheckliste entsprechend reduziert wurde. Zu bemerken ist hier die aufwendige Vorgehensweise zur Erstellung der Verhaltenscheckliste. Zehn trainierte Beobachter bewerteten im Verlauf eines 2-tägigen AC 120 Teilnehmer auf acht verschiedenen Dimensionen. Die Studie bezieht sich auf die Bewertung von zwei verschiedenen Gruppenaufgaben. Zu den acht Dimensionen wurden Verhaltensweisen gesammelt, die dazu führten, dass die Beobachter einen Kandidaten besser oder schlechter bewerteten. Auf diese Weise entstand eine Liste von 348 Verhaltensweisen. Im Anschluss wurden die Verhaltensweisen den Beobachtern erneut vorgelegt und von ihnen den einzelnen Dimensionen zugeordnet. Bei einer 80-prozentigen Übereinstimmung wurde eine Verhaltens-

weise in die endgültige Verhaltenscheckliste aufgenommen, die schließlich noch 273 Verhaltensweisen enthielt. Diese Verhaltenscheckliste wurde in einem weiteren AC mit abgewandelten Aufgaben eingesetzt, bei dem 235 Teilnehmer bewertet wurden. Die Besonderheit der Verhaltenscheckliste ist darin zu sehen, dass den Beobachtern die anspruchsvolle Aufgabe, im Verlauf des Gruppengeschehens zu entscheiden, welche Verhaltensweisen für die Zieldimensionen relevant sind und welcher Zieldimension sie zuzuordnen sind, durch die Vorarbeiten bei der Entwicklung der Verhaltenscheckliste abgenommen wurden. Mit der Erreichung eines Koeffizienten für diskriminante Validität, der unter dem Koeffizienten der konvergenten Validität liegt, ragt das Ergebnis dieser Studie deutlich aus der Vielzahl der Studien zur Konstruktvalidität heraus.

- **Unabhängigkeit und Beobachtbarkeit der Dimensionen:** Kleinmann et al. (1995) haben in 20 Assessment Centern mit 115 Teilnehmern und 15 geschulten Beobachtern den Einfluss der Unabhängigkeit und der Güte der Beobachtbarkeit auf die konvergente und diskriminante Validität untersucht. Sie vermuten, dass der Umstand der zeitgleichen Erfassung mehrerer Anforderungsdimensionen eine bedeutsame Rolle bei der Erklärung der geringen diskriminanten Validität von AC-Dimensionen spielen könnte.

„Werden bei einer Gruppendiskussion Durchsetzungsfähigkeit und Argumentationsvermögen erfasst, so sind diese beiden Konstrukte bei zeitgleicher Messung nicht unabhängig voneinander.“ (Kleinmann et al. 1995, S. 23)

Sie kamen zu dem Ergebnis, dass die Unabhängigkeit der Dimensionen (bewertet nach einem Expertenrating) einen signifikanten Einfluss auf die Höhe der diskriminanten Korrelationskoeffizienten hat. Die Beobachter gaben dann je Teilnehmer für Anforderungsdimensionen unterschiedliche Ratings ab, wenn die Anforderungsdimensionen unterschiedliches Verhalten ermöglichten. Darüber hinaus geben sie zu bedenken, dass aus der differentiellen Psychologie bekannt ist, dass ohnehin viele Persönlichkeitsmerkmale miteinander korrelieren. Ihre zweite Hypothese, dass die Güte der Beobachtbarkeit einer Anforderungsdimension in verschiedenen Übungen sich auf die konvergente Validität auswirkt, konnten sie lediglich für die Dimension Durchsetzungsfähigkeit bestätigen.

- **Transparenz der Anforderungsmerkmale für die Teilnehmer:** Wie weiter oben bereits geschildert, hat Lattmann die Forderung nach transparenten Beurteilungsmerkmalen gestellt. Kleinmann (1991) hat mit einer Studie die empirische Bestätigung für die Berechtigung dieser Forderung geliefert. Bei den Teilnehmern, die in beiden Übungen der Studie die identischen Anforderungsmerkmale richtig erkannt hatten, war die konvergente Validität ihrer Beurteilungen signifikant höher als bei Teilnehmern, die die Anforderungsmerkmale nur in einer oder in gar keiner Übung richtig erkannt hatten.

1.5 Formen und Anwendungsfelder von Gruppenaufgaben

Zunächst ist zwischen den beiden wesentlichen Anwendungsfeldern der Diagnose und des Trainings zu unterscheiden. Gruppenverfahren werden im Trainingsbereich mit den verschiedensten Zielsetzungen und Inhalten eingesetzt. Dieser Anwendungsbereich soll hier

nicht vertieft werden. Im Weiteren beziehen sich die Ausführungen auf Aufgaben, die zu diagnostischen Zwecken eingesetzt werden.

Unter dem Begriff der Gruppenverfahren wird eine Reihe von unterschiedlichen Aufgabenformen subsumiert. Eine Übersicht aus der Literatur zusammen zu stellen, wird dadurch erschwert, dass die Form der Aufgabe nur in wenigen Fällen im Mittelpunkt des Interesses steht. Im Zusammenhang mit der AC-Literatur werden die Aufgaben vor allem in nicht-situative und situative Aufgaben unterschieden, die dann sowohl in Einzelarbeit als auch in der Gruppe bearbeitet werden. Thematisch können Gruppenaufgaben gezielt zum Umgang mit den verschiedensten Fachthemen konzipiert werden oder es wird die Diagnose der sozialen Interaktionen in den Vordergrund gestellt, während die inhaltliche Thematik reines Mittel zum Zweck darstellt. Die am häufigsten zu findenden Differenzierungen von Gruppenaufgaben sind: Gruppendiskussion (führerlos, mit oder ohne Rollenvorgabe, mit expliziter Führungsrolle), Rollenspiel, Planspiel, Konstruktionsübung, Präsentation. Diese Auflistung verdeutlicht, dass die Differenzierung nach eher pragmatischen Kriterien vorgenommen wird und die Abgrenzungen auch verschwimmen können, bzw. nicht eindeutig sind.

Maukisch (1989) stellt eine Möglichkeit vor, diagnostische Instrumente der Personalbeurteilung in Gruppen (Prädiktorklassen) einzuteilen, die sich stärker an wissenschaftlichen denn an pragmatischen Kriterien orientiert. Anhand dieser Systematik lässt sich die Vielzahl der Ausgestaltungs- und Einsatzmöglichkeiten von Gruppenverfahren veranschaulichen.

- **Inhalt und Konzeptweite des angezielten diagnostischen Gegenstandes** bzw. Konstruktes: Hier können zunächst sozial-interaktive Zielsetzungen von fachlich-inhaltlichen Zielsetzungen unterschieden werden. Es ist bei der Bewertung darauf zu achten, dass sich diese beiden Bereiche nicht gegenseitig beeinflussen. Gerade für die Auswahl von Führungspersonal ist diese Trennung jedoch nur theoretisch vollständig aufrecht zu erhalten. Wie ausgeprägt ein Teilnehmer z.B. in der Lage ist, Durchsetzungsfähigkeit zu zeigen, hängt in der Interaktion mit seinen Mitbewerbern immer auch von den inhaltlichen Argumenten ab, die diskutiert werden.
- **Art und Inhalt der Primärdaten:** In Gruppenaufgaben sind dies in der Regel Verhaltensbeobachtungen, sie können aber auch ergänzt werden durch quantifizierbare Ergebnisse der Gruppenarbeit sowie Selbst- und Peerratings der Teilnehmer.
- **Zahl, Art und Aufeinanderfolge vermittelnder Medien:** Hier sind der Phantasie fast keine Grenzen gesetzt. Insbesondere in den mittlerweile meist computergestützten Planspielen werden außerordentlich komplexe Simulationen geschaffen, in denen die Teilnehmer agieren müssen. Aber auch in Konstruktions- oder Kreativitätsaufgaben, die in Gruppen bearbeitet werden, kann und wird alles verwendet, was den pragmatischen Anforderungen der Durchführbarkeit genügt. Eine weitere Möglichkeit stellt der Einsatz von Rollenspielern dar. Dies können sowohl andere Teilnehmer als auch extra geschulte Personen ohne diagnostischen Auftrag sein. Häufig wird aus Effizienzgründen in Assessment Centern den Beobachtern auch eine Doppelfunktion als Rollenspieler und Beurteiler zugewiesen.
- **Art und Weise der Transformation der Daten:** Zu diesem Punkt gibt es in der AC-Forschung bereits einige Ansätze. Es wurde deutlich, dass es einen Unterschied macht,

ob die Urteile der Beobachter vor oder nach einer Diskussion vergeben werden, bzw. ob die Urteile jeweils nach den einzelnen Aufgaben oder am Ende des AC dimensionsweise vergeben werden (Silverman et al., 1986; Kleinmann, 1997). Weitere wichtige Fragen zu diesem Kriterium betreffen die Intransparenz des Urteilsprozesses, also wie aus den Einzelitems, bei Gruppenverfahren in der Regel beobachtete Verhaltensweisen, schließlich ein quantifiziertes Endresultat gefunden wird, und welchen Einfluss hierauf unterschiedlich gestaltete Beobachtungs- und Bewertungsbogen nehmen können.

„Die Forschungen zur Urteilsbildung haben gezeigt, dass die Integration und Kombination multipler Ausgangsdaten bei der Bildung von Urteilen und Vorhersagen über Menschen meist sehr simplen additiven Regeln folgt, dass nur sehr wenig Ausgangsinformation genutzt wird, dass die Beurteiler in ihrem Vorgehen nicht sehr konsistent sind, ihre eigenen Strategien und Gewichtungen nicht angemessen wiedergeben können und dass es bedeutende Unterschiede in den Strategien der Informationsnutzung auch bei langjähriger gleichartiger Erfahrung der Experten gibt (Brehmer, 1987, zitiert nach Maukisch, 1989).“

Bezüglich der Ermittlung eines Endresultates hat sich die statistisch-psychometrische Datenkombination der freien „klinischen“ Urteilsbildung für die Voraussage von Außenkriterien in der Regel als überlegen erwiesen.

Maukisch (1989) vermutet, dass mit einer Forschung, die sich an Prädiktorklassen orientiert, die nach einer solchen Systematik gebildet werden, ein größerer Informationsgehalt erzielen ließe, als dies bisher mit den üblichen Meta-Analysen erreicht wird.

„Die bislang üblichen Vergleiche großer pragmatischer Gruppen haben wenig Informationswert, weil in ihnen die oben angedeuteten Dimensionen vermengt und überlappt sind und die interne Varianz im Grunde keinen echten Vergleich erlaubt. ... Bei präziser mehrdimensionaler Klassifikation und Einbeziehung ökonomischer Kriterien könnten sich überraschende Ergebnisse zeigen bzw. ein jedenfalls differenzierteres Bild der Prognosekraft und Verwendbarkeit der verfügbaren Prädiktoren, in dem sich die bisherigen „Sieger-“ und „Verlierer“-positionen relativieren würden.“ (Maukisch, 1989, S. 253)

Es scheint jedoch keine Studien zu geben, die sich dem Thema der Gruppenaufgaben vor dem Hintergrund einer solchen oder vergleichbaren Systematik angenommen haben. Es ist im Gegenteil häufig der Fall, dass die Begriffsklärung bezüglich der Gruppenaufgaben über ein eher alltagssprachliches Verständnis nicht hinausgeht und eine Beschreibung und Konkretisierung der Rahmenbedingungen bezüglich der o.g. Kriterien nicht stattfindet.

Aus diesem Grund sollen die gängigen Formen der Gruppenaufgaben im Folgenden kurz beschrieben werden. Nach den üblichen pragmatischen Kriterien werden vor dem Hintergrund einer diagnostischen Anwendung die Vor- und Nachteile skizziert sowie definitorische Unklarheiten angesprochen.

1.5.1 Gruppendiskussion

Die Aufgabenbezeichnung Gruppendiskussion stellt den am weitesten gefassten Begriff dar. Bei der Verwendung dieses Begriffs wird offen gelassen, ob es sich um eine abstrakte oder konkrete Themenstellung handelt, ob es Rollenvorgaben gibt, ob eine Diskussionsleitung

eingesetzt wird oder nicht, ob das Thema kompetitiv oder kooperativ bearbeitet wird, ob eine Einigung erzielt werden soll, ob eine Problemlösung erarbeitet werden soll oder ob es eine Musterlösung gibt. Und schließlich können auch ein Rollenspiel, ein Planspiel oder eine Konstruktionsübung Phasen der Gruppendiskussion enthalten.

Historisch gesehen wurden führerlose Gruppendiskussionen zunächst in Forschungszusammenhängen eingesetzt, um Fragen zur Entwicklung von Führung zu untersuchen. Bass (1954) gibt eine Übersicht über die ersten Anwendungen der führerlosen Gruppendiskussion mit diagnostischer Zielsetzung, die in den 20´er und 30´ Jahren im militärischen Kontext stattfanden. Die schnelle Ausbreitung führt er auf die administrativen Vorteile für eine größere Anzahl von Teilnehmern sowie auf die hohe Akzeptanz im Vergleich zu Papier-Bleistift-Tests zurück. Zudem wuchs zu dieser Zeit die Erkenntnis, dass Führung nicht nur von individuellen Voraussetzungen abhängt, sondern wesentlich auch von situativen Einflüssen. Während es einerseits als Vorteil wahrgenommen wird, dass das Instrument Gruppendiskussion inhaltlich mit den unterschiedlichsten Themen gefüllt werden kann und somit für unterschiedliche Zielgruppen einsetzbar ist, stellt Bass schon 1954 die geringe Standardisierung des Instruments hinsichtlich Anzahl der Teilnehmer, Dauer der Diskussion, Art der Problemstellung (wenn überhaupt), Anweisung an die Teilnehmer, Anzahl der Beobachter und Inhalt der Beobachtungen fest. Diese Feststellung besitzt nach wie vor Gültigkeit.

Bass fasst den Kenntnisstand, der die Basis für die weit verbreitete Anwendung der Gruppendiskussion darstellte, wie folgt zusammen: Leaderless group discussion (LDG)

„... have some validity as assessments of the tendency to display successful leadership, i.e., leadership potential. Other evidence suggests that the successful leadership behaviour observed in the LDG concerns primarily initiation of structure rather than consideration of the welfare of others“ (Bass, 1954, S. 488).

Eine Frage, die immer wieder vor allem im Zusammenhang mit Massendiagnostik relevant wird, ist die Frage nach Ergebnisverzerrungen durch Vorinformationen der Teilnehmer. Bass und Klubeck (1954) haben in einer Studie untersucht, in wie weit ein kurzes Coaching vor der Gruppendiskussion zu Veränderungen führt. Dabei stellten sie fest, dass die Teilnehmer, die im ersten Durchlauf ein relativ gutes Ergebnis erzielt hatten, signifikant von einem kurzen Coaching profitieren konnten, wenn auch die absolute Veränderung nicht sehr groß war. Teilnehmer, die im ersten Durchgang dagegen eher schlecht abgeschnitten hatten, zeigten durch das Coaching überhaupt keine Veränderung ihrer Ergebnisse.

Ein weiteres Problem, vor dem Beurteiler immer wieder stehen, ist der Einfluss der Redeanteile. Bass (1954) zitiert einige Studien, die die Korrelation zwischen Redezeit und positiver Bewertung untersuchten. Die Ergebnisse sind alle relativ hoch und schwanken zwischen .65 und .96. Er sieht das jedoch nicht problematisch, sondern als eine logische Konsequenz, setzt man voraus, dass im Grunde jeder Redebeitrag einen Versuch darstellt, das Gruppengeschehen zu beeinflussen, also Führung auszuüben.

Ausgehend von diesen „Wurzeln“ in der Führungsforschung wird die Gruppendiskussion mittlerweile zur Diagnose der verschiedensten sozial-interaktiven Merkmale eingesetzt. Neben der Führungsdimension sind hier ohne Anspruch auf Vollständigkeit Soziale Kompetenzen, Durchsetzungsverhalten, Konfliktverhalten, Planungsverhalten, systematisches Denken

und Handeln, Zielstrebigkeit, Ausdrucksfähigkeit und Argumentationsverhalten zu nennen. Eine ähnlich fundierte Grundlage, wie sie mit den Arbeiten von Bass für das Konstrukt der Führung vorliegt, sucht man jedoch für die Diagnose der weiteren Konstrukte vergeblich.

1.5.2 Rollenspiel

Auch die Aufgabenform des Rollenspiels weist die zur Gruppendiskussion beschriebene Vielzahl von Variationsformen auf. Bezüglich der Rollenvorgaben reicht die Bandbreite von einer schlichten Zuteilung einer Führungsrolle in einer Gruppendiskussion, zu der es keine weiteren Hinweise gibt, wie diese Führungsrolle auszufüllen sei, bis zu Rollenspielen mit detaillierten Rollenvorgaben für alle Beteiligten. Ob die erst genannte Form tatsächlich als Rollenspiel im engeren Sinne zu bezeichnen ist, sei dahingestellt. Wesentlich im Sinne der Diagnostik ist, dass die Zuweisung von Rollen die soziale Interaktion und Gruppendynamik wesentlich beeinflusst und daher bei der Beschreibung von Gruppenaufgaben ein wichtiges Differenzierungskriterium darstellt.

Eine weitere definitorische Unklarheit gibt es bezüglich der Anzahl der beteiligten Personen. Häufig in Assessment Centern eingesetzte Rollenspiele beziehen sich auf Dialogsituationen zwischen zwei Personen, wie z.B. ein Verkaufsgespräch oder ein Mitarbeitergespräch. Entsprechend der sozialpsychologischen Definition einer Gruppe wären diese Rollenspiele nicht mehr den Gruppenaufgaben zuzurechnen. Da für solche Dialogsituationen jedoch vor dem diagnostischen Hintergrund die gleichen Zielsetzungen, Probleme und Diskussionspunkte relevant sind wie für Aufgabenformen mit mehr als zwei Personen, wird eine solche Abgrenzung in diesem Zusammenhang als nicht zielführend betrachtet.

Die Vorteile des Rollenspiels im diagnostischen Zusammenhang werden in einer stärkeren Realitätsnähe und damit einer größeren Inhaltsvalidität gesehen. Für Personen, die im beruflichen Kontext in Interaktion treten, sind die Rollen und die an sie geknüpften Erwartungen in der Regel relativ klar, so dass mit einem Rollenspiel eine höhere Realitätsnähe erreicht werden kann als mit einer führerlosen Gruppendiskussion. Die entscheidende Frage lautet: In wie weit ist das im Rollenspiel gezeigte Verhalten äquivalent zum Verhalten in der realen Arbeitssituation? Torgrud und Holborn (1992) (zitiert nach Karkoschka, 1998) berichten von widersprüchlichen Ergebnissen zu dieser Frage. Sie führen diese widersprüchlichen Ergebnisse auf die sehr unterschiedlichen Untersuchungsdesigns der Studien zurück. Als wesentliche Einflussfaktoren auf die externe Validität bezeichnen sie Verfahrenshinweise, Durchführungsbestimmungen und Rollenspielpartner bzw. -format.

Ein weiterer Vorteil des Rollenspiels besteht in der besseren Möglichkeit zur Standardisierung der Aufgabe durch die Gestaltung der Rollen. Die Summe der Verhaltensmöglichkeiten der Teilnehmer ist durch die Rollenvorgabe eingeschränkt und damit nicht mehr so variabel wie in der Gruppendiskussion. Die Gestaltung der Rollen stellt den Diagnostiker aber auch vor Probleme. Für die Entwicklung von Rollenspielen mit mehreren Teilnehmern ist darauf zu achten, dass die einzelnen Rollen die gleichen Chancen bieten, das Zielverhalten der Aufgabe zu zeigen. Eine noch höhere Standardisierung kann in dyadischen Rollenspielen mit einem explizit eingewiesenen Rollenspieler und einem Teilnehmer erreicht werden. Es wurde jedoch auch gezeigt, dass die Attraktivität oder Ausstrahlung eines Rollenspielers das Ver-

halten des zu beurteilenden und damit auch die externe Validität des Rollenspiels beeinflusst. Teilnehmer, die mit einem unkooperativen Rollenspieler agierten, wurden als weniger sozial kompetent eingestuft als diejenigen, die mit einem neutralen Rollenspieler in Interaktion traten (Steinberg et al. (1982) zitiert nach Karkoschka, 1998).

Als letzter Aspekt in der Diskussion dieser Aufgabenform soll die Fähigkeit der Teilnehmer zur Rollenübernahme thematisiert werden. Wenn diese Fähigkeit zwischen den Teilnehmern unterschiedlich stark ausgeprägt ist, werden die diagnostischen Ergebnisse verzerrt. Dann bestimmt z.B. nicht mehr die unterschiedlich ausgeprägte soziale Kompetenz die Verhaltensvarianz, sondern die unterschiedlich ausgeprägte Fähigkeit zur Rollenübernahme. Aus diesem Grunde sollten die Rollen so gestaltet sein, dass sie möglichst gut an die bisherige Lebenswelt der Teilnehmer anschließen. Besonders deutlich wird diese Problematik in Auswahlprozessen für eine Ausbildung. In diesem Fall haben die Teilnehmer noch keine Erfahrung in den Rollen, die sie später nach der Ausbildung einnehmen sollen. Hier kann der Vorteil des Rollenspiels, die größere Realitätsnähe, nicht greifen, bzw. wird durch den Nachteil des verzerrenden Effekts der Fähigkeit zur Rollenübernahme konterkariert.

1.5.3 Planspiel

Eine sehr ausführliche Auseinandersetzung mit dem Thema Planspiel ist unter dem Titel „Planspiele im Personal- und Organisationsmanagement“ von Thomas Geilhardt und Thomas Mühlbradt (1995) herausgegeben worden. Dort ist folgende Definition zu finden:

„Ein Planspiel ist eine konstruierte Situation, in der sich eine/mehrere Person(en) in/an einem diskreten Modell nach vorgegebenen Regeln verhalten, wobei gezeigtes Verhalten systematisch festgehalten und nach einem explizierbaren Kalkül bewertet werden kann. Ein Planspiel beinhaltet damit immer eine Simulation, ist aber nicht mit dieser gleich zu setzen.“ (Geilhardt, 1995, S. 49)

Diese Aufgabenform ist also nicht zwangsläufig eine Gruppenaufgabe, sondern kann auch so gestaltet werden, dass sie von einzelnen Personen bearbeitet wird. Bei gruppenorientierten Planspielen kann die Interaktion unmittelbar zwischen den Teilnehmern stattfinden oder mittelbar über einen Computer oder andere Medien. Kennzeichnend ist gemäß dieser Definition der relativ hohe Standardisierungsgrad des Aufgabenablaufs, der durch das diskrete Modell und die Planspielregeln gegeben ist.

Während der Ursprung des Planspiels im Schachspiel gesehen wird (ca. 800 vor Christus in Persien) (Rohn, 1995), sind insbesondere computergestützte Planspiele heute in einem breiten Anwendungsfeld des Personal- und Organisationsmanagements eine akzeptierte Methode. Kastner (1995) beurteilt computergestützte Planspiele als hilfreiche Instrumente, um Systemsimulationen durchzuführen, anhand derer die Besonderheiten des systemischen Denkens erfahrbar gemacht und eingeübt werden können.

Als diagnostisches Instrument werden Planspiele bisher jedoch ausnahmslos im Rahmen von anderen Verfahren eingesetzt (Geilhardt & Mühlbradt, 1995). Die Autoren empfehlen darüber hinaus,

„...Planspiele vorerst nur als Medium (z.B. um Gruppendiskussionen in einem Assessment Center auf einer motivierenderen und realistischeren Basis stattfinden

zu lassen) in einem Auswahlkontext zu verwenden.“ (Geilhardt & Mühlbradt, 1995, S. 12)

Damit ist gemeint, nicht die Zielerreichung im Planspiel als Bewertungsgrundlage zu verwenden, sondern wie in einer herkömmlichen Gruppendiskussion die gezeigten Verhaltensweisen der Teilnehmer als diagnostische Daten heranzuziehen. Es stellt auch in den relativ stark standardisierten Abläufen eines Planspiels eine Schwierigkeit dar, diagnostisch saubere Schlussfolgerungen von Aufgabenergebnissen auf Personenmerkmale zu ziehen und den Einfluss des Situationsfaktors zu bewerten:

„Besonders bei Planspielen ist die Gefahr zu beachten, die Erfolgs- und Verhaltensdaten, die einen originären Wechselwirkungseffekt der Person-Situations-Dynamik darstellen, einseitig als Indikatoren der Person im Sinne relativ stabiler Persönlichkeitseigenschaften zu interpretieren.“(Fassheber, 1990, S. 495, zitiert nach Fritz & Funke, 1995, S.72)

Zusammenfassend lässt sich festhalten, dass der diagnostische Nutzen aus Planspielen noch zu wünschen übrig lässt angesichts des enormen Aufwandes, der mit der Planspielentwicklung verbunden ist. Ihr wesentlicher Einsatzbereich liegt bisher im Schulungs- und Trainingsumfeld bzw. in der Diagnose von Systemen.

1.5.4 Konstruktionsübung

Die Idee der Konstruktionsübung basiert auf dem Vorbild der Arbeitsprobe der frühen psychologischen Diagnostik. Moede unterscheidet bereits 1930 zwischen Leistungsproben und Verhaltensproben. In der Leistungsprobe wurde die Fähigkeit zu konkreter Arbeitsleistung gemessen, in der Verhaltensprobe wurde versucht, arbeitswichtige Verhaltensweisen und Einstellungen zu erfassen (Moede, 1930, zitiert nach Fay & Heilmann, 1995).

Bei der Konstruktionsübung als Gruppenübung liegt der Beobachtungsschwerpunkt nicht auf der handwerklichen Tätigkeit oder der Qualität des Endproduktes. Die Aufforderung, sich handwerklich zu betätigen dient als Stimulation des diagnostisch relevanten Verhaltens. Dabei kann die Konstruktionsübung zum einen mit „neutralem Stoff“ durchgeführt werden, d.h. die Bearbeitung des Materials setzt keine spezifischen Kenntnisse oder Fähigkeiten voraus. Gängige Beispiele sind Aufgaben, bei denen mit einer vorgegebenen Menge an Papier und Pappe ein möglichst hoher Turm gebaut werden soll, oder eine Vorrichtung, mit der ein Ei den Sturz vom Tisch unbeschadet übersteht. Zum anderen können auch relevante Inhalte der späteren Zielposition zum Gegenstand der Konstruktionsübung gemacht werden. Wenn die Teilnehmer bereits alle über das nötige Wissen verfügen, kann so eine gute inhaltliche Validität der Aufgaben erreicht werden.

Die Besonderheit der Konstruktionsübung ist aus eignungsdiagnostischer Perspektive darin zu sehen, dass insbesondere im Unterschied zu Gruppendiskussionen aber auch zu Planspielen tatsächliches Handeln in der Realität stattfindet (Fay & Heilmann, 1995). Entscheidungen werden nicht nur getroffen, sondern auch umgesetzt, und zwar nicht nur im Rahmen eines Modells, wie beim Planspiel, sondern in der Realität. Dies erhöht jedoch auch die Schwierigkeit, die Bewertung sozial-interaktiven Verhaltens nicht mit der Qualität des Endproduktes der Konstruktionsübung zu vermengen. Wie soll z.B. ein effektives Führungsver-

halten eines Teilnehmers bewertet werden, das aber auf die Umsetzung eines nicht erfolgreichen Konstruktionsvorschlages gezielt hat? Hier ist im Einzelfall darauf zu achten, die Beobachter für diese Problematik zu sensibilisieren und eine schlüssige Bewertungssystematik zu entwickeln, die bezüglich der Inhalte der Konstruktionsübung, des Vorwissens der Teilnehmer und der Anforderungen der Zielposition keine Widersprüche enthält.

1.5.5 Präsentation

Die Aufgabenform der Präsentation wird nicht in jeder Systematik den Gruppenaufgaben zugerechnet. Entscheidend ist auch hier wieder das diagnostische Ziel. Den Gruppenaufgaben ist die Präsentation dann sinnvollerweise zuzurechnen, wenn sie zur Diagnose sozial-interaktiver Verhaltensweisen eingesetzt wird.

Auch wenn die Präsentation in Monologform gehalten wird, handelt es sich um eine Grupsituation, deren inhaltliche Validität nur erreicht werden kann, wenn es tatsächlich in der Situation auch echte Zuhörer gibt.

Der Vorteil dieser Aufgabenform ist in der großen Anforderungsnähe für die Führungskräfteauswahl vielfältiger Zielpositionen zu sehen.

Als nachteilig ist anzumerken, dass gerade Präsentationstechniken und Rhetorik als relativ gut trainierbar gelten. Sinnvoll kann eine solche Aufgabe z.B. dann sein, wenn es darum geht, ein bestimmtes Mindestniveau in diesem Bereich abzusichern wie z.B. die grundsätzliche Fähigkeit, als Einzelperson vor einer Gruppe die vorhandenen Kompetenzen zu aktualisieren. Andere Gründe für die Durchführung einer Präsentationsaufgabe können sein, dass es beim Einsatz auf der Zielposition keine Zeitpuffer für ein evtl. nötiges Training gibt, oder dass die Fähigkeiten in diesem Bereich ein besonders hohes Niveau erfüllen sollen.

Vor dem Hintergrund dieser empirischen und theoretischen Grundlagen wird im nächsten Kapitel das bestehende Verfahren der OPZ beschrieben und analysiert. Im Anschluss wird die Entwicklung des neuen Konzeptes für die Gruppenaufgaben vorgestellt.

2 Analyse des bestehenden und Entwicklung des neuen Verfahrens

2.1 Rahmenbedingungen und Besonderheiten des Verfahrens

Das Verfahren soll dazu dienen, diejenigen Offizierbewerber (OB) auszuwählen, die mit hoher Wahrscheinlichkeit erwarten lassen, sowohl die Ausbildung zum Offizier als auch ein Studium an einer der Bw-Universitäten erfolgreich zu beenden. Es handelt sich also um ein Auswahlinstrument für eine Ausbildung, wobei der Prognosezeitraum sehr weit gefasst ist. Neben einigen allgemeinen Bemerkungen beschränkt sich die weitere Darstellung auf das Gruppensituationsverfahren.

Die Teilnehmer an diesem Verfahren sind in der Mehrzahl zwischen 17 und 22 Jahre alt und damit, verglichen mit den üblicherweise in der Industrie eingesetzten Verfahren zur Führungskräfteauswahl, sehr jung. Viele stehen kurz vor oder nach dem Abitur und können in ihrem bisherigen Lebenslauf nur sehr begrenzt auf Führungserfahrungen verweisen. Ca. ein Drittel der OB hat bereits eine gewisse Dienstzeit, z.B. im Rahmen des Grundwehrdienstes, bei der Bundeswehr absolviert und damit eine erste militärische Sozialisation und Vorbildung erfahren im Vergleich zu den ungedienten OB. Diese bewerben sich für die Offizierslaufbahn direkt nach ihrer Musterung. Dazu kommen in kleineren Anteilen OB, die sich bereits als Feldwebel verpflichtet hatten sowie einige ältere OB, die bereits ein ziviles Studium abgeschlossen haben und einen Quereinstieg versuchen (vgl. für eine Beschreibung der Stichprobe Kap. 4.1.). All diese, von ihren Voraussetzungen her sehr unterschiedlichen OB durchlaufen das gleiche Gruppensituationsverfahren.

Die später zu besetzenden Zielpositionen sind sehr unterschiedlich bezüglich der konkreten Tätigkeiten und Anforderungen. Aufgrund der Vielzahl der Aufgaben, die durch Offiziere erfüllt werden und die sich im Laufe ihres Werdeganges auch noch regelmäßig verändern, kann die Ableitung der Anforderungen nicht tätigkeitsbezogen erfolgen, wie es üblicherweise gefordert wird (Laubsch, 2001). Der gemeinsame Nenner wird in der so genannten „allgemeinen Offizierseignung“ gesehen. Unter diesem Begriff werden die Merkmale subsumiert, über die jeder Offizier, wenn auch in leicht unterschiedlichen Ausprägungen, verfügen sollte, um mit den Anforderungen der Ausbildung und den verschiedenen späteren militärischen Verwendungen erfolgreich umzugehen. Diese Eignungsmerkmale sind in den ministeriellen Bestimmungen zur Durchführung des Auswahlverfahrens festgelegt.

Das Auswahlverfahren wird laufend über das gesamte Jahr hinweg durchgeführt. Im Jahr werden ca. 7000 OB geprüft. Ein großer Anteil der OB informiert sich im Vorlauf ihrer Prüfung mit Hilfe des Internets über das Auswahlverfahren. Damit scheidet Gruppensituationsverfahren (GSV), die auf Musterlösungen zurückgreifen, von vorneherein aus.

Die oben genannten Rahmenbedingungen müssen bei der Analyse und Weiterentwicklung des GSV entsprechend berücksichtigt werden.

Die Aufgaben sollten so gestaltet sein, dass es nicht zu systematischen Verzerrungen zwischen den verschiedenen Bewerbergruppen (z.B. gedient, ungedient) kommen kann. Damit scheidet militärisch geprägte Aufgabenstellungen aus.

Die Aufgabengestaltung kann nicht auf eine Abbildung späterer Tätigkeiten zurückgreifen. Dennoch sollten die Aufgaben eine Affinität zu beruflichen Situationen haben, in denen sich Offiziere grundsätzlich immer wieder befinden. Gleichzeitig sollten die im GSV enthaltenen Aufgaben möglichst gut zur bisherigen Lebenserfahrung der zu einem großen Teil jungen OB passen.

Aufgrund der Massendiagnostik ist es notwendig, über eine Vielzahl an Aufgabenstellungen zu verfügen. Nur so ist zu gewährleisten, dass die OB nicht schon im Voraus alle Details der Aufgabenstellung kennen können und die reine Intensität der Informationsbeschaffung zu deutlichen Verzerrungen zwischen den Bewerbern führen kann. Dazu ist anzumerken, dass OB mit detaillierten Vorinformationen nicht zwangsläufig besser abschneiden. Bass (1954) zitiert aus einer seiner nicht veröffentlichten Studien, dass Gruppen nach einem 15-minütigen Training in der dann folgenden Gruppendiskussion signifikant weniger Verhalten zeigten, das den Führungsdimensionen „initiation of structure“ und „consideration“ zugerechnet werden konnte.

Für die Berufszufriedenheit der Prüfer ist es darüber hinaus sehr hilfreich, wenn sie bei der Vielzahl an Prüfungen die Aufgabenstellung variieren können. Gleichzeitig sollte die Vielzahl der Aufgaben aber möglichst nicht zu Bewertungsunterschieden führen und weder die Bewerbergerechtigkeit noch die Standardisierung des Verfahrens gefährden.

2.2 Das bestehende Verfahren

Das GSV wird in der Regel von vier OB in einer Gruppe gemeinsam absolviert. Die OB werden von drei Beobachtern (Stabsoffizier, Prüffizier, Psychologe) bewertet. Jeder Beobachter bewertet jeden der vier OB. Üblicherweise werden in der allgemeinen Praxis jedem Beobachter maximal zwei Teilnehmer je Station zugeordnet (Laubsch, 2001).

Im Rahmen des vorgefundenen Prüfverfahrens werden den OB im Gruppensituationsverfahren drei Aufgaben in dieser Reihenfolge gestellt: Vortrag, Rundgespräch und Planspiel. Das Rundgespräch entspricht einer führerlosen Gruppendiskussion zu einer abstrakten Themenstellung. Die in der OPZ als Planspiel bezeichnete Gruppenaufgabe entspricht nicht der Definition eines Planspiels, wie sie in Kap. 1.5.3 gewählt wurde, da es kein diskretes Modell und keine expliziten Spielregeln gibt. Vielmehr handelt es sich um die Vorgabe eines kurzen Szenarios, innerhalb dessen die OB dann agieren sollen. Um Begriffsverwirrung zu vermeiden und den in der OPZ eingeführten Sprachgebrauch beizubehalten, wird der Begriff Planspiel für die Szenarios der OPZ in dieser Arbeit weiter verwendet.

Nach einer allgemeinen Einleitung werden die OB vor jeder Aufgabe gesondert eingewiesen. Für das GSV stehen zwei Stunden im Ablauf des Auswahlverfahrens zur Verfügung. Folgende Merkmale werden dort beurteilt:

Vortrag

- Ausdrucksfähigkeit und Argumentationsverhalten
- Urteilsfähigkeit
- Psychische Belastbarkeit

Rundgespräch:

- Führungsfähigkeit und Durchsetzungsverhalten
- Soziale Kompetenz und Kooperation
- Ausdrucksfähigkeit und Argumentationsverhalten
- Urteilsfähigkeit

Planspiel

- Führungsfähigkeit und Durchsetzungsverhalten
- Soziale Kompetenz und Kooperation
- Ausdrucksfähigkeit und Argumentationsverhalten
- Urteilsfähigkeit
- Planungs- und Entscheidungsverhalten

Abbildung 2-1: Das 2003 bestehende Verfahren im Überblick – Methoden und Dimensionen

Die Bewertung der OB in den o. g. Dimensionen erfolgt auf einer 7-stufigen Skala. Dabei stellt 1 den besten und 7 den schlechtesten Wert dar. Die 4 soll die Leistung eines durchschnittlichen Bewerbers des Jahrgangs beschreiben. Weitere Operationalisierungen für die Skala liegen nicht vor. Den Prüfern liegt je OB ein zweiseitiger DIN A 4 Bewertungsbogen vor, auf dem alle Ergebnisse des gesamten AC festgehalten werden. Für das GSV steht je Aufgabe ca. ein fünftel der Seite zur Verfügung, um Beobachtungen für die drei bis fünf Dimensionen festzuhalten und in einer Spalte daneben die Note zu vergeben.

Für die Aufgabe Vortrag stehen 6 Listen mit je 12 Themen zur Auswahl. Die OB haben 30 Minuten Zeit zur Vorbereitung. In der Einweisung wird darauf hingewiesen, dass der Vortrag als Monolog zu halten ist und dass die vorhandene Tafel genutzt werden kann. Der Vortrag darf maximal 10 Minuten dauern, danach wird abgebrochen. Die 10 Minuten müssen aber nicht ausgefüllt werden. Die Themen sind abstrakter Natur und fordern keine persönliche Stellungnahme der OB.

In der Gruppendiskussion werden verschiedene Themen zur Wahl gestellt. Insgesamt stehen 6 Listen mit je 5 Themen zur Verfügung. Jeder OB soll für sich ein Thema aus der eingesetzten Liste auswählen und dieses nennen. Danach soll sich die Gruppe auf ein Thema einigen. Die OB werden angehalten, kontrovers zu diskutieren und nicht zwangsläufig ihre

eigene Meinung zu vertreten. Die Gruppendiskussion wird durch die Prüfer beendet, wenn diese der Meinung sind, dass sie genügend Informationen gesammelt haben. Dies war in den beobachteten GSV häufig nach fünf bis acht Minuten der Fall.

In der Einweisung zum Planspiel werden die OB in der Regel darauf hingewiesen, dass sie hier zu Entscheidungen in der Art „wer macht was?“ kommen sollen. Im Gesamtverfahren stehen unterschiedlichste Planspiele zur Auswahl. Im Rahmen einer Materialsammlung in der OPZ entstand eine Liste von 12 Planspielen. Den Prüfern ist es freigestellt, welches Planspiel sie einsetzen.

Auch das Planspiel wird von den Prüfern beendet, wenn diese der Meinung sind, dass sie genügend Informationen gesammelt haben. Bei den beobachteten Planspielen schwankte die Spielzeit z.T. erheblich und lag zwischen 5 und 20 Minuten.

In der Praxis zeigte sich, dass in den zur Verfügung stehenden 2 Stunden von den Prüfern weitere administrative Arbeiten verrichtet werden, die nicht unmittelbar zum GSV gehören. Während der 30 Minuten, die die OB zur Vortragsvorbereitung nutzen, werden von den Prüfern die später folgenden Interviews vorbereitet. Ein weiteres unterschiedlich großes Zeitfenster entsteht dadurch, dass es für die Gruppendiskussion und das Planspiel keine festgelegte Laufzeit gibt und dadurch, dass die OB für die Vorträge meist nur zwischen vier und acht Minuten benötigen.

2.3 Bewertung des bestehenden Verfahrens

Positiv ist der Einsatz verschiedener Methoden fest zu halten (Vortrag, Diskussion, Planspiel), so dass die OB in verschiedenen Aufgabenformen die Möglichkeit erhalten, sich zu präsentieren. Aufgrund des Alters und der geringen Erfahrung mit Auswahlverfahren sind die OB zu Beginn oft deutlich nervös und angespannt. Gerade in einer Gruppensituation wäre es für solche OB dann schwierig, sich in einer durchgehenden Aufgabe nach einem schlechten Start noch positiv einzubringen. Trotz des relativ geringen Zeitbudgets von 2 Stunden für das GSV spricht dieser Faktor dafür, auch in Zukunft mit mehreren kurzen Aufgabenstellungen zu arbeiten. Da eine Abbildung der späteren Tätigkeit in der Aufgabenstellung aus oben genannten Gründen nicht möglich ist, macht es auch wenig Sinn, eine längere, detailliertere Gruppenaufgabe zu stellen, die z.B. konkrete Arbeitsabläufe simuliert, wie es in vielen anderen AC üblich ist.

Die relativ hohe Zahl möglicher Aufgabenstellungen je Methode ist angesichts der Massendiagnostik ebenfalls notwendig und sinnvoll. Die abstrakte Themenstellung bei Vortrag und Diskussion verhindert, dass es zu Verzerrungen durch eine mehr oder weniger gut gelungene Rollenübernahme auf der Seite der OB kommt. Andererseits wird durch diese Art der Themenstellung im Vortrag und in der Diskussion der Schwerpunkt der Bearbeitung durch die OB und der Bewertung durch die Prüfer auf die intellektuelle Leistung der OB gelegt anstatt auf die soziale Interaktion. Dies ist vor dem Hintergrund, dass es in den anderen Ver-

fahrenden aufwändige Bestandteile zur Erfassung der intellektuellen Leistungsfähigkeit gibt, weniger positiv zu sehen.

Die Unterschiedlichkeit der zur Verfügung stehenden Planspiele ist eine nicht kontrollierte Varianzquelle im Gesamtverfahren. Auch wenn die Inhaltsvalidität der vorhandenen Planspiele nicht explizit untersucht wurde, ist nahe liegend, dass ein Planspiel, in dem der Einzug in eine Wohngemeinschaft besprochen werden soll, einen anderen Aufforderungscharakter, Anforderungsbezug und Realitätsnähe aufweist als das Szenario eines abgestürzten Hubschraubers. Hier gilt es für die Zukunft die Anforderung der Standardisierung mit der Anforderung der Aufgabenvielfalt besser zu vereinbaren.

Während der Beobachtung des bestehenden Verfahrens fiel auf, dass durch den Vortrag der OB zu Beginn des GSV die Gruppendynamik während der Diskussion und des Planspiels zum Teil stark beeinflusst wurde. Hatte ein OB eine überzeugende Vortragsleistung gezeigt, kam es häufig dazu, dass ihm von den anderen OB in den folgenden Aufgaben bereits eine Führungsrolle zugeschrieben wurde.

