



Westfälische
Wilhelms-Universität
Münster

Fachbereich Psychologie und Sportwissenschaften

Medieneinsatz in der Eignungsdiagnostik

Empirische Untersuchung zur Validität eines medial variierten
Personalauswahlverfahrens

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der
Philosophischen Fakultät
der
Westfälischen Wilhelms-Universität
zu
Münster (Westf.)

vorgelegt von

Timm Hüttemann

aus Münster in Westfalen
im Februar 2008

Gutachter:
Prof. Dr. Hanko Bommert
PD Dr. Uwe Peter Kanning

Tag der mündlichen Prüfung: 31. Juli 2008

Dekan: Prof. Dr. Dr. h. c. Wichard Woyke

Referent: Prof. Dr. Hanko Bommert

Korreferent: Dr. Antje Bohne

Für Ilka

Inhaltsverzeichnis

Danksagung

1 Einleitung	1
2 Theoretische Grundlagen	4
2.1 Historische Entwicklung eignungsdiagnostischer Verfahren	
– ein Überblick –	5
2.2 Gütekriterien eignungsdiagnostischer Verfahren	9
2.2.1 Die Objektivität	9
2.2.2 Die Reliabilität.....	10
2.2.3 Die Validität	10
2.3 Konzepte zur Wahrnehmung diagnostischer Verfahren.....	11
2.3.1 Das Konzept der sozialen Validität	11
2.3.2 Das “Model of Applicants’ Reactions to Employment Selection Systems”	14
2.3.3 Annahmen und empirische Befunde zur Bedeutung der sozialen Validität und wahrgenommenen Fairness in der Eignungsdiagnostik	16
2.4 Vorstellung der relevanten Verfahren und ihrer Standards.....	18
2.4.1 Das Assessment Center	19
2.4.2 Das Rollenspiel.....	21
2.4.3 Situational Judgement Tests	22
2.5 Die Einteilung eignungsdiagnostischer Verfahren nach dem Konzept der „Fidelity“	24
2.6 Die Rolle der Medien bei der Entwicklung von eignungsdiagnostischen Verfahren.....	26
2.7 Erwartungen, Forschungsfragen und Ergebnisse einzelner Entwicklungsstufen der computerbasierten Diagnostik	28
2.7.1 Die computerbasierte Auswertung von Paper-Pencil-Verfahren	29
2.7.2 Die Übertragung von Paper-Pencil Verfahren auf den Computer.....	30
2.7.3 Das computerbasierte Adaptive Testen	34
2.7.4 Computerbasierte Problemlöseszenarien.....	37

2.8 Multimediale computerbasierte Eignungsdiagnostik	43
2.8.1 Entstehung und Definition des Begriffs „Multimedia“	43
2.8.2 Implikationen des Multimedia-Gedankens für die Eignungsdiagnostik	46
2.8.3 Antizipierte Vorteile von Multimedia gegenüber herkömmlichen Verfahren	47
2.8.4 Antizipierte Nachteile von Multimedia gegenüber herkömmlichen Verfahren	49
2.8.5 Empirische Befunde zu den Gütekriterien von videobasierten, eignungsdiagnostischen Verfahren.....	51
2.8.6 Zusammenfassende Bewertung von videobasierten Verfahren.....	64
2.9 Videobasierte und herkömmliche Verfahren im Vergleich.....	66
2.9.1 Der Vergleich von medial variierten Stimuluskomponenten	66
2.9.2 Der Vergleich von medial variierten Stimulus- und Responsekomponenten	69
2.9.3 Zusammenfassende Bewertung der Ergebnisse	72
2.10 Implikationen der Ergebnisse zu videobasierten Situational Judgement Tests für weitere Forschungsansätze	73
3 Herleitung der Forschungsfragen und Hypothesen.....	77
4 Methode.....	80
4.1 Einbettung der Methode in ihren Anwendungskontext.....	80
4.1.1 Darstellung des Verfahrens	80
4.1.2 Darstellung des Kriteriums.....	83
4.2 Vorgehensweise der Materialerstellung	85
4.2.1 Auswahl der Face-to-Face Variante	85
4.2.2 Übertragung des Rollenspiels in eine videobasierte Variante	86
4.2.3 Beschreibung des VAK	99
4.2.4 Erstellung der Paper-Pencil Variante	102
4.2.5 Erstellung der Fragebögen zur Erhebung der sozialen Validität.....	103

4.3 Empirische Überprüfung der intendierten Materialspezifika	105
4.3.1 Vorstudie 1: Empirische Prüfung der multimedialen Eigenschaften des VAK	106
4.3.2 Vorstudie 2: Empirische Betrachtung der Konstruktvalidität des VAK	120
4.3.3 Vorstudie 3: Empirische Prüfung der Durchführbarkeit der Verfahren in der Zielpopulation	130
4.4 Hauptuntersuchung.....	137
4.4.1 Stichprobe.....	137
4.4.2 Material.....	137
4.4.3 Versuchsablauf	138
4.4.4 Design.....	140
5 Ergebnisse	143
5.1 Darstellung der deskriptiven Statistiken.....	143
5.1.1 Deskriptive Statistiken für die Face-to-Face Variante	143
5.1.2 Deskriptive Statistiken für das VAK	144
5.1.3 Deskriptive Statistiken für die Paper-Pencil Variante.....	148
5.1.4 Deskriptive Statistiken für das Kriteriumsmaß	148
5.2 Ergebnisse zur kriterienorientierten Validität	149
5.3 Ergebnisse zur sozialen Validität	151
5.3.1 Ergebnisse zur Faktorstruktur der fünf Itemgruppen	151
5.3.2 Ergebnisse zur Faktorstruktur der Fragebögen.....	152
5.3.3 Ergebnisse zur Faktorenstruktur der zusammengefassten Fragebögen.....	159
5.3.4 Ergebnisse der varianzanalytischen Untersuchung	162
6 Diskussion der Ergebnisse.....	166
6.1 Diskussion der deskriptiven Statistiken	166
6.2 Diskussion der Ergebnisse zur kriterienorientierten Validität	167
6.3 Diskussion der Ergebnisse zur sozialen Validität.....	174
6.4 Diskussion der Schwierigkeiten der fünf Phasen des VAK.....	178
6.5 Diskussion der Reliabilität des VAK	178

7 Diskussion der methodischen Vorgehensweise	180
7.1 Kritische Reflexion des gewählten Anwendungskontextes.....	180
7.2 Kritische Reflexion bezüglich der Wahl des Kriteriums	181
7.3 Kritische Reflexion in Bezug auf die Gestaltung der	
verwendeten Materialien	182
7.4 Kritische Reflexion der untersuchten Stichprobe	185
7.5 Kritische Reflexion des Versuchsablaufs	186
7.6 Kritische Reflexion des methodischen Designs.....	186
8 Generelle Diskussion und Ausblick	188
8.1 Integration der Arbeit in den historischen Entwicklungsprozess.....	188
8.2 Der Bewertungsmaßstab.....	189
8.2.1 Bewertung der Objektivität	190
8.2.2 Bewertung der Reliabilität.....	190
8.2.3 Bewertung der Inhaltsvalidität	191
8.2.4 Bewertung der Konstruktvalidität	191
8.2.5 Bewertung der kriterienorientierten Validität.....	192
8.2.6 Bewertung der sozialen Validität	193
8.2.7 Zusammenfassende Bewertung des VAK	193
8.3 Ausblick	195
8.3.1 Implikationen für die Forschung	195
8.3.2 Implikationen für die Praxis	197
9 Zusammenfassung.....	199
Literaturverzeichnis	201
Anhang	212

Abbildungsverzeichnis

Abbildung 2.1:	Das Konzept der sozialen Validität (nach Schuler, 1993, S. 13)	12
Abbildung 2.2:	Das “Model of Applicants Reactions to Employment Selection Systems” (nach Gilliland, 1993)	15
Abbildung 2.3	Die Entwicklungsstufen der computerbasierten Diagnostik	28
Abbildung 4.1:	Graphische Darstellung des Ablaufes eines typischen PES.....	82
Abbildung 4.2:	Die fünf zentralen Phasen eines Kundengesprächs	90
Abbildung 4.3:	Der inhaltlich und qualitativ gestufte Aufbau der Drehbücher	92
Abbildung 4.4:	Drehbuch für die Phase 1 in der niedrigsten Qualitätsausprägung	93
Abbildung 4.5:	Drehbuch für die Phase 1 in der höchsten Qualitätsausprägung	93
Abbildung 4.6:	Setaufbau nach dem continuity-editing-system (nach Beller, 2005)	95
Abbildung 4.7:	Screenshot der Benutzeroberfläche des Programms Avid Xpress Pro 4.6	96
Abbildung 4.8:	Die Ebenen des VAK im zeitlichen Ablauf	99
Abbildung 4.9:	Screenshot des Bearbeitungsbildschirms für die Sequenz 1 des VAK	101
Abbildung 4.10:	Graphische Darstellung des Versuchsablaufes	111
Abbildung 4.11:	Zellenhäufigkeiten bei einer optimalen Zuordnung der Videos aus Sequenz 1	113
Abbildung 4.12:	Zellenhäufigkeiten bei einer empirischen Zuordnung der Videos aus Sequenz 1	113
Abbildung 4.13:	Mittelwerte der Items unter Berücksichtigung der Gruppenzugehörigkeit	116
Abbildung 4.14:	Mittelwerte der Items unter Berücksichtigung des Geschlechts der Teilnehmer	117
Abbildung 4.15:	Graphische Darstellung des Versuchsablaufs (Voruntersuchung 2).....	125
Abbildung 4.16:	Skalenmittelwerte der fünf Dimensionen (Studenten)	128
Abbildung 4.17:	Skalenmittelwerte der fünf Dimensionen (getrennt nach Geschlecht) .	133
Abbildung 4.18	Graphische Darstellung des Versuchsablaufes (1. Tag)	139
Abbildung 4.19	Graphische Darstellung des Versuchsablaufes (2. Tag)	140

Abbildung 5.1:	Scores (Sequenz 1)	145
Abbildung 5.2:	Scores (Sequenz 2).....	145
Abbildung 5.3:	Scores (Sequenz 3).....	145
Abbildung 5.4:	Scores (Sequenz 4).....	145
Abbildung 5.5:	Scores (Sequenz 5).....	145
Abbildung 5.6:	Scores (gesamt).....	145
Abbildung 5.7:	Ergebnisdarstellung der Parallelanalyse (Face-to-Face Rollenspiel)....	153
Abbildung 5.8:	Ergebnisdarstellung der Parallelanalyse (VAK)	155
Abbildung 5.9:	Ergebnisdarstellung der Parallelanalyse (Paper-Pencil)	157
Abbildung 5.10:	Graphische Ergebnisdarstellung der Parallelanalyse (alle Bedingungen)	160
Abbildung 5.11:	Mittelwerte und 95% Konfidenzintervalle der Faktorscores (Faktor 1).....	163
Abbildung 5.12:	Mittelwerte und 95% Konfidenzintervalle der Faktorscores (Faktor 2).....	164
Box 1:	Das Assessment-Center: Eigenschaften, Gütekriterien und Standards....	20
Box 2:	Das Rollenspiel: Eigenschaften, Gütekriterien und Standards.....	21
Box 3:	SJT: Eigenschaften, Gütekriterien und Standards.....	23

Tabellenverzeichnis

Tabelle 4.1:	Ergebnisse der Expertenratings für die fünf Abschnitte des <i>VAK</i>	98
Tabelle 4.2:	Korrelationen zwischen dem <i>VAK</i> und dem IST 2000, KLT und BIP .	126
Tabelle 4.3	Übereinstimmung der 4 Expertenurteile mittels des Bewertungsschlüssels.....	134
Tabelle 5.1:	Darstellung der deskriptiven Statistiken für das Face-to-Face Rollenspiel.....	143
Tabelle 5.2:	Darstellung der deskriptiven Statistiken für das <i>VAK</i>	144
Tabelle 5.3:	Ergebnisse der Reliabilitätsanalyse des <i>VAK</i>	147
Tabelle 5.4:	Darstellung der deskriptiven Statistiken für die Paper-Pencil Variante	148
Tabelle 5.5:	Darstellung der deskriptiven Statistiken für das Kriterium.....	148
Tabelle 5.6:	Darstellung der korrelativen Zusammenhänge zwischen der IHK-Note und den Ergebnissen aus den drei medial variierten Bedingungen.	149
Tabelle 5.7:	Darstellung der korrelativen Zusammenhänge zwischen den Ergebnissen der drei medial variierten Bedingungen	150
Tabelle 5.8:	Ergebnisse der Parallelanalyse (Face-to-Face)	152
Tabelle 5.9:	Strukturmatrix der extrahierten Faktoren (Face-to-Face)	154
Tabelle 5.10:	Ergebnisse der Parallelanalyse (<i>VAK</i>).....	155
Tabelle 5.11:	Strukturmatrix der extrahierten Faktoren (<i>VAK</i>).....	156
Tabelle 5.12:	Ergebnisse der Parallelanalyse (Paper-Pencil).....	157
Tabelle 5.13:	Strukturmatrix der extrahierten Faktoren (Paper-Pencil).....	158
Tabelle 5.14:	Faktorenstruktur über alle Bedingungen laut Parallelanalyse.....	159
Tabelle 5.15:	Strukturmatrix der extrahierten Faktoren für den gesamten Datensatz	161

Danksagung

An dieser Stelle möchte ich mich bei all denen bedanken, die mich während dieser Arbeit begleitet und zu ihrem Gelingen beigetragen haben.

Als erstes gilt mein Dank meinem Doktorvater Prof. Dr. Hanko Bommert, der es mir ermöglicht hat, als wissenschaftlicher Mitarbeiter an seinem Institut zu promovieren. Die Unterstützung des Projekts mit Institutsmitteln und die Einbindung in seine *Forschungsorientierte Vertiefung zur Medienpsychologie* haben wesentlich dazu beigetragen, dass das *Videobasierte Analyseverfahren für Kundenberater (VAK)* in seiner jetzigen Form realisiert werden konnte. Weiterhin bedanke ich mich bei PD Dr. Uwe Peter Kanning, der sich bereit erklärt hat, diese Arbeit zu begutachten. Seine Erfahrungen im Bereich der videobasierten Situational Judgement Tests führten zu wertvollen Hinweisen bei der Ausarbeitung des theoretischen Hintergrundes der vorliegenden Arbeit.

An der Gestaltung des VAK waren viele helfende Hände beteiligt. Besonders danken möchte ich an dieser Stelle Herrn Peter Kemper und seinem Team vom Servicepunkt Film der Westfälischen Wilhelms-Universität Münster. Die professionelle Beratung in allen technischen und filmischen Fragen sowie die grenzenlose Geduld bei der Umsetzung der Drehbücher in tatsächliche Filmsequenzen werden mir unvergessen bleiben. Im diesem Zusammenhang gebührt auch Ron Sparenberg mein herzlicher Dank, der sich bereit erklärte, den Part des Kundenberaters in den Videos zu übernehmen. Er hat es auf unglaubliche Weise geschafft, umgeben von einem fünfköpfigen Filmteam, drei Kameras und einer Vielzahl von Scheinwerfern und Mikrofonen die Anweisungen aus den Drehbüchern nahezu perfekt umzusetzen.

Ebenfalls unverzichtbar war die Mithilfe der vielen Bankexperten, die sich für die Beantwortung von Fragebögen, als Interviewpartner oder im Rahmen der Ratings zur Bestimmung der inhaltlichen Validität der Filmsequenzen zur Verfügung gestellt haben. Ihnen sei hiermit herzlich gedankt - auch wenn an dieser Stelle keine namentliche Erwähnung erfolgen kann. Gleiches gilt für die beiden Teams aus der *Forschungsorientierten Vertiefung zur Medienpsychologie*, welche an der Durchführung der Vorstudien beteiligt waren. Die Organisation des Versuchablaufes gestaltete sich mit ihrer Hilfe ebenso professionell und

reibungslos wie die Rekrutierung und Betreuung der Versuchspersonen. Danke für diese Unterstützung!

Das Vorhaben, die Hauptuntersuchung der vorliegenden Studie in einem realen, personaldiagnostischen Prozess einzubetten, konnte nur durch die Kooperation einer Finanzdienstleistungsakademie realisiert werden. An dieser Stelle möchte ich Herrn Gunter Nittel und Frau Iris Hotfilter ganz herzlich für ihr Engagement als Wegbereiter und für das mir entgegengebrachte Vertrauen danken.

Ein besonderer Dank gilt Maximilian Bruchmann für seine Bereitschaft, mir als kompetenter Diskussionspartner bei methodischen Überlegungen zur Seite zu stehen.

Bei meinem Kollegen Meinald Thielsch bedanke ich mich für das kollegiale Miteinander während unserer Zeit am Institut.

Studiendirektor Norbert Witzke danke ich dafür, dass er sich als Lektor für diese Arbeit zur Verfügung gestellt hat.

Mein letzter Dank gilt meiner Familie und ganz besonders meiner Partnerin Ilka Strauß für ihr Verständnis und die Unterstützung in den vergangenen Jahren. Ich bin glücklich, sie auch für kommende Herausforderungen an meiner Seite zu wissen.

1. Einleitung

Die zunehmende Globalisierung der Wirtschaftsmärkte, verbunden mit einer rasant voranschreitenden Internationalisierung der Konzerne, stellt in Zeiten des Fachkräftemangels und eines immer stärker werdenden Konkurrenzkampfes der Firmen um geeignetes Personal besonders hohe Anforderungen an die verantwortlichen Mitarbeiter¹ im Bereich des HR-Management. Um tatsächlich die besten Arbeitskräfte aus dem Markt rekrutieren und an das eigene Unternehmen binden zu können, bedarf es, in Kooperation mit der Forschung, einer permanenten Weiterentwicklung und Verbesserung der von ihnen eingesetzten Personalauswahl- und Personalentwicklungsverfahren.

Die Anforderungen, welche an moderne Verfahren in diesem Kontext gestellt werden, sind dabei vielschichtig. Als erste wesentliche Voraussetzung müssen sie unter diagnostischen Aspekten über eine hohe Zuverlässigkeit in der Messung von berufsrelevanten Eigenschaften und eine hohe Vorhersagegüte in Bezug auf den tatsächlichen späteren Berufserfolg verfügen. Weiterhin sollten die Verfahren aus Sicht der Probanden als angenehm und fair erlebt werden, da sich die Bewertung und Akzeptanz des Auswahlverfahrens als erster Kontakt mit dem Unternehmen auf die spätere Entscheidung diesem Unternehmen als Mitarbeiter angehören zu wollen, auswirken kann. Durch eine möglichst reale Abbildung der späteren beruflichen Aufgaben in der zu absolvierenden Auswahl-situation sollte dem Bewerber weiterhin die Möglichkeit zur Selbstselektion gegeben werden. Eine spätere Enttäuschung aufgrund von unrealistischen Vorstellungen bezüglich des Arbeitsplatzes kann so bereits in einem frühen Stadium des Auswahlprozesses vermieden werden. Motowidlo, Dunnette und Carter (1990) sprechen in diesem Zusammenhang vom Grad der *Fidelity* eines Verfahrens und unterscheiden die *high-fidelity simulations* als Verfahren mit einer hohen Wiedergabetreue der realen Situation am Arbeitsplatz von den *low-fidelity simulations*, welche über eine geringe Wiedergabetreue verfügen.

Ein möglicher Ansatz, um den oben genannten Forderungen an moderne Auswahlverfahren gerecht zu werden, liegt in der Nutzung von Multimedia. Die gezielte Erweiterung bisheriger Auswahlverfahren um multimediale Elemente und die daraus resultierende höhere Flexibilität in der Itemgestaltung bietet die Möglichkeit Auswahl-situationen realitätsnäher abzubilden, also ihre Wiedergabetreue der realen Situation am Arbeitsplatz zu erhöhen und

¹ Zugunsten einer besseren Lesbarkeit wird in der vorliegenden Arbeit das grammatische Maskulinum verwendet. Es sind jedoch immer beide Geschlechter gemeint.

damit auch eine Verbesserung in Bezug auf die anderen als relevant erachteten Aspekte zu realisieren.

Tatsächlich zeigen Ergebnisse aus empirischen Studien zu medial variierten Situational Judgment Tests (SJT), dass sich gegenüber der traditionellen Paper-Pencil Form dieser Verfahrensart eine Verbesserung der oben genannten Aspekte feststellen lässt, wenn ihre *Fidelity* durch computer- und videobasierte Multimedia-Varianten erhöht wird. Die multimedialen Verfahren werden nicht nur im direkten Vergleich zu ihren inhaltlich gleich gestalteten, herkömmlichen Varianten als angenehmer und fairer von Seiten der Probanden erlebt, sondern liefern auch bessere Vorhersagen in Bezug auf externe Kriterien für den Berufserfolg (siehe z. B.: Chan & Schmitt, 1997; Funke & Schuler, 1998; Kanning, Grewe, Hollenberg & Hadouch, 2006; Lievens & Sackett, 2006; Richmann-Hirsch, Olson-Buchanan & Drasgow, 2000). Trotz dieser positiven Resultate liegen für eine Vielzahl anderer Verfahren in der Personalauswahl und Personalentwicklung derzeit noch keine empirischen Ergebnisse in Bezug auf Verbesserungsmöglichkeiten durch den Einsatz multimedialer Elemente vor. Dabei dürfte gerade für komplexe Verfahren, wie z. B. das Rollenspiel, die für das Unternehmen einen hohen Personal- und Kostenaufwand bedeuten, die Untersuchung einer Weiterentwicklung durch Multimedia interessant sein.

Die empirische Untersuchung der vorliegenden Arbeit wird durchgeführt, um erste Antworten auf die Frage zu erhalten, inwieweit sich die positiven Ergebnisse zu multimedialen SJTs auch für das Rollenspiel als ein häufig eingesetztes, zeit- und kostenintensives Verfahren der Personalauswahl realisieren lassen. Durch die vergleichende Untersuchung eines herkömmlichen Rollenspiels in Face-to-Face Form mit einer videobasierten, multimedialen Version und einer Paper-Pencil Version sollen die in ihrer *Fidelity* variierten Verfahren in Bezug auf mögliche Unterschiede in ihrer kriterienorientierten und sozialen Validität betrachtet werden. Der Schwerpunkt in der Darstellung wird dabei auf die Entwicklung und empirische Betrachtung der multimedialen, videobasierten Verfahrensvariante gelegt. Als Rahmen, in den die empirische Untersuchung eingebettet wurde, diente ein Personalauswahlverfahren im Finanzdienstleistungssektor.

Zu Beginn werden in Kapitel 2 die theoretischen Grundlagen dargestellt, auf denen die Arbeit aufbaut. Neben einer Beschreibung der Entwicklungsgeschichte der psychologischen Eignungsdiagnostik und der für diese Arbeit relevanten Gütekriterien und Standards des Anwendungsfachs wird der Focus auf die Rolle der Medien bei der Weiterentwicklung

und Verbesserung eignungsdiagnostischer Instrumente gelegt. Zu diesem Zweck erfolgt eine Darstellung empirischer Ergebnisse aus Studien im Kontext vorher spezifizierter Stufen des medialen Entwicklungsprozesses in der Eignungsdiagnostik. In Kapitel 3 werden dann, aufbauend auf den in Kapitel 2 beschriebenen Theorien und Ergebnissen, die Forschungsfragen und Hypothesen der vorliegenden Arbeit entwickelt. Kapitel 4 dient der Darstellung der methodischen Umsetzung der Fragestellungen. Hier werden in einem ersten Teil die einzelnen Schritte des Konstruktionsprozesses der in der Studie verwendeten Materialien und die Voruntersuchungen der Instrumente beschrieben. Im zweiten Teil folgt dann die Darstellung der zur Klärung der Forschungsfragen durchgeführten empirischen Studie. In Kapitel 5 werden die resultierenden Ergebnisse berichtet, welche in Kapitel 6 in Bezug auf die Forschungsfragen und vor dem in Kapitel 2 dargestellten theoretischen Hintergrund diskutiert werden. In Kapitel 7 erfolgt dann die Diskussion der methodischen Vorgehensweise zur Beantwortung der Fragestellungen. Abschließend folgt in Kapitel 8 eine generelle Betrachtung der Ergebnisse aus der gesamten Arbeit verbunden mit einem Ausblick auf mögliche weitere Forschungsfragen, die sich aufgrund der Ergebnisse der vorliegenden Studie ergeben könnten. Kapitel 9 beinhaltet die Zusammenfassung der gesamten Arbeit.

2 Theoretische Grundlagen

Bevor auf die in dieser Arbeit formulierten Forschungsfragen und deren empirische Überprüfung eingegangen werden kann, soll in den folgenden Kapiteln zuerst eine Darstellung der theoretischen Grundlagen und empirischen Ergebnisse erfolgen, auf welchen die Fragestellungen aufbauen. Die Darstellung erfolgt mit dem Ziel, den Leser zum einen mit den für die vorliegende Arbeit relevanten Gütekriterien und Verfahren der psychologischen Eignungsdiagnostik² vertraut zu machen und ihm zum anderen die Rolle der Medien bei der Weiterentwicklung und Anwendung von eignungsdiagnostischen Verfahren zu verdeutlichen.

Zu Beginn der Darstellungen wird mit Kapitel 2.1 ein kurzer Überblick über die historische Entwicklung und den aktuellen Stand der psychologischen Diagnostik gegeben, um den Leser in das Themengebiet einzuführen. Hieran schließt sich mit dem Kapitel 2.2 eine Darstellung der in dieser Arbeit relevanten Gütekriterien diagnostischer Instrumente an. Anschließend werden in Kapitel 2.3 zwei Konzepte zur Wahrnehmung von diagnostischen Verfahren seitens der Probanden vorgestellt, die für die empirische Studie der vorliegenden Arbeit von Interesse sind. Kapitel 2.4 schließlich beinhaltet eine Beschreibung derjenigen diagnostischen Verfahren, auf die bei der Entwicklung der dieser Arbeit zugrunde liegenden Forschungsfragen Bezug genommen wird, während Kapitel 2.5 auf das Konzept der *Fidelity* eignungsdiagnostischer Verfahren eingeht.

Die Rolle der Medien bei der Entwicklung eignungsdiagnostischer Instrumente wird anhand eines historischen Überblicks in Kapitel 2.6 zunächst einführend vorgestellt. Dieser Einführung folgt in Kapitel 2.7 eine ausführlichere Darstellung derjenigen Entwicklungsstufen der computerbasierten Diagnostik, deren Ergebnisse sich auf die Fragestellung der vorliegenden Arbeit und die Materialgestaltung ausgewirkt haben. In Kapitel 2.8 wird dann der Fokus auf das Forschungsgebiet der multimedialen, videobasierten Eignungsdiagnostik gelegt, in welches auch die vorliegende Studie einzuordnen ist. Mit Kapitel 2.9 wird ein Aspekt dieses Forschungsgebietes vertieft, nämlich der direkte Vergleich zwischen videobasierten und herkömmlichen Verfahren in Bezug auf deren Gütekriterien und Fairness.

² Unter eignungsdiagnostischen Verfahren werden im Kontext dieser Arbeit vor allem solche Verfahren verstanden, die mit dem Ziel eingesetzt werden, die Passung der Fähigkeiten eines Anwärters oder Bewerbers auf eine Stelle, Rolle oder Aufgabe mit den Anforderungen, die diese definierte Stelle, Rolle oder Aufgabe auszeichnen, zu ermitteln.

Abschließend erfolgt mit Kapitel 2.10 eine zusammenfassende Betrachtung der bisherigen Ansätze in der videobasierten Eignungsdiagnostik sowie die Darstellung der hieraus resultierenden Motivation für die vorliegende Arbeit.

Die folgende Darstellung der historischen Entwicklung von eignungsdiagnostischen Verfahren hat nicht den Anspruch dem Leser einen erschöpfenden Überblick über das Themengebiet zu geben. Vielmehr soll der Focus auf die Darstellung grundlegender Etappen der Entwicklungsgeschichte gelegt werden, welche mit einer Verbesserung eignungsdiagnostischer Instrumente in Bezug auf deren diagnostische Qualität und Fairness gegenüber den Probanden einhergehen. Diese Schwerpunktsetzung wird gewählt, weil auch die vorliegende Arbeit mit dem Ziel durchgeführt wurde, einen Beitrag zur Weiterentwicklung von eignungsdiagnostischen Instrumenten, im Sinne der beiden oben genannten Aspekte, zu leisten. Dem Leser soll so eine erste Einordnung der Studie und ihrer Motivation in das Anwendungsfeld der Eignungsdiagnostik ermöglicht werden.

2.1 Historische Entwicklung eignungsdiagnostischer Verfahren – ein Überblick –

Berücksichtigt man die ersten dokumentierten Verfahren, zeigt sich, dass zwischen den Anfängen eines Einsatzes von eignungsdiagnostischen Verfahren, im historischen China um 2200 vor Christus, und den heutigen modernen Instrumenten des 21. Jahrhunderts über 4000 Jahre an Erfahrungen im Umgang mit diagnostischen Vorgehensweisen liegen. Trotz dieser langen Zeitspanne hat aber die wesentliche Weiterentwicklung und Optimierung der Verfahren zu ihrer heutigen Form zum größten Teil in diesem und dem vorherigen Jahrhundert stattgefunden. Aus diesem Grund wird auch nur dieser vergleichsweise kurze Zeitraum dargestellt.

Der neuzeitliche Beginn einer regelgeleiteten Messung von menschlichen Fähigkeiten wird von den meisten Quellen übereinstimmend mit der Gründung des psychologischen Laboratoriums in Leipzig im Jahre 1879 durch Wilhelm Wundt (1832-1920) angegeben (z. B. Gregory, 1996; Scheurer & Jäger, 1999). Die Arbeiten von Wundt und seinen Zeitgenossen stellen insofern den Beginn einer neuen Ära dar, als dass hier der Versuch unternommen wurde, ein objektives und replizierbares Maß für die Fähigkeiten eines Probanden zu gewinnen, welches durch einen systematischen, experimentellen Versuchsaufbau erhoben

wurde (ausführlich siehe Gregory, 1996; Scheurer & Jäger, 1999). Durch diesen neuen Ansatz, der sich in seiner frühen Form mit der Messung von Reaktionszeiten auf Stimulusreize als Indikatoren für die Denkgeschwindigkeit eines Probanden beschäftigte, war nicht nur die Basis geschaffen, die Leistungen eines Probanden objektiv abzubilden, sondern auch die Leistungen von mehreren Probanden miteinander vergleichen zu können (Gregory, 1996).

Die weiteren wesentlichen Entwicklungsstufen in Bezug auf die Qualität und Fairness der heute angewendeten diagnostischen Verfahren werden zur besseren Übersicht im Folgenden, in Anlehnung an Gregory (1996), in tabellarischer Form dargestellt. Für eine ausführliche Darstellung der historischen Entwicklung eignungsdiagnostischer Verfahren sei der interessierte Leser z. B. auf Barthel (1989), Gregory (1996) oder Jäger und Petermann (1999) verwiesen.

- (1884) Sir Francis Galton fügt erstmals die einzelnen, teilweise sehr zeitaufwendigen Verfahren seiner Zeitgenossen zu einer zeitsparenden Testbatterie zusammen, die es ermöglichte, die Daten von hunderten von Probanden in relativ kurzen Zeitabständen zu erheben. Diese Methode wird von J. McKeen Cattell in Amerika weiterentwickelt, der 1890 eine Sammlung von 10 Tests veröffentlicht und damit für die Ausbreitung der Anwendung von Eignungstests sorgt.
- (1901) Clark Wissler, zeigt erstmals, dass die Ergebnisse von Testscores aus den bisher üblichen Verfahren keine signifikanten Korrelationen mit dem Studienerfolg der Probanden aufweisen. Dies führt zu einer Neuausrichtung der Testaufgaben. Der Aspekt der Validität eines Verfahrens als wichtiges Kriterium wird erkannt.
- (1904) Binet und Simon entwickeln, aufgrund eines Erlasses des französischen Unterrichtsministeriums, den ersten modernen Intelligenztest zur Erfassung allgemeiner Intelligenz bei Kindern. Durch diesen soll eine systematische Zuweisung von benachteiligten Kindern zu speziellen Förderschulen erfolgen. Die Eignung von Probanden wird erstmals explizit unter Berücksichtigung einer möglichen Benachteiligung der Testteilnehmer geprüft.

- (1912) Hugo Münsterbergs Arbeiten über „Psychologie und Wirtschaftslehre“ bilden die Grundlage für eine stärkere Nachfrage und die damit verbundene Ausbreitung von Fähigkeitstests zur Prognose von späteren Leistungen im Beruf.
- (1916) Lewis Terman entwickelt aus dem Test von Binet und Simon den Stanford-Binet-Test und schafft ein diagnostisches Verfahren, das erstmals den Begriff „IQ“ mit Testresultaten in Verbindung bringt, die durch ein standardisiertes Verfahren erhoben werden und mit einer Normstichprobe in Vergleich gebracht werden können. Ein fairer Vergleich einzelner Probanden miteinander ist gegeben.
- (1917) Robert Yerkes entwickelt den Army Alpha und Army Beta Test zur Klassifizierung von Armeeanwärtern im 1. Weltkrieg. Erstmals kommt damit ein nonverbaler Test zum Einsatz, der die Testfairness für Anwärter mit einer fremden Muttersprache erhöhen soll.
- (1918) Robert Woodworth entwickelt mit dem „Personal Data Sheet“ den ersten modernen Persönlichkeitsfragebogen. Das „Personal Data Sheet“ soll solche Armeeanwärter identifizieren, die den psychischen Belastungen des 1. Weltkrieges nicht gewachsen sind.
- (1939) Mit der Einführung der Taylor-Russel Tafeln wird dem Anwender von Auswahlverfahren die Möglichkeit gegeben, den Nutzen des Einsatzes eines Verfahrens unter Berücksichtigung der Gütekriterien sowie der Selektionsquote (Anzahl der ausgewählten Personen) und der Grundquote (Anzahl an geeigneten Probanden in der Population) zu bestimmen.
- (1951) Lee Cronbach entwickelt mit Cronbachs alpha einen Koeffizienten, der es dem Anwender ermöglicht, die Reliabilität oder Zuverlässigkeit eines Instrumentes zu bestimmen.

- (1957) Cronbach und Gleser entwickeln ein Nutzenmodell, welches die Relevanz der Entscheidung, das Resultat aus der Entscheidung und die Art der Entscheidung selbst mit einschließt und nicht mehr nur auf die Genauigkeit des Messinstruments fokussiert.
- (1959) Campbell und Fiske veröffentlichen mit der Multitrait-Multimethod-Matrix ein Verfahren zur Ermittlung der Konstrukt-Validität eines diagnostischen Verfahrens. Die Gültigkeit für die Annahme, ein bestimmtes Konstrukt mit dem Verfahren zu messen, kann so bestimmt werden.
- (1971) Gerichtsbeschlüsse in den USA legen fest, dass nur solche diagnostischen Verfahren für die Personalauswahl eingesetzt werden dürfen, die einen klaren Bezug zum Beruf haben.
- (1978) Durch die „Uniform Guidelines on Employee Selection“ werden Richtlinien formuliert, die die Benachteiligung von Probanden aufgrund ihres Geschlechts, ihrer Nationalität oder ethnischen Abstammung verhindern sollen.
- (1985) Die American Psychological Association (APA) veröffentlicht zusammen mit anderen Organisationen die „Standards for educational and Psychological Testing“ und schafft damit eine umfassende Richtlinie für den verantwortungsvollen Umgang mit diagnostischen Verfahren.
- (2002) Mit der DIN 33430 werden in Deutschland Qualitätskriterien und Qualitätsstandards für berufsbezogene Eignungsbeurteilungen formuliert und darüber hinaus Qualifikationsanforderungen für die an der Eignungsbeurteilung beteiligten Personen genannt.