Das Rundgespräch wurde als der Bestandteil des GSV bewertet, der am wenigsten zu einer Differenzierung der OB beitragen konnte. Die abstrakte Themenstellung führt zwar dazu, dass häufig alle vier OB etwas zum Thema beitragen können. Durch die Prüfungssituation besteht jedoch wie in allen Bestandteilen des AC auch hier die Problematik der sozialen Erwünschtheit. Vor diesem Hintergrund führt die geringe Identifikation der OB mit der abstrakten Thematik und ihr Bemühen, mit ihren Äußerungen das Wertemodell der Bw möglichst gut zu treffen, zu einer sehr künstlichen Situation, in der für die Prüfer relativ wenig Unterschiede zwischen den OB zu erkennen sind.

Die Freiheit der Prüfer, sowohl das Rundgespräch als auch das Planspiel dann abzubrechen, wenn sie der Meinung sind, genügend Informationen gesammelt zu haben, führt auf Bewerberseite gegebenenfalls zu bedeutenden Ungerechtigkeiten und im Gesamtprozess zu nicht kontrollierbaren Einflüssen auf die Auswahlentscheidung. Die Möglichkeiten, sich in einer Gruppe zu positionieren, sind bei Spielzeiten, die sich in einigen Fällen um bis zu 15 Minuten unterscheiden, nicht mehr vergleichbar.

Der Beobachtungsprozess stellt sich als ausgesprochen anspruchsvoll dar, da die Prüfer jeweils vier OB beobachten und bewerten müssen. Während im Rundgespräch vier Dimensionen bewertet werden, sind es im Planspiel sogar fünf. Dies geht bei einer lebhaften Gruppenaktivität an die Grenzen der Wahrnehmungs- und Verarbeitungskapazität und wirkt sich negativ auf die Beobachtungsqualität aus. Zudem sollte beachtet werden, dass diese Leistung zum Tagesgeschäft gehört und nicht nur in einem einmaligen AC erbracht werden muss.

Es liegt für das GSV keine Operationalisierung der Dimensionen und keine Strukturierungshilfe zur Beobachtung vor. Dies ist insbesondere angesichts der relativ häufigen Fluktuation

der Prüfer und des fehlenden Beobachtertrainings als problematisch zu bewerten. Die neuen Prüfer werden durch ein Training „on the job“ in die Abläufe der OPZ eingeführt.

Der Beobachtungs- und Bewertungsbogen bietet jeweils für die drei Aufgaben ca. ein fünftel eines DIN A 4 Blattes Platz. Auf diesem Raum finden die Beobachtungen zu drei, vier, bzw. fünf Dimensionen Platz. Der Prozess der Urteilsfindung ist daher wenig transparent und nachvollziehbar und es fällt den Prüfern im Falle einer kontroversen Diskussion bei der Urteilsfindung in der Kommission schwer, nicht mit allgemeinen Eindrücken zu argumentieren, sondern sich auf konkrete Beobachtungen zu beziehen.

2.4 Theoretische und organisationsspezifische Fundierung der Konstrukte

Es gibt in der Assessment Center Literatur wenig fundierte Hinweise darauf, nach welchen Kriterien die Zuordnung zwischen Prüfkriterien und Prüfmethoden gestaltet werden sollte, bzw. wie Prüfmethoden entwickelt werden sollten. Das sogenannte „Missing-Link“-Problem wurde schon mehrfach in dieser Arbeit angesprochen. Schuler gibt mit der folgenden Abbildung einen der wenigen Hinweise aus übergeordneter Perspektive. Abb. 2.2. von Schuler zeigt eine Systematik, in der die Ableitung eines Anforderungsprofils, die Entwicklung von eignungsdiagnostischen Verfahren und die Konstruktion von Berufserfolgskriterien in Beziehung zueinander gesetzt werden. Diese können sich auf Leistung, Verhalten oder Eigenschaften beziehen. Um ein **schlüssiges System der Personalauswahl** zu entwerfen, ist es wichtig, dass die Instrumente der Arbeitsanalyse, die Eignungsverfahren und die Erfolgskriterien sich jeweils auf die gleichen Zielsysteme beziehen. Ist dies nicht der Fall, wird durch jede „Übersetzungsleistung“ systematisch eine Quelle der Ungenauigkeit im Verfahren installiert.

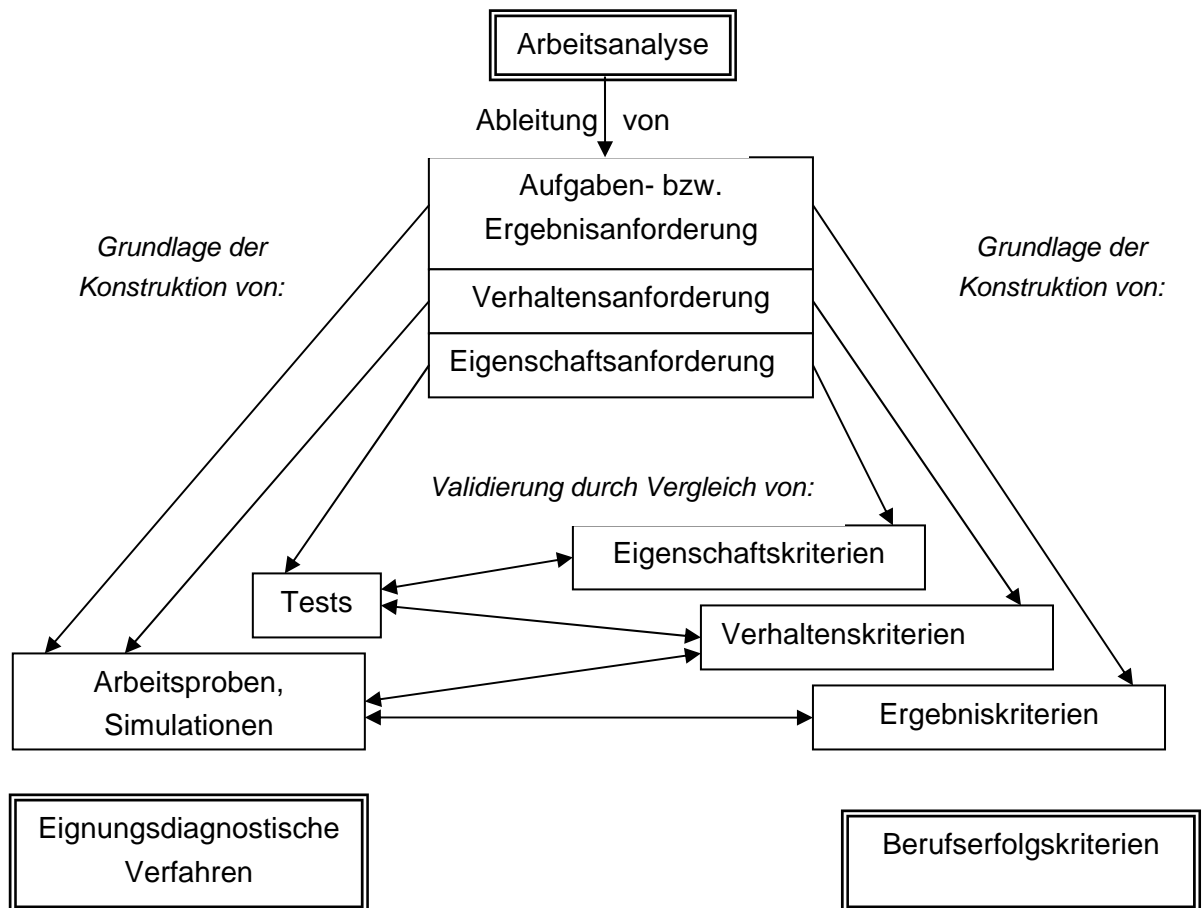


Abbildung 2-2: Systematik der Personalauswahl nach Schuler (1996).

In der Assessment Center Praxis wird die Bewertung in Gruppenaufgaben in der Regel anhand von Verhaltensanforderungen und/oder anhand der erreichten Gruppenleistung vorgenommen. Bei der Bewertung anhand von Verhaltensanforderungen wird diese häufig als Note für bestimmte Dimensionen zusammengefasst. Bedeutsam werden Schulers Überlegungen insbesondere bei der Entwicklung der Berufserfolgskriterien. Diese Thematik sprengt jedoch den Rahmen dieser Arbeit.

2.4.1 Auswahl der Zieldimensionen für die Gruppenaufgaben

Das Anforderungsprofil für Offizierbewerber, die allgemeine Offizierseignung, setzt sich aus allen drei Zieldimensionen zusammen. Bei der Konstruktion der Gruppenaufgaben in der OPZ wurde darauf geachtet, dass die **Bewertung anhand von Verhaltensanforderungen** vorgenommen wird. Grundsätzlich wären auch leistungsbezogene Dimensionen denkbar gewesen. Vor dem Hintergrund der Massendiagnostik in der OPZ erscheinen diese aber für eine Prüfung in Gruppensituationen als weniger geeignet. Wie in Kap. 2.1. geschildert, sind

Gruppenaufgaben mit Musterlösungen wenig geeignet und die Erfassung individueller Leistungsmerkmale ist in Einzelaufgaben valider möglich.

Eine theoretische Fundierung der Konstrukte, die im GSV bewertet werden sollen, ist Voraussetzung für die Durchführung einer inhaltlich sinnvollen Konstruktvalidierung und der Einordnung in ein nomologisches Netzwerk.

Aber es ist für ein praxisrelevantes AC nicht möglich, sich ausschließlich an einzelnen Theorien zu orientieren, da alle bisherigen Theorien zu sozial-interaktivem Verhalten die Realität immer nur ausschnitthaft zu erklären vermögen und daher der Komplexität des Gruppengeschehens nicht gerecht werden können.

„... assessment centers are not grounded in any particular theories of leadership or management“. (Norton, 1977, S. 450, zitiert nach Fennekels, 1987, S. 33)

Es wird stattdessen in der Regel der Weg beschritten, anhand von Anforderungsanalysen der Zielpositionen die relevanten Anforderungen und entsprechenden Übungen fest zu legen.

Es wurde weiter vorne schon beschrieben, dass die klassische Anforderungsanalyse im Falle der Offizierbewerbersauswahl nicht sinnvoll eingesetzt werden kann. Ausgangsbasis für die Projektbearbeitung waren die ministeriell vorgegebenen Eignungsmerkmale. Durch den seit 1994 erweiterten Auftrag der Bundeswehr haben sich die beruflichen Anforderungen an die Offiziere in den letzten 10 Jahren jedoch deutlich verändert. Zusätzlich zur traditionellen Aufgabe der Landes-/Bündnisverteidigung und zum Auftrag der humanitären Hilfeleistung/Katastrophenhilfe entstand mit dem Entscheid des Bundesverfassungsgerichtes vom 12.07.1994 das Aufgabengebiet, rechtmäßige Einsätze zu friedenserhaltenden Maßnahmen im Ausland durchzuführen.

Um diese Entwicklungen aufzunehmen und angesichts knapper Projektressourcen wurden Interviews mit erfahrenen Offizieren geführt, um das bestehende Anforderungsprofil zu aktualisieren. In den Interviews wurde unter anderem qualitativ und mit Hilfe offener Fragestellungen Informationen zur Veränderung des Anforderungsprofils erhoben, um sicher zu stellen, dass keine wesentliche Dimension in den bisherigen Überlegungen vernachlässigt worden war. Darüber hinaus wurde den Offizieren die folgende Liste mit Anforderungsdimensionen vorgelegt und unter anderem eine Gewichtung der Prüfungsrelevanz abgefragt.

Abbildung 2-3: Liste möglicher Anforderungsdimensionen für die Offizierbewerberauswahl:

Gewissenhaftigkeit	Soziale Kompetenz
Leistungsbereitschaft	Führungsbefähigung
Psychische Belastbarkeit	Durchsetzungsverhalten
Anpassungsfähigkeit	Planungs- und Entscheidungsverhalten
Soziales Vergleichsverhalten	Teamfähigkeit
Bewältigungsverhalten	Risikokompetenz
Handlungsorientierung	Urteilsfähigkeit
Lageorientierung	Sprachliche Ausdruckfähigkeit
Kampfbereitschaft	Denkfähigkeit
Berufsvorstellungen	Kognitives Potenzial
Konzentrationsfähigkeit	Körperliche Leistungsfähigkeit u. Belastbarkeit
Reaktionsfähigkeit	Psychische Funktionsstörungen
Vermutetes Lernpotenzial	Soziale Anpassungsstörungen
Mechanisch-technisches Verständnis	Gewaltneigung
Elektrotechnische Kenntnisse	Dogmatismus
Räumliches Vorstellungsvermögen	Neigung zur Straffälligkeit
Rechenfertigkeit	
Mathematikkenntnisse	

Bei diesen Anforderungsdimensionen handelt es sich um Eigenschaften, kognitive und körperliche Leistungsmerkmale, Kompetenzen, Kenntnisse und Devianzmerkmale, wobei eine Zuordnung nicht immer eindeutig möglich ist. Im Verlauf der Interviews ergab sich zu dieser Auflistung kein Ergänzungsbedarf.

Als ein wesentliches Ergebnis der Interviews zur Aktualisierung des Anforderungsprofils sei an dieser Stelle die **gesteigerte Bedeutung der sozialen Kompetenzen im weiteren Sinne** genannt. Insbesondere waren dies **Führungskompetenzen, Sozialkompetenz und Teamfähigkeit**.

Eine weitere Dimension, der eine gleich bleibend hohe Bedeutung für die allgemeine Offizierseignung zugemessen wurde, ist Planungs- und Entscheidungsverhalten. Auf die Problematik der Bewertung von Entscheidungsverhalten wird im Folgenden noch weiter einge-

gangen. Die Dimension **Planungsverhalten** wurde als eine Zieldimension für das Gruppensituationsverfahren festgelegt. Die zunehmende Komplexität der beruflichen Anforderungen für Offiziere führen dazu, dass auch Planungstätigkeiten immer häufiger in Gruppen bewältigt werden müssen. Das Konstrukt ist in diesem Falle nicht schwerpunktmäßig im Sinne einer intellektuellen Leistungsfähigkeit zu verstehen, sondern zielt auf Verhaltensanforderungen im Interaktionsprozess, die für eine erfolgreiche Planung in der Gruppe wichtig sind. Ein weiterer Grund für die Aufnahme dieser Dimension in die Gruppenstation bestand darin, dass abgesehen vom Bewerberinterview im gesamten Auswahlverfahren kein weiteres geeignetes Instrument zur Diagnose von Planungsverhalten enthalten war. Es bestand darüber hinaus Einigkeit, dass Planungstätigkeiten sowohl im Rahmen der Ausbildung als auch später in den verschiedenen Verwendungen eine wesentliche Rolle spielen, es sich also um eine tätigkeitsnahe Anforderung handelt.

Eine umfassende Darstellung der Ergebnisse der Interviews ist für die Fragestellung dieser Arbeit wenig zielführend. Die Diskussion der daraus abgeleiteten Konsequenzen für die Zuordnung einzelner Dimensionen zu den zukünftigen Prüfmethode soll hier im Weiteren ausschnittshaft erfolgen, insofern sie für die Auswahl der Dimensionen für das GSV relevant sind.

Es ist eine Vielzahl weiterer Konstrukte denkbar, die im Rahmen von Gruppensituationen bewertet werden können. Einige, die bei der Aktualisierung des Anforderungsprofils auch hohe Prioritäten erhalten haben, wie z.B. Durchsetzungsfähigkeit oder Handlungs- bzw. Lageorientierung, werden durch die Erhebung mit einem neuen Fragebogeninstrument im Auswahlverfahren berücksichtigt. Für diese Dimensionen konnte davon ausgegangen werden, dass sie zufrieden stellend mit Hilfe eines Fragebogens erfasst werden können (vgl. Kuhl (1983) und Hossiep, Paschen (1998)).

Für Dimensionen, die auf die **kognitive und intellektuelle Leistungsfähigkeit** zielen, wie z.B. Konzentrationsfähigkeit oder Mathematikkenntnisse, gilt, dass sie grundsätzlich auch im Rahmen von Gruppenaufgaben erhoben werden können, der Gruppenprozess und die soziale Interaktion hierfür aber nicht zwingend notwendig sind, sondern vielmehr eher zu einer Verschlechterung der Standardisierung, der Objektivität, Reliabilität und Validität führen. Diese Dimensionen werden in der OPZ mit Hilfe eines computergestützten Testverfahrens erhoben.

Eine Fähigkeit, die nicht nur für Führungskräfte der Bundeswehr eine zunehmende Bedeutung erlangt hat, ist der **Umgang mit komplexen Systemen**. Die Anforderungen an Führungskräfte haben sich in den letzten Jahrzehnten rasant verändert. Seit Anfang der 80´er Jahre gibt es Konzepte, auch in der Führungskräfteauswahl und -entwicklung dem sich beschleunigenden Wandel und der zunehmenden Dynamik unserer Umwelt durch Integration der Erkenntnisse aus der Systemtheorie besser gerecht zu werden. Führungstätigkeit sollte verstanden werden als das Wirken in Systemen:

„Wir dürfen nicht länger die Probleme isolieren, um sie zu lösen. Wir müssen zuerst die Gesamtzusammenhänge erkennen und können dann einzelne Probleme genauer betrachten, näher an das Bild herantreten. Dies aber immer nur in einer permanenten Bewegung des Heran- und Zurücktretens, eben in einem Handlungssystem des Erkennens, Diagnostizierens, Planens, Sinngabens und Zielesetzens.“ (Kastner, 1991, S. 44)

Grundsätzlich ist davon auszugehen, dass systemisches Denken auch im Erwachsenenalter lernbar ist. Dies ist jedoch ein aufwändiges Vorhaben, denn es gilt, eingefahrene Denkmuster und Werthaltungen zu verändern (vgl. Kreuzig, 1991). Neben der herkömmlichen Intelligenzdiagnostik gibt es entsprechend Bemühungen, interindividuelle Unterschiede in Strategien des Planens und Entscheidens in komplexen Systemen fest zu stellen. Diese Unterschiede lassen sich nach Dörner einer so genannten „operativen Intelligenz“ zuschreiben, der Fähigkeit in komplexen Situationen Planungs- und Entscheidungsstrategien zu entwickeln, kontrolliert einzusetzen und aufgrund von Rückmeldungen wieder zu verändern (vgl. Putz-Osterloh, 1991).

Die Diagnose geschieht in der Regel mit Hilfe computersimulierter Planspiele. Auf diese Weise ist es möglich, die komplexen Anforderungen standardisiert abzubilden. Nicht Computer unterstützte situative Verfahren können eine solche Standardisierung nicht leisten und sind daher aus Objektivitäts- und Fairnessgründen für die Diagnose von Verhalten in komplexen Systemen nicht gut geeignet.

Zur Erfassung dieser individuellen Strategien gab es bei der Bundeswehr konkrete Projekte, (vgl. Melter (1995)), die sich jedoch nicht in der Praxis der Offizierbewerbersauswahl niederschlagen haben (zur Problematik Computer gestützter Planspiele vergleiche auch Funke (1995)). Auch Kastner sieht den Nutzen Computer gestützter Planspiele vielmehr in der Möglichkeit, spezifische Systemeigenschaften zu simulieren und im Umgang mit diesen Simulationen das systemische Denken erfahrbar und trainierbar zu machen (Kastner, 1995).

Das Konstrukt der **Risikokompetenz** weist für die Auswahl von Offizierbewerbern auf den ersten Blick eine hohe Relevanz auf. Risikokompetenz kann je nach Ausgestaltung eine deutliche Facette sozialer Kompetenzen enthalten. Einen wesentlichen Bestandteil von Risikokompetenz stellt die Beschaffung und Bewertung von Informationen dar. Dies kann zum einen verstanden werden als vornehmlich intellektuelle Fähigkeit, ausreichend Informationen zu beschaffen und angemessen zu verarbeiten und auf dieser Basis mehr oder weniger risikoaffine oder risikoaverse Entscheidungen zu treffen. Bei einem derart gelagerten Konstruktverständnis bewegt man sich jedoch stärker im Bereich der Entscheidungsfindung bzw. des Vorgehens bei komplexen Problemlösungen und die Beurteilung findet auf einer inhaltlichen Ebene der Entscheidungen statt.

Zum anderen ist dieser Prozess der Informationsbeschaffung und -bewertung in der Regel aber gerade auch ein Prozess der sozialen Interaktion, in dem Gruppenprozesse eine wesentliche Rolle spielen, es im Extremfall zu Phänomenen des Group-Think kommen kann, aus dem viel extremere und riskantere Entscheidungen resultieren, als dies bei einer individuellen Entscheidung der Fall gewesen wäre. Die Abbildung und Auswertung dieser Prozes-

se im Rahmen einer Gruppenaufgabe ist daher nur bedingt zur Diagnose der individuellen Risikokompetenz geeignet, da von einem sehr starken Einfluss des nicht standardisierbaren Gruppenprozesses auf die jeweilige individuelle Urteilsbildung der Teilnehmenden ausgegangen werden muss.

Ein weiterer Grund, der gegen die Aufnahme der Risikokompetenz als Zieldimension der Gruppensituationsverfahren spricht, ist der Aspekt der Trainierbarkeit einer solchen Kompetenz. Die Risikoabwägungen, die ein Offizier später in seinem beruflichen Leben bewältigen muss, weichen bezüglich der emotionalen Belastungen und der inhaltlichen Rahmenbedingungen erheblich von den bereits vorhandenen Alltagserfahrungen der jungen Offizierbewerber ab. Im Laufe ihrer Ausbildung werden die Offiziere sowohl inhaltlich als auch mental intensiv und z.T. mit speziellen Trainings auf derartige Entscheidungssituationen und den Umgang mit den für ihren Beruf spezifischen Risiken vorbereitet (vgl. Fröhling. & Rausch (1996)).

Es wurde daher vor dem Hintergrund dieser theoretischen und organisationspezifischen Überlegungen entschieden, das Konstrukt der Risikokompetenz nicht als soziale Kompetenz mit dem Schwerpunkt des Verhaltens in Gruppen sondern vielmehr als individuelle Disposition aufzufassen und zu berücksichtigen. Das Konstruktverständnis orientiert sich zum einen (Facette der Risikobereitschaft) an der Disposition Sensation Seeking (Erlebnissuche), die 1979 von Zuckerman beschrieben wurde. Gemeint ist das individuell unterschiedliche Bedürfnis nach abwechslungsreichen, neuen und komplexen Eindrücken und der zugehörigen Bereitschaft, um solcher Eindrücke willen physische und soziale Risiken in Kauf zu nehmen. Dieses Konstrukt ist außerdem nahe an der Dimension „Offenheit für Erfahrungen“ aus dem Big Five Konzept anzusiedeln. Eine zweite Konstruktfacette (Risikoverhalten) fragt nach der adäquaten Einschätzung von Risiken und der Art und Weise, wie Handlungen abgeleitet werden. Für das Auswahlverfahren der Offizierbewerber wurde vorgeschlagen, dieses Konstrukt durch eine Selbstauskunft der Bewerber im Rahmen des neu entwickelten Fragebogens zu bewerten.

Aus dem **Konzept der Big Five** wurde das Konstrukt **Gewissenhaftigkeit** als relevant für die Bewerberbewertung in der OPZ betrachtet. Ähnlich wie für das Konstrukt der Risikokompetenz leitet sich auch für das Konstrukt der Gewissenhaftigkeit die Bedeutung für das Auswahlverfahren unmittelbar aus den besonderen Aufgabenstellungen im Rahmen des Offizierberufes ab. Über alle möglichen späteren Verwendungen hinweg ist davon auszugehen, dass Offiziere in der Ausübung ihrer Aufgaben, insbesondere im Umgang mit Waffen und den ihnen anvertrauten Menschen, besondere Sorgfalt walten lassen müssen.

Die Dimensionen des „Big Five“-Konzepts wurden in der Literatur häufig als mögliche Konstrukte für den Einsatz im AC diskutiert und untersucht. Grundsätzlich ist ihr Einsatz in der Personalauswahl der Kritik ausgesetzt, dass sie als allgemeine Persönlichkeitsdimensionen eine relativ geringe Nähe zu den jeweiligen Arbeitsanforderungen aufweisen. Insbesondere für das Merkmal Gewissenhaftigkeit ergaben sich in Metastudien jedoch durchaus akzeptable Werte, wobei auf deutliche Wechselwirkungen mit der Art der Zielposition hingewiesen wird (vgl. Staufenbiehl, Rösner (1999)).

Für die übrigen Dimensionen des Big-Five-Konzeptes (Verträglichkeit, Extraversion und Neurotizismus) gilt, dass über eine grundsätzlich wünschenswerte durchschnittliche Ausprägung dieser Dimensionen hinaus, keine begründeten Hypothesen über den Zusammenhang zwischen diesen Kriterien und einem späteren Berufserfolg aufgestellt werden können. Die ideale Ausprägung der Extraversion ist für einen Offizier, der als Ausbilder tätig ist, als unterschiedlich anzunehmen zu der idealen Ausprägung für einen Jet-Piloten, so dass ohne eine verwendungsspezifische Anforderungsanalyse für dieses Konstrukt keine Zusammenhang zwischen Ausprägung und Berufserfolg vermutet werden kann.

Nach der Systematik von Schuler (s. S. 45) ist die Diagnose einer allgemeinen Persönlichkeitseigenschaft wie Gewissenhaftigkeit mit dem Instrument einer Gruppenaufgabe wenig sinnvoll, sie sollte vielmehr mit Hilfe von Tests vonstatten gehen. Entsprechend wurde diese Dimension in den im Projektverlauf neu entwickelten Fragebogen aufgenommen und nicht als Konstrukt für die Gruppenaufgaben ausgewählt.

Neben den Bemühungen, Instrumente der Personalauswahl mit dem Ziel verbesserter Gütekriterien zu verändern, gibt es auch Überlegungen und Entwicklungen in Forschung und Praxis, um den gesamten Prozess der Personalauswahl besser an die sich immer schneller verändernden Umweltbedingungen anzupassen. Insbesondere für die Offizierbewerbersauswahl, deren erfolgreiche Teilnehmer eine lange und kostspielige Ausbildungsphase durchlaufen, bevor sie schließlich ihren Beruf ergreifen, könnte die Erfassung von **Lernpotenzial** ein interessantes Kriterium darstellen.

Das Kriterium Lernpotenzial wird üblicherweise in die beiden Facetten Lernbereitschaft und Lernfähigkeit aufgeteilt (vgl. Sarges, 1996). Lernfähigkeit ist dabei nicht über die Erfassung von Intelligenzmerkmalen abzubilden. Es gibt keine allgemeine Lernfähigkeit sondern individuelle Lernleistungen differieren bei unterschiedlichen Aufgabentypen. Der wesentliche Unterschied zwischen einem Lernpotenzial-AC und einem herkömmlichen AC ist darin zu sehen, dass nicht die einmalige Performance Grundlage der Bewertung ist, sondern die Entwicklung der Teilnehmer im Verlauf einer Trainings- bzw. Lernphase beurteilt wird. Ein elaboriertes Lernpotenzial-AC durchläuft üblicherweise fünf Phasen. In der Vorbereitungsphase bekommen die Teilnehmer Material zur Verfügung gestellt, dass sie sich selbstständig erarbeiten. In der zweiten Phase des AC werden dann Kenntnisse, Kompetenzen und Fähigkeiten erfasst. Darauf folgt eine begleitete Lern- und Trainingsphase mit ausführlichem Feedback zur Unterstützung der Lernprozesse. In der vierten Phase werden ein zweites Mal die Zieldimensionen diagnostiziert, um die Lernfortschritte fest zu stellen. In der Abschlussphase erhält jeder Teilnehmer ein ausführliches Feedback über sein Stärken- und Schwächenprofil und es werden individuelle Maßnahmen zur Personalentwicklung vereinbart.

Das Lernpotenzial-AC wird üblicherweise zur Karriereplanung und als Baustein im Rahmen der Personalentwicklung für Führungskräfte eingesetzt statt zur Personalauswahl (vgl. Sarges, 1996). Angst und Stress wirken sich bekanntermaßen nachteilig auf Lernerfolg aus. Für die Bewerber in der OPZ ist die Auswahl-situation mit zum Teil erheblichem Stress verbunden, da es sich um eine Entscheidung handelt, die sich sehr langfristig auf ihre berufliche

Zukunft auswirken wird. Es wäre daher möglich, dass statt Lernpotenzial eher die individuelle Belastungsfähigkeit gemessen wird und dass die individuellen Rahmenbedingungen, wie z.B. berufliche Alternativen und soziale Sicherheit, das Ergebnis stark moderieren würden. Diese kurze Darstellung veranschaulicht zudem, dass ein solches Lernpotenzial-AC sich deutlich aufwendiger darstellt als ein durchschnittliches Auswahl-AC. Für die Auswahl der Offizierbewerber ist diese AC-Form aufgrund der Problematik der Massendiagnostik so nicht anwendbar. Der Ressourcenaufwand wäre vermutlich um ein mehrfaches höher als dies zurzeit der Fall ist. Angesichts der hohen Kosten, die der Bw aber z.B. durch Studienabbrecher entstehen, könnte sich eine Investition in dieser Richtung dennoch als sinnvoll erweisen. Im Rahmen des Projektes, in dem diese Arbeit entstand, ging es jedoch nicht um eine vollständige Erneuerung des Verfahrens, sondern nur um die Analyse und Verbesserung einzelner Module. Die vorangegangenen Überlegungen sind insofern in die Verfahrensänderungen eingegangen, dass die Dimension der Leistungsmotivation, die als ein nahes Konstrukt zur Lernbereitschaft verstanden werden kann, unter anderem Bestandteil des neuen Fragebogens geworden ist.

Die geschilderten Überlegungen zur Auswahl der Dimensionen für das Gruppenverfahren bezüglich der moderierenden Wirkung des Gruppenprozesses und die Frage nach der Trainierbarkeit von Fähigkeiten und Kompetenzen gelten grundsätzlich auch für andere Konstrukte, die üblicherweise mit Hilfe von Gruppenaufgaben diagnostiziert werden, wie Führungs- oder Teamkompetenz. Auch hier bedingen sich die individuellen beobachtbaren Verhaltensweisen und der Gruppenprozess wechselseitig. Die dargestellten Überlegungen zur Auswahl der Zieldimensionen für die Gruppenaufgaben sind daher im Sinne gradueller Unterscheidungen zu verstehen und vor dem Hintergrund eines möglichst effektiven Einsatzes knapper Ressourcen in einer Massendiagnostik.

Gruppenaufgaben sind insbesondere zur Diagnose sozialer Kompetenzen geeignet. Das GSV stellt im Rahmen des gesamten Verfahrens zur Offizierbewerbersauswahl den einzigen Methodenbaustein dar, der die soziale Interaktion direkt beobachtbar macht. Es wäre also wenig effektiv, mit dieser Methode Dimensionen zu bewerten, die auch, bzw. besser mit Hilfe anderer Methoden bewertet werden können. Die Leitfrage für die Auswahl der Dimensionen für das GSV war daher: Für welche Dimensionen ist es elementar, sie anhand der Beobachtung sozialer Interaktion zu bewerten? Der Zusatznutzen aber auch die Schwierigkeit von Gruppenaufgaben im Vergleich zu anderen Verfahren liegt eben gerade in der Beobachtbarkeit komplexer sozialer Interaktionsprozesse.

Aufgrund der geschilderten Ergebnisse aus der Aktualisierung des Anforderungsprofils sowie der Überlegungen, ein insgesamt schlüssiges System der Personalauswahl zu konzipieren und einen effizienten Ressourceneinsatz in der OPZ zu ermöglichen, wurden die folgenden Eignungsmerkmale als Zieldimensionen für die Planspiele festgelegt:

- Führungskompetenz
- Teamkompetenz
- Sozialkompetenz
- Planungsverhalten

Im folgenden Kapitel sollen konstruktrelevante Theorien vorgestellt werden und mit den organisationsspezifischen Anforderungen zu einem nachvollziehbaren Konstruktentwurf verknüpft werden. Dieser bildet die Basis für das Konstruktverständnis und die Operationalisierung. Es werden nur die Dimensionen des zukünftigen Verfahrens behandelt, da die weiteren Dimensionen, die noch im alten Verfahren im Rahmen des GSV bewertet werden, Argumentationsverhalten/Ausdrucksfähigkeit sowie Urteilsfähigkeit, ihren Schwerpunkt stärker auf der inhaltlichen und intellektuellen Ebene der Kommunikation haben als auf der Ebene der sozialen Interaktion.

2.4.2 Führungskompetenz

Zur Erklärung des Phänomens „Führung“ gibt es eine Vielzahl verschiedenster Theorien. Das Handwörterbuch der Führung (Kieser, 1995) liefert einen umfassenden Überblick. Die einzelnen Theorien stellen jeweils spezifische Sichtweisen auf das Thema dar, die bestimmte Aspekte des komplexen Führungsphänomens hervorheben. Daher ist auch die theoriegeleitete Suche nach einer allgemeingültigen Definition von Führung wenig sinnvoll. Gerade für die Gestaltung der Gruppenaufgaben und ein gemeinsames Konstruktverständnis der Beobachter im Rahmen des Auswahlprozesses ist jedoch eine Eingrenzung notwendig. Angesichts der Zielsetzung der Gruppenaufgaben im AC der OPZ, bei jungen Menschen Führungskompetenz im Allgemeinen zu bewerten, wird für eine solche Eingrenzung folgendes Verständnis von Führung zu Grunde gelegt (Weinert, 1998):

- Führung ist ein Gruppenphänomen (das die Interaktion zwischen zwei oder mehreren Personen einschließt)
- Führung ist intentionale soziale Einflussnahme
- Führung zielt darauf ab, durch Kommunikationsprozesse Ziele zu erreichen

Der letzte Punkt verdeutlicht die Problematik, dass die angezielten Konstrukte Führungspotenzial und Sozialkompetenz, bzw. Teamkompetenz, nicht als voneinander unabhängige Konstrukte definiert werden können, da sie alle über die Gestaltung von Kommunikationsprozessen ihren Ausdruck finden. Am deutlichsten wird dies am Beispiel eines Teilnehmers, der sich nicht verbal am Gruppengeschehen beteiligt. In diesem Fall wird er in allen angezielten Konstrukten eine schlechte oder, je nach Vereinbarung, gar keine Bewertung erhalten.

Die relevanten Führungstheorien lassen sich grob vier unterschiedlichen Modellansätzen zuordnen, die in der dargestellten Reihenfolge gleichzeitig eine chronologische Entwicklung widerspiegeln:

Zu den ältesten Führungstheorien zählen die **Eigenschaftstheorien**. Sie konzentrieren sich zur Erklärung des Führungsphänomens auf die Führungsperson und bilden daher häufig noch immer die Grundlage für Auswahlentscheidungen (vgl. Lattmann, 1989). Den Theorien liegt die Annahme zu Grunde, dass Eigenschaften die primären Determinanten von Verhalten sind. Daher suchen sie nach zentralen Eigenschaften, die positiven Einfluss auf den Führungserfolg haben. Eigenschaften werden in diesem Zusammenhang als hypothetische Konstrukte aufgefasst, sie sind Erklärungen und Interpretationen des beobachtbaren Verhaltens.

„Die Konstruktbemühungen sind als erfolgreich zu betrachten, wenn für die so bezeichneten Eigenschaften zeitliche und transssituationale Konsistenz nachgewiesen werden kann. So gesehen sind Konstrukte Indikatoren für Verhaltensbereitschaften mit individuellen Ausprägungen.“ (Delhees, 1995, S. 899)

Für Auswahlentscheidungen, die sich nicht auf konkrete Tätigkeitssimulationen der Zielpositionen stützen können, wie es in der OPZ der Fall ist, bieten Eigenschaftstheorien daher einen hilfreichen theoretischen Rahmen. Im Zuge dieser Theorien wurden beinahe endlose Auflistungen von Führer-Traits zu den Bereichen Fähigkeiten/Fertigkeiten (z.B. Intelligenz), Persönlichkeitscharakteristika (z.B. Selbstvertrauen, Dominanz) und soziale Charakteristika (z.B. Kooperationsbereitschaft) erstellt. Wunderer (2000) fasst inhaltlich zusammen, dass die folgenden Eigenschaften bei statistischen Messungen der Persönlichkeitsstrukturen erfolgreicher Führungskräfte, vor allem im Militär, relativ häufig gefunden wurden, ohne aber ein eindeutiges und stabiles Bild zu ergeben:

- Problemlösefähigkeit, Wortgewandtheit, Wachsamkeit, Urteilkraft
- Zielbestimmtheit, Dynamik, Ausdauer und Konsequenz, Initiative und Leistungsehrgeiz
- Durchsetzungsstärke, Stehvermögen, Selbstvertrauen, Zuverlässigkeit, soziale Sensitivität und Kontaktbegabung, Ausstrahlung (Wunderer, 2000, S. 72)

Delhees bietet angesichts dieser Vielfalt an führungsrelevanten Eigenschaften ein vierdimensionales Kategoriensystem an (Delhees, 1995, S. 902):

1. Prädispositionen der Einflussbefähigung (z.B. Dominanz, Durchsetzungsfähigkeit, Selbstvertrauen, Bedürfnis nach sozial ausgerichteter Macht)
2. Soziale und interpersonelle Fertigkeiten (z.B. Kooperationsbereitschaft, Interaktionskompetenz, Gruppensteuerung)
3. Merkmale der Aufgaben- und Zielorientiertheit (z.B. Initiative, Ehrgeiz, Durchsetzungsfähigkeit)
4. Prädispositionen der Informationsverarbeitung und –evaluation (z.B. Intelligenz, Entscheidungsfähigkeit, Urteilsvermögen)

Eine zweite Gruppe von Führungstheorien lässt sich unter dem Begriff der **verhaltensorientierten Ansätze** zusammenfassen. Sie bemühen sich um die Isolation typischer und transsituativ stabiler Verhaltensweisen, sowie um die Untersuchung ihrer Effekte auf bestimmte Zielkriterien. In den sog. Iowa-Studien wurden die Auswirkungen von autoritärer und demo-

kratischer Führung verglichen, in den Michigan-Studien wurde zwischen einem produktions- und einem personen-zentrierten Führungsstil differenziert. Der dritte bekannte Ansatz der Ohio-University differenzierte zwischen der Führungsstildimension „initiation of structure“, also der Definition und Strukturierung der Arbeitsrollen, und „consideration“, also der Aufmerksamkeit und Rücksichtnahme der Führungskraft gegenüber den Erwartungen, Gefühlen und Bedürfnissen der Mitarbeiter (vgl. Wunderer, 2000). Zentrales Ergebnis dieser Theorien ist, dass es den besten Führungsstil nicht gibt. Je nach Aufgabenart und Gruppenzusammensetzung ergaben sich unterschiedliche Auswirkungen auf z.B. Produktivität, Arbeitszufriedenheit, Entscheidungseffizienz etc.

Diese Theorien können für das AC der OPZ zur Strukturierung der Beobachtungen durchaus hilfreich sein. Jedoch können sie nicht zur Eingrenzung des Konstruktes Führungspotenzial herangezogen werden. Denn auch von der Anforderungsseite her ist es nicht möglich, den idealen Führungsstil zu bestimmen. Die späteren Offiziere müssen sowohl in der Lage sein, in bestimmten Situationen aufgrund ihrer Position autokratisch zu führen als auch bei der Zusammenarbeit in Gruppen die Entscheidungsprozesse eher demokratisch und partizipativ zu gestalten. Gleiches gilt für die Dimensionen der Personenorientierung und der Aufgabenorientierung.

Die dritte Gruppe der Führungstheorien basiert auf **Kontingenzansätzen**. Das bekannteste Modell wurde von Fiedler (1964) entwickelt. Die Kontingenzmodelle berücksichtigen, dass Führungserfolg in starkem Maße von Rahmenbedingungen abhängig ist. Bei diesen Ansätzen geht es um die optimale Koordination von personenabhängigen Führungsstilen mit situativen Parametern.

Schreyögg unterscheidet vier Theorie-Arten, die auf dem Kontingenzansatz basieren (Schreyögg, 1995, S. 994):

1. den Moderator-Ansatz, der die Situation als intervenierende Variable studiert (Fiedlers Kontingenzmodell)
2. den situationsanalytischen Ansatz, der sein Hauptaugenmerk auf eine Systematik der Situationserkundung richtet (Entscheidungsmodell von Vroom/Yetton)
3. den kognitiv-instrumentellen Ansatz, der den Führungsstil selbst als Situationselement begreift (Weg-Ziel-Theorie von House)
4. den Interaktions-Ansatz, der Situation und Führung nicht als unabhängige, sondern sich gegenseitig beeinflussende Variable sieht (situationale Führungstheorie von Hersey/Blanchard)

Diese Führungstheorien sind nur zum Teil hilfreich für einen Konstruktentwurf zur Führungskompetenz. Die Situation der Gruppenaufgabe, die sich auf nur sehr kurze Zeit erstreckt, und in einer Gruppe stattfindet, in der sich die einzelnen Gruppenmitglieder untereinander gar nicht oder erst seit einem Tag kennen, unterscheidet sich stark von den üblicherweise angenommenen Settings vieler Führungstheorien, in denen Führungskraft und Mitarbeiter sich über längere Zeit kennen lernen und über einen nennenswerten Zeitraum hinweg zusammen

arbeiten. Entsprechend scheiden z.B. die situative Reifegrad-Theorie von Hersey/Blanchard oder die Weg-Ziel-Theorie von House aus.

Die ersten beiden Theoriearten lenken das Augenmerk dagegen stärker auf die Bedeutung der Situationsgestaltung in den Gruppenaufgaben. Insbesondere vor dem Hintergrund des Moderator-Ansatzes wird daher bei der Weiterentwicklung der Gruppenaufgaben versucht, mit Hilfe situativer Strukturierungsmerkmale über die verschiedenen Planspielaufgaben hinweg eine Standardisierung der Situation zu erreichen (vgl. Kap. 2.6.2 und 2.6.3). So soll der Situationseinfluss auf das Führungsverhalten der OB über die verschiedenen Gruppenaufgaben hinweg besser kontrolliert werden.

Die vierte Gruppe bilden die **transaktionalen Führungstheorien**. Sie berücksichtigen explizit den Prozesscharakter von Führungsbeziehungen und die wechselseitige Beeinflussung zwischen Führungskräften und Geführten (Hollander, 1995; Zalesny & Graen, 1995). Vor diesem theoretischen Hintergrund macht es kaum Sinn, ein Konstrukt wie Führungspotenzial in immer wechselnden Gruppen diagnostizieren zu wollen, da das Führungsverhalten wesentlich vom Verhalten der anderen Gruppenmitglieder abhängt. Eine Standardisierung der diagnostischen Situation wäre entsprechend nur dadurch zu erreichen, dass ein Teilnehmer in einer Gruppe agiert, die aus ausgewiesenen Rollenspielern besteht. Dies ist jedoch gerade für eine Massendiagnostik, wie sie in der OPZ durchgeführt wird, nicht zu leisten.

Ein weiteres Problem für die Auswertung dieser Ansätze besteht darin, dass die beiden wichtigsten Theorien aus dieser Gruppe, das Idiosyncrasiekreditmodell von Hollander und die Vertical-Dyad-Linkage-Theory nach Graen sich explizit auf längerfristige Beziehungen zwischen Führungskräften und ihren Mitarbeitern beziehen. Bei Graen lässt sich allerdings ein Hinweis finden, dass Führungskräfte in sog. unreifen Beziehungen, dazu zählen auch zeitlich kurze Beziehungen, auf ihren individuellen generellen Führungsstil zurückgreifen. Dies stützt die Annahme, dass ein Rückschluss von der kurzen diagnostischen Gruppensituation auf individuelle transsituational stabile personengebundene Anteile des Führungsverhaltens möglich und zulässig ist.

Eine Sonderstellung nehmen die **Attributionstheorien** ein. Sie sind nicht originär der Führungsforschung zuzurechnen, für die Diagnose von Führungskompetenz aber von Bedeutung. Attributionstheorien beschäftigen sich allgemein mit den Prozessen sozialer Wahrnehmung unter dem Gesichtspunkt der Ursachenzuschreibung. Bezogen auf das Führungsphänomen gibt es zwei attributionstheoretische Forschungsrichtungen. Die eine konzentriert sich auf die Attributionen, die Untergebene in Bezug auf das Verhalten ihrer Vorgesetzten vornehmen, die andere beschäftigt sich mit den Attributionen der Vorgesetzten in Bezug auf das Verhalten der Untergebenen (Mitchell, 1995). Erstere ist für den diagnostischen Prozess von großer Bedeutung, denn Attributionsprozesse laufen nicht nur bei jedem Gruppenmitglied ab, sondern insbesondere auch bei den Beobachtern. Auch dieser Theorieansatz gibt keine Hinweise für eine inhaltliche Ausgestaltung des Konstruktes Führungskompetenz. Vielmehr wird deutlich, dass im Rahmen eines individuellen Informationsverarbeitungsprozesses darüber entschieden wird, ob von einer bestimmten Verhaltensweise oder einem erzielten Ver-

haltenseffekt einer Person darauf geschlossen wird, dass dieser Person Führung zugeschrieben wird. Es ist also nur dann eine zufrieden stellende Konstruktvalidität zu erreichen, wenn u.a. diese Zuschreibungsprozesse bei den einzelnen Beobachtern möglichst gleich verlaufen.

Für eine theoretisch gestützte inhaltliche Abgrenzung des Konstruktes Führungskompetenz im Rahmen einer relativ kurzen Gruppensituation verbleiben im Wesentlichen die Erkenntnisse der Eigenschaftstheorien. Während Führungspotenzial im allgemeinen eher als eine Eigenschaft aufgefasst wird, also als ein Personenmerkmal, das eine gewisse Konstanz über Situationen und Aufgaben hinweg aufweist, wird unter Führungskompetenz eher ein Profil spezifischer erlern- und trainierbarer Verhaltensfertigkeiten verstanden. Da Führungspotenzial nicht direkt beobachtbar ist, muss im diagnostischen Verfahren der Rückgriff auf Führungskompetenz erfolgen.

In Anlehnung an das Kategorienschema von Delhees wurde daher für jede der Kategorien eins bis drei stellvertretend eine Unterdimension für das Konstrukt Führungskompetenz gewählt. Die vierte Kategorie beschreibt eher intellektuelle Personenmerkmale und wurde für das Gruppenverfahren nicht berücksichtigt. Intellektuelle Personenmerkmale werden in anderen Bereichen des AC in höherer Qualität erfasst, als dies in einer Gruppenaufgabe möglich wäre. Die Auswahl der Unterdimensionen für das Konstrukt Führungskompetenz wurde mit dem Ziel einer möglichst guten Beobachtbarkeit und vor dem Hintergrund der geringen Führungserfahrung der Teilnehmer getroffen. Stellvertretend für die Kategorie „Prädispositionen der Einflussbefähigung“ wurde die Unterdimension **„Überzeugung“** ausgewählt, für die Kategorie „Soziale und interpersonelle Fähigkeiten“ die Unterdimension **„Strukturierung der Gruppenarbeit“** und für die Kategorie „Merkmale der Aufgaben- und Zielorientiertheit“ die Unterdimension **„Initiative“**. Die Operationalisierung der Unterdimensionen ist der Verhaltenscheckliste im Anhang zu entnehmen.

Die Auswahl verdeutlicht, dass das Konstruktverständnis sich auf eine eher aufgabenorientierte Auffassung des Führungsstils stützt. Die Facette der Mitarbeiterorientierung wird eher über die Konstrukte Sozial- bzw. Teamkompetenz erfasst. Zum einen ist von den Teilnehmern hinsichtlich der Auswahl-situation nur ein begrenztes Maß an „Mitarbeiterorientierung“ zu erwarten, zum anderen wird auf diese Weise eine bessere Abgrenzung der Konstrukte angestrebt.

Da in den Aufgaben des GSV keine Führungsrolle vergeben wird und keine Positionsmacht vorhanden ist, ist von den OB insbesondere demokratisches Führungsverhalten zu erwarten, denn sie sind für die Durchsetzung einer Entscheidung immer auf die explizite oder implizite Zustimmung der anderen Gruppenmitglieder angewiesen. Aus diesem Grunde wurde die Teildimension „Überzeugung“ als Stellvertreter für Präpositionen zur Einflussbefähigung gewählt. Dies entspricht auch stärker den bisherigen Erfahrungen der Teilnehmer. Nur wenige der jungen OB haben bereits Erfahrung mit Situationen, in denen sie aufgrund ihres Status

die alleinige Entscheidungsbefugnis besaßen und einen autokratischen Führungsstil erproben konnten.

2.4.3 Sozialkompetenz

Der Begriff der Sozialen Kompetenz entstand im Zuge der Forschungen zur Sozialen Intelligenz. Thorndike (1920) war einer der Ersten, der sich mit Fragen Sozialer Intelligenz beschäftigte. Er war der Meinung, dass Soziale Intelligenz etwas anderes sei als die allgemeine Intelligenz. Er definierte **Soziale Intelligenz** als

„...the ability to understand and manage men and women, boys and girls – to act wisely in human relations“. (Thorndike, 1920, S. 228)

Seine Definition enthält also eine kognitive und eine handlungsorientierte Komponente. Diese sehr allgemeine Definition wurde in der Intelligenzforschung aufgegriffen, wobei der Schwerpunkt auf die kognitiven Voraussetzungen sozialen Verhaltens gelegt wurde. In diagnostischen Arbeiten in der Tradition der klassischen Intelligenzforschung wird die soziale Intelligenz meist mit sehr abstrakten Problemlöseaufgaben erfasst. In mehreren Studien zeigte sich jedoch, dass diese Messungen weder mit Selbst- noch mit Fremdbeobachtungen sozialen Verhaltens korrelieren, aber hohe Zusammenhänge zur allgemeinen Intelligenz aufweisen (vgl. Schmidt, 1995).

Kastner (1999) fasst soziale Intelligenz ähnlich auf wie seinerzeit Thorndike. Er versteht unter sozialer Intelligenz nicht nur die rein kognitiven Fähigkeiten, sondern auch aktives Verhalten in der sozialen Interaktion. Er differenziert soziale Intelligenz in verschiedene Fähigkeitsbereiche:

- Fähigkeit zur Diagnose sozialer Beziehungen (z.B. soziale Wahrnehmung)
- Fähigkeit, sich einzulassen und zu öffnen (z.B. Offenheit)
- Fähigkeit, gezielt auf andere zu wirken (z.B. Charisma)
- Fähigkeit, zu kommunizieren (z.B. Zuhören)

Sein Konzept der sozialen Kompetenz enthält neben der sozialen Intelligenz eine explizite ethische Komponente, die soziale Verantwortlichkeit. Hier differenziert Kastner nach drei Handlungsbereichen:

- Indirekt-passives Handeln (s. zurückhalten, tolerieren, respektieren)
- Indirekt-aktives Handeln (andere beobachten unser Handeln und können so sozial erwünschtes Handeln erlernen)
- Direkt aktives Handeln (Verantwortungsübernahme für andere)

Die folgende Abb. stellt in einem Vier-Felder-Schema die möglichen Typisierungen Sozialer Kompetenz anhand der beiden Facetten Soziale Intelligenz und Soziale Verantwortlichkeit dar.

+	A	B
Soziale Intelligenz	C	D
-	-	+
	Soziale Verantwortlichkeit	

Abbildung 2-4: Facetten der Sozialkompetenz nach Kastner (1999)

Aus der Perspektive der Personalauswahl ergibt sich aus dieser Klassifizierung im Abgleich mit der Unternehmenskultur eine eindeutige Priorität für die Beurteilung von Offizierbewerbern. Der ideale Kandidat entspricht Typ B: Er verfügt über hohe soziale Intelligenz und über hohe soziale Verantwortlichkeit. Insbesondere für die Organisation Bundeswehr sind Vertreter der A- und C-Kategorie als problematisch zu beurteilen, da einer gewissen Wertebindung und charakterlichen Eignung eine besondere Bedeutung im Rahmen des Anforderungsprofils für Soldaten zugeschrieben wird. Typ D ist daher bei einer gewissen Mindestausprägung Sozialer Intelligenz den Typen A und C in der Auswahl vorzuziehen.

In wie weit eine valide Messung der Dimension Soziale Verantwortlichkeit im Rahmen eines Auswahlprozesses möglich ist, ist kritisch zu beurteilen, da die Problematik des sozial erwünschten Antwortverhaltens im Bereich der Messung von Werthaltungen und Einstellungen nicht zu vernachlässigen ist. Hier wäre vermutlich die Erhebung mit Hilfe eines „behaviour description“ Interviews empfehlenswert. Bei Einstellungen und Werten handelt es sich um nur schwer veränderbare handlungsleitende Grundsätze, die entsprechend in den zu schildernden erlebten Episoden sichtbar werden sollten.

Die oben geschilderten vier Fähigkeitsbereiche der Sozialen Intelligenz werden von Kastner noch in insgesamt 28 Aspekte weiter aufgefächert (vgl. Kastner, 1999). Für die Personalauswahl ergibt sich daraus für die Diagnose ein nicht mehr zu bewältigendes Konstruktverständnis. Hiermit steht Kastner jedoch nicht allein. Dies ist vielmehr ein Phänomen, das die verschiedenen Konzepte zur Klärung des Konstruktes Soziale Kompetenz gleichermaßen aufweisen.

Die folgende Tabelle von Wegner (1989) gibt einen Eindruck über die Vielzahl der Begriffe, die in diesem Zusammenhang bearbeitet werden, zugeordnet zu den zwei Bereichen, die schon Thorndike als wesentliche Komponenten dieses Konstrukts berücksichtigt hatte.

Tabelle 2-1: Begriffe im Zusammenhang mit Sozialer Intelligenz (Wegner, 1989)

Soziale Intelligenz	
Soziale Kognition	Sozialverhalten
Rollenübernahme	Soziale Geschicklichkeit
Empathie	Soziale Kompetenz
Übernahme einer sozialen Rolle	Verhaltenseffektivität
Soziale Einsicht	Führungsverhalten
Soziale Sensibilität	Konfliktmanagement
Sensitivität für individuelle Differenzen	Interpersonale Fähigkeiten
Differenzierung sozialer Reize	Diplomatische Fähigkeiten
Empfänglichkeit für soziale Reize	Soziale Partizipation
Einfühlungsvermögen	Interaktionskompetenz
Beurteilungsfähigkeit	Fähigkeit zur Selbstdarstellung
Selbstüberwachung	
Soziale Kontrolle	

Wie komplex das Konstrukt Soziale Kompetenz verstanden wird, lässt sich auch an dem Erklärungsmodell von Meichenbaum et al. (1981) ablesen. Ihr Modell berücksichtigt drei Komponenten, die sie wie folgt bezeichnen:

- Overt behaviors: Die Fähigkeit, eine Vielzahl verschiedener, sozial angemessener Verhaltensweisen zur Zielerreichung einzusetzen, wird als wesentlicher Bestandteil Sozialer Kompetenz erkannt. Es wird also von Verhaltensweisen auf Soziale Kompetenz geschlossen.
- Cognitive processes: Zu dieser Modell-Komponente werden die Gedanken, die Art und Weise der Informationsverarbeitung sowie Gefühle und Erwartungen im Zusammenhang mit sozialen Situationen gezählt.
- Cognitive structure: Mit dieser Komponente beschreiben sie grundlegendere individuelle motivationale oder affektive Strukturen, die die anderen beiden Komponenten wesentlich beeinflussen.

Diese Komponenten werden als interdependent aufgefasst und zudem als beeinflussbar durch die jeweiligen situativen Besonderheiten. Dieses Konzept beschäftigt sich jedoch aus rein theoretischer Perspektive mit dem Konstrukt Soziale Kompetenz. Die Autoren geben keinerlei Hinweise auf Möglichkeiten zur Diagnostik.

Kanning unternimmt den Versuch, die Vielzahl der Dimensionen sozialer Kompetenz aus der Literatur zu bündeln (Kanning, 2003). Über einen mehrstufigen Prozess der Sichtung, Sortie-

zung, Zuordnung und Rückgriff auf etablierte Konzepte der Psychologie kommt er dabei zu folgendem Resultat:

Tabelle 2-2: Drei Gruppen sozialer Kompetenzen nach Kanning (2003)

Perzeptiv-kognitiver Bereich	Motivational-emotionaler Bereich
Selbstaufmerksamkeit - Direkt - Indirekt Personenwahrnehmung Perspektivenübernahme Kontrollüberzeugungen - Internal - External Entscheidungsfreudigkeit Wissen	Emotionale Stabilität Prosozialität Wertepluralismus
Behavioraler Bereich	
Extraversion Durchsetzungsfähigkeit Handlungsflexibilität Kommunikationsstil - Unterstützung (fordern & gewähren) - Bewertung - Einflussnahme - Expressivität - Zuhören Konfliktverhalten - Verwirklichung eigener Interessen - Berücksichtigung der Interessen anderer Selbststeuerung - Verhaltenskontrolle im soz. Kontext - Selbstdarstellung	

Er beschränkt sich dabei ausschließlich auf allgemeine soziale Kompetenzen. Allgemeine soziale Kompetenzen sind solche, die keinerlei Spezifizierung auf bestimmte Situationen aufweisen. Dies gilt beispielsweise für Extraversion oder Perspektivübernahme. Über spezifische soziale Kompetenzen verfügen nur solche Menschen, die entsprechende Lernerfahrungen vorweisen. Dennoch erscheint die Anzahl der aus der Literatur gewonnen Dimensionen noch immer recht groß. Kanning bezeichnet die Liste aus Tab. 2-2 als

„eine auf qualitativem Weg gewonnene Quintessenz der derzeitigen Diskussion und Konzeptbildung im Forschungsbereich der sozialen Kompetenzen...“ (Kanning, 2003, S. 22).

Im Anschluss an diese Bestandsaufnahme hat Kanning die Kompetenzen empirisch gebündelt und auf ihren tatsächlichen Bedeutungsgehalt für die Gestaltung sozialer Interaktionen

untersucht. Anhand einer ersten Studie mit einer sehr spezifischen Stichprobe von Polizeibeamten, für die eine gewisse Nähe zur Stichprobe der Offizierbewerber angenommen werden kann, konnte die Kompetenz aus Tab. 2-2 zu fünf gut interpretierbaren Faktoren zweiter Ordnung integriert werden:

- soziale Wahrnehmung
- Verhaltenskontrolle
- Durchsetzungsfähigkeit
- Soziale Orientierung
- Kommunikationsfähigkeit

Dennoch kommt auch Kanning zu der Schlussfolgerung:

„... dass es bislang keine allgemein akzeptierte oder gar empirisch fundierte Taxonomie sozialer Kompetenzen gibt. Dies spiegelt sich auch in der Diagnostik wider.“ (Kanning, 2003, S. 22)

Trotz der theoretischen Komplexität dieses Konstruktes wird in der Literatur die Meinung vertreten, dass sich die Soziale Kompetenz eines Individuums durchaus durch die reine Beobachtung seines Verhaltens bestimmen lässt (Schuler & Barthelme, 1995). Die Soziale Kompetenz äußert sich in der Interaktion und setzt die Wahrnehmung und die Analyse der Situation voraus. Schließlich ist das Verhalten der Diagnose leichter zugänglich als die kognitiven Prozesse und Strukturen, oder die handlungsleitenden Werte und Einstellungen.

Für die Diagnose Sozialer Kompetenz im Rahmen beruflicher Auswahlverfahren sind daher insbesondere die **verhaltensorientierten Komponenten des Konstruktes** von Interesse..

Auch auf der Verhaltensebene wird das Konstrukt nicht eindimensional definiert. Hinsch und Pfingsten (1982) definieren Soziale Kompetenz als

„... die Verfügbarkeit und Anwendung von kognitiven, emotionalen und motorischen Verhaltensweisen, die in bestimmten sozialen Situationen zu einem langfristig günstigen Verhältnis zwischen positiven und negativen Konsequenzen führen“ (zitiert nach Hornke et al. 1993, S. 306).

Argyle (1969) bezeichnet Soziale Kompetenz als

„... die Beherrschung interaktiver Fähigkeiten und ihre Anwendung in verschiedenen sozialen Situationen“ (zitiert nach Lotz, 1984, S. 29)

Auch diese zwar verhaltensorientierten, aber sehr globalen Definitionen sind für Diagnosezwecke nicht sehr hilfreich. So kommen verschiedene Autoren zu der Schlussfolgerung, dass es angemessener zu sein scheint, mehr die einzelnen sozialen Fertigkeiten in den Mittelpunkt der Betrachtung zu stellen, als von einem globalen Konstrukt soziale Kompetenz auszugehen (Karkoschka, 1998).

Ein weiterer Grund, der im vorliegenden Zusammenhang dafür spricht, sich auf soziale Fähigkeiten zu beschränken, liegt darin, dass soziale Kompetenz immer nur situationsspezifisch beurteilt werden kann.

„Die Berücksichtigung der Situationsspezifität ist für die Bestimmung sozial kompetenten Verhaltens eine zentrale Voraussetzung. ... Deswegen ist eine ausführliche Arbeitsplatzanalyse unabdingbare Voraussetzung zur Messung sozialer Kompetenz.“ (Karkoschka, 1998, S. 22)

Das Konstrukt der sozialen Kompetenz beinhaltet in allen Konzepten immer einen Aspekt der Wirkungsbeurteilung, sei es soziale Akzeptanz oder ein günstiges Verhältnis zwischen positiven und negativen Konsequenzen. Daraus leitet sich die Bedeutung der Situationsspezifität ab. Bei der Beurteilung von Fähigkeiten kann dagegen von einer gewissen Übertragbarkeit ausgegangen werden. Hat eine Person eine Fähigkeit in einer Situation gezeigt, ist diese Fähigkeit zunächst grundsätzlich vorhanden und es kann erwartet werden, dass diese Fähigkeit auch in anderen Situationen eingesetzt werden kann.

Auch Schuler und Barthelme (1995) sprechen nicht von Sozialer Kompetenz, sondern von **sozialen Fähigkeiten und Fertigkeiten**. Sie nennen z.B.:

- Kommunikative Fähigkeiten
- Kooperations- und Koordinationsfähigkeit
- Teamfähigkeit
- Konfliktfähigkeit oder Konfliktmanagement

Für die Eignungsdiagnostik ergibt sich die letztendlich die Notwendigkeit, sich auf die je nach Anforderungsprofil wichtigsten Facetten zu konzentrieren, und nicht ein aus vielen einzelnen Facetten aggregiertes Gesamturteil, die soziale Kompetenz, in den Mittelpunkt der Betrachtung zu stellen (Karkoschka, S. 27)

Für die Qualität der Diagnose im Auswahlverfahren der OPZ ist es darüber hinaus entscheidend, die Beobachter in den Gruppenaufgaben nicht durch zu viele Konstrukte zu überlasten und die Konstrukte so trennscharf wie möglich zu definieren.

Bei der Auswahl der Fähigkeiten zur Beurteilung sozialer Kompetenz der Offizierbewerber wurde berücksichtigt, dass es sich um allgemeine Fähigkeiten handeln sollte, nicht um spezifische. In fast allen Konzepten zur sozialen Kompetenz ist **Kommunikationsfähigkeit** als wesentliche Facette enthalten. Die Fähigkeit, zu kommunizieren, bei Kastner als Facette der sozialen Intelligenz und auch bei Kanning ist Kommunikationsfähigkeit einer der fünf wesentlichen Faktoren.

Fydrich und Bürgener (1999) haben eine Ratingskala für Soziale Kompetenz vorgestellt, die sich auf verschiedene Dimensionen der verbalen und nonverbalen Kommunikation beschränkt. Die ermittelten Gütekriterien zur Reliabilität und Validität werden als die bislang besten unter den publizierten Ratingsystemen bezüglich der Sozialen Kompetenz hervorgehoben. Die Fähigkeit zu kommunizieren bietet sich daher für die Diagnose in der OPZ an.

Ein weiterer Grund für die Fokussierung auf die Facette der Kommunikation liegt im relativ geringen Alter der OB. Soziale Fähigkeiten und Fertigkeiten gelten grundsätzlich als erlern- und trainierbar und sind in den beruflich relevanten Bereichen im jungen Erwachsenenalter

noch nicht sehr ausgeprägt. Es ist ein sehr anspruchsvolles Unterfangen, bei so jungen Bewerbern in einer Auswahl-situation soziale Fähigkeiten und Fertigkeiten zu diagnostizieren. Die Differenzierungsmöglichkeiten sind noch nicht sehr groß. Hinzu kommt, dass für die Gruppenverfahren nur wenig Zeit zur Verfügung steht. Viele Facetten sozialer Interaktion entwickeln sich aber erst im Verlauf von längeren Zeitspannen der Gruppeninteraktion und sind daher in Planspielen von ca. 10 Minuten Dauer kaum valide zu beobachten.

Auf der Basis dieser Überlegungen sowie aufgrund der überragenden Bedeutung der Kommunikation, wie sie z.B. auch bereits in Argyles Definition deutlich wird, wird daher für das Auswahlverfahren der OPZ **die Kommunikationsfähigkeit als die zu diagnostizierende Facette ausgewählt**. Auf der VCL wird das Konstrukt Soziale Kompetenz entsprechend mit Items zur **verbalen** und **nonverbalen Kommunikation** operationalisiert (s. Anhang)

2.4.4 Teamkompetenz

Das Erfolgskriterium des Auswahlverfahrens für Offizierbewerber ist die erfolgreiche Absolvierung der Offizierausbildung und eines Studiums. Hierfür wurde in der Anforderungsaktualisierung als eine wesentliche Voraussetzung Teamfähigkeit benannt. Aus diesem Grunde wurde Teamfähigkeit als eine weitere soziale Fähigkeit bestimmt, die in der Gruppenstation des Auswahlverfahrens diagnostiziert werden soll.

Ähnlich wie bei dem Konstrukt Soziale Kompetenz sind mit der reinen Benennung jedoch die inhaltliche Ausgestaltung des Konstrukts Teamkompetenz und die Operationalisierung für die Diagnose noch nicht geklärt. Eine inhaltliche Konkretisierung im Rahmen der Anforderungsaktualisierung war aus Ressourcengründen leider nicht möglich, eine Ableitung aus Tätigkeiten ist, wie schon geschildert, aufgrund der vielfältigen späteren Verwendungen nicht sinnvoll. Die Konkretisierung erfolgte daher auf der Basis etablierter Konzepte zur Teamfähigkeit unter Berücksichtigung der organisatorischen Rahmenbedingungen.

Üblicherweise wird in der Literatur ein Team als eine sich entwickelnde Gruppe von Personen aufgefasst, die über längere Zeiträume und verschiedenste Aufgaben hinweg möglichst synergetisch zusammenarbeiten (Kastner, 1999). Teams werden in der Regel eingesetzt, um komplexe Problemstellungen effektiver bearbeiten zu können, als dies einer Einzelperson möglich wäre. Charakteristisch für Teamarbeit ist, dass die Ergebnisse nicht mehr einzelnen Teammitgliedern zurechenbar sind. Aus diesem Grunde ist eine wesentliche Bedingung für erfolgreiche Teamarbeit, dass die Teammitglieder auf Konkurrenzverhalten untereinander verzichten zugunsten der gemeinsamen Zielerreichung. Kooperatives und integratives Verhalten sind daher zwingende Voraussetzungen für Teamfähigkeit.

Schuler et al. (1993) haben die wichtigsten Teilanforderungen von Teamfähigkeit wie folgt definiert:

- Gegenseitiger Informationsaustausch
- Gegenseitige Hilfe und Unterstützung
- Kooperation
- Offenes Ansprechen von Problemen
- Partizipation von betroffenen Kollegen und Abteilungen
- Durchsetzungsvermögen
- Einhalten anerkannter Verhaltensregeln

Dieses Konzept wurde als Basis für die inhaltliche Ausgestaltung des Konstrukts genutzt. Da die Prüfer in der OPZ in jeder Gruppenaufgabe mehrere Konstrukte bewerten, ist es wesentlich für die Beobachtungsqualität, dass für jedes Konstrukt nur zwei oder drei Konstruktfacetten ausgewählt werden. Ein Kriterium war hierbei die Aussicht auf gute Beobachtbarkeit in einer Gruppenaufgabe, die nur ca. 10 Minuten dauert. In leichter Abwandlung zu Schulers Benennung waren dies:

- „**Unterstützung**“ geben und suchen,
- „**Partizipation**“ der anderen Gruppenmitglieder,
- Eigene „**Interessen und Erwartungen**“ nennen, als wesentliche Basis für Kooperation.

Um gegenseitigen Informationsaustausch bewerten zu können, müssten Gruppenaufgaben entwickelt werden, die verschiedene Rollen (mit vorbereitetem Informationsmaterial) für die Teilnehmer vorsehen. Eine solche Gruppenaufgabe ist in der Durchführung relativ zeitaufwendig, da die Teilnehmer sich diese Informationen zunächst aneignen müssen, bevor sie aktiv werden können. Zudem ist die Entwicklung relativ aufwendig, da die einzelnen Rollen so gestaltet sein müssen, dass alle Teilnehmer die gleichen Chancen erhalten, das gewünschte Verhalten zu zeigen (vgl. Kap. 1.5.2).

Ein offenes Ansprechen von Problemen ist realistisch erst zu erwarten, wenn die Gruppenmitglieder ein Mindestmaß an Vertrauen zueinander aufgebaut haben und sich in der Gruppe relativ wohl fühlen. Beides wird durch die kurze Bearbeitungszeit sowie die Auswahl-situation erschwert.

Die Dimension Durchsetzungsvermögen wird bereits im Fragebogen erfasst.

Die Einhaltung anerkannter Verhaltensregeln kann sich in der gegebenen Situation nur auf die üblichen Kommunikationsregeln, wie z.B. Andere ausreden zu lassen oder in der Diskussion nicht laut zu werden, beziehen, da es keine gruppenspezifischen Vereinbarungen zu weiteren Regeln gibt. Diese Facette wird in diesem Verfahren bereits über die Bewertung der kommunikativen Fähigkeiten erfasst.

Der Schwerpunkt zur Erfassung des Konstruktes Teamfähigkeit wird also auf Verhaltensweisen gelegt, die elementar sind für kooperative und integrative Gruppenprozesse.

2.4.5 Planungsverhalten

Das in Assessment Centern bewertete Konstrukt der Planungskompetenz unterscheidet sich theoretisch deutlich von den bisher genannten Konstrukten. Im Gegensatz zu Führungs- und Sozialkompetenzen wird der Schwerpunkt üblicherweise auf interne kognitive Prozesse gelegt. Planung und Problemlösung spielen im Kontext der allgemeinen Handlungsregulation auf das engste zusammen (Funke & Fritz, 1995). Die Konstrukte „komplexes Problemlösen“ und „systemisches Denken“ standen, basierend auf den Szenarien von Dörner (1976, 1989), im Mittelpunkt einer Diagnostik, die sich auf computersimulierte Planspiele stützte (vgl. auch Kap. 2.4.1). Dies mag eine sinnvolle Fragestellung für die Auswahl für spezifische Positionen sein.

Im Zusammenhang mit den Gruppenaufgaben in der OPZ liegt der Schwerpunkt des diagnostischen Interesses zur Planungskompetenz jedoch nicht auf den internen Prozessen der Bewerber, sondern auf ihrem sichtbaren **Planungsverhalten in der Gruppe**. Die intellektuellen Grundlagen der Bewerber werden an anderer Stelle im Verfahren getestet und die organisationsspezifischen methodischen Vorgehensweisen zur Planung werden die Offiziere im Laufe ihrer Ausbildung noch lernen. Aus diesem Grunde geht es auch nicht um das Resultat der Planungsbemühungen, wie es z.B. im Rahmen einer Postkorb-Übung bewertet wird. Dort wird aus den Resultaten der Planungsbemühungen eines einzelnen Bewerbers auf seine Planungskompetenz geschlossen. In der Gruppenaufgabe geht es vielmehr um eine prozessorientierte Bewertung des Planungsverhaltens in der Gruppe.

Planen ist „Probearbeiten“ (Dörner, 1989, S. 235). Damit Planung in der Gruppe effektiv sein kann, braucht es vor allem **Transparenz**, also eine Kommunikation der Gruppenmitglieder über ihre Planungsüberlegungen. Aus diesem Grunde wurde die Teildimensionen „**Strukturierung der Aufgabe**“ für die Bewertung ausgewählt. Mit dieser Dimension wird bewertet, ob die OB in der Gruppe sinnvolle Prioritäten ansprechen und transparent in Teilschritte zerlegen können, so dass die Voraussetzungen entstehen, Verantwortlichkeiten zu klären. Eine zweite Teildimension wurde „**Inhaltliche Weiterentwicklung**“ genannt. Die Besonderheit einer Planung in der Gruppe gegenüber einer Planung als Einzelleistung ist, dass ein grundsätzlich größeres **Kreativitätspotenzial** zur Verfügung steht. Dies kann jedoch nur positiv genutzt werden, wenn nicht eine einzige genannte Idee ausschließlich weiter verfolgt wird, sondern die Gruppenmitglieder ihre Kreativität einbringen, in dem sie neue Ideen nennen und sich nicht an der Diskussion eines einzigen Punktes festhalten. Dies waren die leitenden Überlegungen für die auf der Checkliste zu findenden Operationalisierungen zur Dimension Planungsverhalten.

2.4.6 Unabhängigkeit bzw. Abhängigkeit der Konstrukte

Eine Kritik, die häufig im Zusammenhang mit Gruppenaufgaben formuliert wird, ist der Vorwurf, die Konstrukte, die im Rahmen einer Aufgabe bewertet werden, seien voneinander nicht unabhängig. Die Forderung nach unabhängigen Konstrukten entstammt der klassischen Testtheorie und ist insbesondere für die Konstruktvalidität und hier für die Bestimmung

der diskriminanten Validität von elementarer Bedeutung. Es stellt sich in der Praxis der Personalauswahl anhand von Verhaltensbeobachtung jedoch als schwierig heraus, dieser berechtigten Forderung in vollem Umfang gerecht zu werden.

Im klassischen Testverfahren bei einem normierten Test kann die Ausprägung eines Merkmals unmittelbar, ohne Interpretationsprozesse, durch den Testleiter aus den Testwerten abgeleitet werden und die Merkmale werden jeweils einzeln erhoben. In der Gruppenaufgabe wird das Teilnehmerverhalten während einer Übung auf mehreren Dimensionen gleichzeitig beobachtet und bewertet, so dass man nicht von unabhängig voneinander erhobenen Variablen sprechen kann. Für die These, dass sich die zeitgleiche Erfassung mehrerer Dimensionen negativ auf die diskriminante Validität auswirkt, gibt es zwar nach Kenntnis der Autorin bis jetzt noch keine empirische Überprüfung, jedoch scheint sie plausibel, denn:

Eine erschöpfende Variation in den Verhaltensweisen einer Übung seitens der Teilnehmer ist aufgrund der zeitgleichen Beobachtung der dimensionsrelevanten Verhaltensweisen durch die Beobachter nur schwer realisierbar. (Kleinmann et al., 1995, S.23)

Die auf der Hand liegende Lösung für diese Problematik, in einer Gruppenaufgabe jeweils nur eine Dimension zu beobachten und zu bewerten, also die Anforderungsdimensionen sequentiell in verschiedenen Übungen einzusetzen, ist aus Effizienzgründen in der betrieblichen Praxis der Personalauswahl meist nicht möglich.

Neben der testtheoretischen Argumentation gibt es darüber hinaus inhaltliche Argumente, warum die Prüfpraxis der Forderung nach unabhängigen Konstrukten im Rahmen von Gruppenaufgaben nur unvollständig gerecht werden kann. In den Gruppenaufgaben werden die Dimensionen über Verhaltensbeobachtungen und damit über Kommunikation erfasst. D.h. jedes Konstrukt kann nur über die Vermittlung des Kommunikationsverhaltens der Teilnehmer erfasst werden, ganz nach dem Grundsatz „man kann nicht nicht kommunizieren“. Einzelne Verhaltensweisen müssen für den Bewertungsprozess also den jeweiligen Dimensionen zugeordnet werden und dieser Zuordnungsvorgang kann je nach Zieldimension mehr oder weniger eindeutig erfolgen.

Kleinmann et al. (1995) haben untersucht, in wie weit sich die Abhängigkeit von Dimensionen auf die diskriminante Validität auswirkt. Die Dimensionen in ihrer Studie waren: Einfühlungsvermögen, Beharrlichkeit, Kreativität, Entscheidungsfähigkeit, Analytische Fähigkeiten, Planung/Kontrolle, Überzeugungsvermögen und Durchsetzungsvermögen. Die eingesetzten Aufgaben waren Gruppendiskussion, Rollenspiel, Computersimulation und Präsentation. Die Abhängigkeit bzw. Unabhängigkeit der Dimensionen wurde durch ein Expertenrating dreier Diplom-Psychologen festgesetzt. Zwei Dimensionen sollten als abhängig gelten, wenn ein systematischer Zusammenhang zwischen ihnen angenommen wurde. Hier setzt auch eine Kritik an dieser Studie an. Eine theoriegestützte Ableitung der Einordnung, ob Dimensionen als eher abhängig oder eher unabhängig voneinander eingestuft werden, wäre eine interessante Ergänzung dieser Studie gewesen. Die oft sehr pragmatisch gestaltete Zuordnung von Dimensionen zu AC-Aufgaben hätte eine wichtige theoretische Entscheidungsgrundlage be-

kommen können. Sind zwei Dimensionen erfahrungsgemäß abhängig, sind sehr ähnliche Bewertungen und eine geringe diskriminante Validität nur konsequent. Entsprechend konnte gezeigt werden, dass die Unabhängigkeit/Abhängigkeit der Anforderungsdimensionen als Moderatorvariable für die diskriminante Validität wirkt.

Grundsätzlich ist festzuhalten, dass eine Forderung nach möglichst gegen Null gehenden diskriminanten Korrelationskoeffizienten praktisch kaum zu realisieren und theoretisch nur wenig sinnvoll scheint, da aus der Differentiellen Psychologie bekannt ist, dass viele Persönlichkeitsmerkmale miteinander korrelieren.

In der vorliegenden Studie ist zu vermuten, dass die Dimensionen Sozialkompetenz und Teamkompetenz eine höhere Abhängigkeit aufweisen als die übrigen Dimensionen untereinander. Hier wurde zum einen versucht, durch die VCL dennoch eine differenzierte Beobachtung zu ermöglichen. Die Zielsetzung geht auf der Grundlage des Anforderungsprofils und des Aufbaus des restlichen Prüfverfahrens eher in die Richtung einer Vertiefung der Beobachtung des kooperativen Verhaltens mit der Dimension Teamkompetenz, dass sicher nicht als unabhängig von der Fähigkeit zur Kommunikation (Dimension Sozialkompetenz) angenommen werden kann. Sollte die Praxis erweisen, dass auch nach einem Beobachtertraining keine differenzierte Bewertung dieser beiden Dimensionen möglich ist, wäre der Verzicht auf die Dimension Sozialkompetenz zur Entlastung der Prüfer eine sinnvolle Option.

2.5 Empfehlungen zur Verbesserung des Verfahrens

Auf der Basis des bisher Beschriebenen und dem dargestellten Stand der Forschung wurden die folgenden Empfehlungen zur Verbesserung des GSV erarbeitet.

2.5.1 Allgemeine Empfehlungen

- **Veränderung der Reihenfolge der GSV-Methoden**, also Verschiebung des Vortrages an das Ende des GSV. So wird jedem OB nach der Arbeit in der Gruppe noch einmal die Gelegenheit gegeben, vor der Gruppe und relativ unabhängig von der bisherigen Gruppendynamik eine Einzelleistung zu zeigen.
- **Reduzierung und Fokussierung der zu beobachtenden Dimensionen** auf jeweils drei pro Methode, um eine bessere Beobachtungsqualität zu ermöglichen. Die Dimensionen Argumentations- und Ausdrucksfähigkeit und Urteilsfähigkeit sollen bei der Bewertung der Planspiele entfallen, um eine Fokussierung auf die Dimensionen zu ermöglichen, für die das GSV den elementaren Prüfbaustein darstellt, nämlich Führungskompetenz, Sozial- und Teamkompetenz sowie Planungsverhalten.
- **Veränderung der Aufgabenstellung und der Beobachtungsmerkmale für den Vortrag**. Formulierung eines konkreten Szenarios, in dem der OB mit Hilfe vorgegebener Informationen eine Entscheidung fällen soll, weitere Schritte beschließen und beides dann vor

seinen Mitbewerbern vertreten und begründen soll mit dem Ziel, diese von seiner Entscheidung zu überzeugen. Das eigentliche anforderungsnahe Erkenntnisziel dieses Aufgabenteils, nämlich ob der OB grundsätzlich in der Lage ist, vor einer Gruppe eine führende Rolle zu übernehmen, wird so stärker fokussiert als bei der bisherigen Aufgabenstellung. Im Weiteren wird auf die Veränderungen zur Aufgabe „Vortrag“ nicht weiter eingegangen, da diese nicht im Zentrum der Fragestellungen dieser Arbeit stehen.

- **Entwicklung einer Beobachtungshilfe in Form einer Verhaltenscheckliste (VCL).** Diese soll nicht nur der allgemeinen Verbesserung der Beobachtungsqualität dienen, sondern auch ein einheitlicheres Verständnis der Dimensionen über die Prüfgruppen hinweg ermöglichen und so die Objektivität des Verfahrens erhöhen. Weiter stellt sie eine Möglichkeit dar, die Trennung von Beobachtung und Bewertung zu unterstützen.

Darüber hinaus wäre es wünschenswert, die Zusammenstellung der Gruppen nach Alter und Vorbildung homogener zu gestalten. Dies ist im Rahmen der aktuellen Organisationsform jedoch nicht im Gesamtverfahren zu verankern. In der Organisation der einzelnen Prüfräten, damit werden die OB bezeichnet, die zeitgleich durch das zweieinhalb tägige Verfahren laufen, wird bereits versucht, diesem Faktor Rechnung zu tragen.

2.5.2 Empfehlungen für die Planspiele

- **Ersatz der Gruppendiskussion durch ein zweites Planspiel,** um eine stärkere Differenzierung der OB zu ermöglichen. Wie in Kap. 2.3 beschrieben, ergab die Analyse des bestehenden Verfahrens, dass die Bewertung der OB im Rundgespräch deutlich weniger differenziert ausfiel als im Planspiel. Die Aufgabenstellung im Rundgespräch, zu einer abstrakten Themenstellung zu diskutieren, ließ offensichtlich nur wenig authentisches Gruppengeschehen aufkommen. Die Art der Themenauswahl (jeder OB wählt für sich ein Thema, danach entscheidet die Mehrheit) ließ die OB mit unterschiedlichem Engagement in die Diskussion starten. Der Hinweis an die OB, dass sie nicht zwangsläufig ihre eigene Meinung vertreten sollen, führte zu deutlichen Bemühungen, sozial erwünschte Meinungen zu äußern. Dies verstärkte die Künstlichkeit der Situation über das für eine Auswahlaufgabe normale Niveau hinaus und es kam zu wenig engagiertem Diskussionsverhalten. Die Diskussionen verliefen selten tatsächlich kontrovers. Die OB setzten sich in der Diskussion weniger mit ihren Mitbewerbern auseinander als mit den Anforderungen durch die Auswahl-situation. Die Bandbreite der gezeigten sozialen Interaktionen war daher oftmals gering. Durch die Vorgabe eines kurzen Szenarios, wie es im Planspiel erfolgt, tritt die Auswahl-situation stärker in den Hintergrund. Das Planspielszenario bietet mehr Verhaltensanreize, es müssen mehr Informationen verarbeitet werden als durch eine abstrakte These. Bei der Aufforderung zur Diskussion im Rundgespräch ist kein klares Ziel in der Aufgabe enthalten, dies vermindert vermutlich ebenfalls authentisches Verhalten, da das Ziel der OB, eine positive Bewertung zu erreichen, im Vordergrund bestehen bleibt und nicht durch ein klares Aufgabenziel relativiert wird, wie es im Planspiel vorgegeben wird.

- **Entwicklung neuer Planspielaufgaben**

Neue Planspielaufgaben sollten einerseits eine bessere Standardisierung des Gesamtverfahrens über die Prüfgruppen hinweg ermöglichen. Es sollte keinen Einfluss auf die Erfolgchancen der OB haben, welches Planspiel die Prüfer einer Prüfgruppe auswählen. Dies ist bei der Verschiedenartigkeit der bestehenden Planspiele nicht gegeben (s. Anhang). Aufgrund der Massendiagnostik müssen andererseits ausreichend Planspiele vorhanden sein, bzw. darüber hinaus in regelmäßigen Abständen neue zu entwickeln sein, ohne den Anspruch einer möglichst hohen Standardisierung aufzugeben. Dies ist für die Berufszufriedenheit der psychologischen Prüfer ebenso wichtig wie für die Beobachtungsqualität und die Chancengleichheit der Bewerber.

Die Eignung einer Gruppenaufgabe für einen Auswahlprozess hängt entscheidend von ihrer Anforderungsnähe und von ihrer Nähe zum Erfahrungshintergrund der Bewerber ab. Für den Auswahlprozess der OPZ bedeutet dies, dass eine Vielzahl von Planspielen eine möglichst einheitliche Anforderungsnähe aufweisen sollte sowie dass es nach Möglichkeit keine Bewerbergruppen geben sollte, deren Erfahrungshintergrund systematisch eine größere Nähe oder Distanz zu den Szenarios aufweist.

Eine im militärischen Zusammenhang nahe liegende Aufgabenart mit vermutlich hoher Anforderungsnähe sind strategische Aufgabenstellungen. Insbesondere für Trainings- und Ausbildungszwecke, aber auch für Auswahlzwecke im Karriereverlauf ist diese Aufgabenart sicher gut geeignet.