Die oben skizzierten Entwicklungsstufen in der Eignungsdiagnostik zeigen, welcher immense Fortschritt durch die Arbeiten der Forscher und die Anforderungen seitens Institutionen wie der APA für die Situation der heutigen Anwender und vor allem der Probanden erzielt werden konnte. Aus einfachen Reaktionszeitmessungen zur indirekten Bestimmung

von mentalen Fähigkeiten hat sich ein Anwendungsfeld entwickelt, das über Verfahren verfügt, die in der Lage sind komplexe Fähigkeiten und Eigenschaften zu messen.

Durch DIN-Normen, Standards und Gütekriterien werden strenge Anforderungen an diese Verfahren und ihre Anwender formuliert, um eine hohe Qualität und Fairness des diagnostischen Prozesses zu gewährleisten. Sie dienen weiterhin der Qualitätssicherung im Entwicklungsprozess, da auch neue Verfahren diesen Gütekriterien und Standards gerecht werden müssen, wenn sie in der Praxis erfolgreich sein wollen.

Um es dem Leser zu erleichtern, die Forschungsfragen der vorliegenden Arbeit sowie die Ergebnisse ihrer empirischen Überprüfung nachzuvollziehen, werden in Kapitel 2.2 die in diesem Kontext zentralen Gütekriterien vorgestellt. In Kapitel 2.3 erfolgt eine Darstellung der in dieser Arbeit relevanten Verfahren und ihrer derzeit gültigen Standards.

2.2 Gütekriterien eignungsdiagnostischer Verfahren

Die in dieser Arbeit zugrunde gelegten Definitionen beziehen sich auf das Werk von Lienert und Raatz (1999), welches als ein Standardwerk im Bereich der klassischen Testtheorie angesehen werden kann (vergleichend siehe: Amelang & Schmidt-Atzert, 2006; Amelang & Zielinski, 2004; Fisseni, 2004; Krone & Hock, 2007; Kubinger, 2006).

Nach Lienert und Raatz (1999) werden mit der *Objektivität*, der *Reliabilität* und der *Validität* die drei zentralen Gütekriterien vorgestellt, die es bei der Konstruktion und Bewertung von diagnostischen Verfahren zu berücksichtigen gilt.

2.2.1 Die Objektivität

Unter der Objektivität eines Verfahrens verstehen Lienert und Raatz (1999) den Grad, in dem die Ergebnisse eines Tests unabhängig von der Person des Untersuchers sind. Ein Test wäre demnach vollkommen objektiv, wenn verschiedene Untersucher mit dem gleichen Verfahren bei denselben Probanden zu gleichen Ergebnissen gelangten. Bei der Betrachtung der Objektivität eines Verfahrens werden mit der Durchführungs-, Auswertungswertungs- und Interpretationsobjektivität drei Arten unterschieden (Lienert und Raatz, 1999, S. 7ff).

2.2.2 Die Reliabilität

Die Reliabilität eines Verfahrens gibt nach Linert und Raatz (1999) den Grad der Genauigkeit an, mit dem ein z.B. Test ein bestimmtes Persönlichkeits- oder Verhaltensmerkmal misst. Sie wird auch als Zuverlässigkeit des Messinstrumentes bezeichnet. Da zur Bestimmung der Reliabilität eines Verfahrens unterschiedliche methodische Zugänge existieren, welche zu verschiedenen Koeffizienten bei einem Verfahren führen können, wird im Rahmen dieser Arbeit neben den berichteten Kennwerten immer auch die jeweilige Methode dargestellt.

2.2.3 Die Validität

Als drittes zentrales Gütekriterium nach Linert und Raatz (1999) gibt die Validität oder Gültigkeit eines Tests den Grad der Genauigkeit an, mit dem dieser Test dasjenige Persönlichkeitsmerkmal, das er messen oder vorhersagen soll, tatsächlich misst oder vorhersagt. Ein Test ist demnach vollkommen valide, wenn seine Ergebnisse einen unmittelbaren und fehlerfreien Rückschluss auf den Ausprägungsgrad des zu erfassenden Persönlichkeitsmerkmals zulässt. (Lienert und Raatz, 1999, S. 10ff). Auch bei der Validität unterscheiden die Autoren verschiedene Arten. Im Kontext der vorliegenden Arbeit sind vor allem die inhaltliche Validität und die kriterienbezogene Validität von Bedeutung, weshalb diese beiden Arten näher erläutert werden.

Ein Verfahren ist inhaltlich valide (content validity), wenn seine Elemente so beschaffen sind, dass sie das zu erfassende Persönlichkeitsmerkmal oder die in Frage stehende Verhaltensweise repräsentieren. Der Test selbst stellt also das optimale Kriterium für das Persönlichkeitsmerkmal oder die Verhaltensweise dar. Inhaltliche Validität wird einem Verfahren in der Regel durch ein Rating von Experten als „Konsens von Kundigen“ zugebilligt (Lienert und Raatz, 1999, S. 10ff).

Die kriterienbezogene Validität (criterion validity) eines Verfahrens bezeichnet das Ausmaß, mit welchem die aus ihm resultierenden Ergebnisse einer Stichprobe von Probanden mit einem so genannten Außenkriterium positiv korreliert sind (Lienert und Raatz, 1999, S. 10ff).

Im Fall von Personalauswahlverfahren ist besonders von Interesse, inwieweit sich aus den Probandenleistungen im Verfahren Prognosen für zukünftige Leistungen in der Zielposition

machen lassen. Da die Prognose zukünftiger Leistungen von Bewerbern ein zentrales Anliegen der Personaldiagnostik darstellt, nimmt die kriterienorientierte Validität in diesem Anwendungskontext eine besondere Stellung im Vergleich zu den übrigen Gütekriterien ein (hierzu siehe Walsh und Bezt, 2001). Auch im Falle der vorliegenden Arbeit liegt ein Schwerpunkt bei der Bewertung der Gütekriterien auf dem Aspekt der kriterienorientierten Validität.

2.3 Konzepte zur Wahrnehmung diagnostischer Verfahren

Neben den Gütekriterien als Bewertungsgrößen für die Qualität eines eignungsdiagnostischen Verfahrens stellt auch die Wahrnehmung und Bewertung des Verfahrens durch die Probanden einen wichtigen Qualitätsmaßstab dar. Im folgenden Kapitel sollen zwei zentrale Konzepte vorgestellt werden, welche sich mit dem Aspekt der wahrgenommenen Fairness eines Verfahrens durch den Probanden bzw. dessen Erleben einer Testsituation als akzeptable, soziale Situation beschäftigen. Das *Konzept der sozialen Validität* nach Schuler und Stehle (1983, 1985) bezieht sich auf den zweiten Aspekt und findet vor allem im deutschsprachigen Raum Anwendung. Das im amerikanischen Anwenderkreis populäre *Model of Applicants' Reactions to Employment Selection Systems* von Gilliland (1993) fokussiert stärker den Aspekt der wahrgenommenen Fairness eines Verfahrens. Beide Modelle stellen den theoretischen Bezugsrahmen für die zweite Fragestellung der vorliegenden Arbeit dar.

2.3.1 Das Konzept der sozialen Validität

Das erstmals von Schuler und Stehle (1983, 1985) eingeführte Konzept der sozialen Validität wird von den Autoren als „... eine Art Sammelkonzept derjenigen Aspekte verstanden, die eine eignungsdiagnostische Situation zu einer, für den Probanden, akzeptablen sozialen Situation machen.“ (Schuler & Stehle, 1983, S. 35). Die Autoren gehen weiterhin davon aus, „dass Diagnostik in einem sozialen Kontext stattfindet und dass die Fortentwicklung psychologischer Methoden auch eine Anpassung an die Veränderung sozialer Realitäten zu sein hat.“ (Schuler & Stehle, 1983, S. 35). Das Konzept der sozialen Validität stellt einen Versuch dar, einen Handlungsleitfaden für die Entwicklung und Durchführung von diagnostischen Verfahren vorzustellen, der „ergänzend zu der technisch-empirischen Validität

im testtheoretischen Sinne“, den sozialen Aspekt der diagnostischen Situation betont (Schuler & Stehle, 1983, S. 35).

Das Konzept in seiner aktuellen Form (Schuler, 1993), welches in Abbildung 2.1 graphisch dargestellt wird, beinhaltet die vier Komponenten *Information*, *Partizipation*, *Transparenz* und *Kommunikation*, die nach Aussage der Autoren für die soziale Qualität eines eignungsdiagnostischen Prozesses als relevant angesehen werden können. Diese werden im Folgenden ausführlicher vorgestellt.

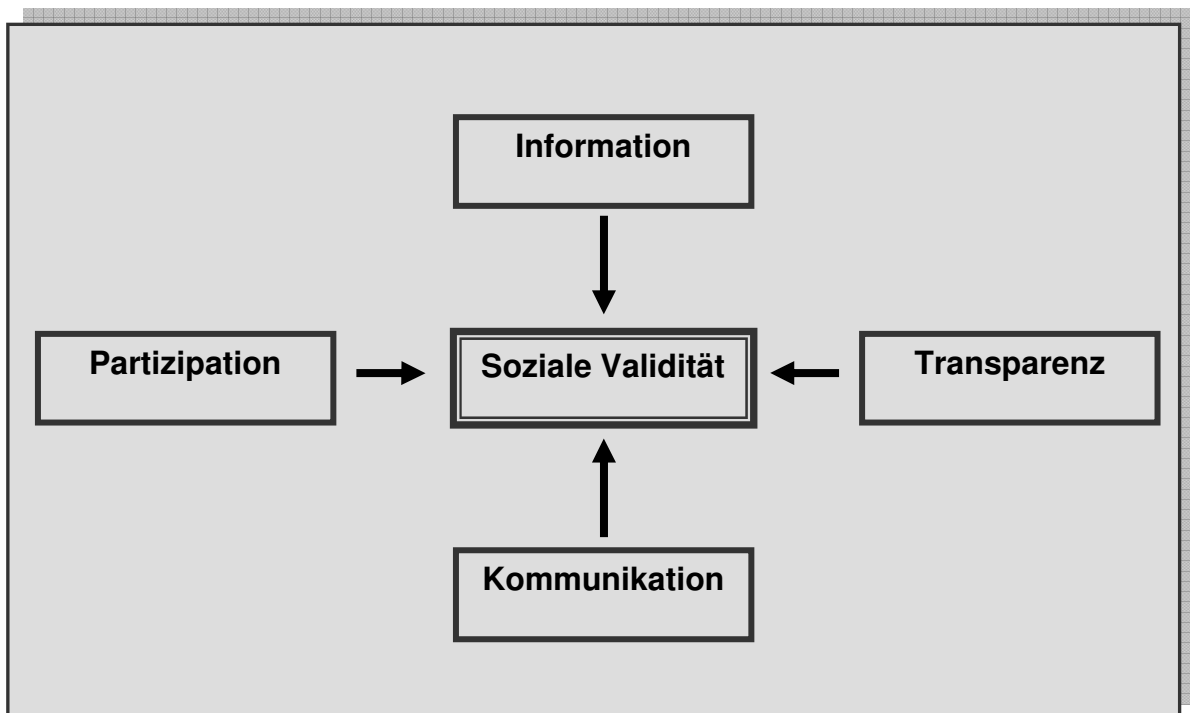


Abbildung 2.1: Das Konzept der sozialen Validität (nach Schuler, 1993, S. 13)

Information

Die Informationskomponente beinhaltet die Idee, dass der Bewerber, schon während er das Auswahlverfahren durchläuft, in der Lage sein sollte, sich ein Bild über die Anforderungen zu machen, die im späteren Job auf ihn zukommen können. Für den gesamten diagnostischen Prozess wird weiterhin eine Vermittlung von Informationen über die Unternehmenskultur und -politik sowie zu Themen wie den individuellen Weiterbildungsmöglichkeiten und Karrierechancen gefordert (Schuler, 1993, S. 13).

Partizipation

Bei der Komponente Partizipation werden zwei Aspekte unterschieden. Partizipation im engeren Sinne wird als Beteiligung bei der Entwicklung, Auswahl und Durchführung des eignungsdiagnostischen Verfahrens bzw. der diagnostischen Instrumente verstanden. Als weiterer Aspekt bezieht sich die Partizipation auf die Kontrolle, die ein Proband während des Auswahlprozesses über die Situation ausüben kann. Dies beinhaltet sowohl die Kontrolle über das eigene Verhalten und das Verhalten und die Entscheidungen der relevanten, am Prozess beteiligten Personen als auch die Kontrolle einer möglichen negativen Einflussnahme durch andere (Schuler, 1993, S. 13).

Transparenz

Als weitere Komponente bezeichnet die Transparenz eines Verfahrens den Grad, in welchem ein Proband die Zielsetzung erkennen oder ableiten kann, mit der ein diagnostischer Prozess durchgeführt wird. Die Transparenz bezieht sich in erster Linie auf die Auswahl-situation, in welcher es dem Probanden möglich sein sollte zu erkennen, welche Personen am Auswahlprozess beteiligt sind, welche Rolle sie in dem Prozess spielen und über welche Kompetenzen sie verfügen. Im Weiteren sollte der Proband darüber aufgeklärt sein, welches die an ihn gestellten Anforderungen bezüglich seines Verhaltens sind. Auch sollte dem Probanden die Wichtigkeit und Relevanz des eingesetzten Verfahrens genauso ersichtlich sein, wie ihm die Maßstäbe und Bewertungskriterien bei der Urteilsfindung transparent gemacht werden sollten. Als letzter Punkt sollten die Bewertungsgrundlagen für das gezeigte Verhalten transparent sein, um dem Teilnehmer eine Selbstbewertung seiner Leistungen zu ermöglichen und damit den Prozess der Eigenselektion zu fördern (Schuler, 1993). Werden diese Aspekte der Transparenz berücksichtigt, lassen sich Erlebnisqualitäten negativer Art wie Unsicherheit in der Testsituation, Bewertungsängste und ein wahrgenommenes Machtgefälle zwischen Diagnostiker und Proband vermeiden (Schuler, 1993, S. 13f).

Kommunikation

Mit der Kommunikation als letzter Komponente des Konzeptes wird das Ausmaß beschrieben, in welchem ein Proband Informationen über seine Leistung in dem von ihm absolvierten Verfahren zurückgemeldet bekommt. Die Rückmeldung sollte offen, ehrlich und wahrheitsgetreu erfolgen und Entwicklungspotentiale aufzeigen können. Weiterhin sollten die

Inhalte nachvollziehbar und unterstützend formuliert werden, um die Selbsteinsicht des Probanden zu fördern (Schuler, 1993, S. 14).

Das Modell von Schuler und Stehle (1983) wird von den Autoren als eine Sammlung von relevanten Einflussgrößen angesehen, die eine Testsituation aus Sicht des Probanden zu einer akzeptablen, sozialen Situation werden lassen. Schuler (1993, S. 14) weist allerdings darauf hin, dass es sich bei dem Ansatz eher um eine Heuristik handelt als um ein empirisch prüfbares Modell und sich bisherige empirische Untersuchungen lediglich mit Teilaspekten des Konzeptes befassen.

2.3.2 Das “Model of Applicants’ Reactions to Employment Selection Systems”

Mit seinem *Model of Applicants’ Reactions to Employment Selection Systems* stellt Gilliland (1993) ein Konzept vor, das stärker auf den Aspekt der wahrgenommenen Fairness fokussiert und systematisch auf einem theoretischen Rahmenkonzept aufbaut.

„Although few attempts have been made to empirically assess reactions to selection procedures, even less attention has been given to developing a theoretical model of these perceptual processes ... A natural theoretical orientation that can be used to organize this area of research is that of organizational justice theories” (Gilliland, 1993, S. 695).

Als Grundlage für sein Modell wählt Gilliland theoretische Modelle aus der Organisationspsychologie, die sich auf Aspekte wie Arbeitszufriedenheit, Gleichberechtigung und Mitarbeiterbedürfnisse beziehen. Die zentralen Charakteristika des hieraus resultierenden Modells, welches in Abbildung 2.2 graphisch dargestellt ist, werden im Folgenden näher beschrieben.

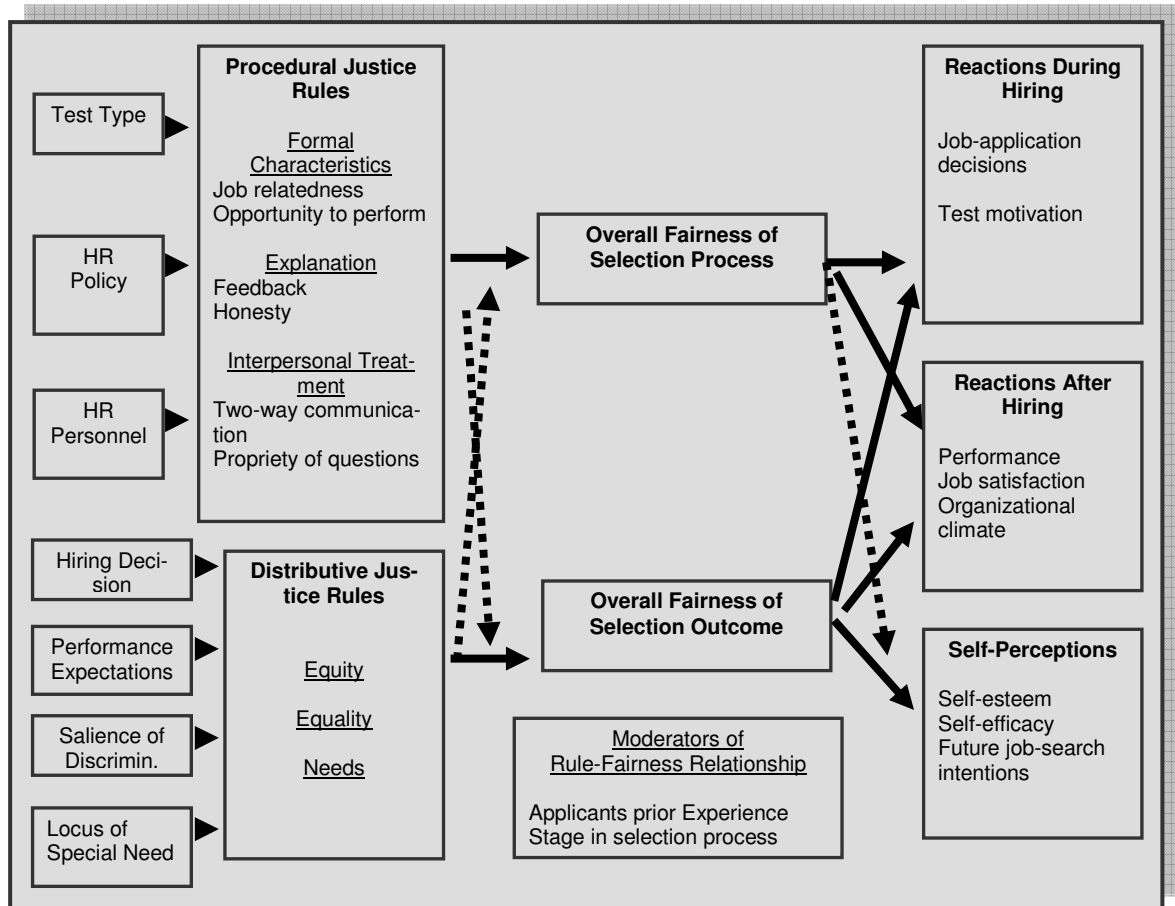


Abbildung 2.2: Das "Model of Applicants Reactions to Employment Selection Systems" (nach Gilliland, 1993)

Das Modell von Gilliland (1993) unterscheidet zwei Aspekte in der Bewertung eines Auswahlprozesses durch den Probanden, die wahrgenommene Fairness des Prozesses selbst und die wahrgenommene Fairness des Prozessergebnisses. Wie der Prozess oder das Prozessergebnis bewertet werden, hängt davon ab, in welchem Ausmaß die Bewertungsmaßstäbe, die ein Proband an eine Auswahl-situation anlegt, positiv bestätigt oder verletzt werden. Hier unterscheidet Gilliland Bewertungsmaßstäbe oder -regeln, die sich auf die Gerechtigkeit in Form einer Gleichbehandlung von Probanden beziehen, so genannte **Distributive Justice Rules**, und Bewertungsmaßstäbe, die sich auf die Gerechtigkeit eines Verfahrens, als Auswahlverfahren, beziehen, die so genannten **Procedural Justice Rules** (Gilliland, 1993, S. 696). Beide Arten von Regeln werden ihrerseits von den Rahmenbedingungen der jeweiligen Auswahl-situation, wie beispielsweise der Art der verwendeten Tests, dem personalpolitischen Vorgehen und den agierenden Mitarbeitern sowie den individuellen Erwartungen des Probanden an die Fairness und Gleichberechtigung im Verfahren und seinen Bedürfnissen beeinflusst. Neben den Bewertungsregeln berücksichtigt das Modell

als weitere Einflussfaktoren auch die früheren Erfahrungen eines Probanden mit Auswahlverfahren und die Stufe des Auswahlprozesses, in welcher sich ein Proband befindet.

Der dargestellte Bewertungsprozess und die damit verbundene empfundene Fairness von Prozess und Ergebnis beeinflusst laut Gilliland erstens das Verhalten während des Auswahlprozesses, wie zum Beispiel die Testmotivation, zweitens das Verhalten nach dem Auswahlprozess wie der erbrachten Arbeitsleistung und der Arbeitszufriedenheit und drittens die Einstellung zur eigenen Person und die weiteren Überlegungen bei der Jobsuche (Gilliland, 1993, S. 701).

2.3.3 Annahmen und empirische Befunde zur Bedeutung der sozialen Validität und wahrgenommenen Fairness in der Eignungsdiagnostik

Unabhängig davon, ob das eher einfache Modell von Schuler und Stehle oder aber der komplexe Ansatz nach Gilliland betrachtet wird, machen die theoretischen Annahmen über die Rolle der Wahrnehmung von Auswahlverfahren deutlich, dass bei der Generierung und Weiterentwicklung von diagnostischen Verfahren neben der Einhaltung der traditionellen testtheoretischen Gütekriterien auch die Wahrnehmung eines Verfahrens durch den Probanden berücksichtigt werden sollte.

Aus Sicht des Anwenders mag die Forderung, dass ein Verfahren die Fähigkeiten eines Probanden zuverlässig misst und über eine hohe prädiktive, kriterienorientierte Validität verfügt, weiterhin an oberster Stelle stehen, sollte aber nicht als erschöpfendes Kriterium bei der Auswahl eines Verfahrens gelten. Zahlreiche Beiträge aus der aktuellen Forschung zu diesem Thema zeigen vielmehr, dass auch die Bewertung des Verfahrens durch den Teilnehmer als ein wichtiges Auswahlkriterium gelten sollte.

In der aktuellen Literatur (Kanning, Grewe, Hollenberg & Hadouche, 2006) werden vor allem zwei Aspekte genannt, warum eine Berücksichtigung der sozialen Validität³ von Bedeutung ist; diese werden im Folgenden dargestellt. Der erste Aspekt bezieht sich auf den Einfluss der wahrgenommenen sozialen Validität auf die erbrachte Leistung im Verfahren. So weisen einige Autoren darauf hin, dass die Bewertung eines diagnostischen Verfahrens durch einen Probanden insofern einen Einfluss auf dessen Performance haben kann, als dass seine Leistung von Bewertungsängsten oder einer ablehnenden Haltung gegenüber

³ Im Folgenden wird dieser Terminus vereinheitlichend verwendet, wenn es um die Wahrnehmung von Auswahlverfahren in der Eignungsdiagnostik geht, und bezieht auch den Aspekt der wahrgenommenen Fairness nach Gilliland (1993) mit ein.

dem Verfahren beeinflusst werden können (Arvey, Strickland, Drauden & Martin, 1990; Chan & Schmitt, 2004; Chan, Schmitt, DeShon, Clause & Delbridge, 1997).

Schmitt und Ryan (1992) konnten in diesem Zusammenhang empirisch zeigen, dass sich die kriterienorientierte Validität des „School and College Ability Tests“ (SCAT) in Abhängigkeit von der Einstellung der Probanden zu diesem Instrument, gemessen mit der „Test Attitude Survey“ (TAS), dahingehend veränderte, dass die kriterienorientierte Validität des SCAT bei Probanden mit einer niedrigen Motivation und negativen Einstellungen zum Test signifikant schlechter ausfiel (Schmitt & Ryan, 1992, S. 633).

Der zweite in der Literatur genannte Aspekt bezieht sich auf den Einfluss, den die soziale Validität eines Verfahrens auf die Außenwirkung des Unternehmens haben kann, welches das Verfahren zu eignungsdiagnostischen Zwecken einsetzt (Chan & Schmitt, 2004; Kravitz, Stinson & Chavez, 1996; Rynes, 1993; Rynes, Bretz & Gerhart, 1991; Thornton III, 1993). Danach könnten diagnostische Verfahren, zum Beispiel in der Personalauswahl, auch als Instrumente des Personalmarketings verstanden werden und somit, bei einer positiven Wahrnehmung des Probanden, auch zu einem positiven Bild des gesamten Unternehmens führen. Bauer, Truxillo, Paronto, Weekley und Campion (2004) konnten in diesem Zusammenhang zeigen, dass sich signifikante Zusammenhänge zwischen der Einschätzung von Screening-Instrumenten in der Personalauswahl und der Einschätzung der Fortschrittlichkeit des Unternehmens ergaben, welches diese Instrumente einsetzte (Bauer et al, 2004, S.142).

Gilliland (1993) fügt diesen eher wirtschaftlichen Aspekten noch eine ethische und eine rechtliche Komponente hinzu. Er nennt in diesem Zusammenhang die Verantwortung, die ein auswählendes Unternehmen gegenüber den Bewerbern hat, da ein Auswahlverfahren immer auch eine Auswirkung auf das Selbstwertgefühl und die Einschätzung der eigenen Fähigkeiten haben kann. Aus einer arbeitsrechtlichen Perspektive weist er weiterhin darauf hin, dass sich auch die Wahrscheinlichkeit, dass ein abgewiesener Proband arbeitsrechtliche Schritte einleitet, durch ein faires Verfahren deutlich verringert (Gilliland, 1993, S. 695), was für ein Unternehmen, neben dem Imagegewinn, auch eine enorme Kostenersparnis darstellen kann.

Die ausführliche Vorstellung der Konzepte und Ergebnisse zur sozialen Validität soll deren hohe Relevanz im Kontext der Eignungsdiagnostik darlegen und die Notwendigkeit vor Augen führen, bei der Konzeption neuer Verfahren auch deren soziale Validität einer genauen Betrachtung zu unterziehen. Im Fall der vorliegenden Arbeit liegt der Focus auf der Fragestellung, inwieweit die soziale Validität eines Verfahrens durch die Art der medialen

Darbietungsform dieses Verfahrens beeinflusst wird. Aus diesem Grund wird auf eine ausführliche, generelle Darstellung der empirischen Befunde zur sozialen Validität verzichtet; der interessierte Leser sei hier auf z. B. Schuler und Stehle (1983), Schuler (1993), Gilliland (1993), Gilliland und Hale (2005) oder Ployhart und Harold (2004) verwiesen. Stattdessen werden für jedes Kapitel der vorliegenden Arbeit, welches eine Betrachtung des Einflusses von medialen Weiterentwicklungen auf die Eigenschaften eignungsdiagnostischer Verfahren zum Inhalt hat, die relevanten Ergebnisse zu Aspekten der sozialen Validität gesondert vorgestellt.

In den vorangegangenen Kapiteln über die Anforderungen an diagnostische Verfahren, wurden zwei Themengebiete vorgestellt, die in den empirischen Fragestellungen der vorliegenden Arbeit Berücksichtigung finden. Zum einen sind dies die relevanten testtheoretischen Gütekriterien, die es bei der Entwicklung und Durchführung von eignungsdiagnostischen Verfahren zu beachten gilt, und zum anderen zwei Konzepte, die sich mit der empfundenen Fairness bzw. der sozialen Validität von Auswahlverfahren seitens der Probanden befassen. Nachdem diese Konzepte eingeführt sind, werden im folgenden Kapitel die personaldiagnostischen Verfahren dargestellt, welche für die Forschungsfragen der vorliegenden Arbeit relevant sind.

2.4 Vorstellung der relevanten Verfahren und ihrer Standards

Die übergeordnete Fragestellung der vorliegenden Arbeit bezieht sich auf den Aspekt, inwieweit sich die mediale Variation in der Darbietungsform eines eignungsdiagnostischen Verfahrens sowohl auf dessen Gütekriterien im klassischen Sinne als auch auf dessen soziale Validität auswirkt.

Im Rahmen der vorliegenden Studie sind dabei drei Verfahren von Relevanz, die sowohl in Bezug auf ihren eignungsdiagnostischen Hintergrund als auch in Bezug auf ihre Gütekriterien und Standards vorgestellt werden sollen. Als erstes Verfahren wird das Assessment-Center vorgestellt, da es als eignungsdiagnostisches Rahmenkonzept für die empirische Untersuchung der vorliegenden Studie diene. Bei dem zweiten vorzustellenden Verfahren handelt es sich um das Rollenspiel als dasjenige Verfahren, welches einer medialen Manipulation unterzogen werden soll und im Zentrum der empirischen Untersuchung steht. Drittens wird die Verfahrensgruppe der so genannten Situational Judgement Tests vorgestellt, da ein großer Teil der Forschung zur medialen Variation von Testverfahren im Kontext

dieser Verfahrensart stattgefunden hat und die Fragestellungen der vorliegenden Arbeit aus den bisherigen Ergebnissen zu diesem Forschungsgebiet entwickelt werden.

2.4.1 Das Assessment-Center

Assessment-Center werden in Deutschland seit den 70er Jahren zu personaldiagnostischen Zwecken eingesetzt und haben sich nach Kanning (2004) inzwischen sowohl in der Wirtschaft als auch in Bereichen des öffentlichen Dienstes als simulationsorientierte Verfahren fest etabliert. Eine Definition, die alle wesentlichen Aspekte dieses Ansatzes beinhaltet, gibt Fennekels (1987).

„Das Assessment-Center Verfahren ist ein systematisches und flexibles Verfahren zur kontrollierten und qualifizierten Feststellung von Verhaltensleistungen und –defiziten, das von mehreren Beobachtern gleichzeitig für mehrere Teilnehmer, in Bezug auf vorher festgelegte Übungen und bestimmte Anforderungen, vornehmlich zur Personalauswahl und -weiterentwicklung von vielen Personalentwicklungsabteilungen in Großunternehmen mit Erfolg und steigender Tendenz eingesetzt wird“ (Fennekels, 1987, S. 10).

Der Vorteil dieses multiplen Vorgehens gegenüber einem Einzelverfahren liegt in der Tatsache begründet, dass die Bewertung der Fähigkeiten jedes Teilnehmers im Idealfall in mehreren Übungen erfolgt und durch mehrere geschulte Beobachter anhand von fest definierten Beobachtungskriterien bzw. Verhaltenshinweisen für die einzelnen Situationen erfolgt. Dadurch soll die Auftretenswahrscheinlichkeit von systematischen Beobachtungsfehlern vermindert und die Chance auf ein reliables und valides Ergebnis erhöht werden (Fisseni, 2004). Die Ergebnisse einer Metaanalyse nach Schmidt und Hunter (1998) zeigten für das AC ein kriterienbezogene Validität von .37. Dieser Wert wird durch die Ergebnisse der Metaanalysen von Thornton, Gaugler, Rosenthal und Bentson (1992) mit einem über alle Studien gemittelten Koeffizienten von .37 und von Arthur, Day, McNelly und Edens (2003) mit Koeffizienten von .25 - .39 bestätigt. Damit nimmt das Assessmentcenter im Vergleich zu anderen Verfahren der Personalauswahl eine mittlere Position ein. In Bezug

auf ihre soziale Validität sind Assessment-Center nach Schuler und Stehle (1983) und Schuler (1993) psychologischen Tests überlegen.

Eine zusammenfassende Darstellung einzelner Quellen (Kanning, 2004; Höft & Funke, 2006; Kleinmann, 2003; Kleinmann 1997; Sünderhauf, Stumpf & Höft, 2005; Fisseni, 2004; Schuler & Stehle, 1983) bezüglich der Gütekriterien und Standards für das Assessment-Center zeigt Box 1.

Assessmentcenter:

Kombination mehrerer Übungen, in welchen vorher definierte Eigenschaften und Verhaltensweisen beobachtet werden

Dauer:

Zwischen einem und drei Tagen, in seltenen Fällen bis zu einer Woche

Validität:

Die Studien zur kriteriumsbezogenen Validität weisen Koeffizienten um .37 auf, hohe soziale Validität

Nachteile:

Hohe Kosten, hoher organisatorischer Aufwand

Standards:

- Anforderungsanalyse zur Bestimmung der Merkmalsdimensionen
- Einsatz von geschulten Rollenspielern
- Beobachtertraining
- Verhaltensverankerte Ratingskalen
- Begleitung durch einen professionellen Moderator
- Beobachterkonferenz zwecks Integration der einzelnen Urteile zu einem Gesamturteil
- Individuelles Feedback für die Teilnehmer bezüglich ihrer Leistungen

Box 1: Das Assessment-Center: Eigenschaften, Gütekriterien und Standards

2.4.2 Das Rollenspiel

Nach Höft und Funke (2006) handelt es sich bei eignungsdiagnostischen Rollenspielen um Simulationen einer Interaktionssituation von zwei Personen im innerbetrieblichen Kontext oder in einer Kundensituation. Je nach Aufgabenstellung kann es sich dabei um eine kooperative oder aber konfrontative Aufgabe für den Probanden handeln. Bei den Interaktionspartnern handelt es sich entweder um zwei Probanden oder aber um einen Probanden und einen geschulten Rollenspieler. Ziel der Übung ist es primär das Interaktionsverhalten eines Probanden mit seinem Übungspartner zu beobachten (Kanning, 2004) beziehungsweise zu ermitteln, wie gut sich ein Proband in der simulierten Situation bewährt (Höft & Funke, 2006). Im Kontext der Personaldiagnostik wird das Rollenspiel in den meisten Fällen nicht als alleinstehende Übung eingesetzt. Vielmehr wird es in ein komplexeres Verfahren eingebettet, das mehrere Übungsansätze miteinander verbindet, wie das oben beschriebene Assessment-Center. Eine zusammenfassende Darstellung einzelner Quellen (Höft & Funke; 2006; Kanning, 2004; Kleinmann 1997, 2003) bezüglich der Gütekriterien und Standards für das Rollenspiel zeigt Box 2.

Rollenspiel:

Interaktionsübung für 2 Personen

Dauer:

Ca. 10 -15 Minuten

Validität:

Kriteriumsbezogene Validitätskoeffizienten um .30 bei Beobachterübereinstimmungen (Reliabilität) zwischen .57 und .80.

Vorteile:

Hohe inhaltliche Validität, hohe soziale Validität

Nachteile:

Hohe Kosten, hoher zeitlicher Aufwand, hoher Personalaufwand, Möglichkeit von Beobachterfehlern, mangelnde Standardisierung bei ungeschulten Rollenspielern

Standards:

- Einsatz von geschulten Rollenspielern und Beobachtern
- Keine strengen Rollendefinitionen für den Probanden

Box 2: Das Rollenspiel: Eigenschaften, Gütekriterien und Standards

2.4.3 Situational Judgement Tests

Nach McDaniel, Finnegan, Morgenson, Campion und Braverman (2001) finden die im amerikanischsprachigen Raum als Situational Judgement Tests (SJT) bezeichneten situativen Tests bereits seit 1920 Verwendung. Die von den Autoren vorgeschlagene Definition bezieht sich auf die klassische Form von situativen Tests, die als Papier- und Bleistiftform durchgeführt werden.