Für die vorliegende Auswahl-situation sind mit strategischen Aufgabenstellungen jedoch zwei Schwierigkeiten verbunden. Zum einen verfügen die OB über systematisch unterschiedliches Vorwissen zu militärischen Vorgehensweisen, je nachdem ob es sich um gediente oder ungediente Bewerber handelt. Zum anderen sind strategische Aufgaben in der Regel darauf angelegt, dass die Güte der inhaltlichen Bearbeitung ein wesentliches Bewertungskriterium darstellt. Entsprechend gibt es so etwas wie eine oder mehrere Musterlösungen. Aufgabenstellungen mit Musterlösungen sind für die Massendiagnostik der OPZ jedoch nicht gut geeignet, da sich Bewerber mit Kenntnis der Lösungen deutliche Vorteile verschaffen können. Zudem verschärft sich bei Gruppenaufgaben, die eine Bewertung der inhaltlichen Lösung vorsehen, das Problem, ein gleichmäßiges Schwierigkeitsniveau über die verschiedenen Aufgaben hinweg zu gewährleisten. Erfahrungsgemäß fällt es den militärischen Prüfern ohnehin schwer, den Schwerpunkt in der Bewertung der Gruppenaufgaben auf Verhaltensweisen zu legen und sich von der inhaltlichen Qualität der Äußerungen nicht übermäßig in ihrer Bewertung beeinflussen zu lassen. Eine strategische Aufgabenstellung würde diese Bewertungsvermischung vermutlich noch verstärken. Auch der Einfluss der Gruppenzusammensetzung auf die Bewertung des einzelnen Bewerbers, ohnehin ein zu beachtender Kritikpunkt bei dem Einsatz von Gruppenaufgaben zu Auswahlzwecken und unter dem Gesichtspunkt der Standardisierung, wird bei der Bewertung von inhaltlichen Kriterien erhöht. Es besteht dann die Gefahr, dass inhaltlich starke individuelle Lösungsvorschläge einzelner Bewerber die Bewertung der Verhaltensdimensionen überstrahlen.

Um angesichts der Tätigkeitsvielfalt in den späteren Zielpositionen dennoch Aufgaben zu entwickeln, die ein gleichmäßiges Niveau der Anforderungsnähe aufweisen, wurde auf die Standardisierung von bestimmten Situationsmerkmalen zurückgegriffen. Hierzu wurden Situationsmerkmale ausgewählt, bei denen von einer möglichst hohen Anforderungsnähe ausgegangen werden kann. Dies war zum einen der **Umgang mit einer Krisensituation**, zum anderen die möglichst kooperative Einigung über den **Umgang mit einer knappen Resource innerhalb des eigenen Teams**. Es wurde in Rücksprache mit dem Prüfungspersonal angenommen, dass diese beiden Situationstypen in allen späteren Offiziersverwendungen von Relevanz sind.

Es könnte durchaus sinnvoll sein, um diesen Ansatz der Aufgabengestaltung über die Augenscheinvalidität hinaus auch systematisch abzusichern, eine entsprechend gestaltete Anforderungsanalyse über die verschiedenen Verwendungsbereiche der Offiziere hinweg durchzuführen. Der Einsatz der Critical Incident Technic würde sich hier anbieten. Eine solche Untersuchung könnte weitere allgemein relevante Situationsklassen hervorbringen, die zur Weiterentwicklung anforderungsnaher Gruppenaufgaben genutzt werden können.

Eine weitere Leitidee für die Entwicklung der neuen Planspielaufgaben war es, durch die Gestaltung der Situationsmerkmale und eine entsprechend angepasste und standardisierte Diskussionsinstruktion insbesondere Verhaltensweisen der jeweiligen Zieldimensionen zu begünstigen. Es ist bekannt, dass sich insbesondere die Art der Zielverknüpfung der Gruppenmitglieder (kompetitiv oder kooperativ) auf das Interaktionsverhalten auswirkt. Echte kompetitive Aufgabenstellungen sind für die Auswahl der OB wenig sinnvoll, da die Bundeswehr in der Offiziersausbildung besonderen Wert auf die Entwicklung eines kooperativen Führungsstils legt.

Becker-Beck (1997) hat die Auswirkungen der Veränderungen der Diskussionsinstruktion auf die Interaktion unter einer anderen Fragestellung untersucht. Das Szenario für die Gruppendiskussionen bildete ein Ökologie-Ökonomie-Dilemma ab und die Gruppen wurden dazu aufgefordert, zu einer einheitlichen Entscheidung zu kommen. Die Diskussionsinstruktion wurde in Variante A so formuliert, dass die Gruppenmitglieder ihren eigenen Standpunkt dezidiert darstellen und ihn bei der Formulierung der Entscheidung möglichst weitgehend durchsetzen sollten. In Variante B sollten die Gruppenmitglieder die Standpunkte und Argumente der anderen Gruppenteilnehmer möglichst gut verstehen und bei der Formulierung der Entscheidung mit einbeziehen. Die Analyse des Gruppeninteraktionsprozesses mit Hilfe eines abgewandelten Symlog-Schemas ergab folgende Ergebnisse:

„Während bei den Gruppen mit „Verstehen“-Instruktion eine Tendenz besteht, auf kooperative Leistungsakte wiederum mit Kooperation zu reagieren, ist dieses Reaktionsmuster bei den „Durchsetzen“-Gruppen nicht überzufällig häufig. In den Gruppen mit „Durchsetzen“-Instruktion besteht die Tendenz, Konflikte intensiver auszutragen: Akte der Abgrenzung und Disziplinierung werden hier überzufällig häufig von Akten der Verweigerung/Entwertung gefolgt, was in der anderen Versuchsbedingung nicht der Fall ist. Bei den „Verstehen“-Gruppen besteht dagegen die Tendenz, auf Akte der Abgrenzung und Disziplinierung überzufällig häufig mit

Öffnung und Zustimmung zu reagieren, was auf den sozialen Prozess des Konfliktvermeidens und Nachgebens hindeutet.“ (Becker-Beck, 1997, S. 234)

Vorausgesetzt, dass die Rahmenbedingung der Auswahl-situation solche Wirkungen der Diskussionsinstruktion nicht moderiert, wird angenommen, dass das Szenario und die Diskussionsinstruktion in den Planspielen zum Krisenmanagement ähnliche Verhaltenstendenzen hervorruft wie in Variante A, während in den Planspielen zum Ressourcenengpass häufiger Verhaltensweisen wie in Variante B entstehen sollten. Auf diese Weise können die jeweiligen Zieldimensionen in den beiden Planspielarten besser beobachtet werden,

Zudem sollte durch die Entwicklung standardisierter Situationsmerkmale auch über die verschiedenen Planspiele hinweg ein vergleichbarer Rahmen für ihre Bewertung gewährleistet werden. Karkoschka stellt fest, dass man sozial kompetentes Verhalten nicht unabhängig von der Situation, in der die soziale Interaktion statt findet, betrachten kann (Karkoschka, 1998). Gleiches ist seit langem fester Bestandteil der Führungsforschung. Durch die gezielte Gestaltung der Situationsmerkmale in den beiden neuen Planspielgruppen wird versucht, den moderierenden Einfluss der Situation in der Bewertung der sozialen Verhaltensweisen zu verringern.

2.6 Die neuen Gruppenaufgaben und die Verhaltenscheckliste

In diesem Kapitel werden die Produkte der Weiterentwicklung des Verfahrens auf der Basis der bisher beschriebenen theoretischen und organisationsspezifischen Grundlage beschrieben. Es wurde zum einen die erste Version einer Verhaltenscheckliste entwickelt, die über den Zeitraum dieser Untersuchung hinaus im Rahmen des Projektes noch weiter erprobt und verändert wurde. Zum anderen wurden im Verlauf des Projektes insgesamt 20 neue Planspiele entwickelt. Jeweils 10 zur Situation Krisenmanagement und 10 zur Situation Ressourcenengpass. In der Datenerhebung zur vorliegenden Untersuchung wurde jeweils ein Planspiel jeder Bauart eingesetzt.

2.6.1 Entwicklung der Verhaltenscheckliste

Um ein Instrument zu erhalten, das an die Zielgruppe der OB und ihre Verhaltensweisen angepasst ist, wurden die Items der Verhaltenscheckliste (VCL) sowohl bei der Beobachtung des bestehenden Verfahrens gesammelt, als auch aus einer Sichtung der wenn auch kurzen Beschreibungen der Prüfer auf den bestehenden Beobachtungsbogen abgeleitet. Um den Beobachtungs- und Bewertungsprozess zu strukturieren und sowohl reflektierter als auch transparenter zu gestalten, wurden zu den Dimensionen die in Kap. 2.4. hergeleiteten Teildimensionen aufgenommen. Diese Teildimensionen sollen eine Fokussierung auf die wesentlichen Facetten der einzelnen Konstrukte unterstützen und ein über die Prüfgruppen hinweg einheitliches Konstruktverständnis unterstützen.

Die genaue Ausgestaltung der VCL wurde im Verlauf der Erprobung gemeinsam mit den Prüferinnen und Prüfern weiterentwickelt. Diese Veränderungen bezogen sich sowohl auf

einzelne Items als auch auf das Layout. Im Anhang ist die Version der VCL und der Bewertungsbogen beigelegt, die zum Zeitpunkt der Datenerhebung für diese Arbeit im Einsatz waren.

2.6.2 Die Planspielaufgaben zur Situation „Krisenmanagement“

Die neu entwickelten Planspielaufgaben sollen durch die Standardisierung bestimmter Situationsmerkmale zum einen eine bessere Vergleichbarkeit der Auswahl-situationen gewährleisten bei gleichzeitiger Aufgabenvielfalt. Zum anderen wurde die Idee verfolgt, die Szenarien so zu gestalten, dass sie insbesondere für die Zieldimensionen Verhaltensanreize bieten. Aus diesem Grund wurden für die Szenarios „Krisenmanagement“ Situationsmerkmale ausgewählt, die stärkere Verhaltensanreize für Führungsverhalten bieten.

Die neu entwickelten Planspielaufgaben „Krisenmanagement“ weisen alle die im Folgenden genannten Situationsmerkmale auf:

- Das Gruppenziel für die OB lautet, schnelle Entscheidungen zum Krisenmanagement zu treffen.
- Das Szenario verlangt, dass Verantwortung für andere übernommen werden muss.
- Auch in der Realsituation wäre eine Planungs- oder Besprechungsphase möglich. Damit soll eine leichtere Akzeptanz des Szenarios seitens der OB unterstützt werden.
- Eine hohe Quantität der Handlungsalternativen ist im Szenario möglich (informieren, Schaden vermeiden, planen, Risiken abschätzen, Image berücksichtigen etc.) → Es gibt keine Musterlösung.
- Zeitdruck wird durch die Aufgabenstellung und die vorgegebene Spielzeit von 15 Min. erzeugt.
- Eine Tafelskizze zu jedem Szenario schafft standardisierte Rahmenbedingungen und motiviert dazu, den Platz zu verlassen und die Tafel zu nutzen.

2.6.3 Die Planspielaufgaben zur Situation „Ressourcenengpass“

Die Szenarios „Ressourcenengpass“ enthalten Situationsmerkmale, die insbesondere Verhaltensanreize für die Zieldimension Teamkompetenz, wie sie in Kap. 2.5 definiert ist, bieten. Die neu entwickelten Planspielaufgaben „Ressourcenengpass“ weisen alle die im Folgenden genannten Situationsmerkmale auf:

- Das Gruppenziel für die OB lautet, eine von allen getragene Entscheidung herbeizuführen und den Erhalt des Gruppenzusammenhalts zu stützen.
- Das Szenario verlangt, dass Verantwortung für die eigene Gruppe übernommen werden muss.

- Eine begehrte, nicht-teilbare Ressource muss in der OB-Gruppe zugeteilt werden. Die Ressource wurde von allen gleichermaßen verdient, aber nicht alle können profitieren. Somit entsteht ein Konflikt zwischen Eigeninteresse und Gruppenzusammenhalt.
- Zeitdruck wird durch die Aufgabenstellung und die vorgegebene Spielzeit von 15 Min. erzeugt.
- Es gibt keine „einsamen Verlierer“. Bei einer Gruppe von drei OB bekommt nur einer die Ressource, bei vier OB können zwei profitieren;

Das neue Verfahren hat bezüglich der Planspiele folgende Struktur:

Planspiel zur Situation Krisenmanagement

- Führungskompetenz
- Sozialkompetenz
- Planungsverhalten

Planspiel zur Situation Ressourcenengpass

- Teamkompetenz
- Sozialkompetenz
- Planungsverhalten

Abbildung 2-5: Das neue Verfahren im Überblick – Methoden und Dimensionen

3 Methodik der empirischen Studie

3.1 Datenerhebung

3.1.1 Design

Das Design zur Datenerhebung im Rahmen der Erprobung der neuen Verfahren konnte nicht ausschließlich nach Forschungsgesichtspunkten gestaltet werden, da es im laufenden Prüfbetrieb der OPZ durchführbar sein musste. Dies bedeutete zum einen, dass die herkömmlichen Noten während der gesamten Erprobung weiter vergeben werden mussten, um die Bewerbungsgerechtigkeit für den Jahrgang sicher zu stellen. Dies stellte bei der Benotung der neuen Dimensionen eine deutliche Zusatzbelastung für die Beobachter dar. Zum anderen sollten die Bestandteile des neuen Verfahrens schrittweise eingeführt werden, da es aufgrund der angespannten Personalsituation keine Möglichkeit gab, vor die Erprobung ein Training zu schalten. Daher mussten die Prüfer im Erprobungszeitraum die Veränderungen im normalen Prüfalltag bewältigen können. Aus diesem Grunde konnte, anders als in anderen praxisorientierten Untersuchungen, die isolierte Wirkung der Interventionen ohne die „Störvariable“ des sonst üblichen Beobachtertrainings untersucht werden. Vor diesem Hintergrund entstand das folgende Design.

In die Datenerhebung waren vier Prüfgruppen eingebunden, die aus organisatorischen Gründen im Laufe der Erprobung personell unterschiedlich besetzt waren. Es war daher nicht möglich, den Personenfaktor auf Prüferseite zu kontrollieren.

Außerdem nahmen noch zwei externe Beobachter an der Datenerhebung teil. Diese gingen im Verlauf der Untersuchung jeweils gemeinsam in eine der vier Prüfgruppen. Dieses Vorgehen wurde gewählt, um einen Eindruck davon zu bekommen, in wie weit die Veränderungen der Urteilsvergabe durch die Ausbildung und berufliche Sozialisation der OPZ-Prüfer moderiert wird. Es war organisatorisch leider nicht möglich, bei den externen Beobachtern auf eine annähernd gleich große Fallzahl zu kommen wie bei den OPZ-Prüfern, so dass die Ergebnisse nur im Sinne von Tendenzen und Hinweisen zur Interpretation bewertet werden können, aber nur selten Signifikanz aufweisen.

An insgesamt zwölf Terminen wurden in diesen vier Prüfgruppen Daten erhoben. Als zentrales Referenzverfahren für den Persönlichkeitsbereich wurde zusätzlich der BIPTM (Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung) eingesetzt. Dieses Testverfahren wird in Kap. 3.1.2 kurz beschrieben und die Befunde zu den Hauptgütekriterien vorgestellt.

Für die Termine 7 und 8 stehen leider aufgrund einer organisatorischen Panne keine BIPTM - Daten zur Verfügung. Die OB waren nicht zur Codierung der Bögen aufgefordert worden. Die erhobenen Daten konnten nachträglich nicht mehr den GSV-Daten zugeordnet werden.

An den Terminen eins bis sechs wurde das herkömmliche Gruppensituationsverfahren durchgeführt. Die Prüfer durften jedoch das Planspiel nicht frei wählen. Aus den bestehenden Planspielen wurde eines ausgewählt und für diese Erprobungsphase festgelegt, um über die Prüfgruppen hinweg eine einheitliche Ausgangssituation zu gewährleisten.

Bei der Auswahl des Planspieles wurde berücksichtigt, dass es sich um ein gut etabliertes Planspiel handelt, das allen Prüfenden bekannt ist, und dass in diesem Szenario die besonderen Situationsmerkmale der neuen Planspiele nicht enthalten sind. Nach diesen Kriterien fiel die Wahl auf das Planspiel „Reiseplanung“. Für die Bewertung der Untersuchungsergebnisse bedeutet dies, dass die Basislinie relativ hoch angelegt ist, da es sich bei diesem Planspiel, zumindest dem Augenschein nach, um eines der besten vorhandenen handelt.

Um eine Beurteilungsgerechtigkeit im Bewerbervergleich des Einstellungsjahres zu gewährleisten, wurden während der gesamten Erprobungsphase die eignungsrelevanten Beurteilungen nach dem herkömmlichen Bewertungsraster vorgenommen. Diese werden im Weiteren als Urteile „alt“ bezeichnet.

An den Terminen eins und zwei wurde also das herkömmliche Verfahren durchgeführt, mit der einzigen Einschränkung, dass die Prüfer das Planspiel nicht frei aus dem bestehenden Fundus wählen konnten, sondern auf das Planspiel 1: „Reiseplanung“ festgelegt waren.

An den Terminen drei bis sechs wurde die neu entwickelte Verhaltenscheckliste für dieses Planspiel eingesetzt. Die mit Hilfe dieser VCL entstandenen Urteile werden im Weiteren als

Urteile „neu“ bezeichnet. Die Noten nach dem alten Beurteilungsschema wurden dann im Anschluss vergeben.

An den Terminen sieben bis zwölf wurde dann das neu entwickelte Verfahren eingesetzt. Es wurde ein festgelegtes Planspiel mit dem Situationsmerkmal Krisenmanagement eingesetzt sowie ein festgelegtes Planspiel mit dem Situationsmerkmal Ressourcenengpass und vier neue Vortragsthemen. An den Terminen sieben und acht wurden die neuen Planspiele ohne den Einsatz der VCL durchgeführt, an den Terminen neun bis zwölf wurde dann die VCL eingesetzt.

Zur Überprüfung der Hypothesen werden die Daten der 12 Einzeltermine in folgende Blöcke zusammengefasst:

Block 1: Ausgangslage

Block 2: Einsatz der VCL zum Planspiel 1

Block 3: Durchführung der neuen Verfahren

Block 4: Durchführung der neuen Verfahren und Einsatz der VCL

Die folgende Tabelle stellt dieses Design im zeitlichen Ablauf dar.

Tabelle 3-1: Übersicht zum Design der Datenerhebung

Termine	1 u. 2	3 – 6	7 u. 8	9 - 12
Block	1	2	3	4
Aufgaben	Alt	Alt	Neu	Neu
Urteilsverfahren	Alt(ohne VCL)	Neu (mit VCL) im Planspiel	Alt (ohne VCL)	Neu (mit VCL)
BIP™ (Bochumer Inventar zur berufsbezogenen Persönlich- keitsbeschreibung)	X	X		X

Im Rahmen der vorliegenden Arbeit soll die Qualität der herkömmlichen Gruppenaufgaben mit der Qualität der neu entwickelten Gruppenaufgaben verglichen werden. Dabei wird auf die gängigen Gütekriterien der klassischen Testtheorie zurückgegriffen.

Die Intervention bei den neuen Verfahren lässt sich unterteilen in

- den Einsatz der Verhaltenscheckliste und
- den Einsatz der neuen Planspielaufgaben.

Daraus ergeben sich die folgenden Kombinationsmöglichkeiten:

Tabelle 3-2: Untersuchungskonzept

	Ohne Verhaltenscheckliste	Mit Verhaltenscheckliste
Alte Gruppenaufgaben	Block 1	Block 2 (nur im Planspiel)
Neue Gruppenaufgaben	Block 3	Block 4

Um den Einfluss der Verhaltenscheckliste zu untersuchen, werden also die Daten aus dem Block 1 mit denen aus Block 2, sowie die Daten aus Block 3 mit denen aus Block 4 verglichen. Zur Untersuchung der Wirkung der neuen Aufgabenstellung werden die Daten aus Block 1 mit denen aus Block 3, sowie die Daten aus Block 2 mit denen aus Block 4 verglichen.

3.1.2 Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung

Der BIP™ ist ein Fragebogeninstrument zur Selbstbeschreibung. Er wurde aus mehreren Gründen als externes Referenz-Instrument ausgewählt. Erstens deckt er die besonders interessierenden Zieldimensionen Führungskompetenz, Sozialkompetenz und Teamkompetenz ab. Zweitens wurde er für Anwendungen im Personalmanagement konzipiert und nicht für eine klinische Diagnose. Drittens ist er ein aktuelles deutschsprachiges Instrument, das dem heutigen Sprachgebrauch gerecht wird. Dies erschien aufgrund des geringen Alters der OB besonders wichtig. Er erfasst 14 im Berufsleben relevante Persönlichkeitsfacetten mit Hilfe von jeweils 12 – 16 Items:

- Leistungsmotivation (14 Items) LM
- Gestaltungsmotivation (12 Items) GM
- Führungsmotivation (15 Items) FM
- Gewissenhaftigkeit (14 Items) GH
- Flexibilität (14 Items) FL
- Handlungsorientierung (14 Items) HO
- Sensitivität (12 Items) SE
- Kontaktfähigkeit (16 Items) KF
- Soziabilität (15 Items) SO
- Teamorientierung (13 Items) TO
- Durchsetzungsstärke (12 Items) DS
- Emotionale Stabilität (16 Items) ES
- Belastbarkeit (13 Items) BK
- Selbstbewusstsein (16 Items) SB

Die Skalen sind unipolar definiert. Der BIP™ arbeitet mit einer sechs-stufigen Skala („trifft voll zu“ bis „trifft überhaupt nicht zu“). Mit dieser Skala bleibt auch im Falle sozial erwünschten Verhaltens noch eine Varianz zwischen den Teilnehmern erhalten und es wird ihnen eine

Entscheidung abverlangt, da die Skala keinen Mittelpunkt aufweist (vgl. Hossiep, Paschen, 1998)

Der BIP™ ist ein vollständig neu entwickeltes Verfahren. Die Informationen zu seiner externen Validität beziehen sich auf konkurrenz erhobene Daten. Vor dem Hintergrund der Multi-determination der entsprechenden Kriterien liegen die Kennwerte des BIP™ an der oberen Grenze dessen, was durch einen Persönlichkeitsfragebogen erreicht werden kann (für ausführliche Informationen s. Hossiep, Paschen, 1998, S. 69 ff.).

Zur Auswertung werden verschiedene Normierungstabellen angeboten. Im Rahmen dieser Untersuchung wurde der BIP™ mit der 9-stufigen Normierung für unter 24-jährige ausgewertet.

Die Werte für Cronbachs alpha liegen in den für Persönlichkeitstests üblichen Bandbreiten. Die Skalen für Gestaltungsmotivation und Soziabilität weisen mit Cronbachs alpha = .75 den geringsten Wert auf, alle anderen Skalenwerte liegen zum Teil deutlich darüber. Dies spricht dafür, dass die Skalen inhaltlich ausreichend homogen sind. Auch eine Retest-Reliabilitätsmessung ergab zufrieden stellende Ergebnisse (Hossiep, Paschen, 1998, S. 24 ff.)

3.1.3 Stichprobe

Die Datenerhebung wurde zwischen dem 07.07.03 und dem 06.08.03 durchgeführt. An der Datenerhebung nahmen vier der insgesamt 10 Prüfgruppen der OPZ teil. Aus Gründen der Organisation der Datenerfassung in der OPZ ist es nicht möglich für die Beschreibung der Stichprobe auf die Daten der 183 Bewerberinnen und Bewerber aus diesen vier Prüfgruppen zuzugreifen. Es kann jedoch die Gesamtmenge der Bewerber, die in diesem Zeitraum die OPZ durchlaufen haben und aus der die Stichprobe stammt, beschrieben werden (s. Anhang, die folgenden Prozentangaben sind gerundet).

Insgesamt haben im Zeitraum der Datenerhebung 791 OB das Prüfverfahren durchlaufen. Davon sind 81 % männlich und 19 % weiblich.

Das Alter variiert zwischen 17 und 42 Jahren, wobei der Anteil der über 25-jährigen bei rund 4 % liegt. Der Mittelwert für das Alter in der Gesamtgruppe liegt bei 20,85 Jahren, die Standardabweichung bei 2,48.

Von den Bewerbern verfügen 67 % über eine Hochschulreife, 16 % über eine Fachhochschulreife und 12 % über die Mittlere Reife. Knapp 4 % der OB geben an, eine abgeschlossene Berufsausbildung zu haben. Knapp 1 % verteilt sich auf die Rubriken Hauptschulabschluss, abgeschlossenes Studium und fehlende Angaben.

Etwas mehr als ein Drittel der Bewerber (38 %) ist ungedient ohne Vordienstzeit. Der größte Teil der weiblichen Bewerber (88 %) gehört zu dieser Gruppe. Dennoch setzt sich diese Bewerbergruppe noch nicht zu gleichen Teilen aus männlichen (56 %) und weiblichen Bewerbern (44 %) zusammen.

Ein gleich großer Anteil der Bewerber kommt aus der Dienstzeit bei der Bundeswehr in die OPZ (38 %). In dieser Gruppe beträgt der Anteil der weiblichen Bewerber nur gut 3 %. Eine weitere nennenswerte Gruppe stellen die gedienten Bewerber mit Vordienstzeit (17 %). In dieser Gruppe gibt es keine weiblichen Bewerber. Der restliche Teil der Bewerber besteht

aus Bewerbern, die nicht zugeordnet werden konnten (4 %) sowie einigen Sonderfällen wie z.B. Bewerbern aus der Unteroffizierslaufbahn oder Reserveoffizierbewerbern.

Diese Sonderfälle haben durchaus für die Güte der Diagnostik während der Gruppenstation, in der sie auftreten, eine Bedeutung. Hier kommt die Problematik einer Massendiagnostik zum Tragen. Für eine effiziente methodische Gestaltung des Auswahlverfahrens ist diese Gruppe methodisch zu vernachlässigen. In absoluten Zahlen handelt es sich jedoch um über 200 Personen pro Jahr, die mit deutlich differierenden Voraussetzungen als die Masse der Bewerber durch das Auswahlverfahren laufen und einen Anspruch auf ein möglichst faires und valides Verfahren haben, bzw. deren Teilnahme auch die Ergebnisse der mit ihnen geprüften Bewerber möglichst nicht verzerren sollen. Auf diese Problematik soll im Rahmen dieser Arbeit jedoch nicht näher eingegangen werden. In der Praxis ist hier die Erfahrung der Prüfer gefragt, um mögliche Verzerrungen aufgrund dieser Sonderfälle gering zu halten.

Die Zuordnung der Bewerber in die Prüfgruppen, also auch in die vier Prüfgruppen, aus denen die Daten dieser Studie stammen, erfolgt zufällig, so dass von der Gruppe der Gesamtbewerber (791) auf die tatsächliche Stichprobe (183) mit ausreichender Sicherheit geschlossen werden kann. Da mögliche Personeneffekte auf Bewerberseite in der Analyse nicht im Vordergrund stehen und auch nicht kontrolliert werden konnten, soll diese Form der Stichprobenbeschreibung ausreichen.

3.2 Hypothesen der Arbeit

Ein linearer Kausalzusammenhang zwischen den Interventionen und den einzelnen Gütekriterien ist vor dem beschriebenen theoretischen Hintergrund nicht zu erwarten. Die Interventionen werden in Kombination mit einer Reihe weiterer Einflussfaktoren wirksam, die im Rahmen der Datenerhebung kaum kontrolliert werden konnten. Abbildung 3-1 stellt schematisch die wesentlichen Einflussfaktoren auf den Bewertungsprozess in der Gruppensituation dar vom Verhalten des OB über die Beobachtung durch die Prüfer bis hin zur Notenvergabe. Die Wirkzusammenhänge sind dabei vielfach wechselseitig und sich bedingend.

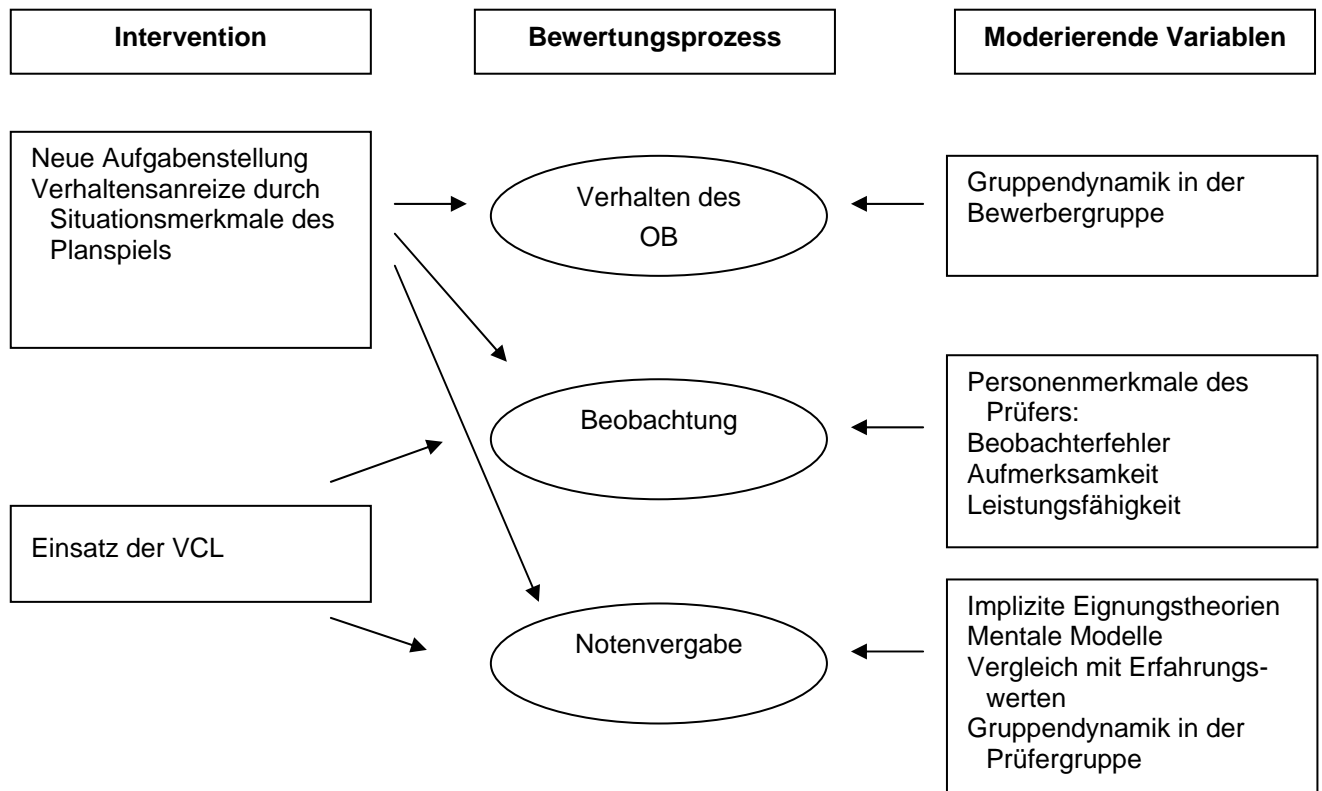


Abbildung 3-1: Wesentliche Einflussfaktoren auf den Bewertungsprozess

Die Abbildung beschränkt sich auf die für diese Untersuchung wesentlichen Einflussfaktoren und hat nicht den Anspruch einer vollständigen Darstellung des Interaktions- und Bewertungsprozesses. Es wird auf der Grundlage der dargestellten theoretischen und empirischen Basis von beiden Interventionen eine positive Wirkung auf die Gütekriterien der Benotung vermutet. Diese Annahme wird im folgenden Kapitel kurz erläutert.

3.2.1 Auswirkung der Interventionen auf den Bewertungsprozess

Es wird angenommen, dass die **Intervention der neuen Aufgabenstellung** sich auf alle drei Phasen des Bewertungsprozesses auswirkt und zu einer Qualitätsverbesserung der Gütekriterien führt. Insgesamt wird auf den verschiedenen Ebenen der Einfluss der moderierenden Variablen durch die Interventionen verringert.

In der Kleingruppenforschung wird zwischen Input, Gruppeninteraktionsprozess und Output unterschieden. Als Input werden dabei Faktoren wie Mitgliedermerkmale, Gruppenzusammensetzung, Aufgabenstellung und Situationsdefinition verstanden (Becker-Beck, 1997). Mitgliedermerkmale und Gruppenzusammensetzung sind im Setting dieser Studie als nicht kontrollierte moderierende Variablen aufzufassen, während die Veränderung der Aufgabenstellung eine der unabhängigen Variablen der Untersuchung darstellt.

Neben der Vermutung, dass die Situationsmerkmale der neuen Planspiele eine direkte Wirkung auf das Verhalten der OB haben, in dem sie mehr oder weniger starke bzw. spezifische Verhaltensanreize bieten, kann davon ausgegangen werden, dass sich die Veränderung der Aufgabenstellung mit den neuen Planspielen in jedem Falle indirekt, und zwar über die Beeinflussung des Gruppeninteraktionsprozesses auf das Verhalten der OB auswirkt. Dies entspricht einer Reduzierung des Einflusses der moderierenden Variable Gruppeninteraktionsprozess.

Der Einfluss der neuen Aufgaben auf die Beobachtungsleistung der Prüfer bezieht sich unter anderem auf ihren Neuigkeitswert. Es ist den Prüfern nicht mehr möglich, sich auf bestimmte Schlüsselverhaltensweisen bzw. inhaltliche Äußerungen der Bewerber zu beschränken, die sich jeder Prüfer im Laufe der häufig jahrelangen Prüferfahrungen mit den bestehenden Planspielen angeeignet hat. Es wird vermutet, dass die neuen Szenarien eingefahrene individuelle Beobachtungsroutinen auflockern und die Aufmerksamkeit der Prüfer erhöhen. Führen die neuen Aufgaben tatsächlich zu einer vermehrten Aktivität der OB hinsichtlich der Zieldimensionen, erhöht sich auch die Chance, dass diese Aktivitäten angemessen beobachtet werden. Die Reduzierung der Zieldimensionen in den Planspielen sollte ebenfalls zu einer Erhöhung der Beobachtungsqualität führen.

Diese Argumentation lässt sich auf den Einfluss der neuen Aufgaben auf die Benotung übertragen. Viele Prüfer haben im Laufe der Jahre Vorlieben für einige der bestehenden Planspiele entwickelt und auch Vorstellungen zum Idealverhalten der Bewerber. Diese mentalen Modelle variieren jedoch relativ stark zwischen den Prüfern und auch den Prüfgruppen. Die Einführung der neuen Planspiele führt dazu, dass im Bewertungs- und damit auch im Benotungsprozess wieder verstärkt und umfassender auf die Beobachtungen zurückgegriffen wird. Die Fokussierung auf weniger Zieldimensionen sollte die Zuordnung der Beobachtungen zu den Zieldimensionen erleichtern und zu einer Erhöhung der Bewertungsqualität führen. Die Standardisierung der Situationsmerkmale sollte eine gleichmäßigere Bewertung der sozialen Verhaltensweisen unterstützen.

Die **Intervention durch den Einsatz der VCL** wirkt positiv auf die Qualität der Beobachtung und der Notenvergabe. Durch die Teildimensionen der Konstrukte wird eine Fokussierung der Beobachtung unterstützt und die Zuordnung der relevanten beobachteten Verhaltensweisen zu den Zieldimensionen erleichtert. Durch die instrumentelle Trennung von Beobachtung und Bewertung wird der Tendenz, früh ein Gesamturteil zu fällen, entgegengewirkt. Der Bewertungsbogen unterstützt die Prüfer dabei, ihre Notengebung an den ausgewählten Teildimensionen zu orientieren und verringert den Einfluss irrelevanter Beobachtungen. Die Definition der Zieldimensionen durch die Subdimensionen und die ihr zugeordneten Items unterstützt ein gleichsinniges Konstruktverständnis zwischen den Prüfern und über die Prüfgruppen hinweg.

Die Wirkung der Interventionen wird anhand der üblichen testtheoretischen Gütekriterien überprüft. Die Bestimmung der prognostischen Validität war weder im Rahmen des Projektes

noch im Rahmen dieser Arbeit möglich. Dies ist begründet durch die lange Zeitspanne, die zwischen der Diagnose der OB und dem Abschluss der Ausbildung liegt. Es ergeben sich die folgenden Hypothesen zu den weiteren Gütekriterien.

3.2.2 Objektivität / Reliabilität

H_0 : Die Beobachterreliabilität ist mit VCL kleiner oder genau so groß als ohne.

Hypothese 1: die Beobachterreliabilität ist mit VCL höher als ohne.

- a) Bei den alten Aufgaben
 - Bei den OPZ-Prüfern
 - Bei den externen Beobachtern
- b) Bei den neuen Aufgaben
 - Bei den OPZ-Prüfern
 - Bei den externen Beobachtern

H_0 : Die Beobachterreliabilität ist bei den neuen Aufgaben kleiner oder genau so groß als bei den alten Aufgaben.

Hypothese 2: Die Beobachterreliabilität ist bei den neuen Aufgaben größer als bei den alten.

- a) ohne VCL
 - Bei den OPZ-Prüfern
 - Bei den externen Beobachtern
- a) mit VCL
 - Bei den OPZ-Prüfern
 - Bei den externen Beobachtern

3.2.3 Differenzierung der Bewerber

H_0 : Die Differenzierung der OB gelingt mit VCL schlechter oder genau so gut als ohne.

Hypothese 3: Die Differenzierung der OB gelingt mit VCL besser als ohne.

- a) Bei den alten Aufgaben
 - Bei den OPZ-Prüfern
 - Bei den externen Beobachtern
- b) Bei den neuen Aufgaben
 - Bei den OPZ-Prüfern
 - Bei den externen Beobachtern

H₀: Die Differenzierung der OB gelingt bei den neuen Aufgaben schlechter oder genau so gut als bei den alten Aufgaben.

Hypothese 4: Die Differenzierung der OB gelingt bei den neuen Aufgaben besser als bei den alten.

a) ohne VCL

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

b) mit VCL

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

3.2.4 Interne Konstruktvalidität

Veränderung im Zusammenhang mit der VCL:

H₀: Die konvergente Validität ist mit VCL kleiner oder genau so groß als ohne.

Hypothese 5

1. Die konvergente Validität ist mit VCL höher als ohne.

Bei den neuen Aufgaben

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

H₀: Die diskriminante Validität ist mit VCL kleiner oder genau so groß als ohne.

2. Die diskriminante Validität ist mit VCL höher als ohne.

a) Bei den alten Aufgaben

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

b) Bei den neuen Aufgaben

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

Veränderung im Zusammenhang mit den Aufgaben:

Hypothese 6:

H₀: Die konvergente Validität ist bei den neuen Aufgaben kleiner oder genau so groß als bei den alten.

1. Die konvergente Validität ist bei den neuen Aufgaben höher als bei den alten.

a) ohne VCL

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

b) mit VCL

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

H₀: Die konvergente Validität ist bei den neuen Aufgaben kleiner oder genau so groß als bei den alten Aufgaben.

2. Die diskriminante Validität ist bei den neuen Aufgaben höher als bei den alten.

a) ohne VCL

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

b) mit VCL

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

3.2.5 Externe Konstruktvalidität

Veränderung im Zusammenhang mit der Verhaltenscheckliste:

Hypothese 7:

H₀: Die externe konvergente Validität ist mit VCL kleiner oder genau so groß als ohne.

1. Die externe konvergente Validität ist mit VCL höher als ohne.

a) Bei den alten Aufgaben

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

b) Bei den neuen Aufgaben

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

H₀: Die externe diskriminante Validität ist mit VCL kleiner oder genau so groß als ohne.

2. Die externe diskriminante Validität ist mit VCL höher als ohne.

a) Bei den alten Aufgaben

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

b) Bei den neuen Aufgaben

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

Hypothese 8:

H₀: Die externe konvergente Validität ist bei den neuen Aufgaben kleiner oder genau so groß als bei den alten Aufgaben.

1. Die externe konvergente Validität ist bei den neuen Aufgaben höher als bei den alten.

a) ohne VCL

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

b) mit VCL

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

H₀: Die externe diskriminante Validität ist bei den neuen Aufgaben kleiner oder genau so groß als bei den alten Aufgaben.

2. Die externe diskriminante Validität ist bei den neuen Aufgaben höher als bei den alten.

a) ohne VCL

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

b) mit VCL

- Bei den OPZ-Prüfern
- Bei den externen Beobachtern

3.3 Statistische Verfahren

Die Auswertung der vorliegenden Studie basiert im Wesentlichen auf den folgenden zwei statistischen Methoden, die im nächsten Kapitel kurz vorgestellt werden sollen. Zur Berechnung der Beobachterreliabilität wird die Intraklassenkorrelation eingesetzt, zur Überprüfung der Konstruktvalidität der Gruppenaufgaben wird auf die Multitrait-Multimethod-Matrix zurückgegriffen.

3.3.1 Die Intraklassenkorrelation als Zusammenhangsmaß und als Varianzaufklärungsmaß

Eine Beurteilung ist dann reliabel, wenn andere Beurteiler mit gleichem Wissensstand zu einem ähnlichen Urteil kommen. Urteilen mehrere Rater ähnlich, also reliabel, dann wird üblicherweise davon ausgegangen, dass von diesen Ratern auch das gleiche Merkmal erfasst wird. Die Einschätzungen der Beobachter in dieser Arbeit liegen auf einer Ratingskala von 1 bis 7 vor, die als intervallskaliert angenommen wird. Daher stellt die Intraklassenkorrelation eine mögliche Methode zur Reliabilitätsbestimmung dar (Wirtz & Caspar, 2002, S. 157).

„Die Intraklassenkorrelation basiert auf dem varianzanalytischen Modell und kann ähnlich interpretiert werden, wie die Produkt-Moment-Korrelation: Ein Wert von 0 indiziert, dass kein Zusammenhang zwischen den Urteilen der Rater besteht. Je mehr sich der Koeffizient dem Wert 1 annähert, desto besser ist der Zusammenhang und desto höher ist die Reliabilität der Urteile“ (ebda).

In fast allen Arbeiten zur Beobachterreliabilität in Assessment Centern wird die Produkt-Moment-Korrelation verwendet. Es gibt jedoch zwei wesentliche Gründe, die für die Verwendung der Intraklassenkorrelation in dieser Arbeit sprechen.

1. Die ICC kann für beliebig viele Rater gemeinsam berechnet werden. Da in der OPZ im Allgemeinen 3 Beobachter die gleichen Teilnehmer bewerten, ist dies ein großer Vorteil gegenüber der Produkt-Moment-Korrelation.
2. Es können verschiedene ICCs berechnet werden, die entweder die Mittelwertsunterschiede zwischen Ratern als Fehlerquelle verrechnen (ICC_{unjust}) oder einen Reliabilitätskoeffizienten liefern, der vom Effekt unterschiedlicher Ratermittelwerte bereinigt ist

(ICC_{just}). Mit welcher ICC die Reliabilität beurteilt werden sollte, hängt von der Art der weiteren Verwendung der untersuchten Daten ab. Für die OPZ gilt, dass abschließend die Bewerber anhand der Daten, die von verschiedenen Prüfern erhoben wurden, in eine Eignungsreihenfolge gebracht werden. D. h. Mittelwertsunterschiede zwischen den Prüfern sind hier als Fehlerquelle anzusehen, die die Zuverlässigkeit der Beurteilungen systematisch beeinträchtigen. Die ICC_{unjust} ist also das richtige Reliabilitätsmaß. Für die ICC_{unjust} werden unterschiedliche Ratermittelwerte nicht korrigiert, daher ist dieser Koeffizient in der Regel niedriger als der der ICC_{just}.