„... we broadly define situational judgement tests as any paper-and-pencil test designed to measure judgement in work settings. Some of these tests can be classified as situational, in that a scenario is described and the respondent must identify an appropriate response from a list of alternatives”
(McDaniel et al. 2001, S. 730)

Bei situativen Tests geht es also darum herauszufinden, wie ein Proband, der mit verschiedenen Situationen aus dem Berufsalltag konfrontiert wird, sich in diesen Situation verhalten würde bzw. welches Verhalten er in den jeweiligen Situation als das am besten geeignete ansieht. Die Bewertung der Probandenantwort erfolgt dabei anhand von vorher definierten, qualitativ abgestuften Musterantworten, die in Form einer verhaltensverankerten Einstufungsskala konstruiert sein sollten (Schuler & Marcus, 2006).

Die kriterienbezogene Validität von situativen Tests liegt nach einer Metaanalyse von McDaniel et al. (2001) bei .34 und damit in einem moderaten Bereich. Dieses Ergebnis wird von aktuellen Studien bestätigt (Chan & Schmitt, 2002; Howard & Choi, 2000; McDaniel & Nguyen, 2001; Oswald, Schmitt, Kim, Ramsay & Gillepsie, 2004). Weiterhin konnten Pereira, Schmitt und Harvey (2001) zeigen, dass sich bei SJT geringere Leistungsunterschiede zwischen Gruppen unterschiedlicher Nationalitäten ergeben als bei kognitiven Leistungstests.

Betrachtet man die Forschungsliteratur der letzten 10 Jahre auf dem Gebiet der SJTs, so zeigt sich eine rasante Entwicklung sowohl in der Art und Weise, wie die einzelnen Situationen aus dem Berufsleben dem Probanden präsentiert werden, als auch in Bezug auf die Gestaltung der Antwortmöglichkeiten. So nehmen neben den Paper-Pencil Varianten und den SJTs in Form von mündlichen Interviews vor allem videobasierte Verfahren einen immer größeren Raum ein.

Eine zusammenfassende Darstellung einzelner Quellen (Kanning, 2004; Schuler & Marcus, 2006; McDaniel et al., 2001; Oswald, Schmitt, Kim, Ramsay & Gillepsie, 2004; Howard & Choi, 2000; McDaniel & Nguyen, 2001; Pareira, Schmitt & Harvey, 2001; Chan & Schmitt, 2002) bezüglich der Gütekriterien und Standards für SJTs zeigt Box 3.

Situational Judgement Tests:

In mündlicher, schriftlicher oder videobasierter Form durchgeführter Test für eine oder mehrere Personen

Dauer:

ca. 30 bis 90 Minuten

Validität:

Die Studien zur kriteriumsbezogenen Validität weisen Koeffizienten zwischen .16 und .53 auf

Vorteile:

Hoher Realitätsbezug, hohe Akzeptanz, geringer Konstruktionsaufwand bei Paper-Pencil Formen

Nachteile:

Keine homogenen Skalen, Verhalten wird nur beschrieben aber nicht umgesetzt

Standards:

- Anforderungsanalyse zur Bestimmung der Merkmalsdimensionen
- Experteninterviews zur Bestimmung der relevanten Verhaltensweisen und der Skalenstufen
- Hohe Retestrelabilität
- Dem Proband sollte ausreichend Zeit für die Erfassung der Situation zugebilligt werden
- Vor allem bei einer mündlichen Präsentation der Situation wird dem Probanden die Möglichkeit zur Nachfrage gegeben

Box 3: SJT: Eigenschaften, Gütekriterien und Standards

2.5 Die Einteilung eignungsdiagnostischer Verfahren nach dem Konzept der „Fidelity“

Nachdem die für die vorliegende Arbeit relevanten Gütekriterien sowie die relevanten Verfahren und ihrer Standards eingeführt wurden, soll dem Leser im Folgenden das Konzept der *Fidelity* nach Motowidlo, Dunnet und Carter (1990) vorgestellt werden. Neben der Darstellung des Konzeptes erfolgt weiterhin eine Unterscheidung der oben beschriebenen Verfahren nach diesem Konzept, da sich hieraus ebenfalls Implikationen für die Herleitung der Fragestellungen im Rahmen der vorliegenden Arbeit ergeben.

Nach Motowidlo, Dunnette und Carter (1990) lassen sich Verfahren nach dem Grad ihrer *Fidelity*, also dem Grad ihrer *Wiedergabetreue* einteilen. Unter *Wiedergabetreue* wird in diesem Zusammenhang das Ausmaß verstanden, in welchem ein Verfahren sowohl mit der Stimulus Komponente seiner Aufgaben als auch mit der Antwortmöglichkeit der Probanden auf diesen Stimulus die reale Situation am Arbeitsplatz abbilden kann.

„Simulations vary in the *fidelity* with which they present a task stimulus and elicit a response“ (Motowidlo, Dunnette & Carter, 1990, S. 640).

Nach Aussage der Autoren ist der höchste Grad an *Fidelity* dann gegeben, wenn das in der Übung geforderte Verhalten exakt dem später am Arbeitsplatz geforderten Verhalten entspricht und das für die Übung verwendete Material sehr realitätsnah gestaltet ist.

„The highest *fidelity* simulations use very realistic materials and equipment to represent a task situation and provide applicants with an opportunity to respond almost exactly as if they were actually in the job situation“ (Motowidlo, Dunnette & Carter, 1990, S. 640).

Auf der anderen Seite des Kontinuums in Bezug auf die Wiedergabetreue eines Verfahrens sehen die Autoren solche Verfahren, die lediglich auf eine verbale Beschreibung von hypothetischen Situationen im Arbeitsleben zurückgreifen und von den Probanden eine Beschreibung ihrer potentiellen Handlungen abfragen.

„At the lower end of the *fidelity* continuum are simulations that simply present a verbal description of a hypothetical work situation, instead of a concrete representation, and that ask applicants to describe how they would deal with the situation, instead of having them actually carry out some action to deal with it“ (Motowidlo, Dunnette & Carter, 1990, S. 640).

Zum Zwecke einer Unterscheidung zwischen Verfahren mit einer hohen Wiedergabetreue und denen mit einer niedrigen Wiedergabetreue führen die Autoren die Bezeichnungen *high-fidelity simulations* und *low-fidelity simulations* ein (Motowidlo, Dunnette & Carter, 1990). Nach Havighurst, Fields und Fields (2003) weisen *high-fidelity simulations* gegenüber *low-fidelity simulations* gewisse Vorteile, aber auch Nachteile auf. Die Vorteile von Verfahren mit einer hohen *Fidelity* seien deren höhere Augenscheinvalidität und die Tatsache, dass sie einen sehr viel realistischeren Einblick in den späteren Berufsalltag bieten könnten, als dies bei *low-fidelity simulations* möglich sei. Die Nachteile lägen in einem höheren Aufwand für Personal und Equipment sowie den sehr hohen Kosten und dem hohen zeitlichen Aufwand. Hier könnten *low-fidelity simulations* punkten, die weniger Equipment und Personal benötigten und somit deutlich weniger Kosten produzierten.

Wendet man die Systematik von Motowidlo, Dunnette und Carter (1990) auf die für diese Arbeit wesentlichen Verfahren an, so ist das Rollenspiel im Rahmen eines Assessment-Centers den *high-fidelity simulations* zuzurechnen, während die Situational Judgement Tests zu den *low-fidelity simulations* zu zählen sind. Dieser Aspekt wird bei der Formulierung der Fragestellungen der vorliegenden Arbeit berücksichtigt werden, wenn es nämlich um die Übertragbarkeit von Ergebnissen zu medial variierten SJTs auf die eignungsdiagnostische Methode des Rollenspiels geht.

Mit Kapitel 2.6 beginnt die Darstellung eines weiteren grundlegenden Aspektes der vorliegenden Arbeit, nämlich die Betrachtung der Rolle der Medien bei der Entwicklung und Optimierung von eignungsdiagnostischen Verfahren. Neben einem historischen Überblick werden in diesem Kapitel einzelne Epochen der Eignungsdiagnostik unter Berücksichtigung der jeweils typischen medialen Einflüsse und der hierdurch bedingten Weiterentwicklungen in diesem Fach dargestellt.

2.6 Die Rolle der Medien bei der Entwicklung von eignungsdiagnostischen Verfahren

Mit Kapitel 2.1 wurde die Entwicklung diagnostischer Instrumente von einfachen Reaktionszeitmessungen bis hin zu modernen und komplexen Verfahren, welche hohen Ansprüchen an Gütekriterien und Fairness genügen, skizziert. Als ein wichtiger, diese Entwicklung beeinflussender Aspekt, wurden in diesem Zusammenhang die zunehmenden Anforderungen in Bezug auf die Gütekriterien, die Akzeptanz und ethische Überlegungen genannt, die von Seiten der Forscher und Anwender an die Verfahren gestellt werden.

Berücksichtigt man den Beitrag der Medien, zu dieser Entwicklung, so zeigt sich, dass sie einen wesentlichen Anteil bei diesem Prozess geleistet und die Forschung stark beeinflusst haben. Dieser Beitrag soll im Folgenden ausführlicher dargestellt werden.

Die Betrachtung der medialen Entwicklung in der Eignungsdiagnostik beginnt bei den meisten Quellen mit der Einführung des Personalcomputers und den sich daraus ergebenden neuen Chancen und Möglichkeiten in der Anwendung und Auswertung von Testverfahren um 1946 (siehe z. B.: Jäger & Petermann, 1999; Gregory, 1996; Hageböck, 1994; Bartram & Hambleton; 2006). Die hierauf folgenden, wesentlichen Entwicklungsschritte werden im weiteren Verlauf in tabellarischer Form in Anlehnung an Gregory (1996, S. 575) dargestellt. Für eine ausführliche Darstellung sei der interessierte Leser auf z. B. Bartram und Hambleton (2006); Fisseni (2004); Gregory (1996); Hageböck (1994) oder Jäger und Petermann (1999) verwiesen.

- (1946) Hanks entwickelt einen analogen Computer zur Auswertung des SVIB, die bis dahin zeitintensive und fehleranfällige Auswertung wird schneller und genauer.
- (1962) Die Mayo Clinic in den USA erfasst zum ersten Mal die Antworten von Probanden auf speziellen IBM Karten. Die Auswertung von Tests kann dadurch erstmals elektronisch und vollständig automatisiert durch einen Computer erfolgen. Die Objektivität der Auswertung wird deutlich erhöht.
- (1971) Das VA Hospital in Salt Lake City etabliert ein Computersystem mit Eingabeterminals, durch welches der komplette diagnostische Prozess, welchen der Patient anhand verschiedener Testverfahren durchläuft, auto-

- matisiert werden kann. Nach Aussage der Anwender konnten damit die Diagnosegutachten in der Hälfte der Zeit und mit dem halben Kapitelaufwand erstellt werden, die eingesparte Zeit kommt den Klienten zugute.
- (70er) Durch den technischen Fortschritt und die zunehmenden Rechnerleistungen wird computerbasiertes Adaptives Testen ermöglicht. Die Fähigkeit eines Probanden kann, nach Aussage der Entwickler, „maßgeschneidert“ und sehr viel schneller bestimmt werden, da sie mittels eines individuellen Itempools und anhand weniger Items gemessen wird.
 - (1986) Die Amerikan Psychological Association veröffentlicht die „Guidelines for Computer-Based Tests and Interpretations und sorgt damit, durch verbindliche Richtlinien, für eine höhere Fairness in der Anwendung von computerbasierten Tests.
 - (90er) Vor allem in Unternehmen werden verstärkt multimediale eignungsdiagnostische Verfahren zur Auswahl von Bewerbern eingesetzt. Simulationen und Animationen lösen einfache Übertragungen von schriftlichen Tests in Computerversionen immer mehr ab. Die Aufgabenstellungen bieten durch den Einsatz von realitätsnahen Videos und Animationen einen guten Einblick in den späteren Arbeitskontext und tragen zu einer Erhöhung der sozialen Validität bei.
 - (2000) Die Weiterentwicklung des Internets macht eine netzbasierte Vorauswahl von Bewerbern möglich. Den Probanden können online komplexe Aufgaben in virtuellen Welten oder multimediale Auswahlverfahren vorgegeben werden. Die neue Technologie bietet den Unternehmen die Chance weltweit die am besten geeigneten Bewerber kostengünstig und effizient zu rekrutieren.

Die überblicksartige Darstellung der Entwicklung von diagnostischen Instrumenten, unter Berücksichtigung des medienbedingten Fortschritts und Einflusses vor allem in den letzten 40 Jahren, zeigt eine rasante und positive Entwicklung für dieses Forschungsfeld. Im folgenden Kapitel sollen die Erwartungen und Fragen von Anwendern und Forschern in den unterschiedlichen Entwicklungsstufen der computerbasierten Diagnostik ausführlicher vorgestellt und die empirischen Ergebnisse bewertet werden.

2.7 Erwartungen, Forschungsfragen und empirische Ergebnisse einzelner Entwicklungsstufen der computerbasierten Diagnostik

Neben einer Betrachtung der Chancen und Vorteile, welche sich durch den Einsatz immer modernerer Technologien in der Eignungsdiagnostik bieten, lag und liegt das Interesse der Forschung zu jeder Zeit auch auf der Entdeckung von möglichen technologiebedingten, kritischen Aspekten für den diagnostischen Prozess. Im Folgenden werden die empirischen Ergebnisse der in Abbildung 2.3 dargestellten Entwicklungsstufen deshalb sowohl in Bezug auf die positiven als auch die negativen Auswirkungen von Medien auf die Eignungsdiagnostik und die hieraus resultierenden Implikationen für die heutige Forschung vorgestellt.

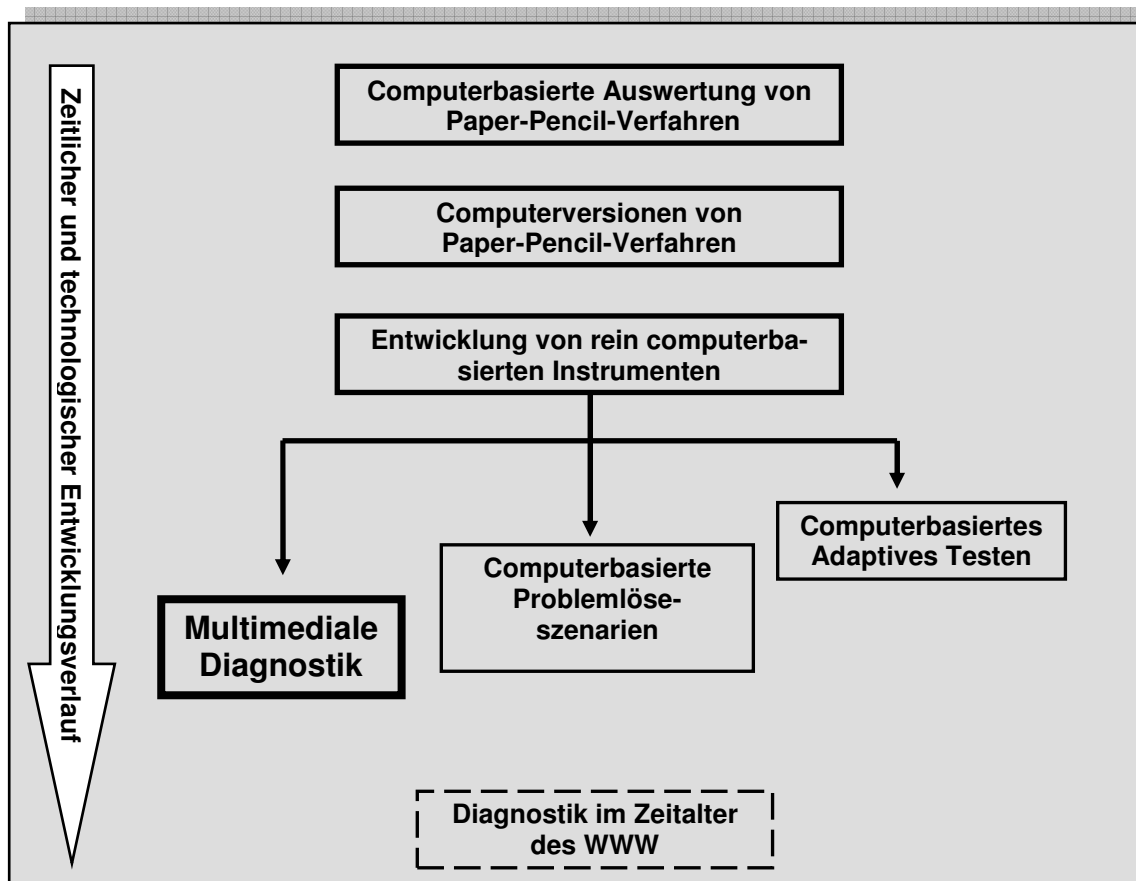


Abbildung 2.3 Die Entwicklungsstufen der computerbasierten Diagnostik

In Abhängigkeit von ihrer Relevanz für die Forschungsfragen der vorliegenden Arbeit erfahren die in Abbildung 2.3 dargestellten Entwicklungsstufen eine unterschiedliche Gewichtung. So wird die erste Stufe *Computerbasierte Auswertung von Paper-Pencil Verfahren* lediglich überblicksartig dargestellt. Sie dient dazu dem Leser einen Einblick in die Anfänge der computerbasierten Diagnostik zu vermitteln und skizziert die generelle Debat-

te um die Vor- und Nachteile des Einsatzes von Computern im diagnostischen Prozess. Die empirischen Ergebnisse der Stufen *Computerversionen von Paper-Pencil-Tests*, *Adaptives Testen* und *Computergestützte Problemlöseszenarien* werden ausführlicher vorgestellt, da sich aus ihnen entweder konkrete Implikationen für die Materialgestaltung der vorliegenden Arbeit ergeben oder Hinweise zu den Grenzen einer Weiterentwicklung durch mediale Aspekte dargestellt werden. Der Schwerpunkt der Darstellung wird auf den Bereich *Multi-mediale Diagnostik* gelegt, da die Untersuchung der vorliegenden Arbeit in diesen Bereich der Eignungsdiagnostik einzuordnen ist. Die der Vollständigkeit halber eingefügte Entwicklungsstufe *Diagnostik im Zeitalter des WWW* wird im Kontext dieser Arbeit nicht behandelt, da dieser Forschungsbereich für die Entwicklung der Fragestellungen der vorliegenden Arbeit nicht relevant ist. Der interessierte Leser sei hier auf die Werke von z. B. Bartram und Hambleton (2006); Kirbach, Montel, Oenning und Wottawa (2004) oder Konradt und Sarges (2003) verwiesen.

2.7.1 Die computerbasierte Auswertung von Paper-Pencil-Verfahren

Der Beginn der computerbasierten Diagnostik mit den ersten elektronischen Auswertungen von Testergebnissen stellte auch gleichzeitig den Beginn einer Diskussion über die Nutzen und Gefahren des Einbezugs von Computern in den diagnostischen Prozess dar. Den berichteten Vorteilen, wie der hohen Zeit- und Kostenersparnis (Klingler, Miller, Johnson & Williams, 1977), wurden Befürchtungen entgegengebracht, der Einsatz von Computern könne den diagnostischen Prozess verarmen und ihn weniger flexible und intuitiv werden lassen bzw. der Computer den Diagnostiker eventuell ganz überflüssig machen (Booth, 1999). Beeinflusst wurde diese Diskussion durch die Debatte um die Vor- oder Nachteile in Bezug auf eine klinische vs. statistische Urteilsbildung, die unter anderem durch Meehl (1954, 1956) zu einer Diagnostiker vs. Computer Debatte wurde. Aus heutiger Sicht lässt sich in Bezug auf die computerbasierte Auswertung von Testergebnissen ein positives Fazit ziehen. Ein Blick in die aktuelle Literatur zeigt den Konsens der Forschung in Bezug auf die vielen positiven Aspekte der computerbasierten Testauswertung, wie eine höhere Auswertungsobjektivität, eine nicht unerheblichen Kosten- und Zeitersparnis, die schnelle Verfügbarkeit der Ergebnisse sowie eine bessere Anschaulichkeit der Ergebnisse für den Probanden durch den Einsatz von Graphiken (Amelang & Zielinski, 2004; Amelang & Schmidt-Atzert, 2006; Booth (1999); Wagner-Menghin, 2003; Gregory, 1999). Auch die

Diskussion um eine klinische vs. statistische Urteilsbildung hat eine neue Richtung genommen. Seit der hohe Nutzen der statistischen Urteilsbildung in Überblickarbeiten, wie z. B. bei Dawes, Faust und Mehl (1989) mit annähernd 100 vergleichenden Studien, gezeigt werden konnte, wird der Einbezug des Computers in den diagnostischen Prozess als Chance gesehen diesen Prozess zu optimieren, ohne sich dabei ausschließlich auf den Computer zu stützen oder den Diagnostiker außen vor zu lassen.

Booth (1999) weist in diesem Zusammenhang darauf hin, dass die Programmierung von Auswertungsalgorithmen für die computerbasierte Diagnostik eine klare Strukturierung der Auswertungsvorgänge und einen transparenten Urteilsprozess mit klaren Entscheidungsregeln erfordere. Der Einsatz des Computers verlangt also vom Diagnostiker, sein Vorgehen zu präzisieren, zu systematisieren, sowie den Urteilsprozess transparent zu machen und damit vorhandene Strategien zu überdenken und eventuell zu optimieren.

Nach einem solchen Analyseprozess kann der Nutzen des Computers dabei sowohl in der regelgeleiteten computerbasierten Auswertung von vorher erhobenen Daten, nach festen empiriebasierten Routinen, als auch in einer systematischen Begleitung des Diagnostikers durch den diagnostischen Prozess entlang allgemein akzeptierter Regeln und Entscheidungswege bzw. wissenschaftlich abgesicherter Strategien liegen (Wagner-Menghin, 2003). Ausführlich zu dem zweiten Aspekt siehe auch Hageböck (1994), welcher mit DIASYS 1 ein computerbasiertes Diagnoseinstrument für die Anwendungen in schulpyschologischen Beratungsstellen vorstellt, das den oben beschriebenen Aspekt berücksichtigt.

Als Fazit aus der Forschungsdebatte soll festgehalten werden, dass der Einbezug eines Computers nur dann gewinnbringend in Bezug auf die genannten Vorteile erscheint, wenn diesem Einsatz eine sorgfältige Analyse des diagnostischen Prozesses bzw. dessen, was es zu messen gilt, vorausgeht.

2.7.2 Die Übertragung von Paper-Pencil Verfahren auf den Computer

Auf die Nutzung des Computers zur reinen Auswertung von Testverfahren folgte, bedingt durch die Möglichkeiten des technologischen Fortschritt und der Einführung von immer leistungsfähigeren und kostengünstigeren PCs, die Übertragung von Paper-Pencil Verfahren in eine computerbasierte Version.

Die Vorteile einer solchen Übertragung liegen aus Sicht der Forscher dabei sowohl in einer höheren Durchführungs- und Auswertungsobjektivität als auch in einer deutlichen Verkür-

zung des diagnostischen Prozesses und damit in einer Reduzierung der Kosten (siehe z. B. Volz-Sidiropoulou, 2004). Da sich die Testversionen aber durch Aspekte wie das Format der Items (schriftlich vs. bewegte Bilder), der Art der Antworteingabe und den generellen Umgangsmöglichkeiten mit dem Testmaterial stark unterscheiden, wurde von Seiten der Forschung darauf hingewiesen, dass sich die ökonomischen Vorteile der computerbasierten Testung nur dann rechtfertigen lassen, wenn der Übertragung der Paper-Pencil Testverfahren in ihre computerbasierte Version eine Überprüfung der Äquivalenz der beiden Versionen folgt (APA, 1986; Mead & Drasgow, 1993; Schwenkmezger & Hank, 1993). Für diese werden in der Literatur üblicherweise drei Aspekte unterschieden, die aktuell im Focus der Forschung stehen und im Folgenden erläutert werden (siehe z. B. Klinck, 2002; Volz-Sidiropoulou, 2004).

Die *psychometrische Äquivalenz* als erster Punkt fordert, dass die Testgüte des ursprünglichen Tests auch im neuen Medium erhalten bleiben soll. Sie ist gewährleistet, wenn sich erstens die Verteilung und die Mittelwerte der Items beider Tests nicht unterscheiden und sich zweitens die Rangfolgen der Personenscores nicht entscheidend verändern, da in diesem Fall eine Korrektur der Verletzung der ersten Forderung über z. B. eine lineare Transformation erfolgen kann (zu diesem Aspekt siehe Mead & Drasgow, 1985, S. 451).

Die *erfahrungsbezogene Äquivalenz* beinhaltet den Aspekt eines möglicherweise unterschiedlichen, subjektiven Erlebens der beiden Darbietungsformen eines Tests durch den Kandidaten und damit einhergehende Unterschiede in der emotionalen Bewertung (Klinck, 2002, S. 13f.).

Die *populationsbezogene Äquivalenz* bezieht sich darauf, wie sich die Fairness der Testverfahren in Abhängigkeit vom eingesetzten Medium für bestimmte Probandengruppen gleicht. Konkret wird betrachtet, inwieweit durch das computerbasierte Verfahren Teilnehmer eines bestimmten Geschlechts, Alters oder Bildungsstandes genauso bevorzugt oder benachteiligt werden, wie dies in der Paper-Pencil Version der Fall wäre (Volz-Sidiropoulou, 2004, S. 288, vergleichend hierzu siehe auch Klinck, 2002, S. 13f.).

Dass die kritische Betrachtung der Äquivalenz von Testverfahren, mit unterschiedlicher medialer Basis, immer noch einen hohen Stellenwert einnimmt, zeigt sich darin, dass sie nach wie vor in der Forschung zur computerbasierten psychologischen Diagnostik einen Schwerpunkt bildet.

Im Folgenden soll der aktuelle Stand der Forschung zu diesen drei Aspekten der Äquivalenz von Paper-Pencil Testverfahren und ihren PC-Versionen überblicksartig vorgestellt werden. Für eine ausführliche Darstellung sei der interessierte Leser z. B. auf die umfas-

sende Überblicksarbeit von Klinck (2002), die Metaanalyse von Mead und Drasgow (1993) oder die Übersicht von Volz-Sidiropoulou (2004) verwiesen, deren Ergebnisse hier zusammenfassend dargestellt werden.

Ergebnisse zur psychometrischen Äquivalenz

In Bezug auf die psychometrische Äquivalenz zeigt sich aus Sicht der Forschung ein uneinheitliches Bild. So kommen Mead & Drasgow (1993) in ihrer Metaanalyse zu dem Ergebnis, dass bei zeitbegrenzten Powertests eine hohe Chance auf eine Äquivalenz der beiden medial variierten Testformen besteht, wenn der Test sorgfältig konstruiert wurde und sowohl über eine zeitliche Begrenzung verfügt als auch über Aufgaben unterschiedlicher Schwierigkeit. Allerdings weisen die Autoren gleichzeitig darauf hin, dass nicht von einer generell gültigen Annahme der Äquivalenz von Powertests ausgegangen werden kann. Vielmehr lasse sich die Äquivalenzannahme nur bei Tests rechtfertigen, deren Übersetzung in eine computerbasierte Form anhand eines sorgfältigen Prozesses vorgenommen werde. (Mead & Drasgow, 1993, S. 456). Dieser Aspekt wird auch von Klinck (2002, S. 15) mit einer Auflistung von Studien belegt, die auf Äquivalenzprobleme bei Powertests hinweisen. Volz-Sidiropoulou (2004, S. 288) kommt bei der Bewertung einzelner Studien sogar zu dem Ergebnis, dass bei Intelligenz- und Leistungstests nicht von einer psychometrischen Äquivalenz ausgegangen werden kann, wenn diese von der Paper-Pencil Version in eine computerbasierte Version übertragen werden.

In Bezug auf Speedtests kommen Mead und Drasgow (1993) im Rahmen ihrer Metaanalyse zu dem Ergebnis, dass die Ergebnisse sehr stark von der Art der Präsentation beeinflusst werden und ein Vergleich von Ergebnissen aus beiden medialen Varianten nicht ohne weiteres möglich scheint. Die Autoren sehen eine mögliche Erklärung in den unterschiedlichen Anforderungen an die motorischen Fähigkeiten der Probanden bei der Bearbeitung von Paper-Pencil Verfahren und einer computerbasierten Form. Sie gehen weiterhin davon aus, dass sich die Unterschiede verringern oder eliminieren lassen würden, wenn die computerbasierte Version sich ähnlich gestalten ließe wie die Paper-Pencil Version (Mead & Drasgow, 1993, S. 453).

Ergebnisse zur erfahrungsbezogenen Äquivalenz

In Bezug auf die erfahrungsbezogene Äquivalenz zeigen einzelne Studien, zusammengefasst nach Volz-Sidiropoulou (2004) und Klinck (2002), eine hohe Akzeptanz von computergestützten Leistungstests und eine Präferenz für diese Art der Testvorgabe gegenüber der

Paper-Pencil Version. Allerdings konnten Richman-Hirsch, Olson-Buchanan und Drasgow (2000) zeigen, dass eine Übertragung der Paper-Pencil Version in eine computerbasierte Variante nicht automatisch zu besseren Ergebnissen in der Bewertung führt. Sie erhoben Urteile von 131 Managern, die das Conflict Resolution Skills Assessment (Olson-Buchanan, Drasgow, Moberg, Mead, Keenan & Donovan, 1998) entweder in Paper-Pencil Form, in Form einer einfachen Übertragung auf den Computer oder in einer multimedialen Variante bearbeiteten. Es zeigte sich, dass nur letztere als valider, fairer, moderner und kürzer wahrgenommen wurde als die anderen beiden Varianten. Die einfache Übertragung in den Computer dagegen wurde nicht besser bewertet als die Papierform.

Auch für Persönlichkeitstests wird von der Tendenz berichtet, dass die Computervarianten von Fragebögen mindestens so positiv bewertet werden wie die ursprünglichen Versionen auf Paper-Pencil Basis und sich die Bereitschaft seitens der Probanden, sich offen und ehrlich zu äußern, erhöht (Juraska und Drasgow, 2001; Millstein, 1987; Volz-Sidiropoulou, 2004).

Ergebnisse zur populationsbezogenen Äquivalenz

Studien, die sich mit dem Vergleich der Testwerte von Probanden unterschiedlichen Alters, Geschlechts oder Bildungsgrades bei der Bearbeitung der computerbasierten- im Vergleich zur Paper-Pencil-Variante beschäftigen, zeigen nach den von Klinck (2002, S. 22 ff.) zusammengefassten Arbeiten eher heterogene Befunde.

Dies gilt auch für Studien, die sich mit dem Effekt von Computervorerfahrung und der Einstellung zum Computer in diesem Kontext auseinandergesetzt haben. Es zeigt sich, dass eine generelle und einheitliche Aussage bezüglich dieser Variable auf die Leistungen der Probanden in den jeweiligen Testversionen nicht möglich ist (siehe Klinck 2002, Volz-Sidiropoulou, 2004). Weiterhin weist Klinck (2002) darauf hin, dass die intellektuelle Leistungsfähigkeit ein beeinflussender Faktor sein könne. Menschen mit einem höheren Intellekt könnten auch über mehr Computererfahrung verfügen und eine positivere Einstellung gegenüber neuen Medien zeigen als solche mit einem niedrigen Intellekt.

Aus den Ergebnissen zur Äquivalenzforschung ergeben sich einige Implikationen für die Übersetzung von herkömmlichen Verfahren in eine computerbasierte Variante, welche auch für die Vorgehensweise bei der Materialerstellung in der vorliegenden Arbeit von Interesse sind. So zeigt sich, dass die Art der Itempräsentation, die Handhabbarkeit des Testmaterials sowie die Sorgfalt bei der Übertragung des ursprünglichen Verfahrens in die

medial variierte, neue Version - und deren hieraus resultierende Qualität - die Kennwerte eines Verfahrens sowie dessen subjektive Wahrnehmung durch die Probanden zu beeinflussen scheinen. Weiterhin zeigen die Studien, dass eine erfolgreiche Konzeptionalisierung einer computerbasierten Verfahrensvariante berücksichtigen sollte, dass die Übersetzung herkömmlicher Items nicht in Form einer einfachen Übertragung vorgenommen werden, sondern alle Möglichkeiten des neuen Mediums zur Verbesserung der Testsituation mit in Betracht gezogen werden sollten. Nur so scheinen die Vorteile einer computerbasierten Version, wie die höhere Durchführungs- und Auswertungsobjektivität und die größere Ökonomie, nicht zulasten der mindestens ebenso wichtigen Gütekriterien und der wahrgenommenen Testfairness zu gehen.

2.7.3 Das computerbasierte Adaptive Testen

Als nächster Schritt der medialen Entwicklung in der Eignungsdiagnostik wird der Forschungsbereich des computerbasierten Adaptiven Testens dargestellt. Der Fokus in diesem Gebiet liegt neben der Betrachtung des Einflusses neuer Technologien auf die Leistungen und das Befinden der Probanden auf einer Betrachtung der Möglichkeiten, welche sich durch den Einsatz des PC bei der Umsetzung von modernen testtheoretischen Ansätzen bieten.

Die entscheidende Rolle des Computers in diesem Anwendungsfeld wird am ehesten deutlich, wenn man sich den Prozess des adaptiven Testens selbst vor Augen führt. Bei einer adaptiven Testvorgabe orientiert sich die Auswahl der Items, die einem Probanden vorgegeben werden, am Leistungsniveau der betreffenden Person. Der gesamte Prozess des Testens ist flexibel und wird vom Probanden durch seine Antworten mitbestimmt. Sowohl die Reihenfolge als auch die Anzahl der zu bearbeitenden Items wird nach jeder Antwort des Probanden anhand von komplexen Algorithmen berechnet (ausführlich siehe z. B.: Kisser, 1999; Rost, 1996; Volz-Sidiropoulou, 2004).

Das Ziel dieses Vorgehens besteht darin, dem Probanden immer dasjenige Item vorzugeben, das den höchsten Informationsgehalt in Bezug auf die Fähigkeit des Probanden liefert. Man spricht in diesem Zusammenhang auch vom „Taylord Testing“, da die Itemvorgabe „maßgeschneidert“ auf den einzelnen Probanden abgestimmt wird. Die Itemvorgabe läuft so lange, bis ein vorher bestimmtes Abbruchkriterium erreicht wird. Die Grundlage für die Berechnung der hierfür notwendigen Item- und Personenparameter bilden nach

Volz-Sidiropoulou (2004, S. 290) „... Algorithmen in Form fester Verzweigungsregeln oder komplizierter psychometrischer Berechnungen“. Da die Klassische Testtheorie bei diesem komplexen Prozess an ihre Grenzen stößt (Kisser, 1999), basieren adaptive Tests auf verschiedenen Modellen der Probabilistischen Testtheorie (ausführlich hierzu siehe z. B. Birnbaum, 1968; Rasch, 1960, Rost, 2004) und nicht auf den Axiomen der Klassischen Testtheorie, wie dies zur Zeit noch für die meisten Tests in Paper-Pencil Form zutrifft (ausführlich hierzu siehe Lienert & Raatz, 1999).

Obwohl auch vor der Nutzung von Computern in diesem Gebiet bereits Methoden entwickelt wurden, die das Adaptive Testen in Paper-Pencil Form ermöglichten, schreibt Kisser (1999) unter Berücksichtigung der oben genannten komplexen Vorgehensweise:

„All dies muss aber rasch geschehen, soll nicht durch lange Wartezeiten zwischen den einzelnen Aufgaben der Proband gelangweilt und der Ökonomiegewinn gefährdet werden. Praktisch sind solche Anforderungen nur bei einer computerunterstützten Testvorgabe erfüllbar“ Kisser (1999, S. 166).

Berücksichtigt man weiterhin den Aspekt, dass auch die Messung der Bearbeitungsdauer eines Items einen Hinweis auf die Fähigkeiten des Probanden liefert, wird eines deutlich. Der Computer ist als unterstützendes Medium in der adaptiven Diagnostik unentbehrlich geworden. Die Einschätzung Kissers (1999) weist gleichzeitig auf mögliche positive Aspekte hin, welche sich die Entwickler und Anwender von computerbasierten adaptiven Testverfahren versprechen. Diese können nach Volz-Sidiropoulou (2004) in testökonomische, messqualitative und motivationale Aspekte unterteilt werden, welche im Folgenden näher erläutert werden.

Unter testökonomischen Gesichtspunkten betrachtet, kommen adaptive Tests, durch die maßgeschneiderte Vorgabe der Aufgaben, in den meisten Fällen mit deutlich weniger Items aus als klassische Testverfahren. In einigen Fällen ließe sich die Itemzahl beim adaptiven Testen um bis zu 50 % reduzieren, ohne einen Verlust der Reliabilität hinnehmen zu müssen (Wild, 1989). Weiterhin sollte sich ein zu messendes Merkmal insofern genauer erfassen lassen, als dass die Verfahren, bedingt durch die zugrunde liegenden probabilistischen Testkonzepte, auf einer homogenen Itemstruktur aufbauen und somit nicht, wie bei klassischen Tests häufig vorkommend, andere Fähigkeitsaspekte mit einbeziehen, die nicht von unmittelbarem Interesse sind (Volz-Sidiropoulou, 2004).

Als Vorteile des adaptiven Ansatzes in Bezug auf die messqualitativen Aspekte lassen sich der hohe Differenzierungsgrad auf allen Stufen des Fähigkeitskontinuums und, bei Tests mit wenigen Items, eine höhere Reliabilität und Validität als bei klassischen Tests gleicher Länge anführen (Volz-Sidiropoulou, 2004).

Bedingt durch den maßgeschneiderten Ablauf von adaptiven Verfahren wird auch unter motivationalen Aspekten ein Vorteil in dieser Art des Testens gesehen. Die Probanden bekommen nur solche Aufgaben präsentiert, die ihrem Fähigkeitsgrad so gut wie möglich entsprechen. Dadurch werden weder Probanden mit einer hohen Fähigkeitsausprägung durch zu leichte Aufgaben unterfordert noch Probanden mit einer niedrigen Fähigkeitsausprägung überfordert und damit demotiviert (Kisser, 1999; Volz-Sidiropoulou, 2004).

Ein kritischer Vergleich der oben genannten theoretischen Vorteile mit empirischen Befunden zum computerbasierten adaptiven Testen wirft allerdings Fragen in Bezug auf die Tragfähigkeit einiger Annahmen auf. Die zentralen hieraus resultierenden Hinweise aus der Literatur sollen im Folgenden zusammenfassend nach Booth, 1999; Kisser, 1999; Moosbrugger, 2004 und Volz-Sidiropoulou, 2004 dargestellt werden.

So führen weniger Items im Test nicht zwangsläufig zu einer Verkürzung der Testzeit, da sich die Probanden mit den einzelnen maßgeschneiderten Items im Schnitt deutlich länger befassen. Weiterhin kann, durch die unterschiedlich wahrgenommenen Schwierigkeiten der Items seitens der Probanden, die regelgeleitete Bestimmung des Abbruchkriteriums sehr kompliziert ausfallen. Auch die Motivation kann beeinträchtigt werden, wenn die Aufgaben zu Anfang des Tests die volle Fähigkeit des Probanden beanspruchen und keine „Eisbrecheraufgaben“ zur Gewöhnung an die Testsituation dargeboten werden. In Bezug auf die Messqualität muss kritisch gesehen werden, dass solche Items, die nicht theoriekonform sind, aus dem Test eliminiert werden müssen. Dies kann zu einer Ausdünnung des Itempools führen, wodurch die Möglichkeit einer Differenzierung, über das gesamte Schwierigkeitskontinuum und die Möglichkeit den Personenparameter auf allen Stufen des Fähigkeitskontinuums exakt zu bestimmen, stark eingeschränkt wird. Dabei ist schon die Konstruktion, Testung und Kalibrierung des ursprünglichen, theoriekonformen Itempools mit einem sehr hohen Aufwand verbunden (ausführliche hierzu siehe Booth, 1999; Kisser, 1999; Moosbrugger, 2004; Volz-Sidiropoulou, 2004).

Aufgrund der dargestellten Kritikpunkte lässt sich aus den bisherigen Erkenntnissen zum computerbasierten Adaptiven Testen festgehalten, dass einer positiven Betrachtung der Chancen, die sich durch die zunehmende Technologisierung und Medialisierung bietet,

immer auch eine kritische, empirische Überprüfung der Gütekriterien und sozialen Validität dieser Verfahren gegenüberstehen sollte. Der Versuch einer optimalen Abbildung probabilistischer Testmodelle durch den Einsatz von Computern mag einen wichtigen Schritt in der Weiterentwicklung der psychologischen Diagnostik darstellen. Er ist aber nur dann Erfolg versprechend, wenn der Focus der Forschung nicht ausschließlich darauf gerichtet wird, dem Anwender von Tests reliable, valide und ökonomische Verfahren an die Hand zu geben. Testtheoretische Betrachtungen sollten immer auch auf Aspekte wie den Motivationsgehalt und die soziale Validität und wahrgenommene Fairness fokussieren, da diese eine Auswirkung auf die Leistung und das Befinden der Probanden haben und damit ebenfalls von hoher Relevanz sind.

2.7.4 Computerbasierte Problemlöseszenarien

Computerbasierte Problemlöseszenarien haben sich aus dem eher grundlagenbezogenen Forschungsgebiet der Kognitionspsychologie entwickelt und finden seit gut 20 Jahren vor allem in der Managementdiagnostik ihren Einsatz (hierzu siehe Funke, 1991; Funke, 1998; Kersting, 2001). Auch für diese Verfahrensgruppe zeigen bisherige Forschungsergebnisse, dass ein gewinnbringender Einsatz von Verfahren, die auf der Verwendung moderner Technologien beruhen, nicht per se garantiert werden kann. Eine sorgfältige Planung bei der Konstruktion der Items sowie eine empirische Prüfung der Gütekriterien sind auch hier wesentliche Voraussetzungen. Im Folgenden wird zuerst eine kurze Definition von computerbasierten Problemlöseszenarien gegeben, bevor konkret auf die Forschungsergebnisse bezüglich solcher Szenarien in der Eignungsdiagnostik eingegangen wird, um sowohl positive Aspekte als auch Entwicklungsfelder dieses Forschungsgebietes aufzuzeigen.

Nach Höft und Funke (2006) versteht man unter computergestützten Szenarios „... spezielle, auf Computern implementierte Aufgaben, die die Kontrolle und Steuerung komplexer dynamischer Systeme oder Prozesse verlangen, welche mit einer mehr oder weniger realitätsnahen Rahmenhandlung versehen sind, z. B. der Leitung eines Unternehmens“ (Höft und Funke, 2006, S. 151).

Charakteristisch für die computerbasierten Szenarien ist dabei, dass sie vom Probanden das Lösen komplexer Probleme erfordern, welche durch die Merkmale *Komplexität*, *Vernetztheit*, *Dynamik*, *Intransparenz* und *Polytelie* definiert sind. Diese werden im Folgenden zu-

sammenfassend nach (Dörner, Kreuzig, Reither & Stäudel, 1983, S. 19 ff.; Funke, 1991, S. 186f. und Kanning, 2004, S. 365 ff.) dargestellt.

Der Aspekt der *Komplexität* bezieht sich auf die Tatsache, dass ein Proband bei seinen Aktionen mehrere Variablen gleichzeitig in Betracht ziehen muss. *Vernetztheit* meint, dass sich die Veränderung einer Variabel auch auf die anderen Variablen des Systems auswirken, ohne dass der Proband diese direkt manipuliert hat. *Intransparenz* bedeutet, dass dem Probanden nicht alle Information über die Wirkungszusammenhänge einzelner Variablen oder die Folgen von Aktionen direkt zugänglich sind und er sich diese Information beschaffen muss bzw. wichtige Erkenntnisse erst im Verlauf des Verfahrens generieren kann. Unter *Dynamik* wird in diesem Zusammenhang verstanden, dass das System sich in Abhängigkeit der Aktionen eines Probanden fortlaufend verändert. *Polytelie* meint, dass dem Probanden mehrere Ziele vorgegeben werden, die teilweise in Gegensatz zueinander stehen können.

Die Anfänge der Erforschung solch komplexer Problemlöseprozesse mittels computerbasierter Szenarien gehen im deutschsprachigen Bereich auf die Arbeiten von Dietrich Dörner und das von ihm konzipierte computerbasierte Szenario „Lohausen“ zurück (ausführlich siehe: Dörner, Kreuzig, Reither & Stäudel, 1983). Dörner, den die Frage beschäftigte, ob die menschliche Denkfähigkeit den Anforderungen, die solche komplexen und nicht genau bestimmten Systeme mit sich bringen, gerecht werden kann, bat in seinem Projekt 48 Probanden, die kleine Stadt „Lohausen“ als Bürgermeister 10 Jahre lang zu regieren, indem sie Entscheidungen trafen und diese einem Computersystem mitteilten.

Aus diesem Ansatz haben sich inzwischen zahlreiche Verfahren entwickelt, die sich vor allem in Aspekten wie der Realitätsnähe der abgebildeten Prozesse zu den realen Prozessen in der späteren Arbeitswelt (Leitung eines Unternehmens vs. Planung einer Mondmission), der Anzahl der zu beachtenden Variablen (von 2 bis zu 2000) oder ihrem Anwendungskontext unterscheiden (Beispiele siehe bei Funke, 1991; Funke, 1995; Kanning, 2004; Kleinmann & Strauss, 1998). Auch die Fragestellungen der Forschung in diesem Bereich sind heterogener geworden und haben sich von den ursprünglichen Fragestellungen Dörners und seiner Kollegen (siehe Dörner, Kreuzig, Reither & Stäudel, 1983, S. 115) auch in Richtung einer anwendungsorientierten Betrachtung weiterentwickelt. Vor allem die Frage nach der Vorhersagemöglichkeit von späteren Leistungen im Beruf, aufgrund der in den Szenarien erbrachten Leistungen eines Probanden, die sich in Kategorien wie dem erreichten Resultat, gemessen z. B. am Umsatz oder der Jahresbilanz des fiktiven Unternehmens, des gezeigten

Frage- und Entscheidungsverhaltens und dem Wissen über Zusammenhänge und Systeme beobachten lässt, nehmen hier einen zentralen Stellenwert ein (Kleinmann & Strauß, 1998). Da in der vorliegenden Arbeit vor allem der Aspekt der Verbesserung des eignungsdiagnostischen Prozesses und seiner Verfahren durch den sinnvollen Einsatz sich immer weiter entwickelnder Medien von Interesse ist, wird zuerst eine Vorstellung der derzeit in der Literatur diskutierten Vor- und Nachteile des Einsatzes von computerbasierten Problemlösenszenarien in diesem Anwendungskontext gegeben. Anschließend wird der Focus auf Forschungsergebnisse zu den Gütekriterien von computerbasierten Problemlösenszenarien gelegt.

Als Vorteile computerbasierter Problemlösenszenarien werden in der Literatur derzeit deren hoher Unterhaltungswert und eine damit wahrscheinlich einhergehende höhere Motivation bei der Bearbeitung genannt. Weiterhin wird auf die hohe Akzeptanz für Verfahren verwiesen, die sehr realitätsnah gestaltet sind, sowie ihre kostengünstige und standardisierte Durchführung und Auswertung hervorgehoben, da es sich um computerbasierte Verfahren handelt (Kanning, 2004). Die Simulationen erlauben es weiterhin sehr komplexe, sich dynamisch verändernde Situationen abzubilden und bieten so die Möglichkeit, realitätsnahe und gegenüber herkömmlichen kognitiven Leistungstests auch neuartige Anforderungen an die Probanden zu stellen (Funke, 1998).

Die derzeit noch existierenden Nachteile und Entwicklungsfelder für den Bereich der anwendungsorientierten computerbasierten Problemlösenszenarien überwiegen allerdings die Vorteile noch deutlich, wie z. B. Funke (1998), Kersting (1999, 2001) zusammengefasst nach Kanning (2004, S. 367 ff.) aufzeigen.

Als erster Punkt ist hier die, je nach Szenario, sehr aufwändige und kostenintensive Entwicklung sowie die Notwendigkeit einer empirischen Überprüfung der Szenarien zu nennen, verbunden mit der Tatsache in entsprechende Hardware investieren zu müssen. Weiterhin konfundieren die Testergebnisse mit dem fachlichen Vorwissen und der Computererfahrung der Probanden, was nur dann nicht als negativ zu bewerten ist, wenn diese Eigenschaften auch für den späteren Berufserfolg relevant sind. In vielen Fällen ist nicht genau bestimmbar, was überhaupt gemessen wird. Mittelhohe Korrelationen zu Intelligenztests und niedrige Korrelationen von Ergebnissen aus verschiedenen Problemlösenszenarien untereinander lassen Zweifel daran aufkommen, dass die Fähigkeit, komplexe Probleme zu lösen ein exakt abgrenzbares, eigenständiges, reliabel und valide bestimmbares Konstrukt darstellt. Viele der Szenarien sind fiktiver Natur und weisen somit eine mangelnde Übereinstimmung mit der späteren realen Situation im Unternehmen auf (siehe auch: Strauß,

Hasselmann & Hasselmann, 1992, S. 534). Sind die in der Simulation zu steuernden Prozesse aber denen in der Realität unähnlich, lässt sich aus dem Ergebnis keine sinnvolle Vorhersage für die spätere Befähigung im Beruf machen, die inhaltliche Validität leidet. Auch der Aspekt einer fehlenden sozialen Dimension kann in diesem Zusammenhang genannt werden. Während in der Testsituation der Proband seine Entscheidung völlig allein als Reaktion auf die Gegebenheiten im System trifft, ist er in der Realität auch von den Entscheidungen anderer Personen im gleichen System abhängig (Funke, 1998). Die per Definitionem festgelegte Vernetztheit der Variablen und die Dynamik der Systeme führen zwangsläufig dazu, dass sich Probanden durch ihre individuellen Entscheidungen im Verlauf des Verfahrens völlig unterschiedlichen Anforderungen ausgesetzt sehen. Ein fairer Vergleich der Probandenleistungen und eine Gleichbehandlung der Teilnehmer in Bezug auf die Testfairness scheinen somit kaum möglich (Funke, 1998). Hinzu kommt das Problem, dass aufgrund der Komplexität einiger Szenarien die darin enthaltenen Anforderungen an den Probanden gar nicht vollständig lösbar sind. Andererseits werden dem Anwender, mittels Logfiles, eine Vielzahl von Variablen zugänglich gemacht, von denen bei weitem nicht alle sinnvolle Maße zur Messung der Fähigkeiten eines Probanden darstellen müssen. Wenn aber keine Musterlösung bzw. keine optimale Lösungsstrategie definiert ist und die Bewertungsmaße keine validen Indikatoren für den Berufserfolg darstellen, fällt nicht nur die Bewertung der Probandenleistungen schwer, sondern auch die Qualität und damit die Glaubwürdigkeit des Messinstruments kann aus Sicht der Probanden angezweifelt werden, da nicht einmal der Testleiter selbst die Musterlösung kennt (Funke, 1998).

Auch in Bezug auf die Gütekriterien, wie Objektivität, Reliabilität und Validität, lassen die Beiträge aus der Forschung zu Problemlöseszenarien generell (z. B. Funke, 1995; Kersting, 1999; Kersting, 2001; Höft & Funke, 2006) ein eher kritisches Urteil zu. Obwohl bei computerbasierten Verfahren zunächst von einer Verbesserung der Durchführungs- und Auswertungsobjektivität ausgegangen werden kann, lässt sich für die Problemlöseszenarien, bedingt durch deren dynamischen und komplexen Charakter, diese Annahme nicht generell halten. Je nach Strategie sieht sich der Proband mit Situationen ganz unterschiedlicher Schwierigkeit konfrontiert, was eine Vergleichbarkeit in Bezug auf die Bearbeitungssituation schwierig, wenn nicht gar unmöglich macht (Höft & Funke, 2006).

Schätzungen bezüglich der Reliabilität von computerbasierten Szenarien gestalten sich nach Höft und Funke (2006) eher schwierig. Während die Bestimmung der Retestreliabilität durch Lerneffekte verzerrt werden kann, trifft man bei der Bestimmung der Paralleltestreliabilität auf das Problem einer mangelnden Äquivalenz der betrachteten Szenarien.

Auch Maße der internen Konsistenz sind aufgrund des dynamischen Charakters der Aufgaben nur schwer bestimmbar. Bedingt durch diese Einschränkungen zeigen sich, je nach Art des Szenarios und der Art der Reliabilitätsbestimmung, sehr heterogene Ergebnisse. So berichten Überblicksarbeiten (siehe z. B. Kleinmann & Strauß, 1998, S. 101; Funke, 1995, S. 224 ff.) für die Retestreliabilität Koeffizienten von $r = .50$ bis $r = .74$, für die Paralleltestreliabilität Koeffizienten von $r = .11$ bis $r = .94$ und für die Paralleltestreliabilität Koeffizienten von $r = .83$ bis $r = .86$.

Bei der Betrachtung der Validität von computerbasierten Problemlöseszenarien zeigt sich ein uneinheitliches Bild, sowohl was die Methodik der Bestimmung als auch die resultierenden Ergebnisse angeht. So berichtet Funke (1995) für den Aspekt der Konstruktvalidität, bestimmt durch die Berechnung der Übereinstimmung zwischen verschiedenen Ergebnismaßen innerhalb eines Szenarios, Korrelationskoeffizienten von $r = .40$ bis $r = .68$ und für Berechnungen des Zusammenhanges von Probandenleistungen in verschiedenen Szenarien Koeffizienten von $r = .30$ bis $r = .70$. Zusammenhänge zwischen Leistungen in Szenarien und Intelligenztests variieren je nach betrachtetem Test und Szenario sogar zwischen Nullkorrelationen und Zusammenhangswerten von $r = .80$ (ausführlich zu einzelnen Studien siehe Funke, 1995, S. 181 ff. oder Kleinmann & Strauß, 1998).

Für den, vor allem für eignungsdiagnostische Verfahren, so wichtigen Aspekt der kriterienorientierten Validität (Kleinmann und Strauß, 1998) werden in den vorliegenden Überblicksarbeiten verhältnismäßig wenig Studien genannt (für eine Übersicht siehe Funke, 1995). Die wenigen Studien zeigen stark unterschiedliche Validitätskoeffizienten, die je nach Szenario und innerhalb eines Szenarios je nach Außenkriterium Werte von $-.58$ bis $.67$ einnehmen können. Eine Interpretation der Ergebnisse und die daraus resultierende Bewertung der prognostischen Validität von computergestützten Problemlöseszenarien wird dadurch erschwert (Funke, 1995; Kleinmann & Strauss, 1998). Kersting (1999, 2001) weist aufgrund der Ergebnisse seiner Studie zur Konstruktvalidität und Kriteriumsvalidität von computerbasierten Problemlöseszenarien weiterhin auf das Problem einer mangelnden inkrementellen Validität der Szenarien gegenüber herkömmlichen Intelligenztests hin.

Fasst man die oben genannten Ergebnisse zusammen, so scheint eine generelle Annahme der Nützlichkeit von computerbasierten Problemlöseszenarien nach dem derzeitigen Stand empirischer Forschung fraglich. Wenn die Szenarien, im Vergleich zu herkömmlichen Tests, weder ein neues, für den Berufserfolg wichtiges Konstrukt messen, noch eine zufriedenstellende inkrementelle Validität aufweisen, muss ihr Einsatz aufgrund der hohen Ent-

wicklungskosten eher kritisch gesehen werden. Die Tatsache, dass solche Verfahren von Seiten der Probanden als positiv beurteilt werden, reicht allein nicht aus.

Sinnvolle Verbesserungsvorschläge für eine zukünftig positive Entwicklung existieren bereits. So weisen Arbeiten aus dem Forschungsgebiet (z. B. Funke, 1998; Funke, 1995; Strauß, Hasselmann & Hasselmann, 1998) darauf hin, dass zukünftige PC-Simulationen vor allem einen hohen Grad an *Fidelity* bezüglich der realen Berufswelt aufweisen sollten. Die Abbildungsgenauigkeit sollte dabei allerdings nicht nur augenscheinlich erreicht werden, sondern sich auch in einer großen Ähnlichkeit in der Tiefenstruktur von Computersystemen und realen Systemen widerspiegeln. Eine gute Passung der Anforderungen, die in einem späteren Job an einen Probanden gestellt werden und denen in den Szenarien setzten, neben einer sorgfältigen Übertragung der realen Situation auf den Computer, eine ebenso sorgfältige Anforderungsanalyse zur Bestimmung der realen Anforderungen voraus (Strauß, Hasselmann & Hasselmann, 1998). Auch Funke (1991) weist auf die noch wenig zufriedenstellenden Ergebnisse bezüglich der Gütekriterien hin. Er betont, dass neben der Notwendigkeit, die den Szenarien zugrunde liegenden mentalen Modelle sorgfältiger zu definieren und ihre Messung durch weitere Differenzierungen zu verbessern, auch die Notwendigkeit einer Standardisierung der Systeme in Bezug auf die Art und Intensität der auftauchenden Probleme besteht. Nur so ließen sich die Bearbeitungsbedingungen für die Probanden zumindest annähernd gleich halten. Strauß, Hasselmann und Hasselmann (1992) schlagen in diesem Zusammenhang vor, dass zukünftige Entwickler eine ausführliche Dokumentation über die Erstellung der Programme sowie das Systemverhalten und die Anforderungsstruktur des Systems zur Verfügung stellen sollten. So ließen sich sowohl die Übereinstimmung mit realen Anforderungen überprüfen als auch optimale Lösungswege und Lösungsstrategien identifizieren.

Die Betrachtung der empirischen Ergebnisse und Verbesserungsansätze zeigt zwei zentrale Aspekte auf, die es zu beachten gilt, wenn neue Technologien in die Entwicklung von diagnostischen Verfahren einbezogen werden und die auch in der vorliegenden Arbeit Berücksichtigung fanden.

Ein Aspekt bezieht sich auf das Bestreben, durch den Einsatz neuer Technologien eine möglichst reale Abbildung der späteren Anforderungen am Arbeitsplatz zu realisieren. Hierfür sollte in einem ersten Schritt die reale Situation und ihre Anforderungen an einen Probanden klar definiert werden. In einem zweiten Schritt sollten dann anhand klarer Übertragungsregeln diese Anforderungen mittels der neuen Technologien in den diagnostischen

Prozess eingebunden werden. Aspekte, wie die vom Probanden empfundene Testfairness oder Gesichtspunkte der sozialen Validität, welche eine zunehmende Bedeutung für die Entwicklung von eignungsdiagnostischen Verfahren gewonnen haben, sollten bei diesem Prozess ebenfalls berücksichtigt werden.

Ein zweiter Aspekt betrifft die durch die hohe *Fidelity* möglicherweise resultierende hohe Komplexität des Lösungsprozesses oder die Abhängigkeit der Items und Lösungswege. Hier gilt es bei dem Einsatz von neuen Technologien abzuwägen, inwieweit die Komplexität oder Vernetztheit einzelner Aspekte in der realen Situation auch durch klare und voneinander unabhängige Itemstrukturen erreicht werden kann.

2.8 Multimediale computerbasierte Eignungsdiagnostik

Bevor mit dem folgenden Kapitel die in der Literatur genannten Vor- und Nachteile des computerbasierten multimedialen Testens beschrieben sowie die empirischen Studien und Ergebnisse zu diesem Themenbereich vorgestellt und einer kritischen Betrachtung unterzogen werden, erfolgt in Kapitel 2.8.1 eine Definition des Begriffs „Multimedia“. Weiterhin wird anhand eines kognitionspsychologischen Modells auf die Rolle der Informationsverarbeitung bezüglich multimedialer Inhalte eingegangen werden, da auch die Theorien und Modelle dieses Forschungsgebietes wichtige Implikationen für die Materialerstellung von multimedialen, videobasierten Verfahren in der Eignungsdiagnostik darstellen. Ein Aspekt, der in Kapitel 2.8.2 ausführlicher dargestellt wird. Auf eine umfassende Darstellung aller in der Literatur in diesem Zusammenhang diskutierten Modelle zur Informationsverarbeitung muss an dieser Stelle allerdings verzichtet werden, da dies zu viel Raum in Anspruch nehmen würde und nicht den Schwerpunkt der vorliegenden Arbeit ausmacht.

2.8.1 Entstehung und Definition des Begriffs „Multimedia“

Der Begriff „Multimedia“ sowie erste multimediale Anwendungen stammen ursprünglich aus dem Bereich der pädagogischen Psychologie. Diese beschäftigte sich mit der Optimierung von Lehr- und Lernmaterialien und -inhalten unter Berücksichtigung der unterschiedlichen Prozesse der Informationsverarbeitung, die bei verschiedenen medialen Darbietungsformen dieser Materialien und Inhalte ablaufen. Im Focus dieser Forschung steht bis heute die Frage, wie eine gelungene Kombination unterschiedlicher Instruktionsmedien realisiert

werden kann, welche die zugrunde liegenden Informationsverarbeitungsprozesse optimal unterstützt. Beflügelt durch die zunehmenden technologischen Möglichkeiten des Computerzeitalters in den 80er und 90er Jahren, führte dies zu Trainingsprogrammen, die eine Integration mehrerer Einzelmedien wie Text, Audioinhalten und Film ermöglichten, also multimedial gestaltet waren (ausführlich hierzu siehe z. B. Weidenmann, 1996). Multimedia, aus aktueller Sicht der Forschung, definiert sich nach Schaumburg & Issing (2004) wie folgt:

„Unter Multimedia wird die computergestützte Integration vormals getrennter Einzelmedien wie Schaubild, Buch, Film, Tonband, Video etc. über eine einzige Nutzerschnittstelle, das Computerterminal, verstanden. ... Aus wahrnehmungspsychologischer Perspektive ist eine solche Integration interessant, da die multimedial präsentierte Information den Nutzer/die Nutzerin gleichzeitig über unterschiedliche Sinnesmodalitäten erreicht. Im Moment sind dies noch vorwiegend der visuelle und der auditive Kanal ...“ (Schaumburg & Issing, 2004, S. 718).

Schaumburg und Issing (2004) nennen weiterhin drei wesentliche Kennzeichen von Multimedia: die Interaktivität, die Multimodalität und die Multicodierung. Das erste Merkmal, die Interaktivität, bezeichnet die Möglichkeit des Nutzers, Eingriffs- und Steuerungsoptionen wahrzunehmen (Haack, 2002). Die Begriffe Multimodalität und Multicodierung gehen auf Weidenmann (1996, 2002) zurück. Dieser bezeichnet die Möglichkeit, unterschiedliche Wahrnehmungskanäle mit einem alle Informationen integrierenden Medium zu erreichen, als Multimodalität der Informationsdarbietung. Von diesem Begriff grenzt Weidenmann (1996) den Aspekt der Multicodierung von Information ab. Dieser bezieht sich auf die unterschiedlichen Symbolsysteme (verbal/piktoral), die zur Darbietung von Informationen genutzt werden, und stellt, laut Weidenmann (1996), ein besonders wichtiges Kennzeichen von Multimedia dar. Entscheidend ist hierbei die Annahme, dass vor allem Informationen, die in unterschiedlichen Symbolsystemen dargeboten würden, von Rezipienten auf verschiedene Art verarbeitet würden, während die Rezeption von Informationen nur im begrenzten Maße von der Modalitätsspezifität beeinflusst sei. So weist die Informationsverarbeitung von gehörter und gelesener Sprache eine weitgehende Übereinstimmung auf, obwohl verschiedene Eingangskanäle involviert seien, während die kognitiven Prozesse bei der Text- und Bildverarbeitung sich ganz wesentlich unterscheiden, obwohl in beiden Fäl-

len visuell wahrgenommen werde (Weidenmann, 1996, S. 338). Weidenmann bezieht sich bei dieser Feststellung auf Modelle zur kognitiven Architektur, wie das *Modell des dualen Code* von Paivio (1971, 1978, 1983), zitiert nach Weidenmann (1996). Solche Modelle gehen nach Weidenmann (1996, S. 338 ff.) davon aus, dass der Rezipient über zwei in ihrer Spezialisierung verschiedene mentale Systeme verfügt ein verbales für die Verarbeitung linguistischer Informationen und ein imaginales für die Verarbeitung von Bildern. Beide Systeme würden zwar qualitativ unterschiedlich arbeiten, aber miteinander in Verbindung stehen und so die einzelnen Informationen auf höherer Ebene integrieren. Würden beide Systeme aktiviert, bedeute dies eine zweifache kognitive Repräsentation, wodurch sich die besseren Gedächtnisleistungen für solche Stimuli erklären ließen (für eine ausführliche Darstellung sei der interessierte Leser auf z. B. Ballstaedt (1990) verwiesen)..

Die Darstellungen Weidenmanns (1996) stehen im Einklang mit Theorien aus der Kognitionspsychologie bezüglich der Verarbeitung von Informationen im kognitiven Apparat. So geht das inzwischen populäre Modell des Arbeitsgedächtnisses von Baddeley (1992) davon aus, dass der menschliche „Arbeitsspeicher“, in den die erfassten Informationen gelangen, unter anderem über zwei spezialisierte Systeme verfügt. Das erste System, die artikulatorische Schleife (articulatory loop), ist nach diesem Modell für die Verarbeitungsprozesse von sprachlichen Informationen zuständig während das zweite System, ein räumlich-visueller Notizblock (visuospatial sketch pad) für die Verarbeitung von Bildern herangezogen wird. Nach den oben dargestellten Modellen und Annahmen über die kognitive Repräsentation unterschiedlich codierter Inhalte sollten multimediale Systeme eine besonders anregende Lernumgebung bieten können, da sie die verschiedenen Codierungssysteme gezielt ansprechen. Allerdings weisen einige Autoren, zusammengefasst nach Thürauf (2004), für den Fall der Wissensvermittlung mit Hilfe von bewegten Bildern über das pictorale Codierungssystem auch auf Probleme hin. So könne nach Strittmatter (1994) die Fülle der mittels Film dargebotenen Informationen zu einer Überlastung der kognitiven Kapazitäten kommen, da das Material vom Rezipienten zuerst segmentiert werden müsse, um es sinnvoll verarbeiten zu können. Gerade bei filmischem Material kann dies nach Singer und Singer (1983) zu einer Überlastung führen, da solche Elemente eine Vielfalt an Informationen und Gestaltungsformen, bedingt durch unterschiedliche Kameratechniken und Montageprinzipien, beinhalten könne. Nach Weidenmann führt dieser Umstand auch zu Forschungsproblemen, da eine systematische Variation des Materials aufgrund der vielfältigen Gestaltungsmöglichkeiten nur schwer zu realisieren ist (Weidenmann, 1996, S. 340 ff.).

Berücksichtigt man die oben genannten Aspekte, muss ein Lern- oder Trainingsprogramm, welches dem Begriff *Multimedia* gerecht werden will, nicht nur gewisse formale Voraussetzungen erfüllen. Die große Herausforderung liegt vor allem darin, eine optimale Kombination aus den anzusprechenden Wahrnehmungskanälen, verwendeten Symbolsystemen und zugelassenen Interaktionsmöglichkeiten zu realisieren. Erst hierdurch kann eine gute mentale Repräsentation der Lerninhalte und damit eine optimale Anwendung des Gelernten in späteren Realsituationen erreicht und eine systematische Variation von Materialien für Forschungszwecke realisiert werden (Weidenmann, 1996, S. 339).

2.8.2 Implikationen des Multimedia-Gedankens für die Eignungsdiagnostik

Im Folgenden soll die Idee einer optimalen mentalen Repräsentation von Inhalten aus dem Trainingskontext auf den Anwendungskontext der psychologischen Eignungsdiagnostik übertragen werden. Für den Fall eines multimedialen eignungsdiagnostischen Verfahrens sollte das Ziel in einer optimalen Gestaltung der Items, im Sinne einer gelungenen mentalen Repräsentation der zu lösenden Aufgabe bestehen. Dahinter steht die Überlegung, dass ein Proband nur dann seine optimale Leistung zeigen kann, wenn ihm nicht nur klar ist, welches Problem es zu lösen gilt, sondern er das Problem auch optimal und in seiner ganzen Komplexität erfassen kann. Erreicht werden könnte dies, parallel zu Ansätzen aus dem Trainingskontext, durch eine sorgfältige Auswahl und Kombination von unterschiedlichen Darstellungskanälen, Symbolsystemen und Interaktionsmöglichkeiten. Berücksichtigt man neben diesen Ansätzen aus der Pädagogischen Psychologie weiterhin die Forderung der eignungsdiagnostischen Anwender nach einer möglichst hohen *Fidelity* der Verfahren, wird eines deutlich. Bei der Generierung von multimedialen Items handelt es sich um einen weitaus komplexeren Prozess als dies bei der Generierung von Paper-Pencil Items der Fall ist. Der hohe Aufwand besteht vor allem darin, dass bei der Generierung multimedialer Items sowohl auditive als auch visuelle Stimuli derart gestaltet sein müssen, dass die Simulation der Realsituation, trotz der Integration von unterschiedlichen Symbolsystemen bei der Darstellung der Items, in sich schlüssig und unter *Fidelity* Gesichtspunkten so optimal wie möglich gelingt. Sich widersprechende Inhalte auf auditiver und visueller Ebene könnten beim Probanden eventuell zu einer Verunsicherung bei der Bearbeitung der Items führen, was den zusätzlichen Nutzen von multimedialen Items in Frage stellen würde. Gleiches könnte auch für die Verwendung von qualitativ schlechtem Videomaterial gelten,

welches die Realsituation nicht überzeugend widerspiegelt (vergleichend siehe Thürauf, 2004). Um eine Verbesserung von eignungsdiagnostischen Instrumenten durch den Einsatz von Multimedia zu erzielen, sollte deshalb der Aspekt der Materialgestaltung von multimedialen Verfahren, nach Meinung des Autors der vorliegenden Arbeit, genauso berücksichtigt werden wie Aspekte der Testfairness oder die Ermittlung der Gütekriterien.

Nachdem die Implikationen des Multimedia-Gedankens für die Generierung eignungsdiagnostischer Verfahren dargestellt und die Bedeutung des Produktionsprozesses von multimedialen Items dargelegt wurde, soll abschließend der Forschungsbereich der multimedialen eignungsdiagnostischen Verfahren als diejenige technologisch-mediale Entwicklungsstufe der Eignungsdiagnostik vorgestellt werden, in welche auch die Studie der vorliegenden Arbeit einzuordnen ist. Hierbei liegt der Schwerpunkt auf Untersuchungen zu multimedialen Situational Judgement Tests, da sich der Großteil der derzeit verfügbaren Literatur zu Multimedia in der Eignungsdiagnostik auf diese Art eignungsdiagnostischer Verfahren bezieht. Auch in diesem Fall beginnt die Vorstellung mit einer Nennung möglicher Vorteile aber auch Nachteile, die sich durch den Einsatz der Medien ergeben können.