Die Produkt-Moment-Korrelation bietet diese Möglichkeiten nicht. Sie ist immer ein adjustiertes Zusammenhangsmaß, das durch Mittelwertsunterschiede zwischen den Beobachtern nicht verändert wird.

Voraussetzung für die Anwendung der ICC ist eine Normalverteilung der Daten. Isolierte Verletzungen der Normalverteilungsannahme können nach Glass, Peckham und Sander (1972) vernachlässigt werden (zitiert nach Wirtz & Caspar, 2002).

Die Berechnungsformel für die ICC hat varianzanalytische Maße zur Grundlage. Für den vorliegenden Fall ist das einfaktorielle Modell der Reliabilität anzuwenden, da die Teilnehmer von jeweils unterschiedlichen Ratergruppen beurteilt wurden. Die Formel für die ICC_{unjust, einfakt.} lautet folgendermaßen:

$$\rho = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\omega}^2}$$

wobei σ_{α}^2 für die Merkmalsvarianz steht und σ_{ω}^2 für die Fehlervarianz. Die Reliabilität ρ ist also definiert als der Anteil der Varianz der wahren Werte an der Gesamtvarianz (Wirtz & Caspar, 2002, S. 168 ff).

Mit der ICC kann daher zum einen die Frage beantwortet werden, wie hoch die Stärke des Zusammenhangs zwischen den Raterurteilen ist. Zum anderen gibt sie an, welcher Anteil der Varianz der wahren Merkmalsausprägung durch die Beurteilungen eines Raters aufgeklärt wird.

Auf die Höhe der ICC_{unjust, einfakt.} wirken zwei Einflussgrößen. Zum einen die Homogenität der Beurteilungen eines Bewerbers durch verschiedene Beobachter, zum anderen der Unterschied zwischen den Mittelwerten der Bewerber. Je stärker sich die bewerbenspezifischen Mittelwerte unterscheiden und je weniger sich die Werte für die einzelnen Bewerber jeweils unterscheiden, desto höher ist der Wert der ICC_{unjust, einfakt.}

Die Signifikanz der ICC_{unjust, einfakt.} hängt davon ab, ob sich die Werte der beurteilten Bewerber in einer einfaktoriellen Varianzanalyse signifikant voneinander unterscheiden. Aus diesem Grund kann es bei der Berechnung der ICC_{unjust, einfakt.} zu kleinen Werten kommen, die dennoch signifikant sind.

3.3.2 Die Multitrait-Multimethod-Matrix (MTMM-Matrix)

Zur empirischen Überprüfung der Konstruktvalidität psychologischer Testverfahren stellt das von Campbell und Fiske (1959) entwickelte Multitrait-Multimethod Untersuchungsdesign das etablierteste Verfahren dar. In diesem Design werden mehrere verschiedene Traits (Konstrukte) mit Hilfe mehrerer unabhängiger Messungen erhoben und miteinander korreliert (s. Kap. 1.3.3). Werden drei Traits jeweils durch drei Methoden erhoben, so können die entsprechenden Interkorrelationen der insgesamt 9 Tests in einer MTMM-Matrix dargestellt werden (vgl. Ostendorf et al., 1986, S.5):

		Methode								
		I			II			III		
Trait		x	y	z	x	y	z	x	y	z
I	x	r_{x1x1}								
	y	r_{x1y1}	r_{y1y1}							
	z	r_{x1z1}	r_{y1z1}	r_{z1z1}						
II	x	r_{x1x2}	r_{x2y1}	r_{x2z1}	↖ Monomethod Blocks ↘					
	y	r_{x1y2}	r_{y1y2}	r_{y2z1}						
	z	r_{x1z2}	r_{y1z2}	r_{z1z2}						
III	x		↗		r_{x3x2}	r_{x3y2}	r_{x3z2}	r_{x3x3}		
	y	↘ Heteromethod Blocks			r_{x2y3}	r_{y2y3}	r_{y3z2}	r_{x3y3}	r_{y3y3}	
	z		↘		r_{x2z3}	r_{y2z3}	r_{z2z3}	r_{x3z3}	r_{y3z3}	r_{z3z3}

Abbildung 3-2: MTMM-Matrix für drei Traits (x,y,z) und drei Methoden (I,II,III)

Campbell und Fiske (1959) schlagen vier Kriterien zur Überprüfung der Konstruktvalidität vor. Das Kriterium 1 steht für konvergente Validität, die Kriterien zwei bis vier für diskriminante Validität.

- Kriterium 1:** Die monotrait Koeffizienten der heteromethod Blocks werden als konvergente Validitäten bezeichnet. Diese sollen signifikant größer 0 sein, z.B. r_{x1x2} , r_{y1y2} , $r_{z1z2} > 0$.
- Kriterium 2:** Die monotrait-heteromethod Korrelationskoeffizienten sollen größer sein als die entsprechenden heterotrait-heteromethod Korrelationskoeffizienten, z.B. $r_{x1x2} > r_{x1y2}$, r_{x1z2} , r_{x2y1} , r_{x2z1} .
- Kriterium 3:** Die monotrait-heteromethod Korrelationskoeffizienten sollen größer sein als die entsprechenden heterotrait-monomethod Korrelationskoeffizienten, z.B. $r_{x1x2} > r_{x1y1}$, r_{x1z1} , r_{x2y2} , r_{z2z2} .
- Kriterium 4:** Die Muster der Traitinterkorrelationen innerhalb der verschiedenen Methodenblöcke (monomethod und heteromethod) sollen gleich sein. Dabei wird der Term „gleich“ von verschiedenen Autoren unterschiedlich interpretiert. So fordert z.B. Alwin (1974, zitiert nach Ostendorf et al. 1986, S. 7) „dass die Rangreihe der Interkorrelationen innerhalb der mono- und heteromethod Matrizen konstant ist.“

Aufgrund der beschriebenen Datenlage ergeben sich in der vorliegenden Studie keine vollständigen 3x3-Korrelationsmatrizen. Die Vergleichsmöglichkeiten, in wie weit die Kriterien eins bis vier sich im Sinne der aufgestellten Hypothesen zwischen den Termin-Blöcken 1 bis 4 unterscheiden, sind daher nicht vollständig vorhanden. Aus diesen Gründen wird sich die MTMM-Analyse auf zwei Kriterien beschränken:

- Die Korrelationen der Urteile zu einem Konstrukt aus verschiedenen Gruppenaufgaben sollen signifikant größer 0 und möglichst groß sein (Kriterium 1, konvergente Validität).
- Die Korrelation der Urteile zu verschiedenen Konstrukten aus einer Gruppenaufgabe soll möglichst klein sein und kleiner als die konvergenten Koeffizienten (Kriterium 3, diskriminante Validität).

4 Ergebnisdarstellung I: Datenmaterial, Reliabilität und Differenzierung der Bewerber

Die Überprüfung der Normalverteilung der Beobachterurteile wurde mit Hilfe des Kolmogorov-Smirnov-Testes mit Lilliefors-Korrektur durchgeführt. Laut diesem Test kann nicht von einer Normalverteilung der Daten ausgegangen werden. Der Signifikanztest für Korrelationen ist jedoch äußerst robust gegenüber Verletzungen der Verteilungsannahme (Bortz, 1999). Ähnliches gilt für die Intraklassenkorrelation (Wirtz & Caspar, 2002), hier können isolierte Verletzungen der Normalverteilung vernachlässigt werden. Aus diesem Grund wurde graphisch mit Hilfe von Histogrammen überprüft, ob die Daten extreme Abweichungen von der Normalverteilung aufweisen, z.B. extrem schief verteilt sind, zu extremen Werten kann es aufgrund der verwendeten 7-stufigen Skala ja nicht kommen. Bei dieser graphischen Überprüfung konnten einige leicht schiefe Verteilungen festgestellt werden, jedoch keine extrem schiefen.

Nach diesen Analyseschritten wurde entschieden, die ICC zu verwenden. So kann der Vorteil genutzt werden, die Übereinstimmung von mehr als zwei Ratern prüfen zu können.

4.1 Deskriptive Darstellung des Datenmaterials

Als Datenmaterial für die Untersuchung stehen die Urteile der OPZ-Beobachter und der externen Beobachter sowie die BIPTM-Daten zur Verfügung. Die Daten verteilen sich wie in Tabelle 4.1. dargestellt auf die verschiedenen, bereits in Kap. 3.1.1. beschriebenen Untersuchungsabschnitte.

Tabelle 4-1: Datenmaterial deskriptiv, Häufigkeiten

	Block 1	Block 2	Block 3	Block 4
N	32	61	30	60
Urteile OPZ Beobachter	92	142	76	124
Urteile externe Beobachter	16	16	16	32
BIP TM	26	45	-	44

N bezeichnet die Anzahl der Teilnehmer, die in den vier Erhebungsprüfgruppen beobachtet und bewertet wurden. Die unterschiedliche Anzahl der Urteile der OPZ-Beobachter resultiert daraus, dass nicht immer alle Prüfgruppen mit drei Beobachtern besetzt waren.

Die nachfolgende Tabelle 4-2. stellt die Mittelwerte der Urteile der OPZ-Beobachter und der externen Beobachter zu den Dimensionen Führung, Planung und Sozialkompetenz und die

jeweilige Standardabweichung dar. Für diese Mittelwertbildung wurden alle Prüfgruppenmittelwerte (also Endurteile je Teilnehmer) herangezogen, die im Rahmen eines Blocks zu einer Dimension gefällt wurden.

Diese Mittelwerte schwanken um den Wert 4,00 herum, wobei der niedrigste Mittelwert aus den OPZ-Urteilen bei 3,81 liegt und der höchste bei 4,34. Von 20 Mittelwerten liegen 11 über 4,00 und 9 unter 4,00. Auffällig ist dabei, dass sämtliche Mittelwerte der Urteile zu Sozialkompetenz unter 4,0 liegen, also tendenziell ein positiveres Bild der Teilnehmer liefern, während die Werte zu den Führungs- und Planungsdimensionen bis auf eine Ausnahme alle über 4,0 liegen, also tendenziell eine etwas schlechtere Bewertung der Teilnehmer durch die OPZ-Prüfer abbilden. Während die Werte zu den Dimensionen Führung und Planung also leicht rechts schief verteilt sind, sind die Werte zur Dimension Sozialkompetenz leicht links schief verteilt.

Die Mittelwerte aus den Urteilen der externen Beobachter liegen ausnahmslos unter denen der OPZ-Beobachter, zum Teil rund 0,5 Punkte, d.h. die externen Beobachter zeichnen ein positiveres Bild der Bewerber als die OPZ-Beobachter.

Die Standardabweichung schwankt zwischen 0,4 und 1,27 für die Werte der OPZ-Beobachter und zwischen 0,44 und 1,54 für die Werte der externen Beobachter. Auffallend ist, dass auch hier die Daten zur Dimension Sozialkompetenz von den Daten zu den anderen Dimensionen abweichen. Die Standardabweichungen sowohl für die Werte der OPZ-Beobachter als auch für die Werte der externen Beobachter sind für die Dimension Sozialkompetenz tendenziell niedriger als für die Werte der Dimensionen Führung und Planung.

Tabelle 4-2: Beobachterurteile deskriptiv 1, Mittelwert und Standardabweichung SD

Konstrukt - Aufgabe - Beobachter	Block 1 ohne VCL		Block 2 mit VCL im PS	
	Mittelwert	SD	Mittelwert	SD
Führung - PS alt - OPZ	4,13	0,79	4,27	1,27
Führung - PS alt - ex. Beo.	3,44	1,18	4,00	1,54
Führung - RG alt - OPZ	3,99	0,61	4,11	0,80
Führung - RG alt - ex. Beo.	3,63	0,95	3,81	1,00
Planung - PS alt - OPZ	4,27	0,79	4,34	1,21
Planung - PS alt - ex. Beo.	3,50	0,76	3,94	1,45
Sozialk. - PS alt - OPZ	3,88	0,48	3,92	0,88
Sozialk. - PS alt - ex. Beo.	3,38	0,58	3,50	0,71
Sozialk. - RG alt - OPZ	3,83	0,40		
Sozialk. - RG alt - ex. Beo.	3,38	0,74		
Konstrukt - Aufgabe - Beobachter	Block 3 ohne VCL		Block 4 mit VCL	
	Mittelwert	SD	Mittelwert	SD
Führung - PS KM - OPZ	4,16	1,07	4,08	1,07
Führung - PS KM - ex. Beo.	3,63	0,99	3,75	0,95
Führung - PS RE - OPZ	4,08	0,98		
Führung - PS RE - ex. Beo.	3,50	1,10		
Planung - PS KM - OPZ	4,26	1,02	4,07	1,02
Planung - PS KM - ex. Beo.	3,44	0,73	3,81	0,54
Planung - PS RE - OPZ			4,00	0,94
Planung - PS RE - ex. Beo.			3,66	1,06
Sozialk. - PS KM - OPZ	3,96	0,64	3,86	0,75
Sozialk. - PS KM - ex. Beo.	3,38	0,44	3,69	0,83
Sozialk. - PS RE - OPZ	3,81	0,62	3,87	0,72
Sozialk. - PS RE - ex. Beo.	3,13	0,83	3,56	0,66
Teamk. - PS RE - OPZ			3,90	0,91
Teamk. - PS RE - ex. Beo.			3,62	0,99

Die nachfolgende Tabelle 4-3 stellt die Mittelwerte der Urteile der OPZ-Beobachter und der externen Beobachter zu den Dimensionen Argumentationsverhalten/Ausdrucksfähigkeit sowie Urteilsfähigkeit und die jeweilige Standardabweichung dar.

Tabelle 4-3: Beobachterurteile deskriptiv 2, Mittelwert und Standardabweichung SD

Konstrukt - Aufgabe - Beobachter	Block 1		Block 2	
	Mittelwert	SD	Mittelwert	SD
Arg. u. Ausdrucksf.- PS alt - OPZ	4,09	0,38	4,13	0,75
Arg. u. Ausdrucksf.- PS alt - ex. Beo.	3,45	0,42	4,00	1,22
Arg. u. Ausdrucksf.- RG alt - OPZ	4,01	0,56	4,20	0,78
Arg. u. Ausdrucksf.- RG alt - ex. Beo.	3,31	0,59	3,75	0,76
Urteilsfähigkeit- PS alt - OPZ	4,15	0,61	4,21	0,89
Urteilsfähigkeit - PS alt - ex. Beo.	3,75	0,46	3,93	0,98
Urteilsfähigkeit - RG alt - OPZ	4,03	0,72	4,20	0,86
Urteilsfähigkeit - RG alt - ex. Beo.	3,75	0,46	4,00	0,53
Konstrukt - Aufgabe - Beobachter	Block 3		Block 4	
	Mittelwert	SD	Mittelwert	SD
Arg. u. Ausdrucksf. - PS KM - OPZ	4,07	0,61	3,97	0,68
Arg. u. Ausdrucksf. - PS KM - ex. Beo.	3,63	0,58	3,69	0,70
Arg. u. Ausdrucksf. - PS RE - OPZ	4,15	0,58	4,01	0,65
Arg. u. Ausdrucksf. - PS RE - ex. Beo.	3,50	0,89	3,44	0,57
Urteilsfähigkeit - PS KM - OPZ	3,96	0,67	3,94	0,73
Urteilsfähigkeit - PS KM - ex. Beo.	3,69	0,26	3,84	0,35
Urteilsfähigkeit - PS RE - OPZ	4,06	0,71	4,03	0,76
Urteilsfähigkeit - PS RE - ex. Beo.	3,69	0,46	3,81	0,40

Die Dimensionen aus Tabelle 4-3 sind nicht auf der VCL enthalten. Die Mittelwerte der OPZ-Beobachter schwanken zwischen 3,97 und 4,21. Die geringste Standardabweichung liegt bei 0,38 und die höchste bei 0,89.

Die Mittelwerte der externen Beobachter schwanken zwischen 3,31 und 4,0. D.h. auch für diese beiden Dimensionen zeichnen die externen Beobachter ein tendenziell positiveres Bild der Bewerber als die OPZ-Beobachter. Die Standardabweichungen schwanken zwischen 0,26 und 1,22.

4.2 Ergebnisse zur Hypothese 1 – Beobachterreliabilität und Verhaltenscheckliste

Hypothese 1: Die Beobachterreliabilität steigt mit dem Einsatz einer VCL.

Zur Überprüfung der Beobachterreliabilität wird, wie in Kap. 3.3.1 begründet, die ICC_{unjust, einfakt.} verwendet. Es werden nur die Bewerber und die über sie gefällten Urteile berücksichtigt, bei denen die Beobachterkommission der OPZ aus 3 Prüfern bestand. Wie in Kap. 3.3.1 beschrieben, hängt die Signifikanz der ICC unter anderem von den Mittelwertsunterschieden der beurteilten Bewerber ab. Bei den relativ kleinen Fallzahlen kann es daher auch zu Artefakten kommen, wenn ein oder zwei Bewerbergruppen deutlich homogener waren als ande-

re. Die ICC-Werte werden in erster Linie deskriptiv als Maß für die Beurteilerübereinstimmung ausgewertet.

Der Vollständigkeit halber und als weitere Informationsquelle werden auch die Ergebnisse zu den Dimensionen dargestellt, die nicht auf der Verhaltenscheckliste enthalten sind: Argumentationsverhalten/Ausdrucksfähigkeit und Urteilsfähigkeit. Werte, die im Sinne der Hypothese ausfallen werden im Weiteren fett gedruckt.

4.2.1 Beobachterreliabilität und Verhaltenscheckliste bei den alten Aufgaben

Die nachfolgende Tabelle 4-4 zeigt die ICC-Werte für die Urteile der OPZ-Beobachter zum alten Planspiel ohne VCL (Block 1) und mit VCL (Block 2):

Tabelle 4-4: ICC-Werte der OPZ-Beobachter zum alten Planspiel ohne und mit VCL

3 OPZ-Prüfer, altes Planspiel	Block 1, ICC _{unjust, einfakt.} Planspiel ohne VCL, n=28	Block 2, ICC _{unjust, einfakt.} Planspiel mit VCL, n=22
Führungsfähigkeit	.64 ***	.66 ***
Soziale Kompetenz	.28 ***	.59 ***
Planungsverhalten	.48 ***	.54 ***
Argumentationsverhalten/Ausdrucksfähigkeit	.17 ***	.56 ***
Urteilsfähigkeit	.50 ***	.42 ***

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Die ICC_{unjust, einfakt.} sind in allen Dimensionen hoch signifikant. Die ICC_{unjust, einfakt.} steigt bei allen drei Dimensionen, die auf der Verhaltenscheckliste enthalten sind. Auch die ICC_{unjust, einfakt.} für Ausdrucksfähigkeit/Argumentationsverhalten steigt deutlich an, obwohl diese Dimension nicht auf der Verhaltenscheckliste enthalten ist. Die ICC_{unjust, einfakt.} für die Dimension Urteilsfähigkeit sinkt dagegen.

Die nachfolgende Tabelle 4-5 gibt die ICC_{unjust, einfakt.} für die Urteile der externen Beobachter zum alten Planspiel ohne VCL (Block 1) und mit VCL (Block 2) wieder:

Tabelle 4-5: ICC-Werte der externen Beobachter zum alten Planspiel ohne und mit VCL

2 externe Beobachter, altes Planspiel	Block 1, ICC _{unjust, einfakt.} Planspiel ohne VCL, n=8	Block 2, ICC _{unjust, einfakt.} Planspiel mit VCL, n=8
Führungsverhalten	.66 *	.85 ***
Soziale Kompetenz	.29	.46
Planungsverhalten	.64 *	.56*
Ausdrucksfähigkeit, Argumentationsverhalten	.30	.78 **
Urteilsfähigkeit	.26	.82 **

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Bei den externen Beobachtern steigt die $ICC_{\text{unjust, einfakt.}}$ in allen Dimensionen bis auf die Dimension Planungsverhalten an. Einige $ICC_{\text{unjust, einfakt.}}$ sind nicht signifikant, darunter die Werte für Soziale Kompetenz.

4.2.2 Beobachterreliabilität und Verhaltenscheckliste bei den neuen Aufgaben

In den Blöcken 3 und 4 wurden 2 neue Planspiele eingesetzt. Die nachfolgende Tabelle 4-6 zeigt die $ICC_{\text{unjust, einfakt.}}$ für die Urteile der OPZ-Prüfer für das Planspiel Ressourcenengpass ohne und mit VCL.

Tabelle 4-6: ICC-Werte der OPZ-Beobachter zum neuen Planspiel RE ohne und mit VCL

3 OPZ-Prüfer, Planspiel Ressourcenengpass	Block 3, $ICC_{\text{unjust, einfakt.}}$ Planspiel ohne VCL, n=16	Block 4, $ICC_{\text{unjust, einfakt.}}$ Planspiel mit VCL, n=18
Führungsfähigkeit	.74 ***	---
Soziale Kompetenz	.42 **	.54 ***
Planungsverhalten	---	.51***
Teamkompetenz	---	.64***
Ausdrucksfähigkeit, Argumentationsverhalten	.44 **	.35 **
Urteilsfähigkeit	.16	.29 *

Signifikanzniveau: *** $p < 0,001$ ** $p < 0,01$ * $p < 0,05$

Für das Planspiel Ressourcenengpass sind die Dimensionen Sozialkompetenz, Planungsverhalten und Teamkompetenz auf der VCL enthalten. Im alten Beurteilungsverfahren ohne VCL werden an der Stelle, die das Planspiel Ressourcenengpass einnimmt, die Dimensionen Führung, Soziale Kompetenz, Ausdrucksfähigkeit/Argumentationsverhalten sowie Urteilsfähigkeit bewertet. Aus diesem Grund ergibt sich nur für die Dimension Soziale Kompetenz eine direkte Vergleichsmöglichkeit. Hier ist ein gesteigener ICC-Wert festzustellen.

Die $ICC_{\text{unjust, einfakt.}}$ für die Urteile der OPZ-Beobachter sind für die Dimensionen, die auf der Verhaltenscheckliste enthalten sind, Soziale Kompetenz, Planungsverhalten und Teamkompetenz, durchgängig signifikant. Bei den Dimensionen Argumentationsverhalten/Ausdrucksfähigkeit und Urteilsfähigkeit ist für die Entwicklung der ICC keine einheitliche Tendenz festzustellen. Der Basis-Wert für Urteilsfähigkeit ist nicht signifikant.

Die folgende Tabelle 4-7 zeigt die ICC-Werte der OPZ-Beobachter für das Planspiel Krisenmanagement.

Tabelle 4-7: ICC-Werte der OPZ-Beobachter zum neuen Planspiel KM ohne und mit VCL

3 OPZ-Prüfer, Planspiel Krisenmanagement	Block 3, ICC _{unjust, einfakt.} Planspiel ohne VCL, n=20	Block 4, ICC _{unjust, einfakt.} Planspiel mit VCL, n=20
Führungsfähigkeit	.83***	.55***
Soziale Kompetenz	.41**	.39**
Planungsverhalten	.73**	.51***
Argumentationsverhalten/Ausdrucksfähigkeit	.58***	.19
Urteilsfähigkeit	.59***	.20

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Alle ICC-Werte der OPZ-Prüfer sind signifikant und mit Einsatz der VCL geringer als ohne. Auch die Werte für Argumentationsverhalten/Ausdrucksfähigkeit und Urteilsfähigkeit sind mit Einsatz der VCL geringer und zudem nicht signifikant.

Die nachfolgende Tabelle 4-8 gibt die ICC_{unjust, einfakt.} für die Urteile der externen Prüfer für das Planspiel Ressourcenengpass ohne und mit VCL wieder.

Tabelle 4-8: ICC-Werte der externen Beobachter zum neuen Planspiel RE ohne und mit VCL

Externe Beobachter, Planspiel Ressourcenengpass	Block 3, ICC _{unjust, einfakt.} Planspiel ohne VCL, n=8	Block 4, ICC _{unjust, einfakt.} Planspiel mit VCL, n=16
Führungsfähigkeit	.73**	---
Soziale Kompetenz	.70*	.47*
Planungsverhalten	---	.74**
Teamkompetenz	---	.68**
Argumentationsverhalten/Ausdrucksfähigkeit	.85**	.27
Urteilsfähigkeit	.74**	.44*

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Auch hier steht für das Planspiel Ressourcenengpass wieder nur ein direkter Vergleichswert für die Dimensionen der VCL zur Verfügung. Für die Dimension Sozialkompetenz ist die IC_{unjust, einfakt.} gesunken. Die ICC-Werte für die Dimensionen Planungsverhalten und Teamkompetenz erreichen mit VCL befriedigende Werte von .74 bzw. .68. Die ICC-Werte für die beiden Dimensionen, die nicht auf der VCL enthalten sind, sind deutlich gesunken. Alle Werte sind signifikant bis auf den Wert für Argumentationsverhalten/Ausdrucksfähigkeit mit VCL. Die folgende Tabelle 4-9 zeigt die ICC-Werte der externen Beobachter für das neue Planspiel Krisenmanagement ohne und mit VCL.

Tabelle 4-9: ICC-Werte der externen Beobachter zum neuen Planspiel KM ohne und mit VCL

Externe Beobachter, Planspiel Krisenmanagement	Block 3, ICC _{unjust, einfakt.} Planspiel ohne VCL, n=8	Block 4, ICC _{unjust, einfakt.} Planspiel mit VCL, n=16
Führungsfähigkeit	.77 **	.81***
Soziale Kompetenz	.02	.70***
Planungsverhalten	.55	.41*
Argumentationsverhalten/Ausdrucksfähigkeit	.69 *	.59**
Urteilsfähigkeit	-.40	-.06

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Die ICC-Werte aus dem Planspiel Krisenmanagement sind für Führungsfähigkeit und Soziale Kompetenz gestiegen, für Planungsverhalten gesunken. Die Ausgangswerte für Soziale Kompetenz und Planungsverhalten sowie die Werte für Urteilsfähigkeit sind nicht signifikant. Für die Dimension Urteilsfähigkeit ergeben sich negative ICC-Werte. Hierzu ist zu beachten, dass laut Wirtz und Caspar (2002) negative ICC-Werte nicht im Sinne eines negativen Zusammenhangs interpretiert werden dürfen, sondern im Sinne eines nicht vorhandenen Zusammenhangs interpretiert werden müssen.

4.2.3 Zusammenfassung zur Hypothese 1: Verhaltenscheckliste und Beobachterreliabilität

Für die alten Aufgaben sprechen die Vergleiche der ICC-Werte der OPZ-Prüfer für eine Bestätigung der Hypothese, dass der Einsatz der VCL die Reliabilität erhöht. Bei den externen Beobachtern weisen 2 von 3 Vergleichen in die Richtung der Hypothese. Interessant ist die Tatsache, dass sich auch der ICC-Wert für die Dimension Argumentationsverhalten/Ausdrucksfähigkeit bei beiden Beobachtergruppen verbessert, obwohl diese Dimension nicht auf der Verhaltenscheckliste enthalten ist.

Bei den neuen Aufgaben ergibt sich ein sehr uneinheitliches Bild. Im Planspiel Ressourcenengpass steigt die ICC für die Dimension Sozialkompetenz bei den OPZ-Prüfern und sinkt bei den externen Beobachtern.

Im Planspiel Krisenmanagement sinken die ICC-Werte der OPZ-Prüfer in allen Dimensionen der VCL, bei den externen Beobachtern steigen die ICC-Werte für die Dimensionen Führung und Sozialkompetenz, während der ICC-Wert für Planungsverhalten sinkt. Daraus ergeben sich keine eindeutigen Hinweise auf einen Zusammenhang des Effektes mit einer bestimmten Dimension oder Beobachtergruppe. Die Hypothese muss für die neuen Aufgaben abgelehnt werden.

Die nicht auf der VCL enthaltenen Dimensionen Ausdrucksfähigkeit und Urteilsfähigkeit weisen ebenfalls in beiden Spielen und bei beiden Beobachtergruppen fast durchgängig sinken-

de ICC-Werte mit Einsatz der VCL bei den neuen Aufgaben auf. 6 von 8 Vergleichswerten zu diesen Dimensionen sind mit Einsatz der VCL geringer.

4.3 Ergebnisse zur Hypothese 2 – Beobachterreliabilität und neue Aufgaben

Hypothese 2: Die Beobachterreliabilität steigt mit dem Einsatz der neuen Aufgaben.

Zur Überprüfung der Beobachterreliabilität in Verbindung mit der Veränderung der Aufgabenstellung werden die ICC-Werte zu den Urteilen der Blöcke 1 und 3 bzw. 2 und 4 gegenübergestellt.

4.3.1 Beobachterreliabilität, alte versus neue Aufgaben ohne VCL

Die nachfolgende Tabelle 4-10 zeigt die ICC-Werte der OPZ-Beobachter beim Rundgespräch und dem neuen Planspiel Ressourcenengpass. Die Urteile wurden jeweils ohne den Einsatz einer VCL ermittelt. Relevant für die Beurteilung der Hypothese sind nur die Dimensionen, die bei den neuen Aufgaben in Zukunft bewertet werden sollen.

Tabelle 4-10: ICC-Werte der OPZ-Beobachter, Rundgespräch versus PS RE ohne VCL

3 OPZ-Prüfer, ohne VCL	Block 1, ICC _{unjust, einfakt.} Altes Rundgespräch n=28	Block 3, ICC _{unjust, einfakt.} Planspiel RE n= 16
Führungsfähigkeit	.44***	.74***
Soziale Kompetenz	.08	.42**
Argumentationsverhalten/Ausdrucksfähigkeit	.30**	.44**
Urteilsfähigkeit	.42***	.16

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Die ICC_{unjust, einfakt.} ist in allen Dimensionen signifikant bis auf den Ausgangswert zur Sozialen Kompetenz und den Vergleichswert zur Urteilsfähigkeit.

Die ICC_{unjust, einfakt.} ist für das neue Planspiel zum Ressourcenengpass bis auf den Wert zur Urteilsfähigkeit durchgängig höher als beim Rundgespräch

Tabelle 4-11: ICC-Werte der OPZ-Beobachter, PS versus PS KM ohne VCL

3 OPZ-Prüfer, ohne VCL	Block 1, ICC _{unjust, einfakt.} Altes Planspiel, n=28	Block 3, ICC _{unjust, einfakt.} Planspiel KM n= 20
Führungsfähigkeit	.64 ***	.83***
Soziale Kompetenz	.28 ***	.41**
Planungsverhalten	.48 ***	.73***
Argumentationsverhalten/Ausdrucksfähigkeit	.17 ***	.58***
Urteilsfähigkeit	.50 ***	.59***

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Der Vergleich zwischen altem Planspiel und neuem Planspiel zum Krisenmanagement zeigt, dass alle Werte für die ICC_{unjust, einfakt.} signifikant und im neuen Planspiel höher sind als im alten.

Die folgenden Tabellen 4-12 und 4-13 zeigen die ICC-Werte für die externen Beobachter bei den alten und neuen Aufgaben. Die Urteile wurden jeweils ohne Verwendung der VCL vergeben.

Tabelle 4-12: ICC-Werte der externen Beobachter, Rundgespräch versus PS RE, ohne VCL

Externe Beobachter, ohne VCL	Block 1, ICC _{unjust, einfakt.} Altes Rundgespräch n=8	Block 3, ICC _{unjust, einfakt.} Planspiel RE n= 8
Führungsfähigkeit	.66*	.73 **
Soziale Kompetenz	.63*	.70 *
Argumentationsverhalten/Ausdrucksfähigkeit	.39	.85 **
Urteilsfähigkeit	.26	.74 **

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Die Ausgangswerte für die Dimensionen Ausdrucksfähigkeit und Urteilsfähigkeit sind nicht signifikant. Im Vergleich zwischen altem Rundgespräch und neuem Planspiel zum Ressourcenengpass sind alle ICC_{unjust, einfakt.} beim neuen Planspiel höher.

Tabelle 4-13: ICC-Werte der externen Beobachter, Planspiel versus PS KM, ohne VCL

Externe Beobachter, ohne VCL	Block 1, ICC _{unjust, einfakt.} Altes Planspiel, n=8	Block 3, ICC _{unjust, einfakt.} Planspiel KM n= 8
Führungsfähigkeit	.66 *	.77 **
Soziale Kompetenz	.29	.02
Planungsverhalten	.64 *	.55
Argumentationsverhalten/Ausdrucksfähigkeit	.30	.69 *
Urteilsfähigkeit	.26	-.40

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Von den 10 Werten sind nur 4 signifikant, zu diesen gehören beide Werte zu Führungsfähigkeit. Im Vergleich zwischen altem Planspiel und neuem Planspiel zum Krisenmanagement ergibt sich ein uneinheitliches Bild. Für die Dimensionen Soziale Kompetenz und Planungsverhalten sind die ICC_{unjust, einfakt.} gesunken, für Führungsverhalten und Ausdrucksfähigkeit/Argumentationsverhalten sind sie gestiegen, während für die Dimension Urteilsfähigkeit mit einem negativen ICC-Wert kein interpretierbares Ergebnis entstanden ist.

4.3.2 Beobachterreliabilität, alte versus neue Aufgaben mit Verhaltenscheckliste

In der folgenden Tabelle 4-14 werden die ICC_{unjust, einfakt.} der Urteile der OPZ-Prüfer bei den alten und neuen Aufgaben dargestellt, die Urteile wurden jeweils mit Einsatz der Verhal-

tenscheckliste ermittelt. Hier kann nur das alte Planspiel dem neuen Planspiel Krisenmanagement gegenübergestellt werden, da zum Rundgespräch keine VCL eingesetzt wurde.

Tabelle 4-14: ICC-Werte der OPZ-Beobachter, altes Planspiel versus PS KM, mit VCL

3 OPZ-Prüfer, mit VCL	Block 2, ICC _{unjust, einfakt.} Planspiel alt, n=22	Block 4, ICC _{unjust, einfakt.} PS KM , n= 20
Führungsfähigkeit	.66	.55***
Soziale Kompetenz	.59	.39**
Planungsverhalten	.54	.51***
Argumentationsverhalten/Ausdrucksfähigkeit	.56 ***	.19
Urteilsfähigkeit	.42 ***	.20

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Die 3 Ausgangswerte der relevanten Dimensionen sind nicht signifikant. Alle ICC aus dem neuen Planspiel Krisenmanagement sind niedriger als diejenigen aus dem alten Planspiel.

Die folgende Tabelle 4-15 zeigt die ICC-Werte der externen Beobachter zu den Urteilen aus dem alten Planspiel und dem neuen Planspiel Krisenmanagement.

Tabelle 4-15: ICC-Werte der externen Beobachter, altes Planspiel versus PS KM, mit VCL

3 OPZ-Prüfer, mit VCL	Block 2, ICC _{unjust, einfakt.} Planspiel alt, n=8	Block 4, ICC _{unjust, einfakt.} PS KM , n= 16
Führungsfähigkeit	.85 ***	.81 ***
Soziale Kompetenz	.46	.70***
Planungsverhalten	.56*	.41*
Ausdrucksfähigkeit	.78 **	.59**
Urteilsfähigkeit	.82 **	-.06

Signifikanzniveau: ***p<0,001 **p<0,01 *p<0,05

Der Wert der ICC_{unjust, einfakt.} für die Dimension Sozialkompetenz ist gestiegen, wobei der Ausgangswert nicht signifikant ist. Alle weiteren Werte der ICC_{unjust, einfakt.} sind gesunken, bzw. für die Dimension Urteilsfähigkeit negativ und damit nicht interpretierbar.

4.3.3 Zusammenfassung zur Hypothese 2: neue Aufgaben und Beobachterreliabilität

Die Ergebnisse zeichnen ein geteiltes Bild. Für den Vergleich der Reliabilität bei alten und neuen Aufgaben ohne Einsatz der VCL bestätigen die ICC-Werte der zukünftig relevanten Dimensionen die Hypothese für die OPZ-Beobachter für beide neue Aufgaben. Für die Urteile der externen Beobachter steigen die ICC-Werte im Planspiel Ressourcenengpass. Im Planspiel Krisenmanagement steigt gerade der ICC-Wert für die Dimension Führung. D.h.

die Reliabilität der externen Beobachter steigt genau in der Dimension, die im Focus dieses Planspiels steht.

Für den Vergleich der Reliabilität bei alten und neuen Aufgaben mit Einsatz der VCL muss die Hypothese abgelehnt werden. Für die Urteile der OPZ-Beobachter zeigt kein ICC-Wert eine Verbesserung an, für die Urteile der externen Beobachter ist nur 1 von 5 ICC-Werten gestiegen.

4.4 Ergebnisse zur Hypothese 3: Differenzierung der OB und VCL

Hypothese 3: Die Differenzierung der Bewerber gelingt mit VCL besser als ohne.

Zur Überprüfung der Hypothesen 3 und 4 wird auf die Tabelle 4-2 (s. S.71) zurückgegriffen. Anhand der Standardabweichung lässt sich beurteilen, wie weit die Prüfgruppen mit ihren Urteilen zu den einzelnen Bewerbern vom Mittelwert abweichen, also wie deutlich sie die Bewerber differenzieren können. Für die Mittelwertbildung wurden alle Urteile, die im Rahmen eines Blocks zu einer Dimension gefällt wurden, herangezogen.

Für die Bewertung der Hypothesen 3 und 4 werden nur die Dimensionen herangezogen, die auch im neuen Verfahren weiterhin bewertet werden, also Führung, Planung und Sozial- bzw. Teamkompetenz. Zur Überprüfung der Hypothese 3 können nur die Werte herangezogen werden, für die der Basiswert ohne VCL und der Vergleichswert mit VCL entstanden ist. Da in Block 2 die VCL nur für das Planspiel eingesetzt wurde und in Block 4 Führung im Ressourcenengpassspiel nicht mit VCL beobachtet wurde, reduziert sich die Anzahl der verwertbaren Daten. Dies gilt entsprechend auch für die Überprüfung des 2. Teils der Hypothese 4 (für eine Darstellung der Gruppenaufgaben und Urteilsdimensionen siehe Tabelle AA im Anhang)

4.4.1 Die Differenzierung der Bewerber gelingt bei den alten Aufgaben mit VCL besser als ohne

Tabelle 4-2 zeigt (s. S. 91), dass die Hypothese für die Urteile der OPZ-Beobachter akzeptiert werden kann. Die Standardabweichung steigt im Vergleich von Block 1 zu Block 2 in allen Dimensionen. Gleiches gilt für die Urteile der externen Beobachter. Auch hier steigt die Standardabweichung in allen drei Dimensionen.

4.4.2 Die Differenzierung der Bewerber gelingt bei den neuen Aufgaben mit VCL besser als ohne

Für die Urteile der OPZ-Beobachter zu den neuen Aufgaben lässt sich die Hypothese nur für die Dimension Soziale Kompetenz bestätigen. Für die Dimensionen Führung und Planung im Planspiel Krisenmanagement bleiben die Standardabweichungen auf dem gleichen relativ hohen Niveau, für die Dimension Sozialkompetenz steigt sie in beiden Planspielen.

Für die Urteile der externen Beobachter muss die Hypothese verworfen werden. Von den 4 Vergleichsmöglichkeiten sinkt die Standardabweichung für die Dimensionen Führung und Planung im Planspiel Krisenmanagement und für die Dimension Sozialkompetenz im Planspiel Ressourcenengpass. Nur die Standardabweichung für die Dimension Sozialkompetenz im Krisenmanagementplanspiel steigt mit dem Einsatz der Verhaltenscheckliste.

4.4.3 Zusammenfassung zur Hypothese 3: Die Differenzierung der OB gelingt mit VCL besser als ohne

Für die alten Aufgaben kann die Hypothese sowohl für die Ergebnisse der OPZ-Beobachter als auch für die der externen Beobachter klar akzeptiert werden.

Für den Einsatz der VCL bei den neuen Aufgaben fallen 2 von 4 Vergleichswerten der OPZ-Prüfer hypothesenkonform aus, die beiden anderen Vergleichswerte verschlechtern sich nicht, sondern bleiben konstant. Für die Ergebnisse der externen Beobachter muss die Hypothese verworfen werden.

4.5 Ergebnisse zur Hypothese 4: Differenzierung der OB und neue Aufgaben

Hypothese 4: Die Differenzierung der Bewerber gelingt mit den neuen Aufgaben besser als mit den alten.

4.5.1 Die Differenzierung der OB gelingt mit den neuen Aufgaben besser als mit den alten (jeweils ohne VCL)

Der Vergleich der Standardabweichungen aus Block 1 und 3 in Tabelle 4-2 zeigt, dass die Urteile der OPZ-Beobachter bei den neuen Aufgaben durchweg eine größere Standardabweichung aufweisen, die Hypothese also akzeptiert werden kann.

Für die Urteile der externen Beobachter fallen von 5 Vergleichswerten nur 2 im Sinne der Hypothese aus. Für die Dimensionen Führung, Planung und Sozialkompetenz sind im Vergleich von altem Planspiel mit neuem Planspiel zum Krisenmanagement die Standardabweichungen gesunken. Im Vergleich Rundgespräch mit Planspiel Ressourcenengpass sind für die Dimensionen Führung und Sozialkompetenz die Standardabweichungen gestiegen. Es ergibt sich kein einheitliches Bild für die beiden neuen Aufgaben.

4.5.2 Die Differenzierung der OB gelingt mit den neuen Aufgaben besser als mit den alten (jeweils mit VCL)

Für die Urteile der OPZ-Beobachter muss die Hypothese verworfen werden. Alle 3 Vergleichswerte sind bei den neuen Aufgaben niedriger.

Für die Urteile der externen Beobachter gilt, dass 2 von 3 Standardabweichungen bei den neuen Aufgaben geringer sind als bei den alten. Lediglich für die Dimension Sozialkompetenz im Vergleich von altem Planspiel mit neuem Planspiel zum Krisenmanagement ist die Standardabweichung gestiegen.

4.5.3 Zusammenfassung zur Hypothese 4: Die Differenzierung der OB gelingt mit den neuen Aufgaben besser als mit den alten

Die Hypothese kann nur für die Urteile der OPZ-Beobachter, die ohne den Einsatz der VCL entstanden sind, akzeptiert werden. In diesem Falle gelingt die Differenzierung der OB mit den neuen Aufgaben besser als mit den alten. Für die Urteile der externen Beobachter und für die Urteile, die mit Einsatz der VCL entstanden sind, muss die Hypothese verworfen werden.

5 Ergebnisdarstellung II: Interne Konstruktvalidierung

Auch die Ergebnisse der internen Konstruktvalidierung werden getrennt nach den beiden Interventionen „Einsatz der VCL“ sowie „Einsatz der neuen Aufgaben“ dargestellt. Die Überprüfung der Hypothesen erfolgt anhand der Logik der in Kap. 3.3.2 vorgestellten Multi-Trait-Multi-Method-Matrix (MTMM). In den folgenden Tabellen zur Konstruktvalidierung gibt n die Anzahl der beurteilten Bewerber an, die den Validitätsberechnungen in den einzelnen Blöcken zugrunde liegen. Der dargestellte Validitätskoeffizient stellt eine Produkt-Moment-Korrelation nach Pearson dar. Auch in diesem Untersuchungsteil sind zum Teil nicht signifikante Korrelationskoeffizienten entstanden. Diese treten insbesondere bei den Daten der externen Beobachter aufgrund der niedrigen Fallzahlen auf. Aus diesem Grund ist eine Verallgemeinerung dieser Ergebnisse zur Konstruktvalidierung von der Stichprobe auf die Grundgesamtheit nur eingeschränkt möglich.