2.8.3 Antizipierte Vorteile von Multimedia gegenüber herkömmlichen Verfahren

Eine ausführliche Zusammenfassung möglicher Vorteile multimedialer Verfahren gegenüber herkömmlichen Papier-Bleistift Verfahren aber auch Herausforderungen für die Forschung und Produktion auf diesem Gebiet geben Olson-Buchanan und Drasgow (2006) in einer aktuellen Übersichtsarbeit zu multimedialen Situational Judgement Tests. Die Autoren weisen in ihrem Artikel darauf hin, dass multimediale Verfahren gegenüber herkömmlichen Verfahren zunächst einmal alle Vorteile aufzuweisen hätten, die computerbasierte Anwendungen generell zu eigen wären. Dies sei vor allem eine vereinfachte und objektivere Durchführung, Auswertung und Interpretation, da die Präsentation des Verfahrens automatisiert ablaufe, die Antworten direkt gespeichert würden und die Ergebnisse unmittelbar berechnet und zur Verfügung gestellt werden könnten (siehe auch Olson-Buchanan, 2001). Neben diesen Vorzügen nennen die Autoren eine Reihe weiterer möglicher Vorteile, die von Seiten der Forschung antizipiert würden. Hiernach zeichnen sich multimediale Verfahren dadurch aus, dass sie von den Probanden als interessanter empfunden würden und eine höhere *face validity* aufwiesen, während herkömmlich Paper-Pencil Tests vor allem von

jüngeren Probanden, als Vertreter des digitalen Zeitalters, als altmodisch und nicht mehr zeitgemäß angesehen werden könnten. Ein weiterer Aspekt sei die höhere *Fidelity*, welche vor allem videobasierte Verfahren gegenüber herkömmlichen Tests aufwies. So könnten durch multimediale, videobasierte Anwendungen besonders wahrheitsgetreue und realitätsnahe Stimuli präsentiert werden, die sogar Schwankungen der Stimmlage, Gesichtsausdrücke und andere nonverbale Signale berücksichtigen könnten. Gerade diese Signale, die in realen Situationen am Arbeitsplatz wichtige Hinweise für ein adäquates Reaktionsverhalten sein können, würden so schon bei der Auswahlituation vom Probanden in die Lösungsfindung mit einbezogen werden müssen. Die Fähigkeit eines Probanden, in sozialen Situationen angemessen zu reagieren, würde so viel realistischer getestet als dies mit *low-fidelity-simulations* ohne Videounterstützung möglich sei. Fricke (1990, S. 464) spricht in diesem Zusammenhang von „true-to-life“-Situationen, die sich hervorragend für die Darstellung von sozialen Situationen eignen, wenn sie in beweglicher, lebensnaher Form als Filmsequenz dargestellt würden. Als ein weiterer Vorteil wird die reduzierte Verwendung von schriftbasiertem Stimulusmaterial beschrieben, welche sich in Richtung einer höheren Testfairness für Probanden aus Randgruppen auswirken könne. Durch die Darstellung sozialer Situationen mittels audio- und videobasierter Items würden vor allem Menschen mit Migrationshintergrund weniger stark benachteiligt, als dies bei einer schriftlichen Darstellung von Situationen und Sachverhalten der Fall wäre. Neben Aspekten wie der höheren Fairness und einer verbesserten inhaltlichen Validität würde von Seiten der Forschung auch die Möglichkeit einer standardisierten Durchführung sowie die Reliabilität multimedialer Verfahren als positiv bewertet. Trotz der medial komplexeren Darstellung von Testsituationen ließen sich die Aufgaben für alle Teilnehmer exakt gleich darstellen und im Vergleich zu Face-to-Face Verfahren könnten unerwünschte Veränderungen z. B. im Verhalten von Rollenspielern, die sich negativ auf die Zuverlässigkeit der Fähigkeitsmessung auswirken können, ausgeschlossen werden (Olson-Buchanan & Drasgow, 2006, S. 254 f.). McHenry und Schmidt (1994, S. 194) weisen darauf hin, dass videobasierte Auswahlinstrumente gegenüber *high-fidelity simulations*, wie Arbeitsproben oder Rollenspielen, eine gute Alternative darstellen könnten, da viele Unternehmen derzeit die hohen Kosten und den großen zeitlichen Aufwand für das Testleiter- oder Beobachtertraining bei solchen Verfahren, trotz deren hohen Validitätswerten, scheuen und deshalb auf einen Einsatz verzichten würden. Smiderle, Perry und Cronshaw (1994) sehen darüber hinaus die Möglichkeit, durch den ergänzenden Einsatz videobasierter Verfahren zu den bereits etablierten Verfahren den gesamten Auswahlprozess zu bereichern und zu verbessern.

2.8.4 Antizipierte Nachteile von Multimedia gegenüber herkömmlichen Verfahren

Neben den vielen Vorteilen, die der Einsatz von multimedialen Verfahren für den diagnostischen Prozess theoretisch bieten kann, werden nach Olson-Buchanan und Drasgow (2006) aber auch eine Reihe von Nachteilen bzw. Herausforderungen für Forschung und Produktion in der aktuellen Literatur angeführt. Wesentliche Nachteile von multimedialen Verfahren gegenüber herkömmlichen Tests seien, vor allem in den Anfängen der Forschung zu diesem Gebiet, die hohen Produktionskosten und die hohen Anschaffungspreise für die notwendige Hardware. So hätten zu Anfang der neunziger Jahre allein die Kosten für eine solche Hardware um die 5.000 \$ betragen. Da sich sowohl die Hardware als auch die Software in einer ständigen Weiterentwicklung befänden, müssten ebenfalls die Ausgaben für das Upgraden von Programmen oder der Wechsel zu einem neuen System berücksichtigt werden. Hinzu kämen die Kosten für die Produktion des Materials, die bei videobasierten Verfahren besonders hoch seien. Dalessio (1994) beispielsweise beziffert die Produktionskosten für Filmsequenzen zum damaligen Zeitpunkt noch auf 2.000 \$ bis 3.000 \$ pro Filmminute (Dalessio, 1994, S. 25) und Jones und DeCotiis (1986) geben die Kosten für ihr Auswahlverfahren für Angestellte im Hotelgewerbe mit 150.000 \$ an.

Während sich die oben genannten Probleme mit der Weiterentwicklung von Software- und Hardwareprodukten und dem starken Preisverfall auf diesem Sektor ein wenig relativiert haben und nach den Erfahrungen des Autors der vorliegenden Arbeit inzwischen eine gute Qualität zu vergleichsweise geringen Kosten erreichbar ist, gibt es Herausforderungen, die weiterhin existent sind und deutlich schwerwiegender erscheinen. Nach Olson-Buchanan und Drasgow (2006) existiert für videobasiertes Material das Problem, dass auch der Inhalt der eingesetzten Filmsequenzen, wie die verwendeten Formulierungen, die Frisuren und die Kleidung der Akteure oder die Büroeinrichtungen nach einigen Jahren veraltet erscheinen können. Im Gegensatz zu Paper-Pencil Verfahren sei eine inhaltliche Anpassung in diesem Fall aber deutlich aufwendiger. Weiterhin weisen die Autoren darauf hin, dass bereits bei der Produktion von multimedialen Items ein erheblich größerer Aufwand in Kauf genommen werden müsse als bei herkömmlichen Tests. Neben den für klassische SJT-Items typischen Konstruktionsschritten wie der Sammlung von kritischen Verhaltensweisen, der darauf basierenden schriftlichen Ausformulierung der späteren Items und einer Voruntersuchung mit nachfolgender Modifizierung der Items folge bei der Gestaltung multimedialer Items der aufwändige Schritt der Übertragung der schriftlichen Inhalte in die Filmsequenzen (siehe auch Drasgow, Olson-Buchanan & Moberg, 1999 sowie McHenry &

Schmitt, 1994). Nach dieser Übertragung müsse weiterhin überprüft werden, inwieweit die intendierte Darstellung inhaltlicher Aspekte durch die Audio- und Videostimuli auch tatsächlich realisiert werden konnten.

„However, an additional task would be to determine whether the video/audio stimuli accurately represent the content as originally intended. Here the test developer should examine the video to determine whether the nuances of vocal tone, inflection, and facial expression conveyed what had originally been intended” (Olson-Buchanan und Drasgow, 2006, S. 256).

Das der letztgenannte Aspekt von besonderer Bedeutung bei der Erstellung multimedialer Items ist und der Produktionsprozess eine sorgfältige Analyse der auditiven und visuellen Informationen beinhalten sollte, welche die entscheidungsrelevanten Botschaften vermitteln, wird neben den Forderungen aus der Praxis (Drasgow, Olson-Buchanan & Moberg, 1999; McHenry & Schmitt, 1994; Olson-Buchanan & Drasgow, 2006) auch durch die in Kapitel 2.7.1 der vorliegenden Arbeit genannten Theorien zur Wahrnehmung und Verarbeitung multimedialer Items unterstrichen. Olson-Buchanan und Drasgow (2006) weisen neben der Notwendigkeit einer weiteren Voruntersuchung des Materials weiterhin auf den Aspekt hin, dass eine möglicherweise notwendige Modifizierung von Items zu diesem Zeitpunkt des Produktionsprozesses sehr aufwendig und kostenintensiv sei (Olson-Buchanan & Drasgow, 2006, S. 256 f.). Sie unterstreichen damit die Bedeutung einer von Anfang an sorgfältigen Produktion der multimedialen Items.

Berücksichtigt man sowohl den Aspekt der relativ hohen Kosten für Hardwarekomponenten, Software und Materialerstellung als auch den erhöhten Aufwand bei der Produktion von multimedialen Items, so muss unter ökonomischen Gesichtspunkten die Forderung gestellt werden, dass sich durch die neuen auf Videotechnologie basierenden Verfahren sowohl unter Fairnessaspekten als auch in Bezug auf relevante Gütekriterien oder dem erreichten Grad an *Fidelity* eine deutliche Verbesserung des diagnostischen Prozesses erzielen lässt, um diesen hohen Aufwand zu rechtfertigen. Ob und inwieweit dies den bisher existierenden Verfahren gelingt, wird in den Abschnitten 2.8.5 bis 2.9.3 anhand ausgewählter Studien in diesem Bereich diskutiert. Die Auswahl berücksichtigt dabei solche Studien, die eine Beschreibung eines einzelnen videobasierten Verfahrens oder eines videobasierten Verfahrens als Bestandteil eines komplexeren Auswahlprozesses in Bezug auf dessen Gütekriterien und soziale Validität bzw. Fairness beinhalten (Dalessio, 1994; Jones & DeCo-

tiis, 1986; Kanning, 2005; Lievens & Coetsier, 2002; Lievens, Buyse & Sackett, 2005; Olson-Buchanan, Drasgow, Moberg, Mead, Keenan & Donovan, 1998; Schuler, Diemand & Moser, 1993; Shotland, Alliger & Sales, 1998; Smiderle, Perry & Cronshaw, 1994; Weekley & Jones, 1997).

Multimediale Verfahren, die als Material lediglich stehende Bilder verwenden, werden in die Darstellung nicht mit einbezogen. Der interessierte Leser sei hier z. B. auf die Arbeit von Etzel (1999) verwiesen, welche neben einer guten Darstellung des Themengebietes ein ausgesprochen aufwendig konstruiertes multimediales Verfahren dieser Kategorie vorstellt.

2.8.5 Empirische Befunde zu den Gütekriterien von videobasierten, eignungsdiagnostischen Verfahren

Seit den frühen 90er Jahren haben sich Forscher und Anwender mit der Entwicklung und empirischen Prüfung von videobasierten, eignungsdiagnostischen Verfahren auseinandergesetzt und versucht, die vielen möglichen Vorteile, die mit dem Begriff Multimedia in Verbindung gebracht werden, zu realisieren. Trotz der beschriebenen Herausforderungen, die es bei der Erstellung von videobasiertem Material zu berücksichtigen gilt, finden sich inzwischen Studien aus den unterschiedlichsten eignungsdiagnostischen Anwendungsbereichen, die sich in Bezug auf die Art ihrer Probanden und die zu messenden Konstrukte als sehr heterogen erweisen. Die Bandbreite der Verfahren reicht dabei von der Messung der Konfliktfähigkeit von Managern (Olson-Buchanan, Drasgow, Moberg, Mead, Keenan & Donovan, 1998) über die Bestimmung sozialer Kompetenzen bei der Personalauswahl von Polizeibeamten (Kanning, 2005) bis hin zur Selektion von Studienbewerbern für die medizinischen Studiengänge an Belgiens Universitäten unter Berücksichtigung der interpersonellen Fähigkeiten der Anwärter (Lievens, Buyse & Sackett, 2005). Die empirischen Ergebnisse aus diesen einzelnen Studien werden im Folgenden zusammengefasst nach Aspekten der Reliabilität, Validität und sozialen Validität bzw. Fairness vorgestellt.

Ergebnisse zur Reliabilität von videobasierten Auswahlverfahren

Ergebnisse, die sich auf die Reliabilität von videobasierten, eignungsdiagnostischen Verfahren beziehen, werden in der verfügbaren Literatur nur selten oder ohne die explizite Nennung von Koeffizienten dokumentiert. Ein Grund hierfür kann in der Tatsache begründet liegen, dass im vorliegenden Anwendungsbereich der Fokus eher auf Aspekte der Vali-

dität und der Verfahrensfairness gelegt wird. Weiterhin handelt es sich bei den meisten videobasierten Verfahren um Anwendungen, die den Probanden mit realitätsnahen und daher multidimensionalen Situationen konfrontieren und von diesem auch ein multidimensionales Antwortverhalten in Richtung einer erfolgreichen Lösung erfordern (Kanning, 2005). Diese Eigenschaft lässt keine hohen Homogenitätswerte der Skalen erwarten (Lievens & Coetsier, 2002; Smiderle, Perry & Cronshaw, 1994) und macht die Berechnung der internen Konsistenz, als Maß für die Bestimmung der Reliabilität, wenig attraktiv, da die Koeffizienten erwartungsgemäß sehr niedrig ausfallen sollten und eine Unterschätzung der wahren Reliabilität darstellen. Auch die alternativ zu bestimmende Retestreliabilität lässt sich nur schwer ermitteln, da die notwendige zweite Messung im eignungsdiagnostischen Kontext aufgrund der Einzigartigkeit videobasierter Items, fehlender Parallelversionen und der Tatsache, dass die Probanden typischerweise nur einmal zum Verfahren eingeladen werden, häufig nicht zu realisieren ist. Die wenigen, verfügbaren Angaben zur Reliabilität, werden im Folgenden zusammengefasst.

In einem für die Personalauswahl von Polizeibeamten entwickelten modularen „Test Sozialer Kompetenzen“ (TSK) berichtet Kanning (2005) für das fünfte Modul, welches auf videobasierten situativen Items basiert, zwar eine unbefriedigende innere Konsistenz aufgrund des multidimensionalen Charakters der Items ($\alpha = .45$), berichtet aber mit einer Retestreliabilität von .71 (N = 70 über 5 - 6 Wochen) ein „weitaus positiveres Ergebnis“ (Kanning, 2005, S. 143). Zu einem ähnlichen Ergebnis in Bezug auf die innere Konsistenz kommen Lievens und Coetsier (2002). Die Autoren berichten für zwei videobasierte Verfahren in einem Auswahlprozess für Studenten Reliabilitätskoeffizienten (Cronbachs alpha) von .41 und .56. Olson-Buchanan et al. (1998) kommen auf eine Interne Konsistenz von lediglich $\alpha = .32$ für ihr Conflict Resolution Skills Assessment (CRSA). Für den Metro Seattle Video Test (MSVT) zur Messung des Berufserfolges von Angestellten im Personennahverkehr berichten Smiderle, Perry und Cronshaw (1994) Koeffizienten nach Cronbachs alpha von .47 bis .65 und bewerten die Reliabilität des Verfahrens damit als nicht ausreichend für den Auswahlprozess. Auch Weekly und Jones (1997) weisen für ein von ihnen entwickeltes videobasiertes Verfahren zur Vorhersage des Berufserfolges von Servicekräften mit Reliabilitäten von .32 und .29 (Cronbachs alpha) auf wenig zufriedenstellende Werte hin.

Da die wenig zufriedenstellenden Ergebnisse zur internen Konsistenz aufgrund der Multidimensionalität der Items als sehr konservative Schätzung der Reliabilität angesehen wer-

den müssen und für die Retestrelabilität noch zu wenige Angaben in der aktuellen Literatur zu finden sind, fällt eine abschließende Bewertung der Reliabilität von multimedialen, videobasierten Items derzeit nach wie vor schwer (vergleichend siehe auch McHenry & Schmidt, 1994). Weitere, wie die von Kanning (2005) berichteten positiven Ergebnisse zur Retestrelabilität wären wünschenswert, um die Zuverlässigkeit von videobasierten Verfahren fair einschätzen zu können.

Ergebnisse zur Inhaltsvalidität von videobasierten Auswahlverfahren

Als ein wesentlicher, von Entwicklern und Anwendern antizipierter Vorteil von videobasierten Verfahren gegenüber herkömmlichen Verfahren wurde in Kapitel 2.8.3 deren höhere *Fidelity* genannt. Der Einsatz von audio- und videobasierten Items in der Testsituation soll diese in Bezug auf die späteren realen Anforderungen des Arbeitsplatzes deutlich realitätsnäher gestalten und damit nicht nur dem Probanden einen realen Einblick in sein späteres Tätigkeitsfeld ermöglichen, sondern auch eine hohe Übereinstimmung zwischen den Anforderungen in der Testsituation und denen in der späteren Praxis gewährleisten. Inwieweit der zweite Aspekt, der die Inhaltsvalidität eines Verfahrens betrifft, für videobasierte, multimediale Anwendungen zutrifft, wird im Folgenden zusammenfassend dargestellt.

Die Frage nach der inhaltlichen Validität von videobasierten SJTs kann für die meisten Studien schon aufgrund der Darstellungen zur Itemgenerierung als positiv beantwortet werden, da sowohl bei der inhaltlichen Konzeption als auch bei der Vortestung der Items auf die Mitarbeit von Experten aus den jeweiligen Praxisfeldern verwiesen wird. So finden nur solche Verhaltensweisen in den Drehbücher für die zu bewertenden Situationen Verwendung, die von der Mehrheit der Experten als relevant für die jeweilige Situation bewertet wurden und auch die inhaltliche Validität der Antwortalternativen erfährt durch Ergebnisse aus Expertenratings eine Absicherung, da bei einer wenig zufriedenstellenden Beurteilerübereinstimmung die Items oder Antworten modifiziert werden. So finden sich in der Literatur zu videobasierten eignungsdiagnostischen Verfahren kaum ergänzende Hinweise zu Aspekten der Inhaltsvalidität, da sie aufgrund der aufwändigen Itemgenerierung vorausgesetzt wird. Untersuchungen, welche über die Darstellung des Konstruktionsprozesses hinausgehen (z. B. Smiderle, Perry & Cronshaw, 1994), bestätigen aber die hohe Inhaltsvalidität videobasierter Verfahren. Die Autoren, welche die Inhaltsvalidität eines bereits existierenden Verfahrens prüften, konnten für die 65 Items, bewertet durch zwei Experten, eine Beurteilerübereinstimmung von 96.92 % berichten, was als absolut zufriedenstellend gelten kann (Smiderle, Perry & Cronshaw, 1994, S. 15).

Ergebnisse zur Konstruktvalidität von videobasierten Auswahlverfahren

Auf die Frage, welche Konstrukte sich mit den videobasierten Verfahren überhaupt erfassen lassen bzw. in der Praxis bisher erfasst wurden, zeigt die zusammenfassende Betrachtung der Forschungsliteratur zur Konstruktvalidität von videobasierten, eignungsdiagnostischen Verfahren eine sehr heterogene Befundlage. Während die in einigen Studien gemessenen Konstrukte als eher divergent zu intellektuellen Fähigkeiten und konvergent zu Konstrukten wie spezifischer sozialer Kompetenz, interpersonellen Fähigkeiten oder verwandten Skalen aus Persönlichkeitstest dargestellt werden (Kanning, 2005; Lievens & Coetsier, 2002; Lievens, Buyse & Sacket, 2005; Olson-Buchanan et al., 1998), weisen andere Studien positive Korrelationen der in videobasierten Verfahren und in z. B. Intelligenztests erbrachten Leistungen nach (Schuler, Diemand & Moser, 1993; Smiderle, Perry & Cronsaw, 1994; Weekly & Jones, 1997). Aus Gründen der Übersichtlichkeit wird an dieser Stelle auf eine erschöpfende Darstellung aller in der Literatur berichteten Koeffizienten und der zur konvergenten und diskriminanten Validierung eingesetzten Verfahren verzichtet. Stattdessen wird ein zentraler Aspekt in Bezug auf die Konstruktvalidität von videobasierten Verfahren vorgestellt, der als mögliche Ursache für die heterogene Befundlage angesehen werden kann. Hierbei handelt es sich um die bei den Items verwendete Instruktionskomponente, welche in den einzelnen Studien nach ganz unterschiedlichen Ansätzen gestaltet ist. Dieser Aspekt soll im Folgenden anhand von zwei ausgewählten Studien aus dem deutschsprachigen Forschungsraum exemplarisch veranschaulicht werden.

Bei den vorgestellten Untersuchungen handelt es sich zum einen um den Artikel von Kanning (2005) zu Ergebnissen eines Verfahrens für die Personalauswahl von Polizeibeamten für den gehobenen Dienst und zum anderen um den Artikel von Schuler, Diemand und Moser (1993), in welchem die Ergebnisse eines videobasierten eignungsdiagnostischen Instruments für die Finanzdienstleistungsbranche berichtet werden.

Der für die Personalauswahl von Polizeibeamten entwickelte modulare „Test Sozialer Kompetenzen“ konfrontiert die Bewerber im Modul 5 mit einer Auswahl von 11 videogestützten, situativen Items, welche die berufsspezifische soziale Kompetenz des Probanden ermitteln sollen. Der Proband wird mit kurzen Filmszenen konfrontiert, die den Berufsalltag eines Beamten im gehobenen Dienst simulieren, und muss im Anschluss aus drei Verhaltensalternativen diejenige auswählen, die seinem Verhalten in einer solchen Situation am stärksten entspricht. Der Fokus der Responsekomponente liegt hier also ganz klar auf dem eigenen Verhalten, welches man in der dargestellten Situation zeigen würde. Kanning

(2005) kommt in seiner Studie zu dem Ergebnis, dass die Leistungen aus dem videobasierten Verfahren zwar signifikant mit den Leistungen aus klassischen situativen Paper-Pencil Modulen des TSK korreliert sind, die sich ebenfalls auf berufsspezifische soziale Kompetenzen ($r = .40$, $p > .01$) bzw. auf berufsspezifisches Wissen ($r = .37$, $p < .01$) beziehen, sich aber keine signifikanten korrelativen Zusammenhänge zu den Modulen „Intelligenz“ ($r = .01$, ns) und den Konstrukten des Moduls „allgemeine soziale Kompetenz“ ($r = .01$ bis $.04$, ns) ergeben.

Auch das von Schuler et al. (1993) beschriebene Verfahren arbeitet mit 11 situativen, videobasierten Items, welche allerdings Beratungssituationen, Teamsituationen und Gespräche zwischen Vorgesetzten und Mitarbeitern im Finanzdienstleistungssektor nachstellen und die Konstrukte „Kundenorientierung“ und „Teamorientierung“ zu erfassen beabsichtigen. Im Gegensatz zum Ansatz von Kanning (2005) wird der Proband hier aber nicht gebeten diejenigen Verhaltensweisen auszuwählen, welche er am ehesten in der dargebotenen Situation zeigen würden, sondern erhält die Aufforderung „Beschreiben Sie das Verhalten der Beraterin!“ und „Wie könnte sich die Beraterin besser Verhalten?“ (Schuler et al. 1994, S. 6). Mit dieser Fragestellung wird die „Soziale Handlungskompetenz“ nach Aussagen der Autoren eher indirekt gemessen und der Aspekt der „sozialen Wahrnehmungskompetenz“ oder „sozialen Analysefähigkeit“ stärker betont. Aus diesem Grund gehen die Autoren in ihrer Studie davon aus, dass sich das eingesetzte videobasierte Verfahren eher konvergent zu Maßen der Intelligenz verhält und sich somit von sozial-interaktiven Verfahren, welche sich zu Intelligenzmaßen eher diskriminant zeigen, auf Konstruktebene unterscheiden sollte (Schuler et al. 1993, S. 5). Die berichteten Ergebnisse bestätigen diese Annahme. Für die Teilskala „Kundenorientierung“ und „Teamorientierung“ des videobasierten Verfahrens ergaben sich Zusammenhänge zu Kognitionsaufgaben von $r = .33$ ($p < .01$) und $r = .32$ ($p < .01$), während die sozial-interaktiven Verfahren wie Gruppendiskussion oder Kundengespräche mit Korrelationen von $r = -.04$ (ns) bis $.10$ (ns) nur sehr niedrige und nicht signifikante Zusammenhänge zum filmbasierten Verfahren aufwiesen.

Dass sich dieser eher geringfügig anmutende Unterschied in den Instruktionen für die Probanden bedeutsam auf die Konstruktvalidität des Verfahrens auswirken kann, selbst wenn die Aufgaben ansonsten inhaltlich absolut identisch gehalten werden, konnten Ployhart und Ehrhart (2003) zeigen. Die Autoren ließen ihre Probanden sechs Testphasen eines klassischen SJT durchlaufen und konfrontierten diese dabei jeweils mit drei Variationen einer *would do* Instruction und mit drei Variationen einer *should do* Instruction. Die *would do* Instruktionen verlangten alle eine Einschätzung des eigenen Verhaltens in den beschriebe-

nen Situationen, während die *should do* Instruktionen auf das in der Situation ideale Verhalten unabhängig vom eigenen Verhalten abzielte. Es zeigte sich, dass die Ergebnisse aus den beiden Instruktionsgruppen innerhalb der Gruppen deutlich höher korreliert waren als zwischen den beiden Instruktionsgruppen, trotz eines inhaltlich exakt gleichen Itemformats, und dass sich für beide Instruktionsgruppen weiterhin Unterschiede in Bezug auf die kriteriumsorientierte Validität ergaben. Ployhart und Ehrhart (2003) weisen aufgrund dieser Ergebnisse darauf hin, dass auch bei der Formulierung von Instruktionen große Sorgfalt walten müsse, da sich diese sowohl auf die kriterienorientierte als auch auf die konstruktorientierte Validität eines Verfahrens auswirken könne.

„ ... researchers and practitioners must give careful consideration to the types of SJT instructions used; failing to do so could influence criterion-related validity and cloud inferences of construct validity.“ [Ployhart & Ehrhart, 2003, S.1]

Auch McDaniel, Hartman und Grubb III (2003) bestätigen mit den Ergebnissen aus ihrer Metaanalyse über Studien zu SJTs, dass die Ergebnisse von Auswahlverfahren dieser Art in Bezug auf die konstrukt- und kriterienorientierte Validität durch die Art der Instruktion beeinflusst wird.

Anhand dieser Ergebnisse lässt sich gut verdeutlichen, dass sich auch der Anwender videobasierter Verfahren vor einem Einsatz genau darüber informieren sollte, ob sich das verwendete Itemformat, und hier vor allem die Art der Instruktion, eher für die Messung kognitiver Leistungen oder für die Messung spezifischer sozialer Kompetenzen eignet. Neben der Berücksichtigung des gewählten Kriteriums sollte die Auswahl weiterhin im Sinne einer sinnvollen Ergänzung der bereits im Prozess etablierten Verfahren unter Berücksichtigung der diesen Verfahren zugrunde liegenden Konstrukte erfolgen, da hierdurch nicht nur die Höhe der kriterienorientierten Validität des gesamten Auswahlprozesses, sondern auch die durch das videobasierte Verfahren erreichbare inkrementelle Validität beeinflusst wird (vergleichend hierzu siehe Lievens, Buyse und Sackett, 2005).

Dass sich die Kombination von videobasierten Verfahren mit herkömmlichen Verfahren, unter Beachtung der im vorherigen Abschnitt genannten Aspekte, durchaus gewinnbringend auf die Vorhersage der Leistungsfähigkeit von Probanden in späteren Realsituationen auswirkt, wird durch Studien aus den unterschiedlichsten eignungsdiagnostischen Anwen-

dungsbereichen bestätigt. So konnten Lievens, Buyse und Sackett (2005) anhand einer Längsschnittstudie zur Vorhersage des Studienerfolgs von Studierenden der Medizin in Belgien zeigen, dass videobasierte SJTs gegenüber kognitiven Fähigkeitstests eine inkrementelle Validität aufweisen, wenn die Leistung von Studierenden in Kursen, welche interpersonellen Fähigkeiten vermitteln, vorhergesagt werden sollte. Die Autoren weisen im Weiteren darauf hin, dass sich der Anteil an zusätzlich aufgeklärter Varianz mit zunehmender Studiendauer sogar erhöhte. Auch Olson-Buchanan et. al. (1998) berichten von einer Verbesserung der Vorhersage der Konfliktfähigkeit von Managern im Beruf durch den Einsatz ihres videobasierten Verfahrens. Die Autoren konnten anhand einer schrittweise durchgeführten Regressionsanalyse, in welcher im ersten Schritt die Ergebnisse aus Tests zur generellen und verbalen Intelligenz und im zweiten Schritt die Werte aus dem videobasierten SJT eingesetzt wurden, eine signifikante Erhöhung der multiplen Korrelation mit einem durch Vorgesetztenurteile erhobenen Kriterium feststellen.

Die berichteten Ergebnisse zeigen, dass der kombinierte Einsatz von videobasierten Verfahren und klassischen, bereits etablierten Instrumenten durchaus als eine Bereicherung der diagnostischen Praxis angesehen werden kann. Der Frage, inwieweit die neuen multimediale Verfahren, wenn sie für sich allein betrachtet werden, einen brauchbaren Beitrag zur Vorhersage von Leistungen in Studium oder Beruf erbringen können, wird im Folgenden mit einer Darstellung der Ergebnisse zur kriterienorientierten Validität nachgegangen.

Ergebnisse zur kriterienorientierten Validität von videobasierten Auswahlverfahren

Im eignungsdiagnostischen Kontext kann die kriterienorientierte Validität als besonders wichtig angesehen werden, da dem Anwender von Auswahlverfahren vor allem daran gelegen ist, aufgrund der Testleistungen Aussagen über die späteren Leistungen und Verhaltensweisen eines Bewerbers in der realen Bewährungssituation machen zu können. Der Nutzen eines Verfahrens und die Chancen, die diesem in der Praxis eingeräumt werden, hängen demnach in einem nicht unerheblichen Maße davon ab, wie hoch der korrelative Zusammenhang zwischen den durch das Verfahren ermittelten Testleistungen und dem anhand spezifischer Kriterien gemessenen Berufserfolg ausfällt (ausführlich siehe Walsh und Betz, 2001). Dass sich videobasierte Verfahren durchaus als praxistauglich in diesem Sinne erweisen können, zeigen die Ergebnisse zur kriterienorientierten Validität, welche im Folgenden vorgestellt werden.

So berichtet Dalessio (1994) für sein videobasiertes SJT Verfahren zur Vorhersage des Berufserfolges von Versicherungsangestellten, signifikante, positive Korrelationen zwischen dem Gesamtestscore und dem dichotomen Kriterium „Verbleib im Unternehmen ja/nein“ ($r = .30$, $p < .01$). Auch Jones und DeCotiis (1986) konnten für ihren SJT, welches für die Messung von Servicekompetenzen bei Hotelangestellten konzipiert wurde, für unterschiedliche Jobkategorien, wie Portier, Barkeeper oder Kellner, positive Zusammenhänge zwischen Verfahrensleistung und Berufsleistung von $r = .16$ bis $r = .71$ mit einem durchschnittlichen Zusammenhang von $r = .38$ ($p < .001$) aufzeigen. Alson-Buchanan et al. (1998) berichten für ihren videobasierten SJT zur Messung der Konfliktfähigkeit von Managern eine Übereinstimmung von $r = .13$ bis $.20$ zwischen Testscores und dem durch Vorgesetztenurteile ermittelten Verhalten der Manager in realen Konfliktsituationen und bewerten diesen geringen positiven Zusammenhang vor dem Hintergrund, dass es sich um den ersten Versuch einer Validierung handelt, als zufriedenstellend (Alson-Buchanan et al. 1998). Weekley und Jones (1997) berichten von zwei videobasierten SJTs, welche zur Vorhersage des Berufserfolges für Aushilfsangestellte in Supermärkten und für Teilzeitkräfte im ambulanten Pflegedienst konzipiert wurden. Die Autoren kommen für das erste Verfahren auf Koeffizienten von $r = .22$ bis $.33$ ($p < .05$) für die Übereinstimmung aus der Testleistung und dem Berufserfolg laut Vorgesetztenurteil und zeigen für das zweite Verfahren sogar Korrelationen von $r = .35$ ($p < .05$) zwischen den Testergebnissen und Berufserfolg laut Vorgesetztenurteil auf. Eher negativ fällt der Bericht von Smiderle, Perry und Cronshaw (1994) aus. Die Autoren, welche Bedienstete im öffentlichen Nahverkehr mit videobasierten Items testeten, die typische Situationen aus dem Arbeitsalltag darstellten, konnten lediglich einen Zusammenhang von $-.12$ ($p < .01$) zwischen den Testergebnissen und der im Arbeitsalltag als Kriterium ermittelten Anzahl der Kundenbeschwerden verzeichnen. Dieses Ergebnis bewerten die Autoren als nicht zufriedenstellend und stellen die derzeitige Tauglichkeit des untersuchten Verfahrens für eignungsdiagnostische Zwecke in Frage.

Fasst man die obigen Ergebnisse zusammen, so lässt sich der Einsatz von videobasierten Verfahren in der Eignungsdiagnostik unter Validitätsaspekten als grundsätzlich positiv bewerten. Die in einigen Studien ermittelten Validitätskoeffizienten zwischen $.30$ und $.40$ liegen im Vergleich zu den nach Schmidt und Hunter (1998) in einer Metaanalyse ermittelten Validitäten herkömmlicher eignungsdiagnostischer Verfahren in einem mittleren Bereich (siehe Schmidt & Hunter, 1998, S. 22). Trotz dieses recht positiven Gesamtergebnis-

ses lässt sich als kritischer Aspekt die recht große Spannweite der berichteten Koeffizienten von $r = -.12$ bis $.71$ anführen, da hierdurch der Einsatz von videobasierten Verfahren nicht generell als gewinnbringend angenommen werden kann. Die Spannweite ist weiterhin bemerkenswert, da es sich bei den beschriebenen Verfahren in allen Fällen um videobasierte SJTs handelte, die zum Großteil vergleichbare Schritte bei der Generierung der Items durchliefen und damit zumindest inhaltlich eine ähnlich valide Abbildung der Realsituation gewährleisten. Da trotz dieser Tatsache eine erhebliche Varianz in der Höhe der kriterienorientierten Validität feststellbar ist, soll ein wesentlicher Aspekt vorgestellt werden, der sich auf die Verfahrensgüte in diesem Sinne auswirkt und den es deshalb bei der Anwendung von videobasierten Verfahren zu berücksichtigen gilt.

Neben den bereits erwähnten unterschiedlichen Konstrukten, die den Verfahren zugrunde liegen, und möglichen Schwankungen in der diagnostischen Güte des Kriteriums sollte als weiterer die Verfahrensgüte beeinflussender Faktor die Art des Bewertungsschlüssel in Betracht gezogen werden, welcher zur Einschätzung der Probandenantworten herangezogen wird. Die in der Literatur angeführten Schlüssel zur Bewertung von Probandenantworten in videobasierten Verfahren, welche nach dem SJT Prinzip konzipiert sind, lassen sich nach McHenry und Schmitt (1994) in zwei Arten unterteilen, das *expert judgement* und das *empirical keying*. Diese werden im Folgenden vorgestellt.

Die Entwicklung eines Bewertungsschlüssels nach dem *expert judgement* Konzept erfolgt nach McHenry und Schmitt (1994, S. 202) in mehreren Schritten. Zuerst werden Experten aus dem betreffenden Tätigkeitsfeld gebeten, die entwickelten Antwortalternativen für jede der mittels Video darzustellenden Situationen in Bezug auf ihre Qualität in eine Rangfolge zu bringen bzw. die beste und die schlechteste Lösungsalternative zu bestimmen. Diese Expertenantworten werden dann auf den Grad ihrer Übereinstimmung hin geprüft. Danach werden die Videos gedreht und zusammen mit den von den Experten bestätigten Antwortalternativen einer weiteren Expertengruppe mit der Bitte vorgelegt, wiederum die beste und die schlechteste Antwortalternative zu bestimmen und alle Antwortmöglichkeiten in eine Rangfolge zu bringen. Stimmen diese Expertenurteile überein, wird der Punkteschlüssel für die Antworten festgelegt. Dieser unterscheidet sich danach, ob die Probanden gebeten werden, eine Rangfolge der Antwortalternativen zu bilden oder die beste und die schlechteste Alternative zu bestimmen.