5.1 Ergebnisse zur Hypothese 5: Konstruktvalidität und VCL

Hypothese 5: Die Konstruktvalidität ist mit VCL größer als ohne

5.1.1 Die konvergente Validität ist mit VCL größer als ohne

Die Möglichkeiten zum Vergleich der konvergenten Validitäten sind aufgrund der Datenlage nur in begrenztem Umfang vorhanden. Es werden Urteile aus verschiedenen Aufgaben zu einer Dimension benötigt, um die konvergente Validität zu bestimmen. D.h. es werden Urteile aus zwei Aufgaben zu einer Dimension benötigt, die beide ohne VCL ermittelt wurden, sowie Urteile aus zwei Aufgaben zur gleichen Dimension, die beide mit VCL ermittelt wurden. Diese Voraussetzung wird lediglich erfüllt für die Urteile zur Sozialkompetenz in Block 3 und 4. Die entsprechenden konvergenten Validitäten für Sozialkompetenz werden in Tab. 5.1 dargestellt.

Tabelle 5-1: Konvergente Validität ohne und mit VCL

	Ohne VCL, Block 3	Mit VCL, Block 4
Sozialkompetenz, PS KM und PS RE, OPZ-Prüfer	.63** n=30	.78** n=52
Sozialkompetenz, PS KM und PS RE, ext. Beobachter	.53 n=8	.80** n=16

(** auf dem Niveau von 0,01 zweiseitig signifikant)

Sowohl für die Urteile der OPZ-Prüfer als auch für die Urteile der externen Beobachter steigt die konvergente Validität für die Dimension Sozialkompetenz mit dem Einsatz der VCL bei den neuen Aufgaben. Der Basiswert für die externen Beobachter ist nicht signifikant.

Die Hypothese wird für die Dimension Sozialkompetenz für beide Beobachtergruppen akzeptiert. Für die weiteren Dimensionen lassen sich aufgrund der beschriebenen Datenstruktur keine konvergenten Validitätskoeffizienten berechnen.

5.1.2 Die diskriminante Validität ist mit VCL größer als ohne

Die folgende Tabelle 5-2 gibt die diskriminanten Validitätskoeffizienten wieder, die sich aus den Urteilen zum Planspiel 1 ohne und mit VCL errechnen lassen. Für das Rundgespräch lässt sich diese Hypothese nicht prüfen, da zum Rundgespräch in Block 2 keine VCL eingesetzt wurde.

Tabelle 5-2: Diskriminante Validitätskoeffizienten, Planspiel alt, ohne und mit VCL

Produkt-Moment-Korrelation	Block 1, PS alt, n=32	Block 2, PS alt m. VCL, n=58
OPZ-Prüfer	r	r
Planungsverhalten - Sozialkompetenz	.41*	.80**
Führungsverhalten - Sozialkompetenz	.46**	.77**
Führungsverhalten - Planungsverhalten	.85**	.90**
Externe Beobachter		
	n=8	n=8
Planungsverhalten - Sozialkompetenz	.89**	.66
Führungsverhalten - Sozialkompetenz	.66	.43
Führungsverhalten - Planungsverhalten	.72*	.80*

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Es zeigen sich für die Koeffizienten der OPZ-Prüfer und die der externen Beobachter zum Teil gegenläufige Ergebnisse. Damit die diskriminante Validität steigt, müssen die Koeffizienten sinken. Dies ist bei den Werten der OPZ-Prüfer, die alle signifikant sind, für keine Korrelation der Fall, d.h. die diskriminante Validität verschlechtert sich mit Einsatz der VCL.

Für die Urteile der externen Beobachter sinken die Koeffizienten mit Einsatz der VCL erwartungsgemäß für die Beziehung zwischen Planungsverhalten und Sozialkompetenz sowie Führungsverhalten und Sozialkompetenz. Die Korrelation zwischen den Urteilen für Führungsverhalten und Planungsverhalten steigt jedoch. 3 der 6 Werte der externen Beobachter sind nicht signifikant.

Die nachfolgende Tabelle 5-3 zeigt die diskriminanten Validitätskoeffizienten aus den Urteilen zum Planspiel Krisenmanagement ohne und mit VCL.

Tabelle 5-3: Diskriminante Validitätskoeffizienten, Planspiel KM ohne und mit VCL

Produkt-Moment-Korrelation	Block 3, PS KM, n=30	Block 4, PS KM mit VCL, n=60
OPZ-Prüfer	r	r
Planungsverhalten - Sozialkompetenz	.69**	.70**
Führungsverhalten - Sozialkompetenz	.77**	.74**
Führungsverhalten - Planungsverhalten	.91**	.88**
Externe Beobachter	n=8	n=16
Planungsverhalten - Sozialkompetenz	.42	.23
Führungsverhalten - Sozialkompetenz	.12	.25
Führungsverhalten - Planungsverhalten	.46	.71**

(** auf dem Niveau von 0,01 zweiseitig signifikant)

Es zeigen sich auch bei der neuen Aufgabe widersprüchliche Ergebnisse für die Entwicklung der Koeffizienten von OPZ-Beobachtern und externen Beobachtern. Für die OPZ-Beobachter sinken zwei von drei Koeffizienten. Die Korrelationen der Urteile von Führungs- und Sozialverhalten sowie von Führungs- und Planungsverhalten werden mit Einsatz der VCL geringer, wobei alle Werte signifikant sind. Insgesamt liegen die Koeffizienten jedoch auf einem deutlich höheren Niveau als diejenigen der Ausgangssituation in Block 1 (vgl. Tabelle 5-2).

Bei den Urteilen der externen Beobachter entwickelt sich nur einer von drei Koeffizienten in der erwarteten Richtung: die Korrelation der Urteile für Planungsverhalten und Sozialkompetenz sinkt. Insgesamt liegen jedoch bis auf eine Ausnahme alle Korrelationen auf einem extrem niedrigen Niveau und deutlich unter den Koeffizienten der Ausgangssituation in Block 1 (vgl. Tab. 5.2.). Nur der Koeffizient für die Korrelation der Urteile zu Führung und Planung mit VCL ist signifikant.

In der folgenden Tabelle 5-4 werden die diskriminanten Validitätskoeffizienten aus dem Planspiel Ressourcenengpass ohne und mit Einsatz der VCL gezeigt. Während nach dem alten Beurteilungsverfahren an dieser Stelle der Gruppenstation Führung und Soziale Kompetenz bewertet werden, werden nach dem neuen Beurteilungsverfahren mit VCL im Planspiel Ressourcenengpass Sozialkompetenz und Teamkompetenz bewertet. Die diskriminanten Koeffizienten beziehen sich daher nicht auf die gleichen Dimensionen.

Tabelle 5-4: Diskriminante Validitätskoeffizienten, Planspiel RE ohne und mit VCL

Produkt-Moment-Korrelation	Block 3, PS RE, n=30	Block 4, PS RE mit VCL, n=52
OPZ-Prüfer	r	r
Führungsverhalten - Sozialkompetenz	.62**	
Teamkompetenz - Sozialkompetenz		.81**
Externe Beobachter	n=8	n=16
Führungsverhalten - Sozialkompetenz	.74*	
Teamkompetenz - Sozialkompetenz		.76**

Sowohl bei den OPZ-Prüfern als auch bei den externen Beobachtern ist der diskriminante Koeffizient in Block 4 höher als in Block 3, wobei der Unterschied bei den externen Beobachtern nur .02 beträgt. Alle Werte sind signifikant.

5.1.3 Zusammenfassung zur Hypothese 5: Die Konstruktvalidität ist mit VCL größer als ohne

Zur Entwicklung der konvergenten Validität kann nur eine auf die Dimension Soziale Kompetenz eingeschränkte Aussage gemacht werden. Für diese Dimension zeigt sich durch den Einsatz der VCL bei den neuen Aufgaben eine Erhöhung der konvergenten Validität bei beiden Beobachtergruppen.

Die Ergebnisse zu den diskriminanten Koeffizienten ergeben ein sehr gemischtes Bild. Bei den OPZ-Beobachtern verschlechtert sich die diskriminante Validität durch den Einsatz der VCL im alten Planspiel, während sich bei den externen Beobachtern zwei von drei Koeffizienten verbessern.

Im neuen Planspiel Krisenmanagement verbessern sich mit dem Einsatz der VCL zwei der drei Koeffizienten der OPZ-Beobachter, bei den externen Beobachtern verschlechtern sich eben diese zwei Koeffizienten. Für das Planspiel Ressourcenengpass bestehen keine direkten Vergleichsmöglichkeiten. Auch hier entwickeln sich aber die diskriminanten Koeffizienten nicht im Sinne der Hypothese.

Daher kann von einer Verbesserung der diskriminanten Validität, die tatsächlich auf einer besseren inhaltlichen Abgrenzung der Konstrukte basiert, durch den Einsatz der VCL nicht ausgegangen werden.

5.2 Ergebnisse zur Hypothese 6: Konstruktvalidität und neue Aufgaben

Hypothese 6: Die Konstruktvalidität ist bei den neuen Aufgaben höher als bei den alten Aufgaben

5.2.1 Die konvergente Validität ist bei den neuen Aufgaben größer als bei den alten Aufgaben

Zur Bestimmung der konvergenten Validität steht wiederum nur ein Datenvergleich zur Verfügung. Und zwar die Korrelation zwischen den Urteilen zu Sozialkompetenz im Rundgespräch und Sozialkompetenz im Planspiel ohne VCL (Block 1) versus die Korrelation der Urteile zu Sozialkompetenz im Planspiel Ressourcenengpass mit den Urteilen zu Sozialkompetenz im Planspiel Krisenmanagement ohne VCL (Block 3). Diese werden in der folgenden Tabelle 5-5 dargestellt.

Tabelle 5-5: Konvergente Validität alte versus neue Aufgaben (jeweils ohne VCL)

	Alte Aufgaben, Block 1	Neue Aufgaben, Block 3
Sozialkompetenz, ohne VCL, OPZ-Prüfer	.57** n= 32	.63** n= 30
Sozialkompetenz, ohne VCL, Ext. Beobachter	.54 n= 8	.53 n= 8

(** auf dem Niveau von 0,01 zweiseitig signifikant)

Für die OPZ-Beobachter lässt sich ein leichter Anstieg des Korrelationskoeffizienten feststellen, für die externen Beobachter ein sehr leichter Rückgang um .01. Die Koeffizienten der externen Beobachter sind nicht signifikant.

5.2.2 Die diskriminante Validität ist bei den neuen Aufgaben größer als bei den alten Aufgaben

Zur Bestimmung der diskriminanten Validität wird zunächst auf die Daten aus den Tabellen 5-2 und 5-3 zurückgegriffen (s. S. 104 f.). Es werden die Daten aus Block 1 mit den Daten aus Block 3 (altes Planspiel versus Planspiel Krisenmanagement ohne VCL), sowie die Daten aus Block 2 mit denen aus Block 4 (altes Planspiel versus Planspiel Krisenmanagement mit VCL) verglichen.

Für den Vergleich der alten mit der neuen Aufgabe ohne VCL zeigt sich für die OPZ-Beobachter durchgängig eine Erhöhung der Korrelationskoeffizienten, also eine Verschlechterung der diskriminanten Validität.

Für die externen Beobachter zeigt sich dagegen durchgängig eine Verringerung der Korrelationskoeffizienten, also eine Verbesserung der diskriminanten Validität mit den neuen Aufgaben.

Für den Vergleich der alten mit der neuen Aufgabe mit VCL zeigt sich für alle Koeffizienten eine Verringerung, sowohl bei den OPZ-Beobachtern als auch bei den externen Beobachtern.

Für die Bewertung des Planspiels Ressourcenengpass steht nur der Vergleich der Urteile zum Rundgespräch mit den Urteilen zum Planspiel Ressourcenengpass ohne Einsatz der VCL zur Verfügung (also Block 1 versus Block 3), da in Block 2 zum Rundgespräch keine VCL eingesetzt wurde. Die entsprechenden Koeffizienten zeigt die folgende Tabelle 5-6. Der Korrelationskoeffizient für die Dimensionen Führung und Sozialkompetenz ist für beide Beobachtergruppen in Block 3 höher als in Block 1.

Tabelle 5-6: Diskriminante Validitätskoeffizienten, Rundgespräch und PS RE

Korrelation r	Block 1, Rundgespräch, n =32	Block 3, PS RE, n=30
OPZ-Prüfer	r	r
Führungsverhalten - Sozialkompetenz	.59**	.62**
Externe Beobachter		
	n = 8	n=8
Führungsverhalten - Sozialkompetenz	.58	.74*

5.2.3 Zusammenfassung zur Hypothese 6: Die Konstruktvalidität ist bei den neuen Aufgaben höher als bei den alten

Zur Entwicklung der konvergenten Validität kann wiederum nur eine auf die Dimension Soziale Kompetenz eingeschränkte Aussage gemacht werden. Für diese Dimension zeigt sich durch den Einsatz der neuen Aufgaben ohne VCL eine Erhöhung der konvergenten Validität bei den OPZ-Beobachtern, während sich der Koeffizient der externen Beobachter um .01 verschlechtert. In einer qualitativen Bewertung überwiegt die Tendenz zur Akzeptanz der Hypothese.

Widersprüchlich sind die Ergebnisse zur diskriminanten Validität. Bei den neuen Aufgaben ohne VCL haben sich nur die Koeffizienten der externen Beobachter beim Planspiel Krisenmanagement verbessert. Die Koeffizienten zum Planspiel Ressourcenengpass und die Koeffizienten der OPZ-Beobachter haben sich verschlechtert.

Beim neuen Planspiel Krisenmanagement mit VCL hat sich die diskriminante Validität bei beiden Beobachtergruppen verbessert im Vergleich zum alten Planspiel mit VCL.

In einer qualitativen Bewertung dieser Ergebnisse kann auf eine Bestätigung der Hypothese für das Planspiel Krisenmanagement geschlossen werden, nicht jedoch für das Planspiel zum Ressourcenengpass.

5.3 Anmerkung zur diskriminanten Validität

Absolut betrachtet weisen die Urteile der OPZ-Beobachter in Block 1, also in der Ausgangssituation die geringsten diskriminanten Korrelationskoeffizienten auf, also die beste diskriminante Validität. In Block 1 bewerten die OPZ-Beobachter die verschiedenen Dimensionen also am differenziertesten. Bei den externen Beobachtern weisen dagegen in Block 3, und dort insbesondere beim Planspiel zum Krisenmanagement die Urteile die geringsten Korrelationen auf, in Block 4 liegen die Koeffizienten nur geringfügig höher. Dies weist auf eine sehr unterschiedliche Wirkung der Interventionen auf die Urteilsfindung bei OPZ-Beobachtern und externen Beobachtern hin.

6 Ergebnisdarstellung III: Externe Konstruktvalidierung

Für die Ermittlung der externen konvergenten Validität wurden aus den Dimensionen des BIP™ die in Abbildung 6-1 dargestellten Dimensionen ausgewählt. Diese weisen laut ihrer Definition die größte inhaltliche Nähe zu den jeweiligen Konstrukten der OPZ auf. Die Ergebnisse des BIP™ zu diesen Dimensionen wurden je Bewerber mit den Prüfgruppenmittelwerten der entsprechenden Dimensionen aus den GSV-Aufgaben korreliert. Für die Dimension Planungsverhalten des GSV liefert der BIP™ keine sinnvolle Entsprechung. Die Koeffizienten stellen Produkt-Moment-Korrelationen nach Pearson dar.

Abbildung 6-1: Dimensionen des BIP™ zur Überprüfung der konvergenten Validität

GSV	BIP™
Führungskompetenz	Führungsmotivation
Sozialkompetenz	Kontaktfähigkeit
Teamkompetenz	Teamorientierung

Für die Überprüfung der diskriminanten Validität werden die Dimensionen des GSV, die auch in Zukunft im Verfahren bewertet werden sollen, mit allen BIP™ -Dimensionen korreliert. Aufgrund der fehlenden BIP™-Daten aus Block 3 kann der Einsatz der Verhaltenscheckliste nur für die alten Aufgaben (Block 1 versus Block 2) untersucht werden, der Einsatz der neuen Aufgaben nur inklusive der VCL (Block 2 versus Block 4).

6.1 Ergebnisse zur Hypothese 7: Externe Konstruktvalidität und VCL

Hypothese 7: Die externe Konstruktvalidität ist mit VCL höher als ohne.

6.1.1 Die externe konvergente Validität verbessert sich mit Einsatz der VCL beim alten Planspiel

Für eine Aussage zum Einsatz der VCL beim alten Planspiel werden die konvergenten Koeffizienten aus den Urteilen der OPZ-Beobachter und den BIP™-Ergebnissen von Block 1 mit denen von Block 2 verglichen (vgl. Tabelle 6-1).

Tabelle 6-1: Externe konvergente Koeffizienten für Prüfgruppenmittelwerte (ohne und mit VCL) der OPZ-Beobachter und BIP™-Ergebnisse

BIP™-Dimensionen	OPZ-Beobachter	
	PS 1, Block 1 , ohne VCL, n=26	PS 1, Block 2 , mit VCL, n=42
Führungsmotivation	Führung: .22	Führung: .48**
Kontaktfähigkeit	Soziale Kompetenz: .10	Soziale Kompetenz: .35*

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Tabelle 6-2 zeigt die konvergenten Koeffizienten aus den Urteilen der externen Beobachter und den BIP™-Ergebnissen

Tabelle 6-2: Externe konvergente Koeffizienten für Prüfgruppenmittelwerte (ohne und mit VCL) der externen Beobachter und BIP™-Ergebnisse

BIP™-Dimensionen	Externe Beobachter	
	PS 1, Block 1 , ohne VCL, n= 7	PS 1, Block 2 , mit VCL, n= 6
Führungsmotivation	Führung: .07	Führung: .54
Kontaktfähigkeit	Soziale Kompetenz: -.96**	Soziale Kompetenz: .34

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Bei den OPZ-Beobachtern sind die Vergleichswerte signifikant, bei den externen Beobachtern nur der Ausgangswert für Soziale Kompetenz.

Sowohl für die Urteile der OPZ-Beobachter als auch für die Urteile der externen Beobachter steigen die Korrelationen mit Einsatz der VCL, und zwar für beide Dimensionen. Die externe konvergente Validität verbessert sich also für die Dimensionen Führungsfähigkeit und Sozialkompetenz mit Einsatz der VCL beim alten Planspiel.

6.1.2 Die externe diskriminante Validität verbessert sich mit Einsatz der VCL

Die Tabellen enthalten im Weiteren alle Dimensionen des BIP™. Die Dimension, die zur Bestimmung der konvergenten Validität dient, ist kursiv gedruckt. Die Daten können im Wesentlichen nur deskriptiv ausgewertet werden, da nur wenige der Korrelationen signifikant werden. In der Beschreibung der Tabellen wird daher nicht weiter auf die Signifikanzen eingegangen.

Dimension Führung

Die folgende Tabelle 6-3 zeigt die Korrelationen zwischen dem Prüfgruppenmittelwert der OPZ-Beobachter für die Dimension Führungsfähigkeit aus dem alten Planspiel und den Ergebnissen des BIP™ in Block 1 und Block 2.

Tabelle 6-3: Koeffizienten für Prüfgruppenmittelwerte (ohne und mit VCL) der OPZ-Beobachter zur Dimension Führung und BIP™-Ergebnisse

BIP™-Dimensionen	OPZ-Beobachter	
	Führung, PS 1, Block 1, ohne VCL, n=26	Führung, PS 1, Block 2, mit VCL, n=42
Belastbarkeit	-.03	.22
Durchsetzungsstärke	.11	.46**
Emotionale Stabilität	.06	.01
Flexibilität	.07	.25
<i>Führungsmotivation</i>	.22	.48**
Gestaltungsmotivation	.08	.46**
Gewissenhaftigkeit	-.11	.17
Handlungsorientierung	.12	.39*
Kontaktfähigkeit	.24	.44*
Leistungsmotivation	-.10	.42**
Selbstbewusstsein	.38	.40
Sensitivität	.16	.38
Soziabilität	.06	-.02
Teamorientierung)	.14	-.08

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Nur die Korrelationen mit den BIP™-Dimensionen Emotionale Stabilität, Soziabilität und Teamorientierung sind gesunken. Alle anderen Korrelationen sind, zum Teil recht deutlich, gestiegen. Die Ergebnisse der OPZ-Beobachter sprechen daher nicht für eine Verbesserung der diskriminanten Validität für die Dimension Führung durch den Einsatz der VCL beim alten Planspiel.

Die folgende Tabelle 6-4 zeigt die Korrelationen zwischen dem Mittelwert der Urteile der externen Beobachter für die Dimension Führungsfähigkeit aus dem alten Planspiel und den Ergebnissen des BIP™ in Block 1 und Block 2.

Tabelle 6-4: Koeffizienten für Urteilsmitelwerte (ohne und mit VCL) der externen Beobachter zur Dimension Führung und BIP™ -Ergebnisse

BIP™-Dimensionen	Externe Beobachter	
	Führung, PS 1, Block 1, ohne VCL, n= 7	Führung, PS 1, Block 2, mit VCL, n= 6
Belastbarkeit	-.16	-.12
Durchsetzungsstärke	-.19	-.50
Emotionale Stabilität	-.16	-.02
Flexibilität	.01	.28
<i>Führungsmotivation</i>	.07	.54
Gestaltungsmotivation	.37	-.34
Gewissenhaftigkeit	-.11	.61
Handlungsorientierung	-.24	.24
Kontaktfähigkeit	-.37	-.02
Leistungsmotivation	-.25	.74
Selbstbewusstsein	.11	-.07
Sensitivität	-.55	-.25
Soziabilität	-.48	-.64
Teamorientierung	-.53	-.88*

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Bei den externen Beobachtern sind die Korrelationen der Urteile zu Führung mit den BIP™-Dimensionen Durchsetzungsstärke, Gestaltungsmotivation, Selbstbewusstsein, Soziabilität und Teamorientierung gesunken. Die 9 übrigen Korrelationen sind gestiegen, so dass auch die Ergebnisse der externen Beobachter nicht für eine Verbesserung der diskriminanten Validität sprechen.

Dimension Soziale Kompetenz

Tabelle 6-5 zeigt die Korrelationen der Urteile der OPZ-Beobachter zur Dimension Soziale Kompetenz mit den BIP™-Ergebnissen aus Block 1 und Block 2.

Tabelle 6-5: Koeffizienten für Prüfgruppenmittelwerte (ohne und mit VCL) der OPZ-Beobachter zur Dimension Soziale Kompetenz und BIP™-Ergebnisse

BIP™-Dimensionen	OPZ-Beobachter	
	Soziale Kompetenz, PS 1, Block 1, ohne VCL, n=26	Soziale Kompetenz, PS 1, Block 2, mit VCL, n=42
Belastbarkeit	-.22	.17
Durchsetzungsstärke	-.02	.47**
Emotionale Stabilität	.19	.07
Flexibilität	.14	.12
Führungsmotivation	-.04	.29
Gestaltungsmotivation	-.40*	.35*
Gewissenhaftigkeit	-.11	.15
Handlungsorientierung	-.21	.28
<i>Kontaktfähigkeit</i>	.10	.35*
Leistungsmotivation	-.08	.30*
Selbstbewusstsein	.22	.40*
Sensitivität	-.01	.40*
Soziabilität	.19	-.03
Teamorientierung	.28	-.09

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Die Korrelationen der Urteile zur Sozialen Kompetenz mit den Ergebnissen der BIP™-Dimensionen Emotionale Stabilität, Flexibilität, Soziabilität und Teamorientierung sind gesunken, alle anderen Korrelationen sind gestiegen. Für die Urteile der OPZ-Beobachter kann daher keine Verbesserung der diskriminanten Validität für die Dimension Soziale Kompetenz angenommen werden.

Die folgende Tabelle 6-6 zeigt die Korrelationen der Urteile der externen Beobachter zur Sozialen Kompetenz mit den BIP™-Ergebnissen aus Block 1 und Block 2.

Tabelle 6-6: Koeffizienten für Urteilsmittelwerte (ohne und mit VCL) der externen Beobachter zur Dimension Soziale Kompetenz und BIP™-Ergebnisse

BIP™-Dimensionen	Externe Beobachter	
	Soziale Kompetenz, PS 1, Block 1, ohne VCL, n=7	Soziale Kompetenz, PS 1, Block 2, mit VCL, n=6
Belastbarkeit	-.62	-.24
Durchsetzungsstärke	-.93**	-.84*
Emotionale Stabilität	-.73	.63
Flexibilität	-.74	.08
Führungsmotivation	-.88**	-.02
Gestaltungsmotivation	-.11	-.29
Gewissenhaftigkeit	-.41	.30
Handlungsorientierung	-.89**	.01
<i>Kontaktfähigkeit</i>	-.96**	.34
Leistungsmotivation	-.87*	.48
Selbstbewusstsein	-.85*	-.11
Sensitivität	-.87*	-.03
Soziabilität	-.43	-.04
Teamorientierung	-.66	-.45

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Nur die Korrelation zwischen dem Urteil zu Sozialer Kompetenz und dem BIP™-Ergebnis zur Gestaltungsmotivation sinkt, alle anderen Korrelationen steigen. Es kann daher nicht von einer Verbesserung der diskriminanten Validität für die Urteile der externen Beobachter zur Sozialen Kompetenz durch den Einsatz der VCL ausgegangen werden.

Dimension Planungsverhalten

Die Tabelle 6-7 gibt die Korrelationen der Urteile der OPZ-Beobachter zur Dimension Planungsverhalten mit den Ergebnissen des BIP™ aus Block 1 und Block 2 an.

Tabelle 6-7: Koeffizienten für Prüfgruppenmittelwerte (ohne und mit VCL) der OPZ-Beobachter zur Dimension Planungsverhalten und BIP™-Ergebnisse

BIP™-Dimensionen	OPZ-Beobachter	
	Planungsverhalten, PS 1, Block 1, ohne VCL, n=26	Planungsverhalten, PS 1, Block 2, mit VCL, n=42
Belastbarkeit	.03	.15
Durchsetzungsstärke	.14	.39*
Emotionale Stabilität	.04	.00
Flexibilität	.04	.15
Führungsmotivation	.16	.39**
Gestaltungsmotivation	-.06	.43**
Gewissenhaftigkeit	.02	.13
Handlungsorientierung	.09	.30
Kontaktfähigkeit	.15	.28
Leistungsmotivation	-.02	.31*
Selbstbewusstsein	.34	.38*
Sensitivität	.10	.28
Soziabilität	.09	-.14
Teamorientierung	.12	-.11

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Nur die Korrelationen zwischen dem Urteil zu Planungsverhalten und den BIP™-Ergebnissen der Dimensionen Emotionale Stabilität, Soziabilität und Teamorientierung sind mit Einsatz der VCL gesunken. Die Hypothese einer verbesserten diskriminanten Validität muss daher für die Urteile der OPZ-Beobachter abgelehnt werden.

Tabelle 6-8 zeigt die Korrelationen der Urteile der externen Beobachter zur Dimension Planungsverhalten mit den BIP™-Ergebnissen aus Block 1 und Block 2.

Tabelle 6-8: Koeffizienten für Urteilsmittelwerte (ohne und mit VCL) der externen Beobachter zur Dimension Planungsverhalten und BIP™-Ergebnisse

BIP™-Dimensionen	Externe Beobachter	
	Planungsverhalten, PS 1, Block 1, ohne VCL, n=7	Planungsverhalten, PS 1, Block 2, mit VCL, n=6
Belastbarkeit	.04	.17
Durchsetzungsstärke	-.63	-.51
Emotionale Stabilität	-.14	.15
Flexibilität	-.53	.56
Führungsmotivation	-.42	.57
Gestaltungsmotivation	-.07	-.13
Gewissenhaftigkeit	-.15	.81
Handlungsorientierung	-.36	.49
Kontaktfähigkeit	-.66	.27
Leistungsmotivation	-.61	.86*
Selbstbewusstsein	-.35	.11
Sensitivität	-.71	.05
Soziabilität	-.11	-.48
Teamorientierung	-.27	-.88*

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Nur die Korrelationen zwischen dem Urteil zu Planungsverhalten und den BIP™-Ergebnissen der Dimensionen Soziabilität und Teamorientierung sind mit Einsatz der VCL gesunken. Die Hypothese einer verbesserten diskriminanten Validität muss daher für die Urteile der externen Beobachter abgelehnt werden.

6.1.3 Zusammenfassung externe Konstruktvalidität und VCL

Die Mehrzahl der Korrelationen zwischen den Prüfurteilen und den BIP™-Ergebnissen ist mit Einsatz der VCL beim alten Planspiel gestiegen. Dies gilt für alle drei Dimensionen, die auf der VCL enthalten sind und für beide Prüfergruppen. Eine Verbesserung der externen diskriminanten Validität durch den Einsatz der VCL kann daher nicht angenommen werden.

Die konvergenten Korrelationen für die Dimensionen Führung und Soziale Kompetenz sind bei beiden Beobachtergruppen gestiegen. Es kann daher nur eine partielle Verbesserung der externen Konstruktvalidität durch den Einsatz der VCL beim alten Planspiel festgestellt werden.

6.1.4 Nomologisches Netzwerk

Die Ergebnisse der Korrelation der GSV-Urteile mit den Ergebnissen des BIP™ können auch im Sinne einer inhaltlichen Konstruktklärung genutzt werden. Im Folgenden sollen in dieser Hinsicht die Ergebnisse der Urteile der OPZ-Prüfer untersucht werden. Auf die Ergebnisse der externen Beobachter wird nicht weiter eingegangen, da aufgrund der geringen Fallzahl eine differenziertere Betrachtung wenig sinnvoll erscheint.

In diesem Zusammenhang fällt zunächst auf, dass für alle drei Dimensionen des Planspiels die Korrelation der OPZ-Urteile mit den BIP™-Dimensionen Emotionale Stabilität, Soziabilität und Teamorientierung mit Einsatz der VCL gesunken ist.

Für die Dimension **Führung** korrelieren ohne Einsatz der VCL nur die BIP™-Ergebnisse der Dimensionen Führungsmotivation (.22) und Selbstbewusstsein (.38) deutlich über Null.

Mit Einsatz der VCL zeigen die BIP™-Dimensionen Führungsmotivation (.48**), Durchsetzungsstärke (.46**), Gestaltungsmotivation (.46**) sowie Kontaktfähigkeit (.44**) und Leistungsmotivation (.42**) einen deutlichen und signifikanten Zusammenhang mit den Urteilen zur Dimension Führung im Planspiel.

Für die Dimension **Soziale Kompetenz** korrelieren ohne Einsatz der VCL nur die BIP™-Ergebnisse der Dimensionen Teamorientierung (.28) und Selbstbewusstsein (.22) deutlich über Null.

Mit Einsatz der VCL zeigen die BIP™-Dimensionen Durchsetzungsstärke (.47**), Selbstbewusstsein (.40*) und Sensitivität (.40*) sowie Kontaktfähigkeit (.35*) und Gestaltungsmotivation (.35*) einen deutlichen und signifikanten Zusammenhang mit den Urteilen zur Dimension Soziale Kompetenz.

Für die Dimension **Planungsverhalten** korreliert ohne Einsatz der VCL nur das BIP™-Ergebnis der Dimension Selbstbewusstsein (.34) deutlich über Null.

Mit Einsatz der VCL zeigen die Dimensionen Gestaltungsmotivation (.43**), Führungsmotivation (.39**) und Durchsetzungsstärke (.39*) sowie Selbstbewusstsein (.38*) einen deutlichen und signifikanten Zusammenhang mit den Urteilen zur Dimension Planungsverhalten.

6.2 Ergebnisse zur Hypothese 8: Externe Konstruktvalidität und neue Aufgaben

Hypothese 8: Die externe Konstruktvalidität ist bei den neuen Aufgaben höher als bei den alten Aufgaben.

Die Hypothese 8 kann nur über den Vergleich von Block 2 mit Block 4 überprüft werden, d.h. über die Korrelation der Urteile zum alten Planspiel mit VCL mit den Ergebnissen des BIP™

sowie über die Korrelation der Urteile zum Planspiel Krisenmanagement mit den Ergebnissen des BIP™.

6.2.1 Die externe konvergente Validität verbessert sich mit Einsatz des neuen Planspiels zum Krisenmanagement unter Einsatz der VCL

Die folgenden Tabellen 6-9 und 6-10 zeigen die konvergenten Korrelationskoeffizienten der Urteile der OPZ-Beobachter bzw. der externen Beobachter aus dem alten Planspiel mit VCL und aus dem neuen Planspiel zum Krisenmanagement mit VCL mit den Ergebnissen des BIP™ aus Block 2 und Block 4.

Tabelle 6-9: Externe konvergente Koeffizienten für Prüfgruppenmittelwerte (altes PS und PS KM, mit VCL) der OPZ-Beobachter und BIP™-Ergebnisse

	OPZ-Beobachter	
BIP™-Dimensionen	PS 1, Block 2 , mit VCL, n=42	PS KM, Block 4 , mit VCL, n=38
Führungsmotivation	Führung: .48**	Führung: .20
Kontaktfähigkeit	Soziale Kompetenz: .35*	Soziale Kompetenz: .08

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Tabelle 6-10: Externe konvergente Koeffizienten für Prüfgruppenmittelwerte (altes PS und PS KM, mit VCL) der externen Beobachter und BIP™-Ergebnisse

	Externe Beobachter	
BIP™-Dimensionen	PS 1, Block 2 , mit VCL, n= 6	PS KM, Block 4 , mit VCL, n=13
Führungsmotivation	Führung: .54	Führung: .55*
Kontaktfähigkeit	Soziale Kompetenz: .34	Soziale Kompetenz : -.24

Für die Urteile der OPZ-Beobachter sinkt die Korrelation zwischen den Ergebnissen aus dem Planspiel zu Führung und Sozialer Kompetenz und den entsprechenden BIP™-Ergebnissen. Für die Urteile der externen Beobachter steigt die Korrelation geringfügig um .01 für die Dimension Führung und sinkt deutlich für die Dimension Soziale Kompetenz. Da drei von vier Vergleichswerten keine Verbesserung der konvergenten Validität anzeigen, muss die Hypothese abgelehnt werden.

6.2.2 Die externe diskriminante Validität verbessert sich mit Einsatz des neuen Planspiels zum Krisenmanagement unter Einsatz der VCL

Die folgende Tabelle 6-11 zeigt die Korrelationen der Urteile der OPZ-Beobachter zur Dimension Führung mit den BIP™-Ergebnissen aus Block 2 und 4.

Tabelle 6-11: Externe Koeffizienten für Prüfgruppenmittelwerte (altes PS und PS KM, mit VCL) der OPZ-Beobachter und BIP™-Ergebnisse

BIP™-Dimensionen	OPZ-Beobachter	
	Führung, PS 1, Block 2, mit VCL, n= 42	Führung, PS KM, Block 4, mit VCL, n= 38
Belastbarkeit	.22	.34*
Durchsetzungsstärke	.46**	.17
Emotionale Stabilität	.01	.12
Flexibilität	.25	.18
<i>Führungsmotivation</i>	.48**	.20
Gestaltungsmotivation	.46**	.06
Gewissenhaftigkeit	.17	.07
Handlungsorientierung	.39*	.17
Kontaktfähigkeit	.44*	.17
Leistungsmotivation	.42**	.16
Selbstbewusstsein	.40	.36*
Sensitivität	.38	.23
Soziabilität	-.02	.13
Teamorientierung)	-.08	-.04

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Die fett gedruckten Korrelationskoeffizienten sind entsprechend der Hypothese, dass der Einsatz der neuen Aufgabe die diskriminante Validität verbessert, gesunken. Nur die Koeffizienten für die BIP™-Dimensionen Belastbarkeit, Emotionale Stabilität, Soziabilität und Teamorientierung sind gestiegen. Die Hypothese kann daher für die Urteile der OPZ-Beobachter zur Dimension Führung im neuen Planspiel zum Krisenmanagement akzeptiert werden.

Die folgende Tabelle 6-12 zeigt die Korrelationen der Urteile der externen Beobachter zur Dimension Führung mit den BIP™-Ergebnissen aus Block 2 und 4.

Tabelle 6-12: Externe Koeffizienten für Urteilmittelwerte (altes PS und PS KM, mit VCL) der externen Beobachter zur Dimension Führung und BIP™-Ergebnisse

BIP™-Dimensionen	Externe Beobachter	
	Führung, PS 1, Block 2, mit VCL, n= 6	Führung, PS KM, Block 4, mit VCL, n= 13
Belastbarkeit	-.12	.40
Durchsetzungsstärke	-.50	.62*
Emotionale Stabilität	-.02	.50
Flexibilität	.28	.64*
<i>Führungsmotivation</i>	.54	.55*
Gestaltungsmotivation	-.34	.26
Gewissenhaftigkeit	.61	-.10
Handlungsorientierung	.24	.34
Kontaktfähigkeit	-.02	.53
Leistungsmotivation	.74	.57*
Selbstbewusstsein	-.07	.63*
Sensitivität	-.25	.58*
Soziabilität	-.64	-.01
Teamorientierung	-.88*	.19

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Nur die Korrelation für die Dimensionen Gewissenhaftigkeit und Leistungsmotivation sind gesunken, alle übrigen Korrelationen sind mit dem Einsatz der neuen Aufgabe gestiegen. Die Hypothese muss daher für die Dimension Führung für die Urteile der externen Beobachter abgelehnt werden.

Die folgende Tabelle 6-13 zeigt die Korrelationen der Urteile der OPZ-Beobachter zur Dimension Soziale Kompetenz mit den BIP™-Ergebnissen aus Block 2 und 4.

Tabelle 6-13: Externe Koeffizienten für Prüfgruppenmittelwerte (altes PS und PS KM, mit VCL) der OPZ-Beobachter zur Dimension Soziale Kompetenz und BIP™-Ergebnisse

BIP™-Dimensionen	OPZ-Beobachter	
	Soziale Kompetenz, PS 1, Block 2, mit VCL, n=42	Soziale Kompetenz, PS KM, Block 4, mit VCL, n= 38
Belastbarkeit	.17	.21
Durchsetzungsstärke	.47**	.21
Emotionale Stabilität	.07	.06
Flexibilität	.12	.02
Führungsmotivation	.29	.01
Gestaltungsmotivation	.35*	.22
Gewissenhaftigkeit	.15	.20
Handlungsorientierung	.28	.10
<i>Kontaktfähigkeit</i>	.35*	.08
Leistungsmotivation	.30*	.09
Selbstbewusstsein	.40*	.19
Sensitivität	.40*	.17
Soziabilität	-.03	-.08
Teamorientierung	-.09	-.26

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Nur die Korrelationen für die BIP™-Dimensionen Belastbarkeit und Gewissenhaftigkeit sind gestiegen, alle übrigen Korrelationen sind gesunken. Die Hypothese kann daher für die Dimension Soziale Kompetenz für die Urteile der OPZ-Beobachter akzeptiert werden.

Die folgende Tabelle 6-14 zeigt die Korrelationen der Urteile der externen Beobachter zur Dimension Soziale Kompetenz mit den BIPTM-Ergebnissen aus Block 2 und 4.

Tabelle 6-14: Externe Koeffizienten für Urteilsmittelwerte (altes PS und PS KM, mit VCL) der externen Beobachter zur Dimension Soziale Kompetenz und BIPTM-Ergebnisse

BIP TM -Dimensionen	Externe Beobachter	
	Soziale Kompetenz, PS 1, Block 2, mit VCL, n= 6	Soziale Kompetenz, PS KM, Block 4, mit VCL, n= 13
Belastbarkeit	-.24	.25
Durchsetzungsstärke	-.84*	.02
Emotionale Stabilität	.63	-.02
Flexibilität	.08	-.38
Führungsmotivation	-.02	-.12
Gestaltungsmotivation	-.29	-.14
Gewissenhaftigkeit	.30	.25
Handlungsorientierung	.01	-.40
<i>Kontaktfähigkeit</i>	<i>.34</i>	<i>-.24</i>
Leistungsmotivation	.48	-.27
Selbstbewusstsein	-.11	.01
Sensitivität	-.03	.12
Soziabilität	-.04	.30
Teamorientierung	-.45	-.47

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Von den 14 Dimensionen des BIPTM weisen 8 eine gesunkene Korrelation zu den Urteilen der externen Beobachter zur Sozialen Kompetenz auf, während 6 eine gestiegene Korrelation aufweisen. Aus diesem Ergebnis lässt sich nur eine sehr leichte Tendenz im Sinne der Hypothese ableiten.

Die folgende Tabelle 6-15 zeigt die Korrelationen der Urteile der OPZ-Beobachter zur Dimension Planungsverhalten mit den BIP^T-Ergebnissen aus Block 2 und 4.

Tabelle 6-15: Externe Koeffizienten für Prüfgruppenmittelwerte (altes PS und PS KM, mit VCL) der OPZ-Beobachter zur Dimension Planungsverhalten und BIPTM-Ergebnisse

BIP TM -Dimensionen	OPZ-Beobachter	
	Planungsverhalten, PS 1, Block 2, mit VCL, n=42	Planungsverhalten, PS KM, Block 4, mit VCL, n= 38
Belastbarkeit	.15	.32*
Durchsetzungsstärke	.39*	.13
Emotionale Stabilität	.00	.12
Flexibilität	.15	.22
Führungsmotivation	.39**	.24
Gestaltungsmotivation	.43**	.03
Gewissenhaftigkeit	.13	.04
Handlungsorientierung	.30	.15
Kontaktfähigkeit	.28	.06
Leistungsmotivation	.31*	.17
Selbstbewusstsein	.38*	.27
Sensitivität	.28	.20
Soziabilität	-.14	-.04
Teamorientierung	-.11	-.04

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Die Korrelationen der BIPTM-Dimensionen Belastbarkeit, Emotionale Stabilität, Flexibilität, Soziabilität und Teamorientierung mit den Urteilen der OPZ-Prüfer zum Planungsverhalten sind gestiegen, während die übrigen 9 Korrelationen der Hypothese entsprechend gesunken sind. Die Hypothese kann daher für die Dimension Planungsverhalten akzeptiert werden.

Die folgende Tabelle 6-19 zeigt die Korrelationen der Urteile der externen Beobachter zur Dimension Planungsverhalten mit den BIP™-Ergebnissen aus Block 2 und Block 4.