Die von McHenry und Schmitt (1994, S. 203) beschriebene Methode des *empirical keying* geht einen anderen Weg. Die videobasierten Items und ihre Antwortalternativen werden

entwickelt und an einer Stichprobe aus der späteren Zielpopulation getestet, indem die Probanden gebeten werden, die beste und die schlechteste Lösung zu kennzeichnen. Im Anschluss an das Testverfahren werden die Probanden für einen Zeitraum von z. B. einem Jahr beobachtet und nach vorher festgelegten Kriterien einer erfolgreichen und eine nicht erfolgreichen Gruppe zugeordnet. Nach diesem Prozess werden die Antwortoptionen des Verfahrens positiv oder negativ gewichtet, je nachdem ob sich signifikant mehr Probanden aus der erfolgreichen Gruppe oder signifikant mehr Probanden aus der nicht erfolgreichen Gruppe für diese Option entschieden haben.

Dass bei ansonsten exakt gleichen Bedingungen schon die Art des Bewertungsschlüssels einen Einfluss auf die kriterienorientierte Validität haben kann, zeigt eine Studie von Weekley und Jones (1997) zur Vorhersage des Berufserfolges von Pflegepersonal. Der Zusammenhang zwischen der per *empirical keying* bestimmten Testleistung und dem erhobenen Kriterium für Berufserfolg fiel mit $r = .35$ ($p < .05$) deutlich höher aus als der Zusammenhang der in Anlehnung an die *expert judgement* Methode bestimmten Probandenleistung und dem Kriterium für Berufserfolg ($r = .14$, $p < .05$). Einen solchen Effekt berichten auch Olson-Buchanan et al. (1998) in Bezug auf die in ihrer Studie verwendete Bewertungsschlüssel *model-based key* und *empirical key* zur Messung der Konfliktfähigkeit von Managern in Stresssituationen. Die Korrelation der Testscores nach dem *model-based key*, welcher einem *expert judgement* entspricht, mit dem betrachteten Kriterium fiel geringer aus ($r = .14$, $p < .01$) als die Korrelation der Testscores nach dem *empirical key* ($r = .26$, $p < .01$) mit dem Kriterium.

Eine detaillierte Beschreibung unterschiedlicher Möglichkeiten des Scorings von videobasierten SJTs findet der interessierte Leser in der aktuellen und umfassenden Überblicksstudie von Bergman, Drasgow, Donovan, Henning und Juraska (2006). Die Autoren verglichen in einer empirischen Studie 11 verschiedene Bewertungsschlüssel für ein videobasiertes Verfahren zur Messung der Führungskompetenz von Mitarbeitern einer amerikanischen Universität. Neben der Tatsache, dass sich die unterschiedlichen Bewertungsschlüssel in einer großen Spannweite an verschiedenen Koeffizienten niederschlugen (-.03 bis .32), kommen Bergman et al. (2006) zu dem Ergebnis, dass sich vor allem die oben bereits erwähnte „empirical keying“ und „expert judgement“ Methoden für die Vorhersage von Berufserfolg eignen (Bergman et al., 2006, S. 229).

Die Ergebnisse zur kriterienorientierten Validität zeigen für den Anwender von videobasierten Auswahlverfahren die Notwendigkeit auf, neben einer sorgfältigen Auswahl des

Kriteriums und der Frage nach dem zugrunde liegenden Konstrukt auch ein Augenmerk auf die Art des Bewertungsschlüssels zu legen. Berücksichtigt man diese Aspekte bei der Planung und Durchführung der Verfahren, erscheint deren valider und gewinnbringender Einsatz durchaus gewährleistet. Nachdem die Ergebnisse zur Validität von videobasierten Verfahren, nach Meinung des Autors dieser Arbeit, als insgesamt zufriedenstellend angesehen werden können, soll im folgenden Absatz, mit der sozialen Validität auf einen weiteren wichtigen Aspekt von eignungsdiagnostischen Instrumenten eingegangen werden.

Ergebnisse zur sozialen Validität von videobasierten Verfahren

Die Bewertung der sozialen Validität von videobasierten Verfahren gestaltet sich derzeit noch als schwierig. Ein Problem liegt in der Tatsache begründet, dass weder in der Literatur zu videobasierten, multimedialen Verfahren noch in anderen Domänen derzeit eine einheitliche Vorgehensweise zur Bestimmung der sozialen Validität von Verfahren aus Sicht der Probanden existiert (siehe hierzu: Ryan & Ployhart, 2000). Vielmehr finden sich, neben den in Kapitel 2.3 der vorliegenden Arbeit vorgestellten Modellen, Beschreibungen zu unterschiedlichsten methodischen Herangehensweisen bei der Erhebung der Probandenperspektive, welche teilweise auf komplexen Modellen aufbauen und teilweise nur einzelne Aspekte vorhandener Modelle abfragen (Arvey, Strickland, Drauden & Martin, 1990; Bauer, Truxillo, Paronto, Weekly & Champion, 2004; Chan & Schmitt, 2004; Chan, Schmitt, DeShon, Clause & Delbridge, 1997; Kravitz, Stinson & Chavez, 1996; Lievens, DeCorte & Brysse, 2003; Macan, Avedon, Paese & Smith, 1994; Ployhart & Harold, 2004; Schmitt & Ryan, 1992; Steiner & Gilliland, 1996; Truxillo, Steiner & Gilliland, 2004; Wiechmann & Ryan, 2003). Ein allgemein anwendbarer Bewertungsmaßstab ist derzeit also nicht verfügbar.

Ein zweites Problem, speziell für den Bereich der videobasierten Eignungsdiagnostik, stellt die aktuell noch sehr begrenzte Anzahl an veröffentlichten Ergebnissen zur sozialen Validität dar. Obwohl die stetig wachsende Zahl an Studien zu diesem Thema auf ein generelles Interesse der eignungsdiagnostischen Forschung an solchen Konzepten schließen lässt, existieren in der derzeit verfügbaren Literatur zu videobasierten Verfahren kaum Beiträge, welche sich explizit auf die von Probanden erfragte Bewertung beziehen. Die wenigen verfügbaren Studien berichten allerdings über positive Resultate. So zeigten z. B. Shotland, Alliger und Sales (1998) mittels einer Befragung von 68 Versicherungsangestellten zu einem neuen videobasierten Auswahlverfahren, dass dieses Verfahren nicht nur in Bezug auf die Ästhetik, die Interaktionmöglichkeiten und die hohe Realitätsgenauigkeit der Szenen

sehr positiv bewertet wurde, sondern dass auch der in einem anschließenden Interview benötigte Zeitraum für die Darstellung des späteren Tätigkeitsfeldes deutlich kürzer ausfiel, da das Verfahren bereits einen guten Einblick in das spätere Berufsfeld bieten konnte. Dieser Aspekt stellt einen der vier Aspekte der sozialen Validität nach Schuler (1985) dar. Auch Chan und Schmitt (1997), Kanning et al. (2006) und Richmann-Hirsch et al. (2000) berichten über sehr positive Bewertungen von videobasierten Verfahren aus Sicht der Probanden. Diese werden, da es sich um vergleichende Studien handelt, in Kapitel 2.8 ausführlicher dargestellt. An dieser Stelle folgt mit dem nächsten Abschnitt eine Darstellung der empirisch ermittelten Fairness videobasierter Verfahren, welche über Leistungsunterschiede zwischen Probandengruppen erhoben wird, die sich in Bezug auf ihr Geschlecht, ihre Bildung oder ihre ethnische Abstammung unterscheiden. Ein Aspekt, dem im Zeitalter der Globalisierung, zusammenwachsender Märkte und multikultureller Konzerne eine immer größere Bedeutung zukommt.

Ergebnisse zur Fairness von videobasierten Verfahren

Die Tatsache, dass die Bestimmung der Fairness von videobasierten, eignungsdiagnostischen Verfahren unter Berücksichtigung des ethnischen Hintergrunds der Probanden in den meisten empirischen Untersuchungen aus den USA einen festen Bestandteil der Methodik darstellt, wirft ein positives Licht auf dieses Forschungs- und Anwendungsfeld. Es zeigt die Bemühung von Forschern und Anwendern, die bisherigen Probleme von kognitiven Eignungstests zur Vorhersage des Berufserfolges, welche häufig aufgrund ihres hohen schriftlichen Anteils zu einer ungerechtfertigten Benachteiligung von Minderheiten führten, durch den Einsatz videobasierter Verfahren zu lösen (siehe hierzu: Olson-Buchanan & Dragow, 2006; Olson-Buchanan et al. (1998); Chan & Schmitt, 1997). Die Annahme, dass sich vor allem interpersonelle Fähigkeiten bei Menschen aus einem fremden kulturellen Hintergrund mittels audio- und videobasiertem Material viel fairer erheben lassen als durch schriftliche Tests, welche ein hohes Leseverständnis voraussetzen, erscheint sinnvoll. Die berichteten Ergebnisse können diese Hoffnungen aber bisher nur teilweise bestätigen.

So konnten Jones und DeCotis (1986) in ihrem videobasierten Test zur Serviceorientierung von Hotelangestellten zwar keine Leistungsunterschiede zwischen weiblichen und männlichen Probanden feststellen, fanden aber Unterschiede in den Testleistungen bei einer Gruppierung nach ethnischer Herkunft. Lievens und Coetsier (2002) fanden für zwei videobasierte Tests im Rahmen der Aufnahmeprüfungen für das Medizinstudium an belgischen Universitäten einen signifikanten geschlechtsspezifischen Unterschied zugunsten der

weiblichen Probanden ($t = -4.53, p < .01$). Dieser Unterschied konnte allerdings für das Gesamtergebnis der Aufnahmeprüfung, welches sich aus einer Kombination von videobasierten und herkömmlichen Verfahren zusammensetzte, nicht mehr beobachtet werden. Olson-Buchanan et al. (1997) berichten über keine geschlechtsspezifischen Unterschiede in den Probandenleistungen von Managern, die das videobasierte *Conflict Resolution Skills Assessments* absolvierten, weisen aber darauf hin, dass sich für eine Bewertung nach dem *empirical keying* ein Vorteil für „African-American“ gegenüber „Whites“ ($t = -2.21, p < .05$) ergab (Olson-Buchanan et al. 1997, S.20). Dieser Vorteil zeigte sich nicht bei einer Bewertung mittels des *model based keying*. In ihrer Studie über Ergebnisse aus dem *Metro Seattle Video Test* zur Erfassung von interpersonellen Fähigkeiten von Servicekräften im öffentlichen Nahverkehr fanden Smiderle, Perry und Cronshaw (1994) keine Hinweise für eine Benachteiligung von Probanden, welche einer Minorität angehörten. Im Gegensatz hierzu berichten Weekley und Jones (1997), sowohl für ihr Verfahren zur Bestimmung der Berufseignung von Aushilfskräften im Einzelhandel als auch für ihr Verfahren zur Bestimmung der Berufseignung von Pflegepersonal, von signifikanten Unterschieden in den Ergebnissen von Gruppen unterschiedlicher ethnischer Abstammung, die zu Lasten der Minoritäten ausfielen.

Die sehr heterogenen Ergebnisse machen deutlich, dass der Anwender von videobasierten Auswahlverfahren nicht automatisch davon ausgehen kann, dass sich diese Art von Verfahren für alle Probandengruppen gleich fair gestaltet. Um einen möglichst hohen Grad an Fairness zu erreichen, sollten Aspekte wie die Höhe des schriftlichen Anteils der Antwortkomponente, die Art des verwendeten Bewertungsschlüssels und die Frage nach einer gelungenen Kombination von multimedialen und herkömmlichen Verfahren in Betracht gezogen werden. Trotz der derzeit noch uneinheitlichen Ergebnislage erscheinen die Bemühungen, mittels Multimedia für eine größere Fairness in Auswahlprozessen zu sorgen, als sehr ersterbenswert.

2.8.6 Zusammenfassende Bewertung von videobasierten Verfahren

Die zusammenfassende Betrachtung der empirischen Ergebnisse zu videobasierten eignungsdiagnostischen Verfahren zeigt ein uneinheitliches Bild. So lassen sich derzeit nur schwierig Aussagen darüber treffen, wie reliabel die Messungen von berufsrelevanten Eigenschaften und Kompetenzen mit dieser Art von Verfahren tatsächlich ausfallen. Weitere Studien, die statt der Ermittlung der problematischen internen Konsistenz ein Retestdesign berücksichtigen, wären hier wünschenswert, auch wenn diese Art der Reliabilitätsbestimmung für diesen spezifischen Anwendungskontext nur sehr schwer zu realisieren scheint.

Deutlich positiver fällt die Bewertung der Inhaltsvalidität aus. Sowohl die hohen Übereinstimmungswerte bei den Expertenratings der videobasierten Items als auch die positiven Urteile der Probanden, als Maß der *face validity*, zeichnen videobasierte Verfahren als besonders inhaltsvalide aus.

Die Frage, welche Art von Konstrukten sich mittels der videobasierten Verfahren am ehesten messen lassen, deutet auf zwei Gruppen hin. Einige Verfahren messen eher interpersonelle Fähigkeiten und spezifische soziale Kompetenzen oder die ihnen zugrunde liegenden Konstrukte weisen positive Zusammenhänge zu Persönlichkeitsfaktoren wie Offenheit auf. Andere Instrumente eignen sich eher dazu Konstrukte abzubilden, welche positive Übereinstimmungen mit kognitiven Leistungen zeigen. Ein beeinflussender Faktor scheint hier die Art der Probandeninstruktion zu sein. Berücksichtigt man diesen Umstand, so kann der gezielte Einsatz videobasierter Instrumente in Kombination mit herkömmlichen Verfahren auch zu einer Verbesserung des gesamten diagnostischen Prozesses im Sinne einer inkrementellen Validität führen.

Anhand der berichteten Koeffizienten zur kriterienorientierten Validität lässt sich zusammenfassend die Aussage treffen, dass sich die Ergebnisse aus videobasierten eignungsdiagnostischen Verfahren durchaus zur Vorhersage von späterem Berufserfolg eignen. Die kriterienorientierte Validität liegt, verglichen mit Koeffizienten anderer Verfahren nach Schmidt und Hunter (1998), mit Koeffizienten, die sich größtenteils zwischen .30 und .40 bewegen, auf einem mittleren Niveau. Allerdings zeigt die große Spannweite der berichteten Koeffizienten auch hier noch einigen Verbesserungsbedarf. Vor allem die Art des zur Bewertung der Probandenergebnisse verwendeten Schlüssels sollte bei der Anwendung von videobasierten Verfahren berücksichtigt werden, da er die kriterienorientierte Validität zu beeinflussen scheint.

Die wenigen Aussagen zur sozialen Validität von videobasierten Auswahlverfahren zeigen ein positives Bild. Allerdings wären weitere Studien zur Wahrnehmung solcher Verfahren durch die Probanden genauso wünschenswert wie eine Vereinheitlichung der derzeit noch sehr vielfältigen Erhebungsformen und Modelle zu diesem Konzept.

Die berichteten positiven Ergebnisse zur Fairness von videobasierten Verfahren im eignungsdiagnostischen Kontext sind sehr erfreulich. Allerdings kann auch hier noch keine verallgemeinernde Aussage getroffen werden, da nicht alle Verfahren eine Benachteiligung von Minderheiten ausschließen. Weitere Bemühungen in diese positive Richtung könnten in der Erforschung von Faktoren liegen, wie dem Anteil an schriftbasiertem Material in der Instruktions- und Responskomponente oder den verwendeten Bewertungsschlüsseln, welche die Fairness zu beeinflussen scheinen.

Die Gesamtbetrachtung zeigt, dass im Bereich der videobasierten eignungsdiagnostischen Verfahren trotz ihrer verhältnismäßig kurzen Entwicklungsgeschichte bereits einige sehr brauchbare Instrumente entwickelt werden konnten. Viele der dargestellten Verfahren können sich in Bezug auf ausgewählte Gütekriterien und Fairnessaspekte durchaus dem Vergleich mit herkömmlichen bereits etablierten Verfahren stellen. Aufgrund der hohen Entwicklungskosten und des nicht unerheblichen technischen Aufwandes erscheinen Bemühungen, weitere Verfahren dieser Art zu entwickeln und vermehrt einzusetzen, aber vor allem dann gerechtfertigt, wenn diese im Vergleich zu kostengünstigeren Paper-Pencil Versionen gleicher Verfahrensarten nicht nur gleichwertig erscheinen, sondern zu einer Verbesserung des diagnostischen Prozesses führen. Um der Frage nachzugehen, ob die videobasierten Verfahren dieser Forderung derzeit gerecht werden können, werden im folgenden Kapitel (2.9) solche Studien vorgestellt, die Ergebnisse aus einem direkten Vergleich von videobasierten und herkömmlichen SJT berichten (Chan und Schmitt, 1997; Funke und Schuler, 1998; Kanning, Grewe, Hollenber und Hadouch, 2006; Lievens und Sackett, 2006; Richmann-Hirsch, Olson-Buchanan und Drasgow, 2000). Allen Studien ist dabei gemeinsam, dass die Items der miteinander verglichenen Verfahren sich lediglich in der Form ihrer medialen Ausgestaltung unterscheiden, inhaltlich aber völlig identisch sind. Durch diese Vorgehensweise lassen sich präzisere Aussagen über mögliche Vor- oder Nachteile der unterschiedlichen medialen Darbietungsform von Items treffen als durch den globalen Vergleich der Gütekriterien von videobasierten und herkömmlichen Verfahrensformen.

2.9 Videobasierte und herkömmliche Verfahren im Vergleich

Die Darstellung der Ergebnisse aus den vergleichenden Studien erfolgt getrennt nach zwei Aspekten der medialen Variation. Im Kapitel 2.9.1 werden drei Studien vorgestellt, in welchen sich die mediale Variation lediglich auf die Stimuluskomponente der verwendeten Items bezieht, während die Responsekomponente konstant gehalten wird. Alle drei Studien weisen dabei einen unterschiedlichen Focus auf. Dieser liegt bei der ersten Studie auf dem Einfluss, den die mediale Variation auf die Vorhersagegüte der unterschiedlichen Verfahrensformen hat, bei der zweiten auf medial hervorgerufenen Unterschieden in der wahrgenommenen sozialen Validität des jeweiligen Verfahrens und bei der dritten Studie auf dem Aspekt, inwieweit sich die Verfahrensfairness in Bezug auf Gruppen verschiedener ethnischer Hintergründe in den Verfahrensvarianten unterscheidet.

Im Anschluss werden in Kapitel 2.9.2 zwei Studien vorgestellt, welche in ihrem Vergleich auch die mediale Variation der Responsekomponente berücksichtigen bzw. unterschiedliche Grade der Interaktivität der Items in die Betrachtung mit einbeziehen. Auch hier liegt der Focus zum einen auf Unterschieden in der kriterienorientierten Validität und zum anderen auf Unterschieden in Bezug auf die soziale Validität.

2.9.1 Der Vergleich von medial variierten Stimuluskomponenten

Bei der ersten vergleichenden Studie, mit dem Focus auf dem Aspekt der Verfahrensfairness, handelt es sich um die Untersuchung von Chan und Schmitt (1997). Die Autoren ließen $N = 241$ (113 schwarze und 128 weiße) Studierende im Fach Psychologie, welche per Zufall auf vier Gruppen aufgeteilt wurden, entweder einen papier- und bleistiftbasierten oder einen videobasierten Situational Judgement Test absolvieren. Beide Verfahren verwendeten inhaltlich gleiche Items und die gleiche Art von Bewertungsschlüssel. Parallel wurden Fragen zur Bewertung des Verfahrens und zur Lesefähigkeit erhoben. Die Autoren kamen zu dem Ergebnis, dass die Leistungsunterschiede zwischen Probandengruppen mit unterschiedlicher ethnischer Herkunft in der videobasierten Testvariante signifikant geringer ausfielen als in der Papier-Version ($R^2 = .04$, $df = 1$, $p < .05$). Gleichzeitig zeigte sich, dass der Zusammenhang zwischen der Lesefähigkeit und den Testergebnissen bei der papier- und bleistiftbasierten Version deutlich höher ausfiel. Auch die Bewertung des Verfahrens durch die Probanden fiel zugunsten der videobasierten Variante aus ($t(237) = 3.87$, p

< .05). In der Bewertung beider Testversionen durch schwarze und weiße Studierende ergaben sich zudem deutlich geringere Unterschiede für die Videovariante, während die PP-Version von weißen Teilnehmern positiver beurteilt wurde. Weiterhin konnte ein positiver Zusammenhang zwischen der Bewertung des Verfahrens und dem erreichten Testscore festgestellt werden.

Die zweite vergleichende Studie, die auf den Aspekt der prädiktiven Validität fokussiert, ist eine von Lievens und Sackett (2006) durchgeführte Untersuchung zu Ergebnissen aus der Aufnahmeprüfung für die medizinischen Fächer an belgischen Universitäten. Die Autoren verglichen die Ergebnisse von 1159 Studenten, die einen videobasierten SJT absolvierten, und 1750 Studenten, die eine inhaltlich gleiche, aber schriftliche Variation des videobasierten Verfahrens bearbeiteten. Weiterhin wurden Daten aus kognitiven Leistungstests sowie einer schriftlichen Arbeitsprobe erhoben. Im Anschluss an das Verfahren konnten die Probanden einen Fragebogen zur wahrgenommenen sozialen Validität ausfüllen, wobei die Bearbeitung dieses Fragebogens auf freiwilliger Basis stattfand. Als Kriterien wurden zum einen die Durchschnittsnoten der Studierenden nach einem Jahr (GPA) und zum anderen die Noten aus Kursen zur Vermittlung von interpersonellen Kompetenzen herangezogen. Die Autoren kommen zu dem Ergebnis, dass der Zusammenhang zwischen dem videobasierten SJT und dem Kriterium „interpersonelle Kompetenzen“ ($r = .34$) signifikant höher ausfiel ($p < .01$) als der Zusammenhang zwischen dem Paper-Pencil SJT und diesem Kriterium ($r = .08$). Für das Kriterium „GPA“ zeigte sich ein gegenteiliger Effekt mit einer signifikant höheren Übereinstimmung ($p < .05$) zwischen dem schriftlichen SJT und dem Kriterium ($r = .18$), als dies bei dem videobasierten Verfahren der Fall war ($r = .11$). Weiterhin berichten die Autoren in Bezug auf das Kriterium „interpersonelle Kompetenzen“ für den videobasierten SJT einen deutlich höheren Anteil an inkrementeller Validität gegenüber den übrigen Prädiktoren (Anteil an zusätzlich aufgeklärter Varianz 11%, $p < .01$) als für den schriftlichen SJT, der keinen weiteren Varianzanteil aufklären konnte. Für den Aspekt der sozialen Validität beider Verfahrensformen konnten die Autoren keinen signifikanten Unterschied entdecken ($d = -.03$). Allerdings wurden beide Arten von SJTs positiver bewertet als die restlichen Verfahren im gesamten Auswahlprozess.

Die dritte Studie (Richman-Hirsch, Olson-Buchanan & Drasgow, 2000) legt den Schwerpunkt auf die Erfassung der Wahrnehmung und Einstellung von Probanden in Bezug auf drei inhaltlich gleiche, aber medial variierte Formen des *Conflict Resolution Skills Assessment*, einem SJT zur Erfassung der Konfliktlösefähigkeit von Managern (Olson-Buchanan et al. 1998). Im Gegensatz zu den beiden anderen Untersuchungen setzten die Autoren ne-

ben einer schriftlichen und einer videobasierten Variante als dritte Variation eine computerbasierte, schriftliche Ausführung des Verfahrens ein. Als Probanden dienten 131 Manager, die den drei Gruppen randomisiert zugewiesen wurden. Im Anschluss an das CRSA wurden die Probanden mittels eines Fragebogens zur „Wahrnehmung“ gebeten, zu Aspekten wie der sozialen Validität und Fairness des Verfahrens Stellung zu nehmen. In einem weiteren Fragebogen „Einstellungen“ wurde ihnen dann die Möglichkeit gegeben, Aspekte wie die Bearbeitungsdauer oder die Modernität des Verfahrens zu bewerten. Die Ergebnisse einer multivariaten Varianzanalyse, jeweils für die Variablen aus den Fragebögen „Wahrnehmung“ ($F(2,127) = 2.74, p < .05$) und „Einstellungen“ ($F(2,127) = 3.22, p < .05$), zeigten eine deutliche Präferenz der Manager für die videobasierte Version gegenüber der schriftlichen und der rein computerbasierten Variante. Die videobasierte Version wurde als inhaltlich valider ($d = .52, p < .01$) und prädiktiv valider ($d = .34, p < .05$) angesehen und auch in Bezug auf die Möglichkeit, einen Einblick in den späteren Berufsalltag zu erhalten, als besser bewertet ($d = .29, p < .05$). Auch die wahrgenommene Fairness fiel für die Videoversion etwas höher aus als für die anderen beiden Varianten. Weiterhin berichteten die Probanden in der Videoversion eine größere Zufriedenheit bei der Bearbeitung des Verfahrens ($d = .44, p < .05$) sowie des gesamten Erhebungsprozesses ($d = .48, p < .05$) und bewerteten das Verfahren als moderner ($d = .29, p < .05$) und kürzer ($d = .60, p < .01$) als dies die Probanden in den beiden anderen Gruppen taten. Bei einem Vergleich der Bewertungen des Paper-Pencil Verfahrens und des schriftlichen, computerbasierten Verfahrens zeigten sich diese Unterschiede allerdings nicht (Richmann-Hirsch et al., 2000, S. 884).

Die Ergebnisse aus allen drei Studien zeigen, dass die videobasierten Verfahrensvarianten den herkömmlichen papier- und bleistiftbasierten SJTs in Bezug auf die gemessene Fairness, soziale Validität und kriterienorientierte Validität überlegen waren. Die Ergebnisse sind umso bemerkenswerter, da sich die medialen Variationen in allen drei Studien lediglich auf die Stimuluskomponente der Items bezogen, während die Responsekomponente für alle Bedingungen in einem einheitlichen Paper-Pencil Format vorgegeben wurden. Bereits diese Art der medialen Weiterentwicklung führt also zu einer höheren *Fidelity* von videobasierten Verfahren gegenüber herkömmlichen SJT und zu den berichteten Verbesserungen des diagnostischen Prozesses. Allerdings werden hier die Chancen, welche sich aus der medialen Variation der Responsekomponente in Bezug auf die *Fidelity* und andere Eigenschaften ergeben könnten, völlig außer Acht gelassen. Nach Schuler und Funke (1998, S. 116) erweisen sich damit die meisten videobasierten SJTs als „stimulus-centered“, da sie

eine hohe „stimulus *fidelity*“, aber eine niedrige „response fidelity“ aufzuweisen hätten. Auch Richmann-Hirsch et al. (2000) bemerken für ihre Studie folgerichtig:

„In addition, the response mode was not manipulated in any way. Examinees could only respond by choosing one of four multiple-choice responses. Consequently, even the multimedia assessment was not a high-*fidelity* simulation of a true workplace interaction.“ (Richmann-Hirsch et al., 2000, S. 885).

Schiff, Arnone und Cross (1994) weisen in diesem Zusammenhang auf die Tatsache hin, dass die Betrachtung der medialen Gestaltung der Responsekomponente von videobasierten eignungsdiagnostischen Verfahren unter *Fidelity*- und Validitätsaspekten ebenfalls von Bedeutung ist. Bestätigt werden diese Annahmen durch die Ergebnisse von zwei Studien (Kanning et al., 2006; Schuler & Funke, 1998), welche den Focus neben einer Variation der Stimuluskomponente vor allem auf die Betrachtung unterschiedlicher Responsekomponenten legen. Diese werden im Folgenden vorgestellt.

2.9.2 Der Vergleich von medial variierten Stimulus- und Responsekomponenten

Ziel der ersten Studie (Schuler & Funke, 1998) war es, den Einfluss einer medialen Variation sowohl der Stimulus- als auch der Responsekomponente auf die kriterienorientierte Validität unterschiedlicher Varianten eines SJTs zu messen. Indem sie unterschiedliche Stimulus- und Responsekomponenten kombinierten, entwickelten die Autoren insgesamt sieben Formen eines inhaltlich gleichen SJT, dessen videobasierte Ausgangsform ursprünglich in Kreditinstituten zur Messung von spezifischen sozialen Kompetenzen im Umgang mit Kunden und Kollegen eingesetzt wurde. Die Stimuluskomponente variierte in den zwei Ausprägungen *Wörtliche Beschreibung der Situation* (oral) und *Videofilm* (video). Die Responsekomponente wurde vierfach variiert und wies die Ausprägungen *Multiple-choice* (mc), *Stichwortartige Beschreibung des Vorgehens* (wcr), *freie mündliche Antwort* (ofr) und *Rollenspiel* (rp) auf. Als Kriterium dienten die gemittelten Ergebnisse aus den beiden SJT-Varianten mit den Eigenschaften „Stimulus: video – Response: rp“ sowie „Stimulus: oral – Response: rp“. Die Zusammenfassung der Ergebnisse dieser beiden Kombinationen erlaubte einerseits eine Ausbalancierung der medialen Variation des Stimulusmaterials und

wurde andererseits gewählt, da die Responsekomponente „Rollenspiel“ als die am besten geeignete Simulation von Verhalten in realen Jobsituationen angesehen wurde (Schuler et al., 1998, S. 117). Als Probanden dienten $N = 75$ Studenten einer deutschen Universität, welche alle sieben Varianten des SJT, verteilt auf drei Termine, absolvierten sowie die Subskalen „Verbal“ und „Informationsverarbeitungskapazität“ eines Intelligenztests.

In Bezug auf die kriterienorientierte Validität konnten die Autoren zeigen, dass sich lediglich für die beiden SJT-Varianten „oral-wcr“ und „video-wcr“ sowie für die Variante „oral-ofr“ signifikante Zusammenhänge ergaben ($r = .36, p < .01, r = .37, p < .01$ und $r = .59, p < .001$), nicht aber für die Variante mit der Responsekomponente „mc“. Weiterhin zeigte sich, dass sich bei konstant gehaltener Stimuluskomponente eine signifikante Verbesserung der Validität mit zunehmender *Fidelity* der Responsekomponente ergab. Ein Vergleich der unterschiedlichen Stimuluskomponenten bei konstant gehaltener Responsekomponente zeigte keinen Unterschied in Bezug auf die Validität.

Aufgrund dieser Ergebnisse kommen die Autoren zu dem Schluss, dass sich vor allem die Responsekomponente auf die kriterienorientierte Validität eines Verfahrens auswirkt. Eine Erhöhung der *Fidelity* der Responsekomponente sollte demnach mit einer Erhöhung der Validität des Verfahrens einhergehen (Schuler et al., 1998, S. 121).

Die Beantwortung der Frage, wie sich die mediale Variation der Stimulus- und Responsekomponente auf die wahrgenommene soziale Validität eines Verfahrens aus Probandensicht auswirkt, war Ziel von zwei Studien, die Kanning et al. (2006) berichten. Die Autoren weisen in ihrer Arbeit darauf hin, dass das Itemformat von SJTs durch eine mediale Variation der Stimulus- und der Responsekomponente in seiner *Fidelity* beeinflussbar sei und betrachten die Interaktivität der Items in diesem Zusammenhang als einen weiteren wichtigen Aspekt. Im Gegensatz zu Schuler (1998), der durch eine Variation der Responsekomponente von minimal interaktionistisch (mc) bis maximal interaktionistisch (rp) die Interaktivität seiner Items manipuliert, wählen Kanning et al. (2006) einen anderen Ansatz, indem sie unterschiedliche Kombinationen von aufeinander abgestimmten schriftlichen und videobasierten Sequenzen zusammenstellen. Die Autoren gehen davon aus, dass neben dem Grad der medialen *Fidelity* vor allem der Grad an Interaktionsmöglichkeiten, welchen die Itemgestaltung eines Verfahrens dem Probanden ermöglichte, sich positiv auf die Wahrnehmung dieses Verfahrens auswirken sollte. So sollten interaktive Items positiver bewertet werden als nichtinteraktive und videobasierte Items positiver als textbasierte Items. Interaktive videobasierte Items sollten in der Bewertung besser als jede andere Art von Items abschneiden (Kanning et al., 2006, S. 170).

In der ersten Untersuchung ließen die Autoren $N = 284$ Polizeibeamte, welche randomisiert auf vier Gruppen verteilt wurden, jeweils eine von vier Varianten eines 14 Items umfassenden SJT bearbeiten. Jede Variante nutzte videobasiertes Stimulusmaterial, welches entweder mit einer textbasierten Responsekomponente, einer videobasierten Responsekomponente oder einer interaktiven Responsekomponente kombiniert wurde. Alle Probanden wurden im Anschluss gebeten einen Fragebogen zur sozialen Validität auszufüllen, welcher die in Kapitel 2.3.1 vorgestellten Faktoren nach Schuler (1983) *usefulness*, *emotional reaction*, *transparency* und *job relatedness* beinhaltet.

Mittels einer 2×2 MANOVA (text/video und interaktive/noninteraktive) mit den vier Faktoren der sozialen Validität als AVs kamen die Autoren zu dem Ergebnis, dass zwar der Grad der Interaktivität einen signifikanten Einfluss auf die Bewertung des Verfahrens durch die Probanden hatte, dies aber nur für den Punkt *usefulness* der Fall war. Weiterhin zeigte sich weder ein Haupteffekt für die Modalität (Text vs. Video) noch ein Interaktionseffekt (Interaktivität \times Modalität). Die Autoren sehen zwei mögliche Erklärungen für die sehr homogene Bewertung der verschiedenen Varianten. Zum einen sei die Variationsvielfalt der unterschiedlichen Itemformate durch den Verzicht auf eine textbasierte Stimuluskomponente eingeschränkt gewesen und zum zweiten hatten die Probanden nicht die Möglichkeit eines direkten Vergleiches der dargebotenen Varianten. Was eine Frage nach dem jeweiligen Referenzsystem aufwirft, welches zur Bewertung herangezogen wurde. Beide Aspekte wurden in Rahmen einer weiteren Untersuchung berücksichtigt (Kanning et al., 2006).

In der zweiten Untersuchung absolvierten 82 Polizeibeamte im Rahmen eines within-subjekt desings sechs Variationen eines 14 Item umfassenden SJT. Neben den bereits in der ersten Studie verwendeten Varianten wurden eine interaktive und eine nichtinteraktive Form in die Untersuchung mit aufgenommen, welche beide ausschließlich textbasiert waren. Weiterhin wurden die Probanden neben der ursprünglichen Bewertung der Varianten noch um ein Ranking der Itemformate nach jetzt fünf Dimensionen der sozialen Validität gebeten.

Die Autoren berichten, dass sich für alle fünf Dimensionen der sozialen Validität signifikante Unterschiede zwischen den Verfahrensvarianten ergaben. Dabei konnte gezeigt werden, dass die videobasierte interaktive Verfahrensvariante, sowohl in der Bewertung als auch im Ranking, den anderen Varianten in mehreren Aspekten überlegen war. Die textbasierte nichtinteraktive Variante schnitt im Ranking am schlechtesten ab. Weiterhin berichten die Autoren, dass sich für diese Studie kein signifikanter Unterschied im Ranking von

interaktiven Items gegenüber nichtinteraktiven Items zeigte, sehr wohl aber ein Effekt zugunsten von videobasierten gegenüber textbasierten Items. Besonders positiv wurden dabei diejenigen Verfahrensvarianten bewertet, bei denen sowohl die Stimulus- als auch die Responsekomponenten aus filmischem Material erstellt worden waren (Kanning et al., 2006). Beide Studien zeigen, dass bei der Gestaltung von Items im eignungsdiagnostischen Kontext neben der Stimuluskomponente auch die Responsekomponente und deren Ausmaß an Interaktivität eine Rolle bei der Bewertung durch die Probanden spielt. Eine sichere Aussage darüber, wie sich die Interaktivität der Items oder deren mediale Gestaltung auf die Wahrnehmung eines Verfahrens durch die Probanden auswirkt, lässt sich aber aufgrund der unterschiedlichen Befunde aus den Studien derzeit noch nicht treffen. Besonders herauszustellen sind allerdings die durchgängig positiven Bewertungen und Rankings derjenigen Varianten, welche sowohl in der Stimulus- als auch in der Responsekomponente Videos verwendeten. Zeigt doch die hohe Akzeptanz der Probanden für diese Varianten, dass sich das Medium Film nicht nur zur Beschreibung von potentiellen Alltagssituationen aus dem Berufsleben unter Akzeptanzgesichtspunkten hervorragend eignet, sondern auch die Erhebung der Probandenantworten mittels Film auf eine hohe Akzeptanz trifft und Textvarianten überlegen zu sein scheint.

2.9.3 Zusammenfassende Bewertung der Ergebnisse

Sowohl die Ergebnisse aus der Studie von Schuler et al. (1998) als auch die Ergebnisse aus der Arbeit von Kanning et al. (2006) haben eines gezeigt: Neben dem gewählten medialen Format von Stimulus- und Responsekomponente wirkt sich auch der Grad an Interaktion, welchen der Proband erfährt, auf die Wahrnehmung und Bewertung von eignungsdiagnostischen Verfahren aus. Ein Aspekt, der bei der Konzeption und Weiterentwicklung von videobasierten Verfahren in Betracht gezogen werden sollte. Weiterhin wurde durch Kanning et al. (2006) erstmals ein Ansatz vorgestellt, bei welchem der Einsatz von Videosequenzen in der Responsekomponente konsequent verwirklicht wurde. Die positiven Ergebnisse dieses Ansatzes zeigen, dass die Möglichkeiten einer Verbesserung des eignungsdiagnostischen Prozesses durch den Einsatz videobasierter Stimuluskomponenten bei weitem noch nicht ausgereizt scheint. Schriftliche Responsekomponenten haben sich im direkten Vergleich zu mündlichen oder videobasierten Responsekomponenten, sowohl was die kriterienorientierte als auch die soziale Validität angeht, als weniger geeignet herausgestellt.

Responsekomponenten, die ein freies Agieren in Form von Rollenspielen ermöglichen, setzen den Einsatz von Beobachtern und eine anschließende Auswertung voraus und erscheinen damit den videobasierten Komponenten unter ökonomischen Aspekten unterlegen. Beide Aspekte sprechen dafür, dass der Einsatz von Videos bei der Gestaltung der Responsekomponente als eine viel versprechende Möglichkeit zur weiteren Optimierung von eignungsdiagnostischen Verfahren in Betracht gezogen werden sollte.

2.10 Implikationen der Ergebnisse zu videobasierten Situational Judgement Tests für weitere Forschungsansätze

Im folgenden Kapitel werden die wichtigsten Ergebnisse aus der empirischen Forschung zu multimedialen, videobasierten SJTs zusammengefasst und mögliche Implikationen für weitere Untersuchungen vorgestellt, die sich aus den bisherigen Ansätzen ableiten lassen.

Anhand der Ergebnisse aus Untersuchungen zu einzelnen videobasierten SJTs konnte gezeigt werden, dass sich diese, im globalen Vergleich mit anderen eignungsdiagnostischen Verfahren, bezüglich ihrer Gütekriterien auf einem mittleren Niveau bewegen, während die von Probanden wahrgenommene soziale Validität und die ermittelte Fairness als hoch angesehen werden kann. Trotz dieser recht zufriedenstellenden Ergebnisse stellt sich das Gesamtbild der ermittelten Koeffizienten und Werte derzeit aber noch als sehr heterogen dar. Ein Grund hierfür scheint die große Vielfalt an methodischen Vorgehensweisen in diesem Forschungsfeld zu sein. Tatsächlich zeigt die Betrachtung der Arbeiten insgesamt, dass sich derzeit noch keine einheitliche Methodik herausgebildet hat. Vielmehr variieren die Untersuchungen auf vielfältige Art und Weise, angefangen von unterschiedlichen Ansätzen bei der Art der Probandeninstruktion, welche sich auf die Konstruktvalidität auszuwirken scheinen, über verschiedene Arten von Bewertungsschlüsseln, die nachweislich die kriterienorientierte Validität beeinflussen, bis hin zu unterschiedlichen Ansätzen und Modellen, die zur Bestimmung der sozialen Validität eines Verfahrens herangezogen werden. Um verlässliche Aussagen über die oben genannten Charakteristika von videobasierten Verfahren machen zu können, wären weitere Untersuchungen wünschenswert, die den Focus auf die Auswirkungen unterschiedlicher Arten der Probandeninstruktion oder des verwendeten Bewertungsschlüssels legen.

Weiterhin erscheint eine Einigung in Bezug auf ein allgemein anerkanntes Modell zur sozialen Validität bzw. eine einheitliche Vorgehensweise in der Bestimmung dieses wichtigen Bewertungsmaßstabes als erstrebenswert, um hier eine bessere Vergleichbarkeit von einzelnen Befunden realisieren zu können.

Auch die Ergebnisse aus den vorgestellten vergleichenden Studien lassen ein erstes positives Fazit in Bezug auf die Vorteile von videobasierten gegenüber herkömmlichen SJTs zu. Sowohl die Ergebnisse zur kriterienorientierten Validität als auch die zur sozialen Validität und Verfahrensfairness zeigten, trotz inhaltlich exakt gleicher Items, für die videobasierten Varianten signifikant bessere Werte auf als für das konkurrierende herkömmliche Verfahren. Allerdings kann hier nur von einem ersten Eindruck die Rede sein, da die Anzahl an Studien, die einen direkten Vergleich anstreben, derzeit noch sehr begrenzt ist. Hier wären weitere Studien wünschenswert, um die ersten positiven Ergebnisse zu bestätigen.

Weiterer Entwicklungs- und Forschungsbedarf liegt in der Gestaltung der Responsekomponenten. Sowohl die vorgestellten Einzelverfahren als auch die videobasierten Verfahrensvarianten in den vergleichenden Studien verwenden zwar videobasierte Stimuluskomponenten, greifen bei den Responsekomponenten aber auf reine Paper-Pencil- oder computerbasierte Varianten eines multiple-choice Fragebogens zurück. Hierdurch wird allerdings die Möglichkeit außer Acht gelassen die *Fidelity* des Verfahrens, durch eine mediale oder verhaltensorientierte Variation der Responsekomponente weiter zu erhöhen. Die Ergebnisse aus den wenigen Studien, die einen solchen Ansatz realisiert haben, weisen aber darauf hin, dass auch die Gestaltung der Responsekomponente einen signifikanten Einfluss auf die soziale und die kriterienorientierte Validität haben kann. Es zeigte sich, dass mit einem höheren Grad an Interaktionsmöglichkeiten bzw. mit dem Einsatz von Videos in der Responsekomponente die Qualität eines Verfahrens in Bezug auf diese beiden Aspekte steigt. Weitere Studien, die bei der medialen Variation neben der Stimuluskomponente auch die Responsekomponente des Verfahrens berücksichtigen, erscheinen aufgrund dieser ersten Ergebnisse wünschenswert.

Als letzter Aspekt soll der Verfahrenstypus betrachtet werden, welcher in den empirischen Studien derzeit dominiert. Alle hier berichteten Ergebnisse beziehen sich auf die empirische Untersuchung von Situational Judgement Tests, deren *Fidelity* mit Hilfe von videobasierten Stimulus- und, in seltenen Fällen, Responsekomponenten erhöht wurde. Wie in Kapitel 2.8.4 dieser Arbeit dargestellt wurde, handelt es sich bei der Erstellung von videobasierten situativen Items um einen sehr viel kosten- und zeitintensiveren Prozess als bei der Erstellung von schriftlichen Items dieses Typus. So stellt die Durchführung von schriftli-

chen SJTs eine vergleichsweise kostengünstige Form der eignungsdiagnostischen Erhebungen dar, da sowohl die Durchführung in Gruppen als auch eine schnelle und standardisierte Auswertung möglich ist. Unter ökonomischen Gesichtspunkten stellt sich damit für diese Verfahrensart die Frage, inwieweit der deutlich höhere Aufwand, der bei der Generierung von videobasierten Items betrieben werden muss, durch die bisher erzielten Ergebnisse gerechtfertigt scheint. Ergebnisse aus Studien, die neben einer Darstellung der Kosten auch den effektiven Nutzen multimedialer Verfahren betrachten, stehen bisher noch aus.

Die positiven Ergebnisse aus Studien zur medialen Weiterentwicklung und Verbesserung von SJTs führen zu der Überlegung, inwieweit das Verbesserungspotential von Multimedia auch für andere Verfahrensarten genutzt werden kann. Weiterhin stellt sich die Frage, ob nicht der Einsatz von videobasiertem Material vor allem für solche Verfahrensarten untersucht werden sollte, welche bereits in ihrer herkömmlichen Form als sehr zeit- und kostenintensiv einzustufen sind. Der hohe Aufwand bei der Erstellung der multimedialen Variante erscheint hier eher gerechtfertigt als bei den SJTs. Beide Überlegungen sollen für den Fall des in Kapitel 2.4.2 dargestellten Rollenspiels weitergeführt werden. Hierzu werden die wichtigsten Charakteristika des Rollenspiels nochmals aufgegriffen.

Bei dem derzeit, vor allem in Kombination mit anderen Übungen, in Assessment-Centern eingesetzte Rollenspiel handelt es sich um ein Verfahren, welches im Vergleich zu SJTs als sehr zeit- und kostenintensiv eingestuft werden kann. Ein großer Vorteil des Rollenspieles liegt in seiner hohen Inhaltsvalidität und seinem hohen Grad an *Fidelity* bezogen auf die vom Probanden auszuführenden Handlungen. Nachteile liegen in der Notwendigkeit des kostenintensiven Einsatzes externer Beobachter und den möglichen Verfälschungen der Probandenergebnisse aufgrund von Beobachterfehlern, welche sich negativ auf die Gütekriterien auswirken können. In diesem Zusammenhang stellt auch die Auswahl des geeigneten Rollenspielpartners ein erfolgskritisches Element dar, da dessen Professionalität entscheidende Auswirkungen darauf hat, ob die jeweiligen Rollenspiele für die einzelnen Probanden eine vergleichbare Schwierigkeit aufweisen und damit für alle Beteiligten gleich fair sind. Weiterhin erscheint die diagnostische Situation, trotz der großen Realitätsnähe des geforderten Verhaltens, durch die Anwesenheit der Beobachter als eher künstlich. Die *Fidelity* der Umgebungsfaktoren in der diagnostischen Situation ist deshalb als eher gering zu bewerten.

Aufgrund der dargestellten Kritikpunkte stellt sich für das Rollenspiel die Frage, inwieweit auch für diese Verfahrensart durch den Einsatz von video- und computerbasiertem Material eine Verbesserung der Gütekriterien und der sozialen Validität sowie der *Fidelity* der Um-

gebungsfaktoren erreicht werden kann, wie dies im Fall der SJTs gezeigt wurde. Eine einfache Übertragung der Erkenntnisse aus den bisherigen Forschungsarbeiten zu SJTs scheint dabei nicht möglich, da es sich bei SJTs und Rollenspielen um Verfahren handelt, welche sich in der Art der Durchführung und dem Grad der *Fidelity* der jeweiligen Originalvariante stark unterscheiden (ausführlich hierzu siehe Kapitel 2.5).

Eine eigenständige Untersuchung der medialen Verbesserung und Weiterentwicklung von Rollenspielen erscheint unter zwei Gesichtspunkten sinnvoll. Unter dem Gesichtspunkt einer Verbesserung der Gütekriterien und der sozialen Validität könnte eine computerbasierte, multimediale Version der Rollenspielübung eine von Beurteilungsfehlern befreite, standardisierte Durchführung gewährleisten, die weder den Einsatz von Rollenspielern noch den von externen Beobachtern voraussetzt. Auch die *Fidelity* in Bezug auf die Abbildung von Umgebungsfaktoren der realen Situation am Arbeitsplatz könnte durch Videos möglicherweise erhöht werden. Unter ökonomischen Gesichtspunkten steht der Aufwand bei der Erstellung der multimedialen Variante in einem guten Verhältnis zum Aufwand, der in der Originalversion entsteht, da es sich beim Rollenspiel bereits in der herkömmlichen Form um ein sehr zeit- und kostenintensives Verfahren handelt.

Trotz der angeführten Aspekte, die für eine Untersuchung der Frage sprechen, inwieweit sich die mediale Variation der Rollenspielübung positiv auf deren Gütekriterien und soziale Validität auswirkt, existiert nach dem Wissen des Autors der vorliegenden Arbeit in der aktuellen Literatur derzeit keine vergleichende Studie bezüglich medial variierten Rollenspielen. Mit der vorliegenden Arbeit wird eine erste empirische Untersuchung vorgestellt, die sich mit dieser Fragestellung auseinandersetzt.

3 Herleitung der Forschungsfragen und Hypothesen

Im vorangegangenen Kapitel konnte in Bezug auf die kriterienorientierte – und soziale Validität von videobasierten SJTs gezeigt werden, dass sich die untersuchten Verfahren im globalen Vergleich zu anderen eignungsdiagnostischen Instrumenten in einem mittleren und damit zufriedenstellenden Bereich bewegten. Vergleichende Studien zwischen Paper-Pencil Verfahren und deren videobasierten Varianten zeigten zudem, dass sich letztere, auch im direkten Vergleich, durch eine höhere kriterienorientierte Validität und soziale Validität auszeichnen konnten. Die positiven Resultate wurden auf die Erhöhung der *Fidelity* der Stimulus- und Responsekomponente zurückgeführt, was zum einen durch die gezielte Kombination von videobasierten Elementen und zum anderen durch den Einsatz von verbalen oder verhaltensorientierten Responsekomponenten realisiert wurde. Vor allem bei Verfahren zur Messung von spezifischen sozialen Kompetenzen scheinen solche Erweiterungen entscheidend zu einer Verbesserung der herkömmlichen Verfahren beizutragen. Neben der empirisch nachgewiesenen Verbesserung einzelner Verfahren durch ihre mediale Weiterentwicklung wurde weiterhin gezeigt, dass sich der kombinierte Einsatz von Verfahren unterschiedlicher medialer Gestaltung positiv auf den gesamten diagnostischen Prozess auswirken kann. Diese viel versprechenden Ergebnisse führen zu der Frage, inwieweit durch den Einsatz von videobasiertem Material auch für zeitaufwendige und kostspielige Verfahren wie das Rollenspiel eine Verbesserung der kriterienorientierten – und sozialen Validität erreicht werden kann.

Aufgrund der in Kapitel 2.5 beschriebenen Charakteristika von SJTs als *Low-Fidelity-Verfahren* und Rollenspielen als *High-Fidelity-Verfahren* sowie der Zugehörigkeit dieser beiden Verfahren zu verschiedenen Kategorien von diagnostischen Instrumenten wurde bereits dargelegt, dass die Erkenntnisse aus den obigen Studien zu medial variierten SJTs nicht ohne Weiteres auf ein Verfahren wie das Rollenspiel übertragbar scheinen. Diese Überlegung soll im Folgenden näher erläutert werden.

Die Verwendung von videobasiertem Material bei der Itemgenerierung bedeutet für *Low-Fidelity-Verfahren* in den meisten Fällen eine unmittelbare Erhöhung ihrer *Fidelity*. Diese führt vor allem bei der Messung von spezifischen sozialen Fähigkeiten zu den berichteten Verbesserungen in der kriterienorientierten und sozialen Validität. Für *Low-Fidelity-Verfahren* in Form der SJTs gilt weiterhin, dass zwischen den Paper-Pencil Versionen und den videobasierten Varianten, bedingt durch die multiple-choice Antwortformate in schrift-

licher Form, kein wesentlicher Unterschied in Bezug auf die Durchführungs- und Auswertungsobjektivität und damit verbundene Störeinflüsse im diagnostischen Prozess besteht. Der Focus bei der Verbesserung und Weiterentwicklung dieser Verfahren liegt damit auf der Erhöhung der *Fidelity*.

Im Gegensatz hierzu kann für das Rollenspiel als *High-Fidelity-Verfahren*, welches ebenfalls spezifische soziale Kompetenzen misst, nur der Versuch unternommen werden, den ohnehin sehr hohen Grad an *Fidelity* durch den Einsatz von videobasiertem Material gleichwertig abzubilden. Die Abkehr von der direkten Interaktion eines Probanden mit einem potentiellen Kunden stellt in einem ersten Schritt eine Verringerung der *Fidelity* der Stimulus- und Responsekomponente dar, da das Betrachten und Bewerten von Videos nur eine Annäherung an eine tatsächliche Interaktion bedeuten kann. Durch den Einsatz von hochwertigen Videosequenzen, die an realen Arbeitsplätzen gedreht werden, lassen sich aber andererseits die Umgebungsfaktoren der Interaktionen zwischen einem Berater und einem Kunden, wie die Räumlichkeiten und Materialien, sehr viel realitätsnäher gestalten als in einer nachgestellten Rollenspielsituation, welche durch die anwesenden Beobachter und die wenig realitätsgetreuen Umgebungskomponenten verfälscht wird. Die Auswahl-situation wird damit in Bezug auf ihre *Fidelity* aufgewertet. Ein zweiter Ansatzpunkt für eine Verbesserung des Rollenspiels liegt in der Erhöhung der Durchführungs- und Auswertungsobjektivität durch den Einsatz des Computers für die Darstellung der multimedialen Inhalte. Da im Falle eines multimedialen Verfahrens auf den Einsatz der externen Beobachter und Rollenspielern verzichtet werden kann, lassen sich mögliche Fehler im diagnostischen Prozess eventuell einschränken, was sich im Umkehrschluss positiv auf die kriterienorientierte Validität auswirken sollte. Wenn die Probanden ihre Ergebnisse nicht mehr in Abhängigkeit von den Bewertungen externer Beobachter oder dem Verhalten eines Rollenspielers wahrnehmen, könnte sich dies positiv auf die soziale Validität bzw. die wahrgenommene Fairness des Verfahrens auswirken, auch wenn ihnen die direkte Möglichkeit der Interaktion vorenthalten bleibt. Im Kontrast hierzu sollte eine Paper-Pencil Version des Rollenspiels, bedingt durch ihren geringen Grad an *Fidelity* und den Verzicht auf eine computergestützte Verfahrensdurchführung und -auswertung, deutlich geringere Werte sowohl in Bezug auf die kriterienorientierte als auch in Bezug auf die soziale Validität aufweisen.

Aus den oben dargestellten Überlegungen ergeben sich für die empirische Untersuchung der vorliegenden Arbeit zwei konkrete Forschungsfragen, die zum einen auf die kriterienorientierte – und zum anderen auf die soziale Validität fokussieren.

Die Fragestellung zur kriterienorientierte Validität bezieht sich auf den unmittelbaren Vergleich der kriterienorientierten Validitäten von drei medial variierten Versionen eines Rollenspiels. Sie lautet:

„Lässt sich durch die Übertragung des *High-Fidelity*-Verfahrens „Rollenspiel“ in eine video- und computerbasierte Variante eine Verbesserung der kriterienorientierten Validität erreichen und wirkt sich die Übertragung in eine Paper-Pencil Variante negativ auf diesen Aspekt aus?“

Aufgrund der vorangegangenen Überlegungen generiert sich aus dieser Fragestellung die folgende Hypothese:

„Durch die Übertragung des *High-Fidelity*-Verfahrens „Rollenspiel“ in eine video- und computerbasierte Variante ergibt sich eine Verbesserung der kriterienorientierten Validität, während sich die Übertragung in eine Paper-Pencil Variante negativ auf diesen Aspekt auswirkt.“

In Bezug auf die soziale Validität und wahrgenommene Fairness wird für den Fall der vorliegenden Arbeit eine Forschungsfrage formuliert, die sich auf den Vergleich des mittels Fragebogen bewerteten Eindrucks der Probanden für die drei Rollenspielvariationen bezieht. Die zweite Fragestellung der vorliegenden Arbeit lautet somit:

„Lässt sich durch die Übertragung des *High-Fidelity*-Verfahrens „Rollenspiel“ in eine video- und computerbasierte Variante eine Verbesserung der sozialen Validität und der wahrgenommenen Verfahrensfairness erreichen und wirkt sich die Übertragung in eine Paper-Pencil Variante negativ auf diesen Aspekt aus?“

Die aus dieser Forschungsfrage resultierende zweite Hypothese lautet:

„Durch die Übertragung des *High-Fidelity*-Verfahrens „Rollenspiel“ in eine video- und computerbasierte Variante ergibt sich eine Verbesserung der sozialen Validität und wahrgenommenen Fairness, während sich die Übertragung in eine Paper-Pencil Variante negativ auf diesen Aspekt auswirkt.“

4 Methode

Im folgenden Kapitel soll die methodische Vorgehensweise zur empirischen Überprüfung der in Kapitel 3 vorgestellten Hypothesen dargestellt werden. Zur besseren Übersicht wird die gesamte Methodik dafür in 4 thematische Abschnitte unterteilt. Kapitel 4.1 beschreibt den Anwendungskontext, in den die methodische Vorgehensweise eingebettet ist, und stellt das in dieser Arbeit zur Validierung der Rollenspielvarianten herangezogene Kriterium vor. Die Erstellung der drei medial variierten Rollenspiel-Variationen und des Fragebogens zur Erhebung der sozialen Validität wird in Kapitel 4.2 dargelegt. Hieran schließt sich mit Kapitel 4.3 eine Darstellung der Voruntersuchungen der Materialien an. In Kapitel 4.4 erfolgt abschließend die Darstellung der Hauptuntersuchung.

4.1 Einbettung der Methode in ihren Anwendungskontext

Bevor auf die einzelnen Schritte, die zur Erstellung der in dieser Arbeit eingesetzten Materialien notwendig waren, eingegangen wird, sollen für den Leser solche Aspekte ausführlicher erläutert werden, die einen Einfluss auf die Materialgestaltung hatten. Zu diesem Zweck wird zuerst das Verfahren beschrieben, in welches die 3 medialen Variationen des Rollenspiels implementiert wurden. Im Anschluss daran wird der Anwendungskontext dargestellt, in welchen das Verfahren eingebettet ist. Weiterhin sollen das zur Validierung herangezogene Außenkriterium sowie sein fachlicher Hintergrund in diesem Kapitel vorgestellt werden.

4.1.1 Darstellung des Verfahrens

Auf der Suche nach einem Kooperationspartner, der bereit und in der Lage war ein geeignetes Verfahren für die empirische Untersuchung der in dieser Arbeit erhobenen Fragestellungen zur Verfügung zu stellen, konnte die Akademie eines Finanzdienstleistungsverbandes gewonnen werden. Die Weiterbildungseinrichtung betreut ca. 70 Institute in den Bereichen Training, Unternehmenssteuerung, Ausbildung und Personalentwicklung und verfügt über eine jahrzehntelange Erfahrung bei der Durchführung von personaldiagnostischen Auswahlverfahren in den betreuten Unternehmen.

Als methodischer Rahmen für die empirische Untersuchung der in Kapitel 3 formulierten Hypothesen diente das *Personalentwicklungsseminar zur Qualifizierung für den Kundenberaterlehrgang* (PES) des oben genannten Institutes. Dieses PES durchlaufen alle Auszubildenden der betreuten Institute in einem Zeitraum von einigen Wochen vor bis einige Wochen nach der Abschlussprüfung zum Bankkaufmann bei der Industrie und Handelskammer (IHK). Der Focus des PES liegt dabei schwerpunktmäßig auf der Ermittlung der *Kundenorientierung* der Teilnehmer, die als spezifische soziale Fähigkeit verstanden werden kann und die Kompetenz im Umgang mit dem Kunden vom Produktverkauf bis zum Beschwerdemanagement umfasst. Aus der Bewertung dieser Kompetenz resultiert dann die Einschätzung, inwieweit der Kandidat für die Betreuung des gehobenen Kundensegmentes geeignet erscheint. Die erbrachte Leistung in diesem Verfahren stellt damit ein wesentliches Kriterium bei der Entscheidung dar, ob der Kandidat eine hochwertige und kostenintensive Ausbildung zur Qualifizierung für eine gehobene Beratertätigkeit durchlaufen soll oder nicht.

Das eintägige Verfahren läuft in Form eines klassischen Assessment-Centers ab. Die Jungangestellten bekommen hier die Gelegenheit ihre Fähigkeiten in Kooperationsübungen, Rollenspielen in Form von Kundengesprächen (KG) und Mitarbeitergesprächen (MG), Vorträgen und Postkorbübungen zu zeigen. Hierbei werden sie sowohl von erfahrenen Kundenberatern und Personalverantwortlichen aus dem eigenen Hause als auch von externen Trainern der Akademie beobachtet, die ihre Einschätzung bezüglich der Qualität des gezeigten Verhaltens anhand von standardisierten Beobachtungsbögen abgeben. Diese Einschätzungen, die mittels einer fünfstufigen Skala auf vorher festgelegten Dimensionen erfolgen, werden dann am Ende des gesamten PES, für jeden der Kandidaten, in einer Beobachterkonferenz abgeglichen und zu einer Beurteilung zusammengefasst. Den Abschluss des Verfahrens bilden dann Feedbackgespräche, die von Trainern der Akademie mit den Kandidaten geführt werden, um Weiterentwicklungspotentiale aufzuzeigen und den weiteren beruflichen Werdegang der Teilnehmer zu diskutieren. Der typische Ablauf eines solchen PES Verfahrens ist in Abbildung 4.1 graphisch veranschaulicht.



Abbildung 4.1: Graphische Darstellung des Ablaufes eines typischen PES

Da das oben beschriebene *PES* vor allem dazu dient, diejenigen Kandidaten zu identifizieren, denen ein hohes Potential für die spätere Kundenbetreuung zugeschrieben werden kann, liegt der Schwerpunkt, bezogen auf die eingesetzten Übungen, ganz deutlich bei den Rollenspielen. Typische Übungen wie die Postkorbaufgabe oder der Kurzvortrag werden eher ergänzend eingesetzt, um bei Kandidaten, die für die Kundenberatung weniger gut geeignet erscheinen, zusätzlich eine Aussage über Fähigkeiten machen zu können, die eher im Backofficebereich einer Bank benötigt werden.

Bei den durchgeführten Rollenspielen wiederum liegt der inhaltliche Schwerpunkt auf der Kundenberatung, so dass in den meisten Fällen die Aufgabenstellung einer typischen Beratungssituation zwischen einem Berater und einem Kunden nachempfunden ist. Die Beobachter sind bei ihrer Bewertung weiterhin angehalten, mögliche sachliche Fehler, wie die Nennung einer falschen Kondition für ein Darlehen oder einen zu hohen Zinssatz für Einlagen, nicht in die Bewertung einzubeziehen, da der Focus der Einschätzung auf der sozia-

len und nicht auf der sachlichen Komponente liegen soll. Eine solche als *Kundenberatung* oder *Kundengespräch* (KG) bezeichnete Übung aus dem oben beschriebenen Verfahren stellt die in den Hypothesen des vorangegangenen Kapitels beschriebene Face-to-Face-Variante eines Rollenspiels dar, welche als Ausgangsvariante für die Erstellung der beiden anderen medialen Variationen diente, auf welche im Kapitel 4.2 ausführlich eingegangen wird.

Durch die Einbindung aller 3 medialen Variationen in das *PES* Verfahren der Akademie konnte die Betrachtung der Auswirkungen einer medialen Variation der Rollenspielübung auf die jeweils resultierende kriterienorientierte Validität sowie die abschließende Bewertung der Varianten durch die Teilnehmer einerseits in einer realen Auswahl-situation und andererseits so kontrolliert wie möglich durchgeführt werden.

Nachdem das Verfahren, welches als methodischer Rahmen für die empirische Erhebung der vorliegenden Arbeit diente, vorgestellt wurde, soll im Folgenden das Kriterium dargestellt werden, welches zur Bestimmung der kriterienorientierten Validität herangezogen wurde.

4.1.2 Darstellung des Kriteriums

Bei dem in der vorliegenden Studie verwendeten Kriterium handelt es sich um die mündliche Note der Abschlussprüfung zum Bankkaufmann vor der Industrie und Handelskammer des zuständigen Bundeslandes (IHK). Das Bestehen dieser Prüfung stellt für den Auszubildenden den offiziellen Abschluss seiner Lehre dar und die in den Prüfungen erbrachten Leistungen werden in einer Urkunde bescheinigt, die mit einem Gesellenbrief in den handwerklichen Ausbildungsberufen vergleichbar ist. Die Überprüfung der erworbenen Fähigkeiten erfolgt mittels einer Vielzahl schriftlicher Tests in Form von Aufsätzen und Multiple-Choice-Fragebögen und einer mündlichen Abschlussprüfung, deren Note im Rahmen der vorliegenden Studie als Kriterium verwendet wurde. Der Ablauf der mündlichen Prüfung und die Ermittlung der Note werden im Folgenden ausführlicher vorgestellt.

Im Rahmen der mündlichen Prüfung bekommt der Auszubildende die Aufgabe einen Kunden, gespielt von einem der anwesenden Prüfer, in Bezug auf eine vorher festgelegte Fragestellung zu beraten. Die Dauer des Beratungsgesprächs ist hierbei auf 20 Minuten begrenzt. Die Leistung des Probanden wird von einem mindestens sechsköpfigen Beurteiler-

team, bestehend aus zwei Arbeitnehmervertretern, zwei Arbeitgebervertretern und zwei Berufsschullehrern bewertet. Dieses Team wird durch den Prüfungsvorsitzenden ergänzt. Bewertet werden zum einen die Verhaltenskompetenz des Probanden, die sich aus den Aspekten Kontaktverhalten, Informations- und Analyseverhalten sowie Verkaufsverhalten zusammensetzt, und zum anderen die Fachkompetenz. Beide Kompetenzbereiche werden anhand einer fünfstufigen Skala eingeschätzt und gehen mit einer Gewichtung von 60% für die Verhaltenskompetenz und 40% für die Fachkompetenz in die mündliche Abschlussnote ein. Die Abschlussnote, die sich aus der Einschätzung aller sechs Prüfer zusammensetzt, wird als Prozentzahl von 0 bis 100 angegeben, wobei für das Bestehen der Prüfung eine Mindestleistung von 50% erbracht werden muss.

Die mündliche Abschlussnote wurde als geeignetes Kriterium angesehen, weil der gemittelte Wert von sechs unabhängigen Expertenratings aus Sicht des Autors der vorliegenden Arbeit als ein objektives Maß der Verhaltens- und Leistungsbeurteilung angesehen werden kann. Lediglich die direkte Erhebung der verkäuferischen Kompetenz anhand von Verkaufszahlen des numerisch ermittelten Grads der Zielerreichung von vorher vereinbarten Verkaufszielen oder die empirische Erhebung der Kundenzufriedenheit mit dem Berater können als verlässlichere Kriterien angesehen werden. Da diese Informationen aber auf der Stufe der Auszubildenden seitens der beteiligten Unternehmen nicht erhoben werden bzw. nicht zur Verfügung gestellt werden konnten, kann die Note der mündlichen Abschlussprüfung vor der IHK als das beste verfügbare Kriterium angesehen werden.

Im Anschluss an die Darstellung des Kriteriums und die Beschreibung des eignungsdiagnostischen Verfahrens, welches als methodischer Rahmen für die vorliegende Untersuchung herangezogen wurde, werden in Kapitel 4.2 die einzelnen für die Erstellung der drei medial variierten Rollenspielvarianten notwendigen Schritte ausführlich vorgestellt.

4.2 Vorgehensweise der Materialerstellung

Die Vorgehensweise bei der Erstellung der eingesetzten Materialien orientierte sich an dem Ziel, die Übertragbarkeit der positiven Ergebnisse aus dem Forschungsgebiet zu medial variierten Situational Judgement Tests auf das Rollenspiel als ein *High-Fidelity-Verfahren* der Personaldiagnostik zu überprüfen. Analog zur Vorgehensweise der empirischen Studien zu SJTs wurde als Basis für die Erstellung der medial variierten Versionen zuerst eine im Kontext dieser Arbeit optimale Face-To-Face Variante der Rollenspielübung ausgewählt. Ausgehend von dieser herkömmlichen Variante wurden dann die multimediale, videobasierte Variante und die Paper-Pencil Variante entwickelt.

4.2.1 Auswahl der Face-to-Face Variante

Bei der Auswahl der Face-to-Face Variante galt es zwei Aspekte zu berücksichtigen. Die Übung musste zum einen den Standards der Akademie genügen und sich zum anderen gut in die beiden anderen medialen Darbietungsformen übertragen lassen. Um diesen Punkten gerecht zu werden, wurde ein Rollenspiel, in Kooperation mit den Verantwortlichen der Akademie, aus deren Aufgabenpool ausgewählt, welches den Probanden die Möglichkeit gab ihr Wissen um Kundenberatung sehr umfassend und flexibel einzubringen, und das auf Grund seiner offenen Aufgabenstellung gut geeignet erschien die beabsichtigte mediale Übertragung in eine multimediale, videobasierte Variante sowie eine Paper-Pencil Variante zu realisieren.

Das ausgewählte Rollenspiel sah vor, dass der Proband, in der Rolle eines Kundenberaters, auf einen Kunden, gespielt von einem trainierten Rollenspieler, trifft, der seine Ausbildung beendet hat und nun einen neuen Job im Geschäftsgebiet des Beratungsinstitutes beginnt. Eine solche Aufgabenstellung, bei der ein Neukunde um Informationen bittet, bietet dem Probanden sehr viele unterschiedliche Ansatzpunkte für die Gesprächsgestaltung und damit die Gelegenheit seine Kompetenzen im Umgang mit dem Kunden deutlich darzustellen. Im Gegensatz zu sehr spezifischen Aufgabenstellungen wie Konfliktgesprächen oder reinen Verhandlung über Konditionen mit einem langjährigen Kunden wird den Probanden in dieser Art der Interaktion mit dem Kunden die Gelegenheit gegeben eine Fülle von Informationen bezüglich seiner Person, der Institution oder den Produkten anzubieten und andererseits viele Informationen über seinen Gesprächspartner zu erheben. Durch diese

komplexe Situation lassen sich zum einen die Kompetenzen im Umgang mit dem Kunden auf vielfältige Weise beobachten und zum anderen wird die *Fidelity* der Übung gegenüber einer sehr spezifischen Aufgabenstellung erhöht, die dem Probanden nur sehr wenig Handlungsspielraum lässt.

Als Information für die Vorbereitungsphase der Probanden wurden, unter dem Gesichtspunkt einer größtmöglichen Flexibilität, nur die Fakten unbekannter Kunde, beendete Ausbildung und neue Arbeitsstelle vorgesehen. Die Vorbereitungszeit wurde nach Maßgabe der Akademie auf 15 Minuten festgelegt. Auch für die Dauer des Rollenspiels wurde eine Zeit von einer viertel Stunde vereinbart. Die Bewertung der Probanden erfolgte auf mehreren von der Akademie festgelegten Dimensionen. Die Vorgabe der beobachteten Dimensionen und der Beobachtungsbögen von Seiten der Akademie wurde unverändert übernommen, da eine Gleichbehandlung der Probanden mit Probanden früherer Jahrgänge gewährleistet bleiben musste. Die resultierende Instruktion für die Probanden findet der interessierte Leser in Anhang (AI).

4.2.2 Übertragung des Rollenspiels in eine videobasierte Variante

Mit den Kapiteln 2.8.2 und 2.8.4 wurde aufgezeigt, dass es sich bei der Generierung von videobasierten, eignungsdiagnostischen Verfahren um einen komplexen und aufwändigen Prozess handelt. Dies galt auch für die Erstellung der videobasierten Rollenspielvariante, deren gesamter Konstruktionsprozess aus 6 Schritten bestand.