Tabelle 6-16: Externe Koeffizienten für Urteilmittelwerte (altes PS und PS KM, mit VCL) der externen Beobachter zur Dimension Planungsverhalten und BIP™-Ergebnisse

BIP™-Dimensionen	Externe Beobachter	
	Planungsverhalten, PS 1, Block 2, mit VCL, n=6	Planungsverhalten, PS KM, Block 4, mit VCL, n=13
Belastbarkeit	.17	.54
Durchsetzungsstärke	-.51	.73**
Emotionale Stabilität	.15	.40
Flexibilität	.56	.54
Führungsmotivation	.57	.56*
Gestaltungsmotivation	-.13	.29
Gewissenhaftigkeit	.81	-.18
Handlungsorientierung	.49	.18
Kontaktfähigkeit	.27	.39
Leistungsmotivation	.86*	.50
Selbstbewusstsein	.11	.49
Sensitivität	.05	.47
Soziabilität	-.48	-.03
Teamorientierung	-.88*	.28

(** auf dem Niveau von 0,01 zweiseitig signifikant, * auf dem Niveau von 0,05 zweiseitig signifikant)

Nur 5 Korrelationen sind hypothesenkonform gesunken, die übrigen sind gestiegen. Die Hypothese, dass sich die diskriminante Validität mit Einsatz der neuen Aufgabe verbessert, muss daher für die Dimension Planungsverhalten für die externen Beobachter abgelehnt werden.

6.2.3 Zusammenfassung: Externe Konstruktvalidität und neue Aufgabe

Die konvergente Validität konnte durch den Einsatz des neuen Planspiels zum Krisenmanagement mit Einsatz der VCL im Vergleich zum alten Planspiel mit Einsatz der VCL nicht verbessert werden.

Die Ergebnisse zur externen diskriminanten Validität unterscheiden sich für die beiden Beobachtergruppen. Für die Urteile der OPZ-Beobachter verbessert sich die diskriminante Validität für alle drei Zieldimensionen des Planspiels, Führung, Soziale Kompetenz und Planungsverhalten.

Für die Ergebnisse der externen Beobachter verschlechtert sich die diskriminante Validität für die Dimensionen Führung und Planungsverhalten, für die Dimension Soziale Kompetenz zeigt sich nur eine sehr leichte Verbesserungstendenz mit Einsatz des neuen Planspiels zum Krisenmanagement.

6.2.4 Nomologisches Netzwerk

Die Urteile der OPZ-Beobachter zur Dimension **Führung** weisen in der Ausgangssituation (Block2, altes Planspiel mit VCL) die höchsten Korrelationen auf mit den BIPTM-Ergebnissen zu Führungsmotivation (.48**), Gestaltungsmotivation (.46**) und Durchsetzungsstärke (.46**).

In der Vergleichssituation (Block 4, neues Planspiel zum Krisenmanagement mit VCL) gibt es nur noch zwei BIPTM-Dimensionen, mit denen die Urteile zur Führung einen deutlichen positiven Zusammenhang aufweisen: Selbstbewusstsein (.36*) sowie Belastbarkeit (.34*).

Die Urteile der OPZ-Beobachter zur Dimension **Soziale Kompetenz** weisen in der Ausgangssituation (Block 2) hohe Korrelationen mit den BIPTM-Ergebnissen zu Durchsetzungsstärke (.47**), Selbstbewusstsein (.40*) und Sensitivität (.40*) sowie Gestaltungsmotivation (.35*) und Kontaktfähigkeit (.35*) auf.

In der Vergleichssituation zeigen die höchsten positiven Zusammenhänge die BIPTM-Dimensionen Gestaltungsmotivation (.22), Belastbarkeit (.21), Durchsetzungsstärke (.21), Gewissenhaftigkeit (.20) und Selbstbewusstsein (.19).

Die Urteile der OPZ-Beobachter zur Dimension **Planungsverhalten** weisen in der Ausgangssituation (Block 2) hohe Korrelationen mit den BIPTM-Ergebnissen zu Gestaltungsmotivation (.43**), Führungsmotivation (.39**) und Durchsetzungsstärke (.39*) sowie Selbstbewusstsein (.38*) auf.

In der Vergleichssituation korrelieren die BIPTM-Dimensionen Belastbarkeit (.32*), Selbstbewusstsein (.27) und Führungsmotivation (.24) am deutlichsten mit den Urteilen zu Planungsverhalten.

7 Zusammenfassende Darstellung der Ergebnisse und Diskussion

7.1 Tabellarische Übersicht der Ergebnisse

In der folgenden Tabelle werden die Resultate aus den Überprüfungen der Hypothesen im Überblick dargestellt. Die Tabelle verdeutlicht, dass eine über die verschiedenen Aufgaben und Beobachtergruppen aggregierte Bewertung der Hypothesen eine Vereinfachung und einen Informationsverlust darstellen, die dem Sachverhalt wenig gerecht werden.

Tabelle 7-1: Zusammenfassende Darstellung der Hypothesenüberprüfung

	OPZ- Prüfer	Ext. Beo.	OPZ- Prüfer	Ext. Beo.	OPZ- Prüfer	Ext. Beo.
Hyp. 1: Die Beobachterreliabilität ist mit VCL höher als ohne.	Alte Aufgaben PS alt		Neue Aufgabe PS RE		Neue Aufgabe PS KM	
Führungsfähigkeit	Ja	Ja	--	--	Nein	Ja
Soziale Kompetenz	Ja	Ja	Ja	Nein	Nein	Ja
Planungsverhalten	Ja	Nein	--	--	Nein	Nein
Hyp. 1 wird	bestät.	bestät.	bestät.	abgel.	abgel.	bestät.
Hyp. 2: Die Beobachterreliabilität ist bei den neuen Aufgaben größer als bei den alten	Ohne VCL RG versus PS RE		Ohne VCL PS alt versus PS KM		Mit VCL PS alt versus PS KM	
Führungsfähigkeit	Ja	Ja	Ja	Ja	Nein	Nein
Soziale Kompetenz	Ja	Ja	Ja	Nein	Nein	Ja
Planungsverhalten	--	--	Ja	Nein	Nein	Nein
Hyp. 2 wird	bestät.	bestät.	bestät.	abgel.	abgel.	abgel.
Hyp. 3: Die Differenzierung der OB gelingt mit VCL besser als ohne.	Alte Aufgabe PS alt		Neue Aufgabe PS RE		Neue Aufgabe PS KM	
Führungsfähigkeit	Ja	Ja	--	--	Nein	Nein
Soziale Kompetenz	Ja	Ja	Ja	Nein	Ja	Ja
Planungsverhalten	Ja	Ja	--	--	Nein	Nein
Hyp. 3 wird	bestät.	bestät.	bestät.	abgel.	abgel.	abgel.
Hyp. 4: Die Differenzierung der OB gelingt bei den neuen Aufgaben besser als bei den alten.	Ohne VCL RG versus PS RE		Ohne VCL PS alt versus PS KM		Mit VCL PS alt versus PS KM	
Führungsfähigkeit	Ja	Ja	Ja	Nein	Nein	Nein
Soziale Kompetenz	Ja	Ja	Ja	Nein	Nein	Ja
Planungsverhalten	--	--	Ja	Nein	Nein	Nein
Hyp. 4 wird	bestät.	bestät.	bestät.	abgel.	abgel.	abgel.

	OPZ- Prüfer	Ext. Beo.	OPZ- Prüfer	Ext. Beo.
Hyp. 5.1: Die konvergente Validität ist mit VCL höher als ohne.	Neue Aufgaben			
Sozialkompetenz	Ja	Ja		
Hyp. 5.1 wird	bestät.	bestät.		
Hyp. 5.2: Die diskriminante Validität ist mit VCL höher als ohne	Alte Aufgabe PS alt		Neue Aufgabe PS KM	
Planungsv. - Sozialkompetenz	Nein	Ja	Nein	Ja
Führungsv. - Sozialkompetenz	Nein	Ja	Ja	Nein
Führungsv. - Planungsverh.	Nein	Nein	Ja	Nein
Hyp. 5.2 wird	abgel.	bestät.	bestät.	abgel.
Hyp. 6.1: Die konvergente Validität ist bei den neuen Aufgaben höher als bei den alten.	Ohne VCL			
Sozialkompetenz	Ja	Nein		
Hyp. 6.1 wird	bestät.	abgel.		
Hyp. 6.2: Die diskriminante Validität ist bei den neuen Aufgaben höher als bei den alten.	Ohne VCL PS alt versus PS KM		Mit VCL PS alt versus PS KM	
Planungsv. - Sozialkompetenz	Nein	Ja	Ja	Ja
Führungsv. - Sozialkompetenz	Nein	Ja	Ja	Ja
Führungsv. - Planungsverh.	Nein	Ja	Ja	Ja
Hyp. 6.2 wird	abgel.	bestät.	bestät.	bestät.
Hyp. 7.1: Die ext. konvergente Validität ist mit VCL höher als ohne.	Alte Aufgabe PS alt			
Führungsverhalten	Ja	Ja		
Sozialkompetenz	Ja	Ja		
Hyp. 7.1 wird	bestät.	bestät.		
Hyp. 7.2: Die ext. diskriminante Validität ist mit VCL höher als ohne.	Alte Aufgabe PS alt			
Führungsverhalten	Nein	Nein		
Sozialkompetenz	Nein	Nein		
Planungsverhalten	Nein	Nein		
Hyp. 7.2 wird	abgel.	abgel.		

	OPZ- Prüfer	Ext. Beo.
Hyp. 8.1: Die ext. konvergente Validität ist bei den neuen Aufgaben höher als bei den alten.	Ohne VCL PS alt versus PS KM	
Führungsverhalten	Nein	Ja
Sozialkompetenz	Nein	Nein
Hyp. 8.1 wird	abgel.	abgel.
Hyp. 8.2: Die ext. diskriminante Validität ist bei den neuen Aufgaben höher als bei den alten.	Ohne VCL PS alt versus PS KM	
Führungsverhalten	Ja	Nein
Sozialkompetenz	Ja	Ja
Planungsverhalten	Ja	Nein
Hyp. 8.2 wird	bestät.	abgel.

Im folgenden werden die wichtigsten Ergebnisse noch einmal dargestellt, in den inhaltlichen Zusammenhang des Veränderungsprozesses in der OPZ gestellt und vor diesem Hintergrund bewertet und diskutiert. Das weitere Kapitel 7 wird anhand der Gütekriterien gegliedert und eine kritische Diskussion der Methodik der Studie wird an den gegebenen Stellen eingeflochten.

7.2 Objektivität

Durchführungsobjektivität

Die Durchführungsobjektivität bezüglich Raum und Zeit der Untersuchungssituation ist für das GSV in der OPZ relativ hoch. Es findet für alle Bewerber zur selben Uhrzeit zwischen 7 und 9 Uhr morgens statt. Die Instruktionen sind sehr klar vorgegeben, so dass diesbezüglich die Unterschiede zwischen den verschiedenen Prüfgruppen zu vernachlässigen sind.

Eine Verletzung der Durchführungsobjektivität stellt die variable Spielzeit des alten Verfahrens dar. Die Chancen für einen Bewerber, das gewünschte Verhalten in der Gruppe zu zeigen, sind während einer 20-minütigen Aktivitätsphase sicherlich deutlich größer als in einer 7-minütigen. Für das neue Verfahren wurden daher feste Aktivitätszeiten für die einzelnen Gruppenaufgaben festgelegt.

Eine weitere Beeinträchtigung der Durchführungsobjektivität entsteht dadurch, dass für eine Hälfte der Bewerber das GSV an ihrem ersten Tag in der OPZ stattfindet, für die zweite Hälfte der Bewerber an ihrem zweiten Tag. Die Bewerber, die am ersten Tag das GSV durchlaufen, müssen daher mit tatsächlich fremden Mitbewerbern in der Gruppensituation agieren, während die Bewerber bis zum GSV am zweiten Tag bereits Gelegenheit hatten, sich kennen zu lernen und erste Kontakte zu knüpfen. Die soziale und vermutlich auch die emotionale

Die Situation der Bewerber ist daher an diesen beiden Tagen im GSV eine deutlich andere. Im Rahmen der bestehenden Organisation des Auswahlverfahrens ist dies aufgrund der nötigen Massendiagnostik jedoch nicht zu ändern. Im Rahmen der Datenerhebung für die vorliegende Studie wurde darauf geachtet, diesen Störfaktor zu kontrollieren. Die Daten wurden durchweg am jeweils ersten Tag der verschiedenen Prüfraten erhoben.

Auswertungs- und Interpretationsobjektivität

Die Auswertungs- und Interpretationsobjektivität bedeutet im Rahmen eines Gruppenverfahrens, dass in denselben Aufgaben verschiedene Beobachter dieselben Beobachtungen denselben Anforderungsdimensionen zuordnen und diese dann gleichsinnig beurteilen. Im Verfahren der OPZ gab es hierzu für die Prüfer nur geringe instrumentelle Unterstützung. Die Operationalisierung der Eignungsmerkmale war sehr allgemein gehalten, so dass von einer hohen Übereinstimmung über verschiedene Prüfgruppen hinweg nicht ausgegangen werden konnte.

Mit der Entwicklung der VCL wurde dieser Punkt konkret bearbeitet. Die Verhaltenscheckliste enthält für jedes Eignungsmerkmal Unterdimensionen, die ein gemeinsames Verständnis der komplexen Anforderungsmerkmale unterstützen sollen. Die Leistung der konkreten Verhaltensweisen je Dimension unterstützt einen gleichsinnigen Zuordnungsprozess von Beobachtungen zur Anforderungsdimension.

In die Erarbeitung der Checkliste waren einzelne Prüfer eingebunden. Zum Zeitpunkt der Datenerhebung für diese Studie war es jedoch nicht möglich, alle Prüfer, die an der Erhebung teilgenommen haben, gleichermaßen mit der VCL vertraut zu machen. Auch wurde vor der Datenerhebung kein Beobachtertraining durchgeführt. Dies ist für eine angemessene Interpretation der Ergebnisse eine wichtige Rahmenbedingung. Die Studie muss daher als isolierte Untersuchung des Einsatzes einer VCL verstanden werden ohne die üblicherweise in diesem Zusammenhang durchgeführten Trainings.

Im Rahmen der weiteren Projektdurchführung nach der Datenerhebung wurde die VCL über einen längeren Zeitraum in Zusammenarbeit mit den Prüfern weiter überarbeitet, es wurden schließlich Beobachtertrainings durchgeführt sowie Supervisionen der direkten Anwendung der neuen Gruppenaufgaben und der VCL. Es wird daher bis zum Projektabschluss eine deutliche Verbesserung der Auswertungs- und Interpretationsobjektivität angenommen.

Die quantitative Bewertung der Veränderungen der Objektivität geht im Assessment Center in der Regel einher mit der Bewertung der Reliabilität. Auf sie wird im nächsten Kapitel eingegangen.

7.3 Reliabilität

Die Reliabilität wurde mit Hilfe der Intraklassenkorrelation überprüft. Im Fokus der folgenden Kapitel werden die Ergebnisse der OPZ-Prüfer stehen, da für sie die Veränderungen nutzbringend sein sollen. Die Ergebnisse der externen Beobachter werden in die Diskussion einbezogen, wenn sich abweichende systematische Effekte erkennen lassen.

Die Ergebnisse zu den Hypothesen 1 und 2 lassen sich wie folgt zusammenfassen:

- Der Einsatz der VCL erhöht die Reliabilität beim alten Planspiel, jedoch nicht bei den neuen Aufgaben.
- Der Einsatz der neuen Aufgaben erhöht die Reliabilität, jedoch nur, wenn sie ohne die VCL eingesetzt werden.

Jede der beiden Interventionen für sich genommen bewirkt eine deutliche Verbesserung der Reliabilität, eine Kombination der beiden Interventionen dagegen bewirkt nur noch eine leichte Verbesserung im Verhältnis zur Ausgangssituation und eine Verschlechterung im Verhältnis zum isolierten Einsatz einer Intervention.

Die absolute Höhe der $ICC_{\text{unjust,einfakt.}}$ schwankt im Laufe der Datenerhebung zwischen .28 (Sozialkompetenz, PS alt, ohne VCL) und .83 (Führungsfähigkeit, PS KM, ohne VCL).

Insgesamt liegen die Werte der $ICC_{\text{unjust,einfakt.}}$ dieser Studie etwas unter den Werten aus anderen Reliabilitätsstudien (vgl. Kap. 1.3.1) Eine Ursache hierfür liegt vermutlich in der Wahl der $ICC_{\text{unjust,einfakt.}}$ als Messgröße. Dadurch, dass die Mittelwertsunterschiede zwischen den Prüfern als Fehler verrechnet werden, fällt sie kleiner aus als die üblicherweise verwendeten Produkt-Moment-Korrelationen.

Ein weiterer Faktor, der sich auf die Höhe der Reliabilität auswirkt, ist das Verfahren der Urteilsfindung. Gibt es eine Aussprache zwischen den Ratern vor der Urteilsvergabe, erhöht sich in der Regel die Reliabilität (vgl. Scholz, 1994, S. 166). Das Verfahren in der OPZ sieht vor, dass jeder Prüfer direkt im Anschluss an die Gruppenaufgaben seine eigenen Noten vergibt, so dass auch dieser Umstand zu einer Erklärung für die etwas niedrigeren Werte beiträgt. Erst im Anschluss werden die Noten in ein gemeinsames Protokoll übernommen, wo sie zu einer Endbeurteilung verrechnet werden (die in dieser Studie nicht erfasst wurde). Treten hier deutliche Diskrepanzen zwischen den Einzelnoten zu Tage, können die Prüfer nach einer Diskussion ihre Einzelnoten gegebenenfalls anpassen.

Eine wichtige Rahmenbedingung für die Beurteilung der Ergebnisse zur Reliabilität ist in der Tatsache zu sehen, dass die Prüfer während der Datenerhebung durchgängig auch Urteile nach ihrem alten Bewertungsverfahren vergeben mussten, um die Bewerbergerechtigkeit innerhalb des Jahrganges zu gewährleisten. Dies bedeutete an den Terminen, an denen die Urteile mit der VCL vergeben wurden, eine deutliche Zusatzbelastung.

Für die Durchführung und Beurteilung des alten Planspiels können die Prüfer auf sehr viel Erfahrung zurückgreifen, so dass die Zusatzbelastung durch die VCL in Block 2 kompensiert werden kann. In Block 4 jedoch kumulieren die Anforderungen, ein neues Planspiel durchzuführen, mit dem erst an maximal 2 Terminen erste Erfahrungen gesammelt wurden (Block 3) und die Anforderung, die Verhaltensweisen der OB nach zwei verschiedenen Verfahren zu bewerten. Zudem kam aus diesem Grund eine wesentliche Verbesserung des neuen Beurteilungsverfahrens während der Datenerhebung nicht zum Tragen, nämlich die Reduzierung der zu bewertenden Merkmale. Nach dem neuen Beurteilungsverfahren mit VCL werden die

Merkmale Ausdrucksfähigkeit/Argumentationsverhalten sowie Urteilsfähigkeit in den Planspielen nicht mehr bewertet, um durch die Reduzierung der Merkmale eine Verbesserung der Beobachtungs- und Beurteilungsqualität zu ermöglichen. Es kann daher für Block 4 eine Überforderung der Aufmerksamkeitsleistung und Informationsverarbeitungskapazität der Prüfer vermutet werden, die sich negativ auf die Reliabilität ausgewirkt hat. Diese Beeinträchtigung entfällt in Zukunft bei einem normalen Einsatz der neuen Aufgaben und der VCL.

Eine weitere Ursache dafür, dass der Einsatz der VCL die Reliabilität bei den neuen Aufgaben nicht erhöht, kann in den auf der VCL enthaltenen Operationalisierungen selbst gesehen werden. Aufgrund begrenzter Projektressourcen wurden die VCL und die neuen Aufgaben parallel entwickelt. Dies bedeutet, dass die Items der VCL vor dem Hintergrund der in Kap. 2.4 beschriebenen theoretischen Fundierung aus Beobachtungen der alten Planspiele der OPZ abgeleitet wurden. Es ist daher möglich, dass sie für die neuen Aufgaben zum Zeitpunkt der Datenerhebung noch nicht die optimale Passung aufweisen.

Interessant ist die Tatsache, dass sich auch der ICC-Wert für die Dimension Argumentationsverhalten/Ausdrucksfähigkeit beim Einsatz der VCL im alten Planspiel verbessert, obwohl diese Dimension nicht auf der Verhaltenscheckliste enthalten ist. Dies könnte ein Hinweis dafür sein, dass die VCL in dieser Studie grundsätzlich weniger über die auf ihr enthaltenen Operationalisierungen wirkt, sondern vielmehr über eine Steigerung der Aufmerksamkeitsleistung, die durch die Handhabung des neuen Instrumentes notwendig wird. Dies würde auch die These der Überforderung der Prüfer in Block 4 durch die Anwendung der beiden neuen Instrumente stützen.

Wie in Kap. 7.2 bereits erwähnt, wurden im weiteren Projektverlauf weitere Erprobungen und Trainings durchgeführt, so dass zum Projektabschluss von einer deutlichen Verbesserung der Reliabilität auch bei Kombination der Interventionen ausgegangen werden kann.

Schließlich ist noch die Auswirkung eines fehlenden Trainings für die OPZ-Prüfer vor dem Zeitraum der Datenerhebung für diese Studie zu diskutieren. Der Vorteil dieser Rahmenbedingung ist darin zu sehen, dass die Auswirkungen der VCL und der neuen Aufgaben isoliert untersucht werden konnten. Dank der Vielzahl der Studien zum AC ist die Aussage, dass ein Beobachtertraining sich positiv auf die Reliabilität auswirkt, als gesichert zu betrachten. Mit Hilfe des Trainings wird ein übereinstimmendes Verständnis der Konstrukte angestrebt sowie die professionelle Beobachtungs- und Beurteilungsfähigkeit geschult. Art und Weise sowie Qualität des Trainings nehmen also einen erheblichen Einfluss auf die Qualität der Diagnostik und wirken in diesem Sinne in vielen Studien als schwer zu kontrollierende Störvariablen bezüglich instrumenteller Veränderungen des AC-Verfahrens.

Ab einem bestimmten Ausmaß der Veränderung der Diagnose-Instrumente ist jedoch, wie diese Studie zeigt, ein Training notwendig, damit die Veränderungen positive Wirkung entfalten können. Dieses Ausmaß scheint dann erreicht zu sein, wenn sich für die Prüfer der inhaltliche Referenzrahmen für die Beurteilung ändert.

Sehr deutlich wird dies an der Entwicklung der ICC-Werte für die Dimension Urteilsfähigkeit. Die Beurteilung dieses Merkmals beruht weniger auf den sozialen Verhaltensweisen der Be-

werber sondern explizit auf einer inhaltlichen Bewertung ihrer Aussagen. Für das neue Planspiel zum Ressourcenengpass entstehen in der Folge für das Merkmal Urteilsfähigkeit sehr geringe ICC-Werte, da die Gruppenaktivität sich um den Ausgleich persönlicher Interessen dreht. Damit sind die Aussagen der Bewerber inhaltlich nur schwer zu bewerten. Im neuen Planspiel zum Krisenmanagement ohne VCL entsteht dagegen ein ICC-Wert von .59 für Urteilsfähigkeit, da sich in dieser neuen Planungsaufgabe die Aussagen wie gewohnt inhaltlich als mehr oder weniger sinnvoll bewerten lassen. In Block 4 dagegen mit Einsatz der VCL zum neuen Planspiel Krisenmanagement entsteht nur noch ein ICC-Wert von .20 für das Merkmal Urteilsfähigkeit. Dieses Merkmal ist auf der VCL nicht enthalten und eine parallele Bewertung innerhalb eines neuen Planspiels, also ohne den gewohnten Referenzrahmen, stellt vermutlich eine Überforderung der Beobachter dar.

Die zwei Faktoren, die bei der Realisierung von Verbesserungen durch Verfahrensänderungen (ohne Training) in dieser Studie eine Rolle zu spielen scheinen sind also zum einen die Beanspruchung von Aufmerksamkeit und Informationsverarbeitung und zum anderen das Ausmaß der Veränderung des Referenzrahmens für die Urteilsbildung.

7.4 Differenzierung

Das Kriterium der Differenzierung der Bewerber gehört zwar nicht zu den Gütekriterien der klassischen Testtheorie, ist jedoch für die Praxis der Personalauswahl von wesentlicher Bedeutung. Das Ziel eines jeden Auswahlverfahrens ist es, die Bewerber bezüglich bestimmter Merkmale differenzieren zu können. Auch wenn für die reine Auswahlentscheidung die Information ausreichend wäre, dass ein OB die Anforderungen erfüllt, bzw. nicht erfüllt hat, ist es im Sinne eines AC anzustreben, die Auswahlentscheidung auf der Basis einer Zusammenschau differenzierter Urteile zu treffen.

Für die Diagnose der sozial-interaktiven Kompetenzen der Bewerber in der OPZ stellt die Differenzierung der Bewerber eine besondere Herausforderung dar. Aufgrund ihres jugendlichen Alters und der geringen oder nicht vorhandenen beruflichen Erfahrungen unterscheiden sich die Bewerber in geringerem Maße bezüglich beruflich relevanter sozial-interaktiver Verhaltensweisen, als dies üblicherweise in AC zur Führungskräfteauswahl der Fall ist.

Die Differenzierung der Bewerber wurde mit Hilfe der Standardabweichung gemessen. Die Ergebnisse zu den Hypothesen 3 und 4 lassen sich wie folgt zusammenfassen:

- Der Einsatz der VCL erhöht die Differenzierung der Bewerber im alten Planspiel und erhöht, bzw. erhält die Differenzierung bei den neuen Aufgaben.
- Der Einsatz der neuen Planspiele erhöht die Differenzierung der Bewerber, jedoch nur, wenn sie ohne VCL eingesetzt werden.

Jede der beiden Interventionen für sich genommen bewirkt eine deutliche Verbesserung der Differenzierung, eine Kombination der beiden Interventionen bewirkt auch eine deutliche Verbesserung im Verhältnis zur Ausgangssituation, sie fällt jedoch geringer aus als bei dem isolierten Einsatz der VCL zum alten Planspiel.

Die Ergebnisse entsprechen in ihren Tendenzen den Ergebnissen zur Reliabilität. Dies ist methodisch nachvollziehbar, da der Mittelwertsunterschied zwischen den Bewerbern die Höhe der ICC beeinflusst, also ein geringer Mittelwertsunterschied bei gleicher Rater-Übereinstimmung zu einer geringeren ICC führt. Sind die Rater nicht in der Lage, die Bewerber zu differenzieren, so führt dies zu einem niedrigen Niveau der ICC. Da die Standardabweichung in Block 4 relativ hoch ist, die ICC in Block 4 jedoch sehr niedrig, liegt die Ursache für die niedrige ICC tatsächlich in einer geringeren Rater-Übereinstimmung als in den anderen Blöcken.

Aus den Tabellen 4-2 und 4-3 ist zu ersehen, dass die Standardabweichungen der Urteile der OPZ-Prüfer für die Dimensionen Argumentationsverhalten/Ausdrucksfähigkeit sowie Urteilsfähigkeit im Durchschnitt unter denen für die Dimensionen Führung, Planung und Sozial- bzw. Teamkompetenz liegen. D.h. es gelingt den Prüfern schlechter, die Bewerber im GSV anhand der Dimensionen Argumentationsverhalten/Ausdrucksfähigkeit sowie Urteilsfähigkeit zu differenzieren. Dies bestätigt die Entscheidung, auf diese Dimensionen im neuen Verfahren zu verzichten.

Tabelle 4-2 zeigt zudem, dass die Urteile aus dem Rundgespräch die niedrigsten Standardabweichungen aufweisen, so dass die Entscheidung, diese Aufgabenart wegen des geringen Differenzierungsvermögens nicht mehr zu verwenden, ebenfalls empirisch bestätigt werden konnte.

Auffällig ist schließlich, dass sowohl bei den alten wie auch bei den neuen Aufgaben die Standardabweichung für die Dimension Soziale Kompetenz in allen Auswertungen deutlich unter den Standardabweichungen der anderen Dimensionen Führung und Planung liegen. Auch die Reliabilitätswerte zur Sozialen Kompetenz weichen häufig nach unten von den übrigen Reliabilitätswerten ab. Bezüglich dieser Dimension scheint das Konstruktverständnis der Prüfer am wenigsten geklärt zu sein und es fällt den Prüfern, unabhängig von Verfahrensänderungen, am schwersten, die Bewerber zu differenzieren. Die Urteile zur neu eingeführten Dimension Teamkompetenz im Planspiel Ressourcenengpass weisen eine deutlich höhere Standardabweichung auf, als die Urteile zur Sozialkompetenz in den verschiedenen Untersuchungsblöcken.

7.5 Konstruktvalidität

Die Konstruktvalidität wurde mit Hilfe der Multitrait-Multimethod-Analyse bewertet. Da die Auswahl der Konstrukte und ihre Verteilung auf die Aufgaben entsprechend der diagnostischen Ziele der OPZ gestaltet wurden, ergab sich für die Konstruktvalidierung jedoch keine 3x3-Matrix, so dass die Bewertungsmöglichkeiten, insbesondere bezüglich der konvergenten Validität, stark eingeschränkt waren. Eine weitergehende kritische Diskussion der Multitrait-Multimethod-Methode wird im Anschluss an die Ergebnisdarstellung folgen.

Die Ergebnisse zu den Hypothesen 5 und 6 lassen sich wie folgt zusammenfassen:

- Der Einsatz der neuen Aufgaben ohne VCL führt zu einer erhöhten konvergenten Validität für die Dimension Sozialkompetenz
- Der Einsatz der VCL bei den neuen Aufgaben führt zu einer weiteren Erhöhung der konvergenten Validität für Sozialkompetenz.
- Weder der Einsatz der VCL noch der Einsatz der neuen Aufgaben gehen mit einer eindeutigen Entwicklung der diskriminanten Validität einher.
- Für die OPZ-Prüfer ist die diskriminante Validität ohne jede Intervention am größten, für die externen Beobachter jedoch bei den neuen Planspielen ohne Einsatz der VCL.

Die Kombination der beiden Interventionen führt zur größten Verbesserung der konvergenten Validität für Sozialkompetenz. Eine Verbesserung der diskriminanten Validität konnte für die OPZ-Prüfer durch die Interventionen nicht erreicht werden, für die externen Beobachter ist aber die diskriminante Validität in Block 3 und 4 deutlich verbessert.

7.5.1 Kritische Diskussion der MTMM-Analyse

In Relation zu anderen Studien sind die diskriminanten Koeffizienten zwischen den Dimensionen Planung und Soziale Kompetenz (.41), sowie Führung und Soziale Kompetenz (.46) aus Block 1 auffallend niedrig, sie beschreiben also eine gute diskriminante Validität. Maukisch berichtet von mehrfach replizierten Ergebnissen in der amerikanischen Forschung, die eine Korrelation verschiedener Dimensionen in der gleichen Übung von im Durchschnitt .70 ergeben haben (Maukisch, 1989).

Die konvergente Korrelation für die Dimension Soziale Kompetenz liegt dagegen sowohl in den alten als auch in den neuen Aufgaben mit Werten zwischen .57 und .78 im Vergleich zu dem von Maukisch angegebenen Durchschnittswert von .31 deutlich höher und steht für eine relativ gute konvergente Validität.

Daraus ergibt sich die Situation, dass der konvergente Koeffizient sowohl in Block 1 als auch in Block 4 leicht über zwei diskriminanten Koeffizienten liegt und damit eine wesentliche Forderung aus der MTMM-Analyse für Konstruktvalidität erfüllt ist, die nur in wenigen AC-Studien erreicht wurde. Ein ähnliches Ergebnis wurde in einer Studie von Kleinmann et al. (1992) beschrieben. Auch bei dieser Untersuchung wurden ausschließlich Ergebnisse von Gruppendiskussionen verwendet, so dass dies eine Ursache für die erhöhte konvergente Validität darstellen könnte. Mögliche Ursachen für die niedrige diskriminante Validität in der Ausgangssituation werden weiter unten noch dargestellt.

Unabhängigkeit der Methoden

Die hohe konvergente Validität kann also damit zusammen zu hängen, dass der sonst in AC auftretende Methodeneffekt weniger wirksam wird. Der Einfluss der Methode ist bei zwei verschiedenen strukturierten Planspielen sicher geringer als bei der Bewertung der Ergebnisse von z.B. einer Postkorbübung und einer Gruppenaufgabe. Von Seiten der Prüfer wurde die

Gruppenaktivität in den beiden verschiedenen Planspielen jedoch durchaus als sehr unterschiedlich beschrieben. Daraus ergibt sich die Frage nach der Unabhängigkeit der Methoden in dieser Studie, wie sie für die Auswertung im Zusammenhang mit der Multitrait-Multimethod-Analyse gefordert wird. Die Verschiedenartigkeit der Aufgabenstellung kann sinnvollerweise nur ein Kriterium für die Unabhängigkeit der Messungen sein. Ein weiteres ist die Zuordnung der Beobachter zu den Teilnehmern. Hier ist kritisch anzumerken, dass in dieser Studie die Urteile zu den zwei Planspielen von den gleichen Beobachtern in direkter zeitlicher Aufeinanderfolge getroffen wurden. Das Ergebnis des ersten Planspiels beeinflusst also zwangsläufig das Ergebnis des zweiten Planspiels. Auch dass die Teilnehmer zwei Mal hintereinander in der gleichen Gruppe agieren, führt aufgrund der Gruppendynamik zu einer Beeinflussung des zweiten durch das erste Planspiel. Eine Unabhängigkeit der Methoden ist daher, unabhängig von der Verschiedenartigkeit der Aufgaben, nur bedingt gegeben. In wie weit es wiederum überhaupt möglich ist, genau das gleiche Konstrukt mit Hilfe zweier unterschiedlicher Testverfahren zu erheben, soll an dieser Stelle nicht weiter diskutiert werden.

Unabhängigkeit der Dimensionen

Zwei der diskriminanten Koeffizienten in der Ausgangssituation in Block 1 sind für die OPZ-Prüfer auffallend niedrig, d.h. sie sprechen für eine relativ gute diskriminante Validität für die Dimensionen Planungsverhalten und Sozialkompetenz sowie Führungsverhalten und Sozialkompetenz. Führungsverhalten und Planungsverhalten können jedoch nicht gut differenziert werden ($r = .85$). Dies wirft die Frage nach der Unabhängigkeit der Dimensionen auf.

Ein erstes Argument, dass grundsätzlich gegen eine völlige Unabhängigkeit von sozial-interaktiven Konstrukten spricht, ist der Umstand, dass alle Konstrukte aus diesem Bereich über das kommunikative Verhalten der Teilnehmer erst beobachtbar werden. Jemand, der nur über ein begrenztes Kommunikationsvermögen verfügt, wird vermutlich in allen Zieldimensionen einer Gruppenaufgabe eher schlecht bewertet werden. Erst die Teilnehmer, die über ein umfangreiches Kommunikationsvermögen verfügen, können dann gegebenenfalls differenziert bewertet werden.

Für die Planspiele in der OPZ stellt sich empirisch die Differenzierung zwischen Planungs- und Führungsverhalten als besonders problematisch dar. Inhaltlich ist dieses Ergebnis sehr plausibel und im Rahmen der Möglichkeiten nur schwer veränderbar, denn aufgrund der geringen Führungserfahrung der Bewerber und der kurzen Spielzeit verschaffen inhaltlich sinnvolle Planungsbeiträge Respekt und Akzeptanz bei den Mitbewerbern. Es gibt in dieser kurzen Spielzeit kaum andere Möglichkeiten, als über inhaltliche Stellungnahmen Führungsverhalten zu zeigen.

Dies deckt sich mit einem Befund von Kleinmann et al. (1995) (vgl. Kapitel 1.4): Er hat empirisch den Zusammenhang nachweisen können, dass abhängige Dimensionen mit einer geringeren diskriminanten Validität einhergehen. Einen weiteren Grund für die grundsätzlich geringe diskriminante Validität im AC vermutet er in der zeitgleichen Messung von Konstrukten, die allerdings in der Praxis Standard ist.

Aussagekraft der MTMM-Analyse

Grundsätzlich ist fest zu halten, dass die von Campbell und Fiske entwickelte und in dieser Studie eingesetzte so genannte „informelle Auswertung“ der MTMM-Matrix keine Ergebnisse hervorbringen kann, die gegen den Zufall abgesichert sind. Auch bietet diese Methode weder objektive Grenzwerte, wie hoch die einzelnen Korrelationskoeffizienten bzw. deren Differenzen sein müssen noch ob, bzw. wie viele einzelne Verletzungen der Kriterien erlaubt sind. Durch die informelle Inspektion von MTMM-Matrizen kann man daher immer nur zu relativen Aussagen über die Validität von Konstrukten gelangen. Trotz dieser Kritik gilt die MTMM-Analyse als etabliertes Verfahren zur Bewertung von Konstruktvalidität. Da es in dieser Studie explizit um den Vergleich zweier Auswahlinstrumente geht, die unter den gleichen Rahmenbedingungen eingesetzt werden, also um relative Aussagen zur Konstruktvalidität, kann die MTMM-Analyse als geeignete Methode bezeichnet werden.

Wird das Verfahren der Konstruktvalidierung, wie in Kap. 1.2.3 beschrieben, als umfassender Prozess verstanden, der theoretisch-psychologische Vorüberlegungen bei der Verfahrenskonstruktion mit der Überprüfung empirischer Ergebnisse verbindet, so ist die Bewertung der Konstruktvalidität in Relation zu einem wie auch immer gearteten „objektiven“ Vergleichswert ohnehin wenig aussagekräftig.

7.5.2 Wirkungsweise der VCL auf die Konstruktvalidität

Der Einsatz der VCL in dieser Studie hat nicht zu einer vollständigen Replizierung der Ergebnisse von Reilly und Henry (1990) geführt, die neben einer Verbesserung der konvergenten Validität auch eine leichte Verbesserung der diskriminanten Validität beschrieben haben. Dafür lassen sich mehrere Gründe anführen:

- Die Entwicklung der VCL bis zum Zeitpunkt der Datenerhebung musste im Rahmen der Projektressourcen mit deutlich weniger Aufwand durchgeführt werden, als dies bei Reilly und Henry der Fall war (vgl. Kap. 1.4).
- Von den Prüfern waren nur einige wenige in die Entwicklung der VCL eingebunden. D. h., es gab vor dem Einsatz der VCL keinerlei Maßnahmen, die ein für alle Prüfer der VCL entsprechendes Konstruktverständnis unterstützt hätten. Die Daten zeigen daher ausschließlich die Wirkung, die mit dem unmittelbaren Einsatz der VCL zusammenhängt.

Interessanterweise geht der Einsatz der VCL bei den alten Aufgaben mit einer deutlichen Verschlechterung der diskriminanten Validität einher, während sich bei den neuen Aufgaben, wenn auch auf einem sehr hohen Niveau, zwei der drei Koeffizienten leicht verbessern (vgl. Tab. 5-3 und 5-3). Eine Ursache für den ersten Effekt könnte darin liegen, dass die Prüfer im gewohnten Verfahren die Konstrukte wegen eines fehlenden Beobachtertrainings mit Hilfe „gewachsener“ impliziter Eignungstheorien bewerten, die sich nicht in den Items der VCL abbilden lassen und sich an inhaltlichen Äußerungen der Bewerber orientieren. Der Einsatz der VCL ist damit keine Unterstützung für die OPZ-Prüfer sondern bewirkt bezüglich des Konstruktverständnisses eher eine Verwirrung. Dafür spricht der Umstand, dass sich bei den

externen Beobachtern zwei der drei diskriminanten Koeffizienten mit Einsatz der VCL deutlich verbessern.

Wenn diese Vermutung zutrifft, lässt sich auch die sehr schlechte diskriminante Validität in den neuen Aufgaben erklären. Die impliziten Eignungstheorien werden von neuen Prüfern im Rahmen eines „Training on the job“ bei der Einarbeitung von den erfahrenen Prüfern übernommen und orientieren sich eher an aufgabenspezifischen inhaltlichen Aussagen der Bewerber und weniger an Verhaltensankern. Damit sind diese Orientierungspunkte nicht auf neue Planspielszenarios übertragbar. Auch diese Erklärung lässt sich mit Hilfe der diskriminanten Koeffizienten der externen Beobachter stützen, die bei den neuen Aufgaben fast durchweg auf einem sehr niedrigen Niveau liegen.

Es wird deutlich, dass der Einsatz der VCL durchaus systematische Effekte bezüglich der diskriminanten Validität hervorruft, diese jedoch durch die Prüfergruppe (OPZ-Beobachter versus externe Beobachter) moderiert werden. Mit Hilfe der beschriebenen Mechanismen lassen sich die zum großen Teil gegenläufigen Ergebnisse der OPZ-Beobachter und der externen Beobachter bezüglich des Einsatzes der VCL erklären.

Es ist daher zu erwarten, dass mit Hilfe der Trainings und Supervisionen für die OPZ-Beobachter dieser moderierende Faktor gemildert werden kann und zum Projektende von einer deutlichen Verbesserung der Wirkung der VCL für die OPZ-Beobachter ausgegangen werden kann.

7.6 Externe Konstruktvalidität

Zur Überprüfung der externen Konstruktvalidität wurde der BIPTM-Fragebogen eingesetzt. Die Ergebnisse zu den Hypothesen 7 und 8 lassen sich wie folgt zusammenfassen:

- Die konvergente Validität für die Dimensionen Führung und Soziale Kompetenz steigt mit Einsatz der VCL beim alten Planspiel, die diskriminante Validität verschlechtert sich.
- Die konvergente Validität ist bei dem neuen Planspiel Krisenmanagement mit VCL schlechter als beim alten Planspiel mit VCL, die diskriminante Validität ist dagegen verbessert.

Weitere Aussagen lassen sich aufgrund der unvollständigen Datenlage des BIPTM (vgl. Kap. 3.1.1) leider nicht treffen.

Zur Verbesserung der konvergenten Validität durch den Einsatz der VCL im alten Planspiel ist anzumerken, dass fast alle Dimensionen des BIPTM eine deutlich höhere Korrelation mit den GSV-Urteilen aufweisen. Gleiches gilt für die Verbesserung der diskriminanten Validität bei Einsatz des neuen Planspiels Krisenmanagement mit VCL. Auch hier sind fast alle Korrelationen zwischen den GSV-Urteilen und den BIPTM-Dimensionen niedriger als beim alten Planspiel mit VCL.

Dies spricht dafür, dass die Verbesserung der konvergenten, bzw. diskriminanten Validität nicht auf eine genauere Erfassung der Konstrukte Führung und Soziale Kompetenz zurück-

zuführen ist, sondern vielmehr auf eine grundsätzliche Beeinflussung des Urteilsverhaltens, die über die Dimensionen hinweg wirksam wird. Weiter verdeutlichen diese Ergebnisse, dass empirisch nicht von einer vollständigen Unabhängigkeit der Dimensionen ausgegangen werden kann.

7.6.1 Nomologisches Netzwerk

In diesem Kapitel werden die Ergebnisse der externen Konstruktvalidierung bezüglich ihrer Aussagefähigkeit im Sinne einer inhaltlichen Konstruktklärung diskutiert. Dabei beschränkt sich die Auswertung der Korrelationen auf systematische Effekte, da einzelne Werte auch auf Artefakte zurückzuführen sein können.

Es ist auffällig, dass für alle drei GSV-Dimensionen, Führung, Soziale Kompetenz und Planungsverhalten, beim alten Planspiel mit Einsatz der VCL die Korrelationen mit den gleichen drei BIPTM-Dimensionen sinken. Die BIPTM-Dimensionen Emotionale Stabilität, Soziabilität und Teamorientierung korrelieren mit Einsatz der VCL geringer mit allen drei Dimensionen des Planspiels. Dies deutet ebenfalls darauf hin, dass der Einsatz der VCL im alten Planspiel zu einer dimensionsunabhängigen Beeinflussung des Urteilsprozesses bei den OPZ-Beobachtern führt.

Zudem erhöht sich für alle drei Dimensionen des Planspiels die Anzahl der BIPTM-Dimensionen, die relativ hoch mit den Planspiel-Dimensionen korrelieren. Ohne Einsatz der VCL korreliert im Wesentlichen nur die Dimension Selbstbewusstsein relativ hoch mit allen drei Dimensionen aus dem GSV. Mit Einsatz der VCL kommen bei allen drei Dimensionen des Planspiels die BIPTM-Dimensionen Durchsetzungsstärke und Gestaltungsmotivation hinzu, sowie Führungsmotivation bzw. Kontaktfähigkeit bei jeweils zwei von drei Planspiel-Dimensionen. Dies stimmt mit den Ergebnissen anderer Studien bezüglich erfolgsrelevanter Persönlichkeitsmerkmale im AC überein (vgl. Kap. 1.3.3). Inhaltlich können diese BIPTM-Dimensionen als allgemeine Voraussetzungen für ein zielorientiertes aktives Verhalten innerhalb einer Gruppe aufgefasst werden. Diese Korrelationsmuster lassen daher nicht darauf schließen, dass der Einsatz der VCL zu einer differenzierteren Bewertung der Planspieldimensionen führt.

Bezüglich des Vergleichs des alten Planspiels mit VCL mit dem neuen Planspiel Krisenmanagement mit VCL zeigt sich eine Umkehrung dieser Effekte. Die drei BIPTM-Dimensionen Emotionale Stabilität, Soziabilität und Teamorientierung korrelieren beim neuen Planspiel deutlich höher mit den Dimensionen Führung und Planung als beim alten Planspiel. Die Anzahl der BIPTM-Dimensionen, die relativ hoch mit den Dimensionen des Planspiels korrelieren, wird geringer. Einen nennenswerten Zusammenhang mit den Planspieldimensionen Führung und Planung weisen nur noch die BIPTM-Dimensionen Selbstbewusstsein und Belastbarkeit auf.

Zur Dimension Soziale Kompetenz gibt es beim neuen Planspiel keine BIPTM-Dimension mehr, die höher als .22 mit dem GSV-Urteil korreliert. Weiter oben wurde bereits die Vermutung angestellt, dass die OPZ-Prüfer zur Dimension Soziale Kompetenz am wenigsten über

ein klares Konstruktverständnis verfügen. Diese Vermutung wird durch die geringen Korrelationen mit allen BIPTM-Dimensionen in Block 4 unterstützt.

Betrachtet man die Items der VCL, so wird deutlich, dass das Konstrukt der Sozialen Kompetenz vollkommen aufgabenneutral bewertet werden kann. Zur Bewertung der Items zu Führung und Planung ist dagegen eine inhaltliche Beurteilung der Aussagen der Bewerber ein Teil der Bewertung. Während die OPZ-Beobachter die inhaltliche Bewertung aus dem alten Verfahren kennen, besaßen sie zum Zeitpunkt der Datenerhebung keine explizite Schulung in der reinen Verhaltensbeobachtung und -bewertung. Während sie im alten Planspiel auch mit Einsatz der VCL noch auf ihren gewohnten Referenzrahmen zur Bewertung der Sozialen Kompetenz zurückgreifen können, ist dies im neuen Planspiel nicht mehr möglich. Daraus entsteht vermutlich die Unklarheit bezüglich des Konstruktes Soziale Kompetenz, die aus den Daten in Block 4 zu erkennen ist.

Bei den externen Beobachtern ist diese Verringerung der Anzahl relativ hoch korrelierender Dimensionen nicht festzustellen, das Muster, welche Dimensionen jeweils relativ hoch korrelieren, bleibt weitestgehend erhalten. Es lässt sich kein systematischer Effekt durch die Kombination der Intervention bezüglich der inhaltlichen Zusammenhänge zwischen den GSV-Urteilen und den BIPTM-Ergebnissen feststellen. Einzelne Ausreißer-Werte können ihre Gründe in der geringen Fallzahl der externen Beobachter haben.

Für beide Beobachtergruppen gilt, dass sich die GSV-Dimensionen Führung und Planung nach ihren Korrelationen mit den BIPTM-Dimensionen nicht differenzieren lassen.

Zusammenfassung

Die Ergebnisse der externen Validierung zeigen, dass sowohl der Einsatz der VCL als auch der neuen Aufgaben systematische Effekte auf die Bewertung der Dimensionen ausüben.

Beim alten Planspiel führt der Einsatz der VCL zu einer Erweiterung des Fokus, die Beobachter erfassen mehrere Persönlichkeitsfacetten über das Merkmal Selbstbewusstsein hinaus. Die Ergebnisse nähern sich den Resultaten anderer Studien an, die ergeben haben, dass Persönlichkeitsmerkmale wie Durchsetzungsstärke, Führungsmotivation und Soziale Kompetenz (im BIPTM u.a. durch die Kontaktfähigkeit erfasst) mittelhoch mit Erfolg im AC korrelieren.

Die Korrelationen aller GSV-Dimensionen mit den BIPTM-Dimensionen Emotionale Stabilität, Soziabilität und Teamorientierung sinken.

Die Kombination der Interventionen, neues Planspiel mit VCL, führt zu einer Umkehrung des ersten Effektes bei den OPZ-Beobachtern. Die Urteile der OPZ-Beobachter korrelieren nur noch nennenswert mit der BIPTM-Dimension Selbstbewusstsein. Für die Dimension Soziale Kompetenz zeigt sich kein nennenswerter Zusammenhang mehr mit einer BIPTM-Dimension. Diese Umkehrung lässt sich bei den externen Beobachtern nicht feststellen, bei ihnen bleibt das Muster der relativ hoch korrelierenden Dimensionen erhalten.

Auch diese Ergebnisse stützen daher die Vermutung, dass bei den OPZ-Beobachtern die Kombination der beiden Interventionen zu einer Überforderung der Aufmerksamkeitsleistung und Informationsverarbeitungskapazität führen, das sie stärker als die externen Beobachter dem alten Verfahren und den dort erlernten Urteilsmechanismen verhaftet sind. Eine zusätzliche Entlastung für die externen Beobachter lag darin, dass sie keine Aufgaben hinsichtlich Durchführung des Verfahrens oder Kommunikation mit den Bewerbern übernommen haben.

7.6.2 Eignung des BIP™ zur externen Validierung

Aus welchen Gründen der BIP™ zur externen Validierung ausgewählt wurde, wird bereits in Kap. 3.1.2 erläutert. Der Einsatz dieses Instruments ist jedoch aus zwei Gründen auch kritisch zu bewerten.

Der erste liegt in der grundsätzlichen Problematik, Persönlichkeitsmerkmale durch Selbstauskünfte in Personalauswahlverfahren zu erheben. Die Teilnehmer versuchen aufgrund der Auswahl-situation die Konstrukte hinter den Fragen zu erfassen und sozial erwünscht zu antworten. Dadurch, dass ihnen das mehr oder weniger gut gelingt, kommt es zu den oben bereits einmal angesprochenen Artefakten, die die Auswertung der Ergebnisse erschweren.

Der zweite Grund liegt darin, dass die Probanden, die den BIP™ ausfüllen, laut Handbuch älter als 20 Jahre sein sollten. Die ist nur bei gut zwei Drittel der OPZ-Bewerber der Fall, wie die Beschreibung der Stichprobe in Kap. 3.1.3 zeigt. Daraus könnten verzerrende Effekte und eine mangelnde Differenzierung der jüngeren Bewerber resultieren, da in jüngerem Alter viele der berufsrelevanten Persönlichkeitsmerkmale noch nicht so stark ausgeprägt sind und mit Hilfe des Fragebogens möglicherweise nicht adäquat erfasst werden können.

Schließlich ergibt sich die theoretische Frage, in wie weit dasselbe Konstrukt erfasst werden kann, wenn es zum einen über eine Selbstauskunft, zum anderen über Verhaltensbeobachtung erhoben wird.

Für alle drei Punkte gilt jedoch, dass die Kritik zwar berechtigt ist und die Ergebnisse vor diesem Hintergrund bewertet werden müssen, es jedoch an einer besseren und in der Praxis einer aktiven Organisation umsetzbaren Alternative fehlt.

7.7 Kontentvalidität

Aussagen zur Kontentvalidität können nur vor dem Hintergrund konzeptioneller Überlegungen gemacht werden.

In der Ausgangssituation bestand das GSV aus einer Themensammlung für die Rundgespräche und einer Themensammlung für die Kurzvorträge sowie verschiedenen Planspielaufgaben. Kurzvortrag und Rundgespräch hatten von ihrer Aufgabenstellung her einen eher intellektuellen Schwerpunkt. Die Themenlisten für diese beiden Aufgaben deckten systematisch verschiedene gesellschaftliche Bereiche ab. Die Aufgabenstellungen für das Planspiel wiesen keine Gemeinsamkeiten auf, außer dass sie für die relativ kurze Spielzeit von maximal 20 Minuten anwendbar waren, keine Hilfsmittel nötig waren und kein spezielles Wissen der Bewerber erforderten.

Bezüglich des Vortrages ist die Inhaltsvalidität schwer zu vergleichen, da sich die Art der Anforderung an die Bewerber deutlich verändert hat. Auch die alte Anforderung, eine Fragestellung intellektuell zu durchdenken und diese Gedankengänge verständlich zu präsentieren, hat für die Offizierbewerbersauswahl sicher seine Berechtigung und kann als inhaltsvalide bezeichnet werden. Die neue Form des Vortrags legt den Schwerpunkt auf die Anforderung, vor dem Hintergrund eines kurzen Szenarios eine Entscheidung zu treffen, diese vor den Mitbewerbern zu vertreten und sie zu überzeugen. Auf diese Weise wird eine sehr typische und erfolgskritische Anforderung an Offiziere abgebildet, so dass auch hier von einer guten Inhaltsvalidität ausgegangen werden kann.

Die beiden neuen Planspielarten unterscheiden sich von ihrer Konzeption her deutlich von der bisher eingesetzten Planspielsammlung. Für die alten Planspiele ist keine einheitliche Beurteilung der Kontentvalidität möglich, da sie die unterschiedlichsten Aufgabenstellungen abbilden (vgl. Kap. 2.3 und Anhang). Die neuen Planspiele wurden so entwickelt, dass sie bestimmte Situationsmerkmale enthalten, denen eine hohe Inhaltsvalidität zugesprochen wird. Dies ist zum einen die Abbildung einer Krisensituation, zum anderen die Darstellung eines Ressourcenengpasses, der Aushandlungsprozesse innerhalb der Gruppe erfordert. In dieser Studie wurde aus den beiden Planspielgruppen nur jeweils ein Planspiel erprobt. Für das gesamte Projekt ist jedoch von einer deutlich höheren Inhaltsvalidität auszugehen, sowie von einer Standardisierung des Gesamtprozesses in dieser Hinsicht, da die Unterschiede zwischen den Planspielen als wesentlich geringer angesehen werden als dies bei den alten Planspielen der Fall ist.

7.8 Ökonomie und Nützlichkeit

Ökonomie und Nützlichkeit sind für die Praxis zwei wesentliche Qualitätskriterien zur Bewertung von Auswahlverfahren.

Zum Kriterium der Nützlichkeit lässt sich bezüglich der Veränderungen der Gruppenverfahren vor allem die erhöhte Differenzierung der Bewerber festhalten. Auf diese Weise stehen für die Feststellung der Eignungsreihenfolge innerhalb eines Jahrgangs differenziertere Informationen zur Verfügung.

Geht man davon aus, dass im Rahmen des gesamten Projektes inklusive der weiteren Entwicklung der VCL und der Beobachtertrainings sich auch die Konstruktvalidität verbessert hat, sind bessere Voraussetzungen dafür geschaffen, die Informationen aus den Gruppenaufgaben als erste Informationsbasis für einen Personalentwicklungsprozess zu nutzen. In diesem Falle sind die Ergebnisse aus dem Auswahlverfahren über die reine Personalauswahl hinaus verwertbar und tragen zu einer größeren Wirtschaftlichkeit des Verfahrens bei.

Aus ökonomischer Perspektive ist besonders positiv festzuhalten, dass die OPZ durch die Standardisierung der Planspiele hinsichtlich der Situationsmerkmale Krisenmanagement und Ressourcenengpass in die Lage versetzt wird, in Eigenregie weitere Planspiele zu entwickeln, ohne Qualitätsverluste durch Verringerung der Standardisierung des Gesamtprozesses befürchten zu müssen. Aufgrund der Massendiagnostik ist die laufende Entwicklung weiterer Aufgabenstellungen eine wichtige Herausforderung, um die Berufszufriedenheit der

Prüfer nicht zu gefährden und die OB nicht mit einem Verfahren zu konfrontieren, dass nach einem gewissen Zeitraum vollständig im Internet beschrieben ist.

Der Zeitaufwand für die Datenerhebung im GSV ist gleich geblieben und wirkt sich daher nicht auf die ökonomische Bewertung aus.

8 Ausblick

In der Arbeit wurde die Frage nach den Auswirkungen instrumenteller Veränderungen in Gruppenaufgaben zur Personalauswahl ohne zusätzliche Beobachtertrainings bearbeitet. Die Interventionen hatten, isoliert eingesetzt, die erwarteten Wirkungen. Sowohl der Einsatz einer neuen Verhaltenscheckliste während der Beobachtung und Bewertung, sowie der Einsatz von zwei neuen, gezielt für die Beobachtung bestimmter Dimensionen entwickelter Planspielaufgaben, führten zu einer Qualitätsverbesserung. Die Kombination beider Neuerungen führte zu sehr gemischten Ergebnissen.

Tabelle 7.1 verdeutlicht, dass die Auswirkungen der instrumentellen Veränderungen variieren je nach beobachteter Dimension, Beobachtergruppe (OPZ-Prüfer bzw. externe Beobachter) und Aufgabenart. Hier eröffnet sich ein weites Feld für weitere Forschungsfragen.

Die genauen Wirkweisen der instrumentellen Veränderungen auf die individuellen Beobachtungs- und Bewertungsprozesse sind ungeklärt. Dies wird deutlich anhand der unterschiedlichen Ergebnisse von OPZ-Prüfern und externen Beobachtern. Die unterschiedlichen individuellen Voraussetzungen, die jeder Prüfer und Beobachter in die Wahrnehmungs- und Bewertungsprozesse einbringt, sind nicht mit Hilfe von instrumentellen Beobachtungs- und Bewertungshilfen standardisierbar. Darauf begründet sich die Notwendigkeit von Beobachtertrainings, um in der Praxis eine einheitliche Qualität bzw. Qualitätsverbesserung bei der Bewertung von Gruppenaufgaben sicher zu stellen. Eine Forschungsstrategie, um dennoch die Wirkungen verschiedener Auswahlverfahren und Instrumente genauer untersuchen zu können, wäre daher die Durchführung standardisierter Beobachtertrainings in Kombination mit verschiedenen Beobachtungs- und Bewertungsinstrumenten.

Die Ergebnisse dieser Arbeit sprechen für einen systematischen Einfluss der Aufgabenstellung auf die Gruppenprozesse und das individuelle Verhalten. Der Gruppeninteraktionsprozess wird im Rahmen der Assessment-Center-Forschung in der Regel als moderierende Variable aufgefasst, die lediglich durch den Einsatz von geschulten Rollenspielern in Gruppenaufgaben zu kontrollieren sei. Die Integration und Intensivierung von Kleingruppenforschung, die den Interaktionsprozess nicht als „Black-Box“ zwischen Input und Output betrachtet, sondern die Zusammenhänge zwischen Aufgaben, bzw. Situationsmerkmalen und dem Verlauf des Interaktionsprozesses untersucht, würde eine sinnvolle Bereicherung für die AC-Forschung darstellen. Ergebnisse aus dieser Forschungsrichtung könnten vor allem für die Entwicklung von Aufgabenstellungen für die Personalauswahl wichtige Hinweise und eine bessere theoretische Fundierung liefern.

Der direkte Vergleich zwischen altem Planspiel und neuem Planspiel zum Krisenmanagement spricht dafür, dass die entsprechende Auswahl bestimmter Situationsmerkmale für eine Gruppenaufgabe die Bewertungsqualität für einzelne Zieldimensionen verbessern kann. Die OPZ bietet aufgrund der Massendiagnostik die seltene Gelegenheit, solche Fragestellungen in einem realen Auswahlverfahren empirisch zu prüfen. Eine Erhebung und Auswertung der

Beobachtungsurteile aus allen 20 neuen Planspielen war im Rahmen dieser Arbeit leider nicht möglich, bietet sich jedoch als eine anschließende Untersuchung an. Die Ergebnisse wären für die weitere Entwicklung von Gruppenaufgaben mit dem Anspruch eines gewissen Standardisierungsniveaus sehr hilfreich.

Das Instrument der Intraklassenkorrelation hat sich als gut geeignete Methode zur Messung der Reliabilität erwiesen. Sie bietet gegenüber der üblicherweise genutzten Produkt-Moment-Korrelation die Vorteile, dass mehr als 2 Rater berücksichtigt werden können und je nach Weiterverarbeitung der Daten die passende Intraklassenkorrelation ausgewählt werden kann. Zudem kann sie bei Beachtung der Mittelwertsunterschiede zwischen den Ratern mehr Informationen über die Rater-Übereinstimmung liefern. Es wäre wünschenswert, dass sie in entsprechenden Studien häufiger zum Einsatz käme, um die Vergleichbarkeit der verschiedenen Ergebnisse in der Forschung zu erleichtern.

Literaturverzeichnis

Alwin, D.F. (1974). Approaches to the interpretation of relationship in the Multitrait-Multimethod Matrix. In H.L. Costner (Ed.), *Sociological Methodology 1973-1974*. San Francisco: Jossey-Bass, 79-105.

Archambeau, D.J. (1979). Relationships among skill ratings assigned in an assessment center. *Journal of Assessment Center Technology*, 2, S. 7-20.

Argyle, M. (1972). Soziale Interaktion. Köln: Kippenheuer & Witsch.

Bass, B.M. (1954). The leaderless group discussion. *Psychological bulletin*, 51, S. 465 – 492.

Becker-Beck, U. (1997). Soziale Interaktion in Gruppen - Struktur- und Prozessanalyse. Opladen: Westdeutscher Verlag.

Bierhoff, H.W. (1986). Personenwahrnehmung : Vom ersten Eindruck zur sozialen Interaktion. Berlin: Springer.

Blaschke, D. (1987). Soziale Qualifikation im Erwerbsleben – Theoretisches Konzept und empirische Ergebnisse. Nürnberg: Beiträge zur Arbeitsmarkt und Berufsforschung.

Bortz, J. (1999). Statistik für Sozialwissenschaftler, 5. Auflage. Berlin: Springer.

Bray, D.; Campbell, R. (1968). Selection of salesmen by means of an assessment center. In *Journal of Applied Psychology*, 52, S. 36-41.

Bray, D.; Grant, D. (1966). The assessment center in the measurement of potential for business management. In *Psychological monographs*, Whole Nr. 625, Vol. 80, Nr. 17.

Brehmer, B. (1987). Social judgement theory and forecasting. In G. Wright, P. Ayton, *Judgemental forecasting*. Chichester: Wiley.

Bycio, P; Alvares, K.M. & Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis. *Journal of Applied Psychology*, 72, S. 463-474.

Byham, W.C. (1970). Assessment Centers for spotting future managers. *Harvard Business review*, July/August 1970, S. 150-160.

Campbell, D.T.; Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, S. 81-105.

Cohen, B.M.; Moses, J.L.; Byham, W.C. (1974). The validity of assessment centers: A literature review. Monograph II. Pittsburgh: Development Dimensions Press.

Crawley, B.; Pinder, R. & Herriot, P. (1990). Assessment centre dimensions, personality and aptitudes. In Journal of Occupational Psychology, 63, S. 211 – 216.

Cronbach, L.J.; Meehl, P.E. (1955). Construct validity in psychological tests. In Psychological Bulletin, 52, S. 281-302.

Delhees, K.H. (1995). Führungstheorien – Eigenschaftstheorie. In Kieser, A. (Hrsg.) Handwörterbuch der Führung, S. 897 – 906. Stuttgart: Schäfer-Poeschel-Verlag.

Dodd, W. (1977). Attitudes toward assessment center programs. In J. Moses & W. Byham (Hrsg.). Applying the assessment center method. New York. S. 161 – 183.

Dörner, D. (1976). Problemlösen als Informationsverarbeitung. Stuttgart: Kohlhammer.

Dörner, D. (1989). Die Logik des Misslingens. Strategisches Denken in komplexen Situationen. Hamburg: Rowohlt.

Domsch, M; Jochum, I. (1989). Zur Geschichte des Assessment-Centers – Ursprünge und Werdegänge. In Lattmann, Ch. (Hrsg.) Das Assessment-Center-Verfahren der Eignungsbeurteilung. Heidelberg: Physica Verlag

Donat, M.; Moser, K. (1989). Die Arbeits- und Anforderungsanalyse als Grundlage der Gestaltung von Assessment Centern. In Lattmann, Charles (Hrsg.) Das Assessment-Center-Verfahren der Eignungsbeurteilung. Heidelberg: Physica-Verlag.

Ebenrett, H.-J.; Kozielski, K.H.; Welcker, I. (2001). Lagebild "Jugend heute". Strausberg: Sozialwissenschaftliches Institut der Bundeswehr.

Fassheber, P. (1990). Planspiele. In W. Sarges (Hrsg.), Management-Diagnostik (S. 490 – 497). Göttingen: Hogrefe.

Fay, E.; Heilmann, K. (1995). Die Konstruktionsübung „Waage“ als Instrument zur Führungskräfte-Diagnostik. In J. Funke & A. Fritz (Hrsg.), Neue Konzepte und Instrumente zur Planungsdiagnostik. Bonn: Dt. Psychologen-Verlag.

Fennekels, G. (1987). Validität des Assessment-Centers bei Führungskräfteauswahl und -entwicklung. Dissertation an der Rheinische Friedrich-Wilhelms-Universität Bonn.

Fiedler, F.K. (1964). A contingency model of leadership effectiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1). New York: Academic Press.

Fisseni, H-J. (1990). Lehrbuch der psychologischen Diagnostik. Göttingen: Hogrefe.

Fisseni, H-J.; Fennekels, G. (1995). Das Assessment Center - eine Einführung für Praktiker. Göttingen: Verlag für Angewandte Psychologie.

Fritz, A.; Funke, J. (1995). Übersicht über vorliegende Verfahren zur Planungsdiagnostik. In J. Funke & A. Fritz (Hrsg.), *Neue Konzepte und Instrumente zur Planungsdiagnostik.* Bonn: Dt. Psychologen-Verlag.

Fröhling, H.-G.; Rausch, W. W. (1996) Zentrum innere Führung (Hrsg.). Entscheiden und Verantworten - Konfliktsituationen in UN-Einsätzen.

Funke, J. (1995). Erforschung komplexen Problemlösens durch computerunterstützte Planspiele: Kritische Anmerkungen zur Forschungsmethodologie. In Geilhardt, T.; Mühlbradt, T. (Hrsg.) *Planspiele im Personal- und Organisationsmanagement.* S. 205-216. Göttingen: Verlag für angewandte Psychologie.

Funke, J.; Fritz, A. (1995). Über Planen, Problemlösen und Handeln. In J. Funke & A. Fritz (Hrsg.), *Neue Konzepte und Instrumente zur Planungsdiagnostik.* Bonn: Dt. Psychologen-Verlag.

Fydrich, T. & Bürgener, F. (1999). Ratingskala für soziale Kompetenz. In Margraf, J. & Rudolf, K. (Hrsg.). *Soziale Kompetenz – Soziale Phobie. Anwendungsfelder, Entwicklungslinien, Erfolgsaussichten.* S. 81 – 96. Baltmannsweiler: Schneider Verl. Hohengehren.

Gatewood, R.; Thornton, G.C. III; Hennessey Jr., H.W. (1990). Reliability of exercise ratings in the leaderless group discussion. In *Journal of Occupational Psychology*, 63, S. 331 – 342.

Gaugler, B.B.; Thornton, G.C. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. In *Journal of applied psychology*, 74, S. 611 – 618.

Geilhardt, T. (1995). Planspiele- Definition und Taxonomie. In Geilhardt, T.; Mühlbradt, T. (Hrsg.) *Planspiele im Personal- und Organisationsmanagement.* S. 45- 56. Göttingen: Verlag für angewandte Psychologie.

Geilhardt, T.; Mühlbradt, T. (Hrsg.). (1995). Planspiele im Personal- und Organisationsmanagement. Göttingen: Verlag für angewandte Psychologie.

Glass, G.V.; Peckham, P.D.; Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. In Review of educational research, 42, S. 237 – 288.

Guion, R.M. (1987). Changing views for personnel selection research. In Personnel Psychology, 40, S. 199-213.

Hinrichs, J. (1969). Comparison of „real life“ assessments of management potential with situational exercises, paper-pencil ability tests, and personality inventories. In Journal of applied Psychology, 53, S. 425-432.

Hinrichs, J. (1978). An eight-year follow up of a management assessment center. In Journal of Applied Psychology, 3, S. 596-601.

Hinsch, R.; Pfingsten, U. (1982). Gruppentraining sozialer Kompetenzen. München: Urban & Schwarzenberg.

Hollander, E.P. (1995). Führungstheorien – Idiosyncrasiekreditmodell. In Kieser, A. (Hrsg.) Handwörterbuch der Führung, S. 926 – 940. Stuttgart: Schäfer-Poeschel-Verlag.

Hornke, L.F.; Schiff, H.B.; Hausen, C. (1993). Training und psychologische Diagnose des Sozial- und Führungsverhaltens anhand videogestützt präsentierter Führungssituationen. In Untersuchungen des Psychologischen Dienstes der Bundeswehr, 28-30, S. 297 – 381.

Hossiep, R.; Paschen, M. (1998). Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP™) – Testmanual. Göttingen, Bern, Toronto, Seattle: Hogrefe Verlag.

Jeserich, W. (1981). Mitarbeiter auswählen und fördern. Assessment-Center-Verfahren. München: Hanser.

Jones, A. (1981). Inter-rater reliability in the assessment of group exercises at a UK assessment centre. In Journal of Occupational Psychology, 54, S. 79 – 86.

Karkoschka, U. (1998). Validität eignungsdiagnostischer Verfahren zur Messung sozialer Kompetenz: empirische Untersuchungen zu den Auswirkungen von Methodenvariationen auf die soziale und kriterienbezogene Validität. Frankfurt am Main; Berlin; Bern; New York; Paris; Wien: Lang.

Kastner, M. (1991). Führung als Denken und Handeln im System. Voraussetzung für unternehmerisches Handeln. In Kastner, M.; Gerstenberg, B. (Hrsg.). Personalmanagement - Denken und Handeln im System. München: Quintessenz.

Kastner, M. (1999). SynEgoismus. Nachhaltiger Erfolg durch soziale Kompetenz. Freiburg: Herder.

Kastner, M. (1995). Systemisches Denken. In Geilhardt, T.; Mühlbradt, T. (Hrsg.) Planspiele im Personal- und Organisationsmanagement. S. 27- 43. Göttingen: Verlag für angewandte Psychologie.

Kieser, A. (Hrsg.) (1995). Handwörterbuch der Führung. Stuttgart: Schäfer-Poeschel-Verlag.

Klauwer, K.J. (1987). Kriteriumsorientierte Tests. Lehrbuch der Theorie und Praxis lehrzielorientierten Messens. Göttingen: Hogrefe.

Kleinmann, M. (1991) Reaktivität von Assessment Centern. In H. Schuler & F. Funke (Hrsg.), Eignungsdiagnostik in Forschung und Praxis, S. 159 – 162. Stuttgart: Verlag für angewandte Psychologie.

Kleinmann, M. (1997). Assessment-Center Stand der Forschung – Konsequenzen für die Praxis. Göttingen: Verlag für angewandte Psychologie.

Kleinmann, M.; Andres, J.; Fedtke, C.; Godbersen, F. (1992). ACs: Auswertungsstrategien und Konstruktvalidität. In L. Montada (Hrsg.), Bericht über den 38. Kongress der Deutschen Gesellschaft für Psychologie in Trier 1992 (Band 1, S. 224-225). Göttingen: Hogrefe.

Kleinmann, M.; Exler, C.; Kuptsch, C.; Köller, O. (1995). Unabhängigkeit und Beobachtbarkeit von Anforderungsdimensionen im Assessment Center als Moderatoren der Konstruktvalidität. In Zeitschrift für Arbeits- und Organisationspsychologie, 39. Jg., 1, S. 22 – 28.

Klubeck, S.; Bass, B.M. (1954). Differential effects of training on persons of different leadership status. In: Human Relations, 7, S. 59 – 72.

Kompa, A. (1989). Assessment-Center und Eliteherrschaft in Organisationen: Analyse und Kritik einer Methode der Führungskräfteauswahl. In Augsburger Beiträge zu Organisationspsychologie und Personalwesen, 5, S. 1-76.

Kratzer, H. (1991). Die Anforderungsanalyse als Voraussetzung für die Ableitung effektiver Eignungsüberprüfungen. In Schuler, H.; Funke, U. (Hrsg.) Eignungsdiagnostik in Forschung Praxis. Stuttgart: Verlag für angewandte Psychologie, S. 260 - 264.

Kreuzig, H. W. (1991). Auf der Suche nach Führungskräften mit Systemdenken. In Kastner, M.; Gerstenberg, B. (Hrsg.). Personalmanagement - Denken und Handeln im System. München: Quintessenz.

Krumbach, P. (1999). Aktuelle Methoden und Einsatzgebiete der Anforderungsanalyse. In Jochmann, Walter. ((Hrsg.), Innovationen im Assessment- Center. Kapitel V. S. 87- 108. Stuttgart: Schäfer- Poeschel Verlag.

Krystofiak, F.; Cardy, R.; Newman, J.; (1988). Implicit Personality and performance appraisal: The influence of trait inferences on evaluations of behaviour. In: Journal of Applied Psychology, 73, S. 515 – 521.

Kuhl, J. (1983). Motivation, Konflikt und Handlungskontrolle. Berlin, Heidelberg, New York, Tokyo: Springer Verlag

Latham, G.; Wexley, K.; Pursell, E. (1975). Training Managers to minimize rating errors in the observation of behaviour. In Journal of Applied Psychology, 60, , S. 550 – 555.

Lattmann, C. (1989). Das Assessment Center-Verfahren als Mittel zur Beurteilung der Führungseignung. In Lattmann, Charles (1989). Das Assessment-Center-Verfahren der Eignungsbeurteilung. Heidelberg: Physica-Verlag.

Laubsch, K. (2001). Assessment-Center und Situatives Interview in der Personalauswahl von Operateuren komplexer technischer Systeme - ein Vergleich unter den Aspekten Konstruktvalidität und Prädiktive Validität. Institut für Luft- und Raumfahrtmedizin Abteilung Luft- und Raumfahrtpsychologie Hamburg

Lienert, G.A.; Raatz, R. (1998). Testaufbau und Testanalyse. Weinheim: Beltz Psychologie Verlagsunion.

Lotz, G. (1984). Stress, Bewältigung und Soziale Kompetenz bei Schülern. Frankfurt a.M.: Lang.

Maukisch, H. (1989). Informationswert und Ökonomie der diagnostischen Prinzipien von Assessment Center Systemen zur Erfassung von Management Potential. In Lattmann, Charles (1989). Das Assessment-Center-Verfahren der Eignungsbeurteilung. Heidelberg: Physica-Verlag.

Meichenbaum, D.; Butler, L.; Gruson, L. (1981). Toward a conceptual model of social competence. In J.D. Wine & M.D. Smye (Hrsg.) Social competence. S. 36-60. New York: Guilford.

Melter, A. (1995). Computerunterstützte Planspiele in den Streitkräften. In Geilhardt, T.; Mühlbradt, T. (Hrsg.) Planspiele im Personal- und Organisationsmanagement. S. 289-304. Göttingen: Verlag für angewandte Psychologie.

Meyer, W.U.; Försterling, F. (1993). Die Attributionstheorie. In D. Frey & M. Irle (Hrsg.), Kognitive Theorien der Sozialpsychologie. S. 175 – 216. Bern: Huber.

Mitchell, T.R. (1995). Führungstheorie – Attributionstheorie. In Kieser, A. (Hrsg.) Handwörterbuch der Führung, S. 848 – 861. Stuttgart: Schäfer-Poeschel-Verlag.

Moede, W. (1930). Lehrbuch der Psychotechnik. 1. Band. Berlin: Springer.

Moses, J. (1971). Assessment center performance and management progress. Washington D.C.

Neidig, R.D.; Martin, J.C. & Yates, R.E. (1979). The contribution of exercise skill ratings to final assessment center evaluations. In Journal of Assessment Center Technology, 2, S. 21-23.

Neidig, R.D.; Neidig, J.P. (1984). Multiple assessment center exercises and job relatedness. In Journal of Applied Psychology, 69 (1), S. 182 – 186.

Neubauer, R. (1989). Implizite Eignungstheorien im Assessment-Center. In C. Lattmann (Hrsg.), Das Assessment-Center-Verfahren der Eignungsbeurteilung. Sein Aufbau, seine Anwendung und sein Aussagegehalt. S. 191 – 222. Heidelberg: Physica.

Neubauer, R. (1995). Führungskräfteauswahl in der Praxis. In Geilhardt, T.; Mühlbradt, T. (Hrsg.) Planspiele im Personal- und Organisationsmanagement. S. 155 – 171 Göttingen: Verlag für angewandte Psychologie.

Norton, S.D. (1977). The empirical and content validity of assessment centers vs. traditional methods of predicting managerial success. Academy of Management Review, 2, S. 442 – 453.

Ostroff, C.; Schmitt, N. (1987). The relationship between content and criterion-related validity indices: An empirical investigation. Unpublished manuscript. Michigan State University.

Putz-Osterloh, W. (1991). Handeln und entscheiden in komplexen Systemen. In Kastner, M.; Gerstenberg, B. (Hrsg.). Personalmanagement - Denken und Handeln im System. München: Quintessenz.

Reilly, R.R.; Henry, H. (1990). An examination of the effects of using behaviour checklists on the construct validity of assessment center dimensions. In Personnel Psychology, Spring 1990, vol. 43, Issue 1, S. 71 – 85.

Robertson, R.; Gratton, L. & Sharpley, D. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. In Journal of Occupational Psychology., 60 (3), S. 187 – 195.

Rohn, W.E. (1995). Ursprung und Entwicklung des Planspiels. In Geilhardt, T.; Mühlbradt, T. (Hrsg.) Planspiele im Personal- und Organisationsmanagement. S. 57- 67. Göttingen: Verlag für angewandte Psychologie.

Russel, C.G. (1987). Person characteristic versus role congruency explanations for assessment center ratings. In Academy of Management Journal, 30, S. 817 – 826.

Sacket, P.R.; Dreher, G.H. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. In Journal of Applied Psychology, 67, S. 401 – 410.

Sarges, W. (1996). Lernpotential-Assessment Center. In: Sarges, W. (Hrsg.): Weiterentwicklungen der Assessment Center-Methode. Göttingen: Verlag für angewandte Psychologie.

Schelten, A. (1980). Grundlagen der Testbeurteilung und Testerstellung. Heidelberg: Quelle und Meyer.

Schmidt, F.L.; Hunter, J.E. (1978). Moderator research and the law of small numbers. In Personnel Psychology, 31, S. 215 – 232.

Schmidt, J. U.; (1995). Psychologische Messverfahren für soziale Kompetenzen. In Seyfried, B. (Hrsg.): Stolperstein Sozialkompetenz. Was macht es so schwierig sie zu erfassen, zu fördern und zu beurteilen. Berichte zur beruflichen Bildung Bd. 179 (S. 117 - 135). Bielefeld: Bertelsmann.

Scholz, G. (1994). Das Assessment Center. Konstruktvalidität und Dynamisierung. Göttingen: Verlag für angewandte Psychologie.

Scholz, G.; Schuler, H. (1993). Das nomologische Netzwerk des Assessment-Centers: Eine Metaanalyse. In Zeitschrift für Arbeits- und Organisationspsychologie, 27, S. 33-44.

Schreyögg, G. (1995). Führungstheorien – Situationstheorie. In Kieser, A. (Hrsg.) Handwörterbuch der Führung. Stuttgart: Schäfer-Poeschel-Verlag.

Schuler, H. (1988). Stand und Perspektiven der Berufseignungsdiagnostik. In H. Methner & A. Gebert (Hrsg.) *Arbeits- und Betriebspsychologie: Verantwortung und Leistung*. S. 39-58. Bonn: Hogrefe.

Schuler, H. (1989). Die Validität des Assessment Centers. In Lattmann, Charles (1989). *Das Assessment-Center-Verfahren der Eignungsbeurteilung*. Heidelberg: Physica-Verlag.

Schuler, H. (1996). Psychologische Personalauswahl. Einführung in die Berufseignungsdiagnostik. Göttingen: Verlag für Angewandte Psychologie.

Schuler, H.; Barthelme, D. (1995). Soziale Kompetenz als berufliche Anforderung. In B. Seyfried (Hrsg.) „Stolperstein“ Sozialkompetenz. S. 77 – 116. Bielefeld: Bertelsmann.

Schuler, H.; Diemand, A.; Moser, K. (1993). Filmszenen. Entwicklung und Konstruktvalidierung eines neuen eignungsdiagnostischen Verfahrens. In *Zeitschrift für Arbeits- und Organisationspsychologie*. 37, S. 3-9.

Schuler, H.; Funke, U. (1989). Berufseignungsdiagnostik. In Roth, E. (Hrsg.) *Enzyklopädie der Psychologie*. Themenbereich D, Serie III, Band 3: Organisationspsychologie S. 281 – 320. Göttingen: Hogrefe.

Schuler, H.; Stehle, W. (1983). Neuere Entwicklungen des Assessment Center-Ansatzes – beurteilt unter dem Aspekt der sozialen Validität. *Psychologie und Praxis*, 27, S. 33-44.

Schwarz, N. (1985). Theorien konzeptgesteuerter Informationsverarbeitung. In D. Frey & M. Irle (Hrsg.), *Motivations- und Informationsverarbeitungstheorien*. S. 269 - 292. Bern: Huber.

Six, U. (1987). Vorurteile. In D. Frey & S. Greif (Hrsg.), *Sozialpsychologie*, S. 365 – 371. (2. erw. Aufl.). München: Psychologie Verlags Union.

Staufenbiehl, T.; Rösler, F. (1999): Personalauswahl. In: Carl Graf Hoyos, Dieter Frey (Hrsg.): *Arbeits- und Organisationspsychologie – Ein Lehrbuch*. Weinheim: Psychologie Verlags Union.

Steinberg, S.L.; Curran, J.P.; Bell, S.; Paxson, M.; Munroe, S.M. (1982). The effects of confederate prompt delivery style in a standardized social simulation test. In: *Journal of Behavioral Assessment*, 4, S. 263 – 272.

Strack, F. (1985). Urteilsheuristiken. In D. Frey & M. Irle (Hrsg.), *Motivations- und Informationsverarbeitungstheorien*. S. 239 – 269. Bern: Huber.

Thorndike, R.L. (1920). Intelligence and its uses. In *Harper's Magazine*, 140, S. 227 – 235.

Thornton, G.C. & Byham, W.C. (1982). Assessment Centers and managerial performance. New York: Academic Press.

Thornton, G.; Gaugler, B; Rosenthal, D. & Bentson, C. (1992). Die Prädiktive Validität des AC – eine Metaanalyse. In H. Schuler & W. Stehle (Hrsg.), Assessment –Center als Methode der Personalentwicklung, S. 36-60. 2. Auflage. Stuttgart: Verlag für angewandte Psychologie.

Torgrud, L.J.; Holborn, S.W. (1992). Developing externally valid role-play for assessment of social skills: A behaviour analytic perspective. In: Behavioral Assessment, 14, S. 245 – 277.

Wegner, S. (1989). Effizientes Konfliktlösen – Ein neuer Ansatz zur Erfassung sozial-intelligenter Lösungsstrategien im Betrieb. Frankfurt: Lang.

Weinert, A.B. (1998). Lehrbuch der Organisationspsychologie. München-Weinheim: Psychologie Verlags Union.

Wirtz, M.; Caspar, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität. Göttingen, Bern, Toronto, Seattle: Hogrefe Verlag für Psychologie.

Wolf, B.; Barell, G.; Hoenle, S. (1995). Assessment Center auf dem Prüfstand - Bewäh- rungskontrolle und Qualitätssicherung am Beispiel eines Unternehmens-ACs. Hamburg: Windmühle GmbH.

Wollowick, H.; McNamara, W. (1969). Relationship of the components of an assessment center to management success. In Journal of Applied Psychology, 53, S. 348-352.

Wunderer, R. (2000). Führung und Zusammenarbeit. Neuwied: Luchterhand.

Zalesny, M.D.; Graen, G.B. (1995). Führungstheorien – Austauschtheorie. In Kieser, A. (Hrsg.) Handwörterbuch der Führung. Stuttgart: Schäfer-Poeschel-Verlag.

Zuckerman, M. (1979). Sensation seeking: Beyond the optimal level of arousal. London: Wiley.

Anhang

Tabelle A A: Beschreibung der Gesamtheit, aus der die Stichprobe stammt

	n	%
Bewerber in zehn Prüfgruppen	791	100,00
Männlich	642	81,16
Weiblich	149	18,84
Alter:		
17	9	1,14
18	95	12,01
19	128	16,18
20	173	21,87
21	134	16,94
22	106	13,40
23-25	114	14,41
26-30	28	3,54
31-35	1	0,13
36-40	2	0,25
> 40	1	0,13
Bildungsabschlüsse:		
Hauptschulabschluss	1	0,13
Mittlere Reife	96	12,14
Fachhochschulreife	130	16,44
Hochschulreife	529	66,88
Abgeschl. Berufsausbildung	31	3,90
Abgeschl. Studium	3	0,38
Ohne Angaben	1	0,13
Diensterfahrung:		
Ungedient ohne Vordienstzeiten	300	37,93
Davon weiblich	131	43,66
Davon männlich	169	56,34
Dienender Bewerber	297	37,55
Davon Weiblich	10	3,37
Davon männlich	287	96,63
Gedienter Bewerber m. Vordienstzeit	138	17,45
Davon männlich	138	100,00
Sonstiger Bewerber	29	3,67
Davon weiblich	7	24,14
Davon männlich	22	75,86
Gezogener Bewerber	16	2,02
Davon männlich	16	100,00
Bewerber aus der Uffz-Laufbahn	7	0,88
Davon männlich	7	100,00
Bewerber mit abgeschlossenem FH-Studium	3	0,38
Davon männlich	3	100,00
Reserveoffizierbewerber	1	0,13
Davon weiblich	1	100,00
Bewerber der Stichprobe in 4 Prüfgruppen	183	23,14