Weiterhin wurden in den Kapitel 2.7.2 bis 2.7.4 aufbauend auf Ergebnissen aus verschiedenen Stufen der medialen Entwicklung von diagnostischen Instrumenten einige Implikationen für die Neu- und Weiterentwicklung von diagnostischen Verfahren durch den Einsatz neuer Medien dargestellt. Diese Implikationen werden aufgegriffen und die aus ihnen resultierende konkrete Vorgehensweise bei der Gestaltung der videobasierten Rollenspielvariante für den jeweiligen Konstruktionsschritt erläutert.

Zu Beginn werden Überlegungen zur konkreten medialen Form der videobasierten Rollenspielvariante dargestellt, die sich aus den in Kapitel 2.9.2 berichteten Ergebnissen ableiten.

Implikationen bisheriger Ergebnisse aus der Forschung für die Konstruktion

Anhand der in Kapitel 2.9.2 vorgestellten Studien (Kanning et al, 2006; Schuler und Funke, 1998) wurde gezeigt, dass die Wahrnehmung und damit die Bewertung von Situational

Judgement Tests seitens der Probanden umso positiver ausfiel, je höher der videobasierte Anteil bzw. der Grad an *Fidelity* in der Item- und Responsekomponente der jeweiligen SJTs gewählt worden waren. Neben diesen empirischen Befunden, die für den Einsatz von videobasierten Stimulus- und Responsekomponenten sprechen, führte die Tatsache, dass auch in der Face-to-Face Variante und der Paper-Pencil Variante der vorliegenden Arbeit die mediale Basis der Responsekomponente derjenigen der Stimuluskomponente entsprach, zu der Überlegung das Rollenspiel in eine zu hundert Prozent videobasierte Variante zu übersetzen.

Da die Rollenspielübung im Gegensatz zu den bisher in der Forschungsliteratur dargestellten SJTs nicht aus einzelnen Fragen bzw. Situationen besteht, die es zu beantworten oder zu bewerten gilt, sondern der Proband eine Aufgabenstellung erhält, die er dann in Interaktion mit einem Rollenspieler lösen muss, bedurfte es, bei der Übertragung dieser Übung in eine videobasierte Variante, eines neuen konzeptionellen Ansatzes.

Das Bestreben, in der Responsekomponente des videobasierten Verfahrens ein ähnlich hohe *Fidelity* zu gewährleisten wie im Face-to-Face Verfahren, führte zu der Überlegung real gestaltete Kundengespräche filmisch abzubilden, die sich in der Qualität der Beratungsleistung des Beraters auf vielfältige Weise unterschieden, und diese in kleinere Sequenzen zu unterteilen. Im Gegensatz zu der einfachen Bewertung einzelner voneinander unabhängiger Situationen, wie dies bei SJTs der Fall ist, sollten die Probanden die Möglichkeit haben aus den qualitativ variierten Sequenzen diejenigen auszuwählen, welche das dem eigenen Verhalten in einem Kundengespräch am nächsten kommende Beraterverhalten zeigten. Auf diese Weise sollte ein komplettes virtuelles Beratungsgespräch absolviert werden. Allerdings wurde aufgrund der in Kapitel 2.7.4 dargestellten Probleme, die sich beim Einsatz von vernetzten, abhängigen Itemformaten ergeben können, eine feste Struktur für die Itembearbeitung für sinnvoll erachtet. Hierdurch sollten für alle Probanden absolut vergleichbare Bedingungen bei der Bearbeitung geschaffen werden.

Die Zielsetzung ein komplettes virtuelles Beratungsgespräch zu konstruieren setzt aber voraus, dass eine genügende Anzahl an für ein Beratungsgespräch relevanten Verhaltensweisen zur Verfügung steht, die in filmisches Material umgesetzt werden können. Während bei Untersuchungen im Kontext der Situational Judgement Tests die Beschreibungen der jeweiligen zu bewertenden Situationen als Grundlage für die Drehbücher dienen können, gab es für die Erstellung von Drehbüchern für die geplanten Gesprächssequenzen keine derartige Ausgangsbasis. Die Analyse der in den Beobachtungsbögen der Akademie vorhandenen Verhaltenshinweise erschien als Grundstock für das Verfassen von detail-

lierten Drehbüchern als nicht ausreichend. Aus diesem Grund wurden Experteninterviews durchgeführt. Ziel dieser Interviews war es eine möglichst erschöpfende Sammlung an Verhaltensweisen zu erhalten, die für das Führen eines erfolgreichen Kundengesprächs als relevant anzusehen sind. Diese Interviews stellten den ersten der 6 Schritte bei der Erstellung der videobasierten Rollenspielvariante dar.

Schritt 1: Experteninterviews

In Kapitel 2.7.4 wurde dargelegt, dass die sorgfältige Analyse der realen Situation und ihrer Anforderungen an einen Probanden der Gestaltung diagnostischer Verfahren mittels neuer Technologien vorausgehen muss. Um die typischen Verhaltensweisen eines realen Beraters in der Interaktion mit einem realen Kunden in den Videos so authentisch wie möglich abbilden zu können, wurden in Anlehnung an die *critical incident technique* (CIT) nach Flanagan (1954) 10 Experten⁴ der Finanzdienstleistungsorganisation zu dem ihrer Meinung nach erfolgsrelevanten Verhaltensweisen in einem Kundengespräch befragt. Die CIT wurde gewählt, da sie als ein zentrales und häufig angewandtes Instrument zur Sammlung von Verhaltenshinweisen in spezifischen beruflichen Situationen im Rahmen einer Anforderungsanalyse angesehen werden kann (McHenry und Schmidt, 1994).

Als erster Schritt wurde ein übergeordnetes Ziel für die Gesamthandlung *Kundengespräch* definiert. Nach dieser Zieldefinition muss es dem Berater durch sein Verhalten gelingen im Verlauf des Kundengesprächs den Kunden von seiner Person und der Institution in Bezug auf Aspekte wie Servicequalität, Beratungskompetenz und angebotene Produkte zu überzeugen und so ein Vertrauensverhältnis aufzubauen.

In einem zweiten Schritt wurde, neben der Festlegung des Expertisekriteriums der zu interviewenden Berater, auch die Art der Datenerhebung bestimmt. Da eine Beobachtung von Beratungsgesprächen mit echten Kunden durch externe, geschulte Beobachter aus rechtlichen Gründen nur sehr schwierig zu realisieren gewesen wäre, fiel die Entscheidung zugunsten einer retrospektiven Datenerhebung mittels Interview aus. Die nach der CIT relevanten Erfolgsfaktoren konnten als gegeben angesehen werden, da die Berater täglich mehrere Beratungsgespräche absolvieren, das übergeordnete Ziel einer langen Kundenbindung verinnerlicht haben und schon um des eigenen Erfolges willen eine genaue und kritische Betrachtung der eigenen Verhaltensweisen in Kundengesprächen durchführen.

⁴ Unter Experten werden im Kontext dieser Arbeit Kundenberater verstanden, die eine abgeschlossene Ausbildung zum Bankkaufmann aufweisen können, den Kundenberaterlehrgang der Bankakademie mit Erfolg absolviert haben und mindestens fünf Jahre praktische Erfahrung in der Beratung von Kunden nachweisen konnten.

Im dritten Schritt, der eigentlichen Erhebung, wurden die Berater in einem 45 minütigen, teilstandardisierten Interview gebeten, zuerst die ihrer Meinung nach wesentlichen Abschnitte im Verlauf eines Kundengesprächs mit den Spezifikationen *unbekannter Kunde*, *beendete Ausbildung* und *Beginn einer neuen Arbeitsstelle* zu benennen. Nach dieser Einteilung sollten sie dann Beispiele dafür geben, welche Verhaltensweisen eines Kundenberaters in den jeweiligen Abschnitten ein gutes und welche ein schlechtes Beratungsgespräch kennzeichnen. Weiterhin wurden sie gebeten die Relevanz dieser Verhaltensweisen für das übergeordnete Ziel einer langjährigen Kundenbindung zu bewerten. Ergänzend zu den Hinweisen bezüglich eines optimalen bzw. negativen Verhaltens in der Beratungssituation wurden die Berater auch zu wichtigen Umgebungsfaktoren wie Räumlichkeiten, Beratungsmaterial und Büroausstattung befragt.

Durch die in Schritt drei geäußerte Bitte an die Experten, die ihrer Meinung nach typischen Abschnitte für ein Beratungsgespräch mit einem Neukunden zu nennen, wurde die Kategorisierung der gesammelten Verhaltensweisen im folgenden Schritt sehr vereinfacht. Da 80% der Befragten in ihrer Einteilung die gleichen in Abbildung 4.2 dargestellten fünf Abschnitte als zentral erachteten und in einem zweiten Schritt die für diese fünf Abschnitte relevanten positiven wie auch negativen Verhaltensweisen nannten, wurde diese Einteilung als Grundlage für die Kategorisierung aller genannten Verhaltensweisen übernommen.

Die Nennungen der Experten konnten nun für jede der fünf Abschnitte miteinander abgeglichen und zusammengefasst werden. Eine weitere Kategorie bildeten diejenigen Verhaltensweisen, die von den Beratern als erfolgsrelevant, aber phasenunspezifisch bezeichnet wurden. Solche Verhaltensweisen bezogen sich auf die Körperhaltung, die Sprache sowie auf Mimik und Gestik. Das Endresultat bildete eine ausführliche und für den jeweiligen Abschnitt eines Beratungsgesprächs spezifizierte Auflistung von erfolgskritischen Verhaltensweisen.

Unter Berücksichtigung der Tatsache, dass sehr ähnliche Verhaltenshinweise für eine Gesprächsphase aus den 10 einzelnen Interviews jeweils nur einmal aufgenommen wurden, ergaben sich für die Phase *Begrüßung am Schalter und Überleitung in ein Beratungszimmer* 21, für die Phase *Kennenlernen des Kunden* 20, für die Phase *Bedarfsermittlung* 25, für die Phase *Beratung, Verkauf und Crossselling* 34 und für die Phase *Verabschiedung des Kunden & Terminvereinbarung* 12 erfolgskritische Verhaltensweisen.

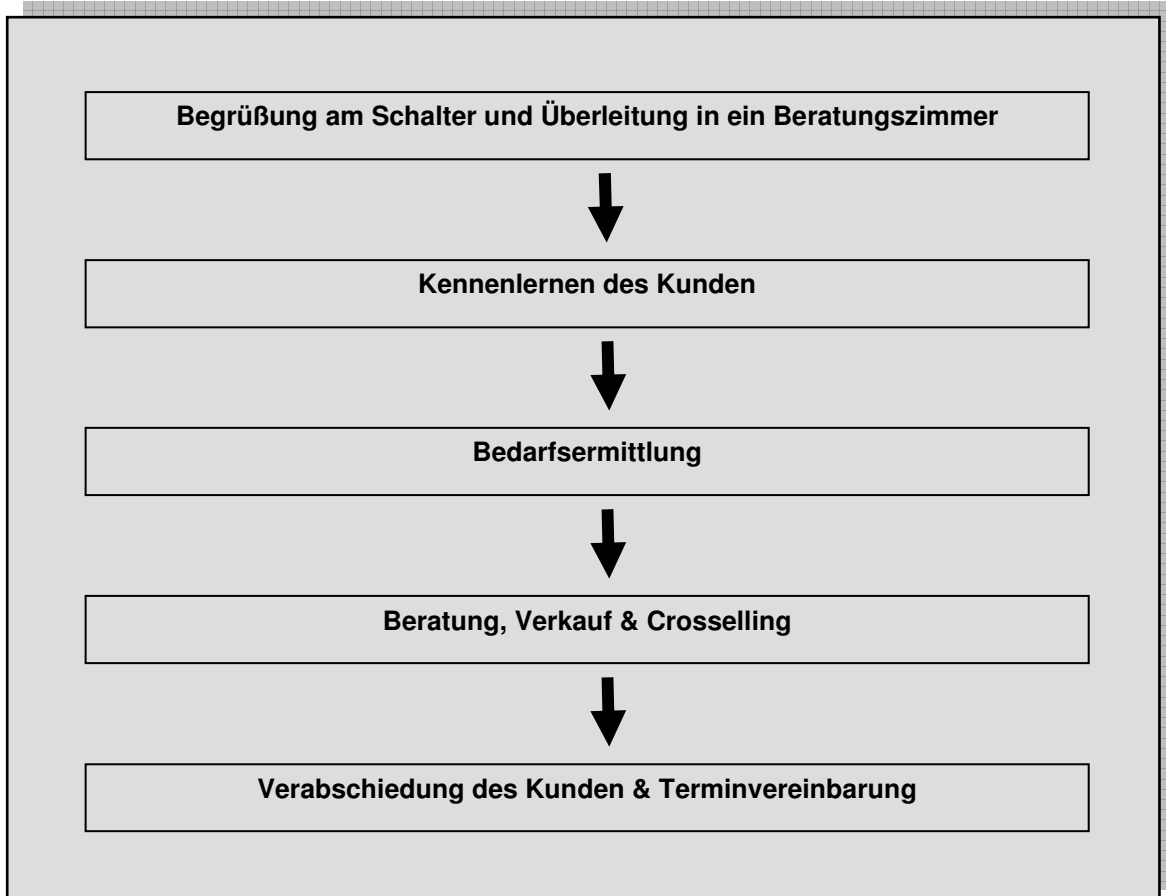


Abbildung 4.2: Die fünf zentralen Phasen eines Kundengesprächs

Nach der Auswertung der ersten Interviewphase wurde, ergänzend zu den erhobenen erfolgsrelevanten Verhaltensweisen, ein Fragebogen vorbereitet, mit dem Ziel zwei weitere für die Videoerstellung wesentliche Aspekte zu klären. Der erste Aspekt bezog sich auf strukturelle Überlegungen. Es sollten Anhaltspunkte dafür gewonnen werden, wie die ermittelten Phasen in Bezug auf ihre Schwierigkeit, ihren zeitlichen Aufwand und ihre Erfolgsrelevanz einzustufen sind. Durch diese Informationen sollte gewährleistet werden, dass die Videos zu den einzelnen Phasen nicht nur inhaltlich, sondern auch in Bezug auf ihre Länge und das Anspruchsniveau einer realen Beratungssituation so gut wie möglich nachempfunden werden konnten. Der zweite Aspekt bestand in der Notwendigkeit eine Basis zu schaffen, die eine Zuordnung der Verhaltenshinweise aus den Beobachtungsbögen der Akademie zu den fünf Phasen des Kundengesprächs und den hierfür mittels Experteninterviews erhobenen relevanten Verhaltensweisen ermöglichte. Um beiden Überlegungen gerecht zu werden, wurden die Experten innerhalb weniger Tage nach der Auswertung des ersten Interviews deshalb gebeten, mittels eines sechs Item umfassenden Fragebogens ihre Einschätzung zu den strukturellen und inhaltlichen Aspekten abzugeben. Eine

Darstellung der Fragen und Ergebnisse in graphischer Form findet der interessierte Leser in Anhang (B).

Die strukturierten Expertenhinweise aus den Interviews, systematisch ergänzt um die in den Beobachtungsbögen der Akademie beschriebenen Verhaltensweisen, bildeten, zusammen mit den Hinweisen zu strukturellen Aspekten der einzelnen Phasen, die Basis für das Verfassen der Drehbücher.

Schritt 2: Verfassen der Drehbücher

Als Grundgerüst für den Aufbau der Drehbücher dienten neben der im Vorhinein festgelegten Handlung – *„Ein potentieller Neukunde, der dem Berater bisher noch unbekannt ist, betritt die Schalterhalle und möchte sich beraten lassen, da er einen neuen Job bei einer Firma im Geschäftsgebiet des Finanzdienstleistungsinstitutes angefangen hat und gerade dabei ist in die Stadt, in der sich auch die Filiale des Finanzdienstleisters befindet, zu ziehen“* – die von den Experten im Interview genannten wesentlichen 5 Phasen eines Kundengesprächs.

Gemäß dem oben genannten Ziel, den Probanden eine ähnlich hohe Flexibilität beim Einsatz ihrer Fähigkeiten zu gewähren, wie in einem normalen Rollenspiel, wurden in einem ersten Schritt für jede dieser fünf Phasen fünf Drehbücher geschrieben, die das Vorgehen des Kundenberaters zwar inhaltlich gleich beschrieben, aber in der Qualität der Beratungsleistung, die der Berater zeigte, Abstufungen vornahmen.

Die fünffach gestufte qualitative Variation der Beratungsleistung in den so resultierenden 25 Drehbüchern wurde durch eine Manipulation derjenigen Verhaltensweisen erreicht, die in den Interviews mit den Experten als kritisch für den Beratungserfolg in der jeweiligen Phase benannt worden waren. Die Benennung der Skalenpunkte orientierte sich an der Bewertungsskala der Akademie, reichte von stark unterdurchschnittlichem (1) bis zu stark überdurchschnittlichem (5) Verhalten und deckte eine große Bandbreite an möglichen Verhaltensweisen ab. Die inhaltliche und qualitative Struktur der Drehbücher zeigt Abbildung 4.3.



Abbildung 4.3: Der inhaltlich und qualitativ gestufte Aufbau der Drehbücher

Um die oben beschriebenen qualitativen Abstufungen, der von den Beratern genannten Verhaltensweisen so genau wie möglich in Verhaltenshinweise für die Darsteller in den späteren Filmsequenzen übersetzen zu können, wurden die Drehbücher mit sehr spezifischen und detaillierten Handlungsanweisungen versehen. In der konkreten Umsetzung bedeutete dies, dass der zu sprechende Text für den Berater und seinen Kunden für jede der 25 Sequenzen in wörtlicher Rede verfasst wurde. Diese detaillierten Beschreibungen des Beraterverhaltens in den Drehbüchern wurden weiterhin um Variationen aus der Kategorie „phasenunabhängige Verhaltensweisen“ mit den Bereichen Sprache, Mimik und Gestik sowie Körperhaltung variiert. Die Hinweise zu den Bereichen Sprache, Mimik, Gestik und Körperhaltung erfolgte im Gegensatz zu den konkreten Verhaltenshinweisen in genereller Form jeweils zu Anfang einer Sequenz.

Abbildung 4.4 zeigt ein solches Drehbuch für die kürzeste Phase „Begrüßung am Schalter und Überleitung in ein Beratungszimmer“ in der niedrigsten Qualitätsstufe während Abbildung 4.5 das Drehbuch für die höchste Qualitätsstufe dieser Phase zeigt.

Begrüßung des Kunden am Schalter und Überleitung in das Beratungszimmer

Ausprägungsstufe 1

Der Kundenberater steht am Schalter und arbeitet am Terminal.
 Der Kunde tritt an den Schalter.
 Der Kundenberater blickt auf, sieht den Kunden ernst an und wendet sich dem Terminal wieder zu.
 Er gibt seine Daten weiter ein (ca. 3 Sekunden) und wendet sich dann dem Kunden zu.
 Sein Gesichtsausdruck bleibt weiter ernst.

KB: „Guten Tag?“
 K: „Guten Tag! Ich würde mich gerne beraten lassen!“
 KB: „Worum geht es dabei?“
 K: „Ich bin neu zugezogen hier und würde mich ganz gerne über den Service bei Ihnen erkundigen!“
 KB: „Gehen Sie schon mal in das Beratungszimmer! Ich gebe nur noch etwas ein und dann komme ich nach!“
 K: Ja, danke.

Der Kundenberater bleibt stehen und gibt weiter Daten in das Terminal ein.
 Der Kunde blickt sich suchend um und geht dann Richtung Beratungszimmer.

K: „Ist es das Zimmer das vorne?“
 KB: „Ja!“ **Schnitt für die Kamera**

Abbildung 4.4: Drehbuch für die Phase 1 in der niedrigsten Qualitätsausprägung

Begrüßung des Kunden am Schalter und Überleitung in das Beratungszimmer

Ausprägungsstufe 5

Der Kundenberater steht am Schalter und arbeitet am Terminal. Der Kunde tritt an den Schalter. Der Kundenberater blickt auf, sieht den Kunden und wendet sich dem Kunden sofort zu. Er blickt freundlich und lächelt.

KB: „Guten Tag! Was kann ich für Sie tun?“
 K: „Guten Tag! Ich würde mich gerne beraten lassen!“
 KB: „Gerne! Um was geht es dabei?“
 K: „Ich bin hier neu zugezogen und würde mich gerne über den Service bei Ihnen vor Ort informieren!“

Der Kundenberater weist mit seinem Arm in Richtung Beratungszimmer und begleitet den Kunden.

KB: „Darf ich Sie in unser Beratungszimmer bitten, dort sind wir ungestört!“
 K: „Danke“

Der Kundenberater begleitet den Kunden durch die Servicehalle.
 Der Kundenberater lässt dem Kunden den Vortritt und folgt ihm in das Beratungszimmer.

KB: „Bitte!“
 K: „Danke sehr!“ **Schnitt für die Kamera**

Abbildung 4.5: Drehbuch für die Phase 1 in der höchsten Qualitätsausprägung

Um eine möglichst objektive Einschätzung darüber zu erhalten, inwieweit die vorgenommenen qualitativen Abstufungen im Verhalten des Kundenberaters sinnvoll und realitätsnah umgesetzt worden waren, wurden die fertigen Drehbücher in einem zweiten Schritt fünf Experten vorgelegt. Diese wurden gebeten das Beraterverhalten in den jeweiligen Drehbüchern auf seinen Realitätsgehalt hin zu überprüfen und in einem zweiten Schritt die zu einer Gesprächsphase gehörenden fünf Drehbücher nach der Qualität der Beratungsleistung, in eine Rangfolge zu bringen. Mit Ausnahme der Phase *Bedarfsermittlung*, wo die Stufen 4 und 5 von zwei Experten im Ranking vertauscht wurden, konnten die intendierten Qualitätsreihenfolgen durch das Expertenurteil zu 100 % bestätigt werden. Die Stufen 4 und 5 der Sequenz *Bedarfsermittlung* wurden dahingehend umformuliert, dass sich alle Experten auf die intendierte Rangfolge verständigen konnten. Nachdem die beabsichtigte Manipulation der Drehbuchinhalte durch das Expertenrating als bestätigt angesehen werden konnte, folgte als nächster Schritt das Drehen der Videos.

Schritt 3: Das Drehen der Videos

In Kapitel 2.8.4 wurde dargestellt, dass die Sorgfalt bei der Übertragung eines Verfahrens in eine medial variierte, neue Version -und deren hieraus resultierende Qualität- die Kennwerte eines Verfahrens sowie dessen subjektive Wahrnehmung durch die Probanden zu beeinflussen scheinen. Das vorrangige Ziel bei der Erstellung der Videosequenzen war es, die Interaktion zwischen einem Berater und einem Neukunden so realitätsnah und qualitativ hochwertig wie möglich abzubilden. Dieses Ziel stellte einige Mindestanforderungen an das Filmteam, den Drehort und die Darsteller.

Für den Dreh der Filmsequenzen wurde deshalb auf die professionelle Hilfe des Service-Punkt Film der Westfälischen Wilhelms-Universität zurückgegriffen. Diese Kooperation bot, neben optimalen technischen Voraussetzungen, auch die Möglichkeit auf fundiertes Wissen in den Bereichen Kameraführung, Ton und Schnitt zurückgreifen zu können. Als Drehort wurde eine echte Finanzdienstleistungsfiliale gewählt und für die Rolle des Beraters in den Videos konnte ein langjähriger Kundenberater einer Finanzdienstleistungsorganisation gewonnen werden. Den Kunden spielte ein erfahrener und geschulter Rollenspieler, der selber über eine Ausbildung zum Bankkaufmann verfügte und die Rolle des Kunden in einer solchen Situation aufgrund seiner Erfahrung authentisch darstellen konnte.

Neben den oben genannten Aspekten hängt die Qualität von videobasiertem Material nicht unwesentlich von der Einhaltung einer Reihe filmischer Konventionen ab, welche auch im Falle der vorliegenden Arbeit berücksichtigt wurden. Eine ausführliche Beschreibung die-

ser Konventionen sowie detaillierte Hinweise zu den für den Dreh verwendeten technischen Komponenten findet der interessierte Leser in Anhang C, da aus Gründen der Übersichtlichkeit auf eine Darstellung an dieser Stelle verzichtet wird. Das aus der Umsetzung der in Anhang C erläuterten Konventionen resultierende Filmset zeigt Abbildung 4.6.

Alle 25 Videosequenzen wurden nach dieser Systematik produziert. Die zur Aufnahme verwendeten Kameraeinstellungen sind hierbei in dem dunkelgrau hinterlegten Bereich lokalisiert, der durch die 180 Grad-Regel vorgegeben wird. Die Position von Kunde (K) und Berater (B) definieren die Handlungsachse. Das am Dreh beteiligte Team bestand, neben den beiden Rollenspielern, aus drei Kameramännern, einer Tonassistentin und einem Skriptgirl.

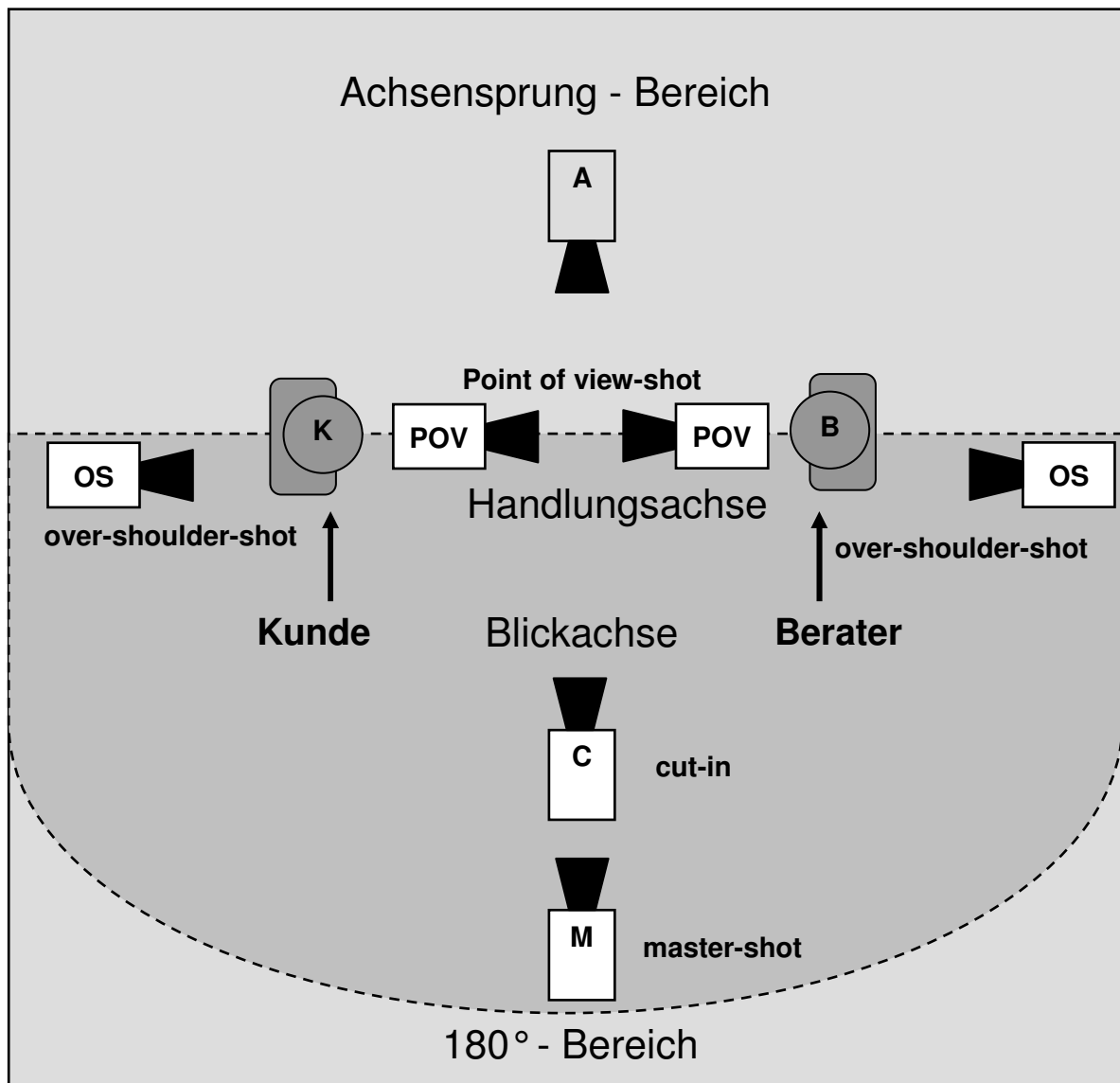


Abbildung 4.6: Setaufbau nach dem continuity-editing-system (nach Beller, 2005)

Schritt 4: Digitalisieren der Bänder und Schnitt

Die abgedrehten Sequenzen wurden zum Zwecke einer besseren Bearbeitung mit Hilfe von zwei Zuspielmaschinen (Sony DSR 1000 DVCam Rekorder und Sony PVW 2800 Beta-camSP Maschine) eingelesen und digitalisiert.

Mit Hilfe der Schnittsoftware Avid Xpress Pro 4.6 sowie der Zusatzsoftware Avid Mojo wurden die Inhalte der jeweiligen Ordner auf einer Compaq W8000 Workstation dann zu fertigen Filmsequenzen zusammen geschnitten. Als Schnittmethode wurde der Group-Clip Modus gewählt. In diesem Modus ist es möglich sich das Film- und Tonmaterial aller drei Kameras gleichzeitig und in Echtzeit anzusehen, was die Auswahl der jeweils geeigneten Sequenz erleichtert, da man den unmittelbaren Vergleich in Bezug auf die Qualität der Sequenzen herstellen kann. Die jeweils am besten geeignete Einstellung wurden dann markiert und die dazugehörige Videospur (V1, V2 oder V3) und Audiospur (A1, A2 oder A3) per Mausclick in eine neue Masterspur für Video und Audio integriert (V4). Das Endresultat ist die fertige Filmsequenz. Die Benutzeroberfläche des Programms Avid Xpress Pro im Group-Clip Modus zeigt Abbildung 4.7.

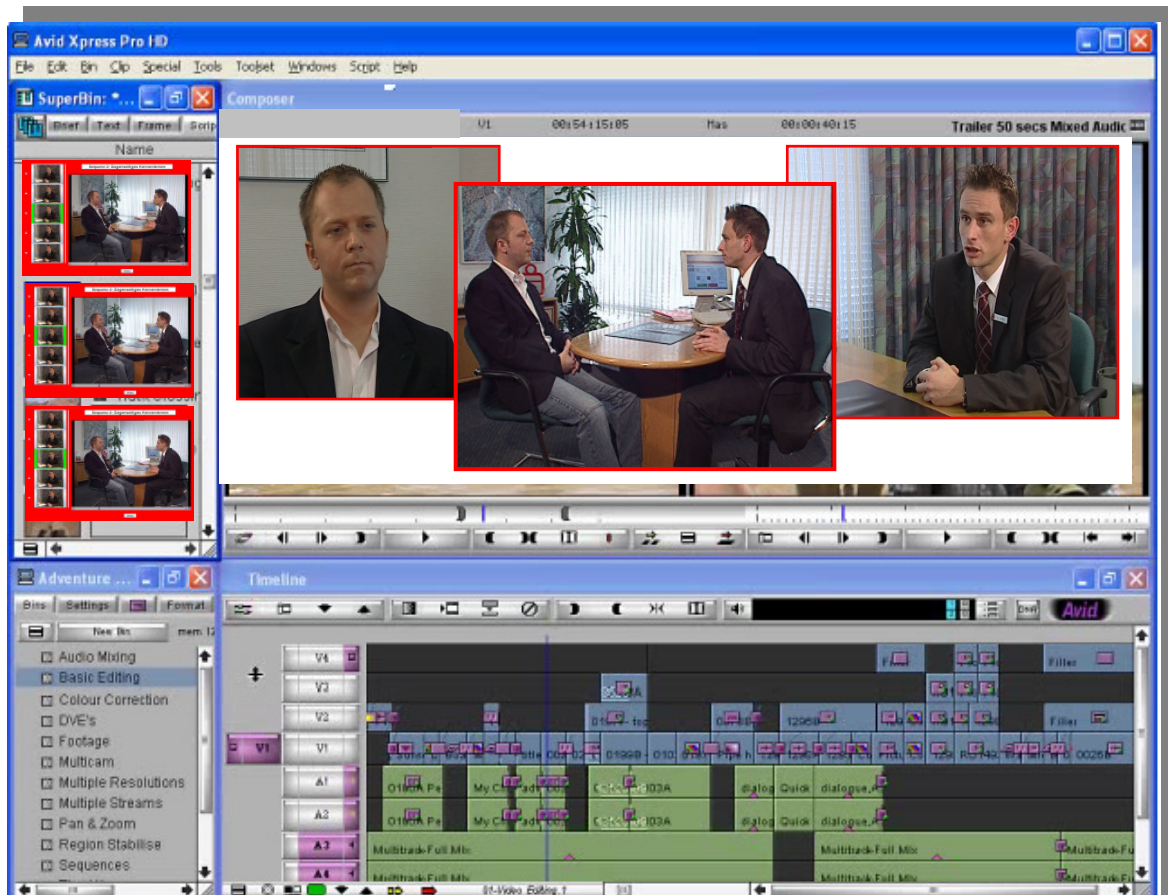


Abbildung 4.7: Screenshot der Benutzeroberfläche des Programms Avid Xpress Pro 4.6

Schritt 5: Erstellen eines Präsentationsprogramms

Um den Probanden die fertiggestellten Videosequenzen, geordnet nach den jeweiligen thematischen Phasen, als computerbasiertes Verfahren präsentieren zu können, bedurfte es eines Programms, welches eine standardisierte Betrachtung und Bearbeitung durch die Probanden sicherstellen sollte. Das Programm, welches den Titel *Videobasiertes Analyseverfahren für Kundenberater (VAK)* erhielt, musste dabei folgende Auflagen erfüllen:

1. Das Verfahren sollte selbsterklärend sein.
2. Das Abspielen der Videosequenzen sollte in hoher Qualität erfolgen und die Videos sollten in einer Größe dargeboten werden, die es den Probanden noch ermöglichte Variationen in der Mimik der Gesprächspartner zu erkennen.
3. Die Darbietung der 5 Videosequenzen pro Gesprächsabschnitt sollte, um Reihenfolgeeffekte zu vermeiden, in randomisierter Form möglich sein.
4. Eine fehlerhafte Bearbeitung durch die Probanden, wie das Übersehen einer Sequenz oder das Vergessen einer Auswahl, sollte durch logische Schleifen und entsprechende Fehlermeldung unmöglich gemacht werden
5. Die von den Probanden generierten Daten sollten sicher und ohne Verlust direkt abgespeichert werden können.

Diese Vorgaben wurden durch eine javabasierte Stand-Alone-Anwendung realisiert. Stand-Alone-Anwendungen werden auf einzelnen Rechner installiert und können lokal auf diesem Rechner ausgeführt werden. Das wichtigste Kriterium für die Entscheidung eine Stand-alone-Anwendung zu nutzen und keine serverbasierte Online-Lösung einzusetzen, war die Forderung, dass alle Kandidaten unter exakt gleichen Bedingungen die Videos bearbeiten sollten. Eine Online-Lösung konnte dieser Forderung auf Grund der unterschiedlichen technischen Möglichkeiten der an der Untersuchung teilnehmenden Unternehmen nicht gerecht werden. Auch die Forderung nach einem reibungslosen Ablauf der Untersuchung und einer sicheren Form der Datenspeicherung sprachen für die Stand-Alone-Version. Die vielen positiven Aspekte einer online durchgeführten Untersuchung, wie die einfache Zugriffsmöglichkeit auf das Programm durch eine hohe Zahl von Probanden oder der Ersparnis von Personal- und Sachkosten (Reips, 2002), mussten zugunsten einer höheren Sicherheit der produzierten Daten und einer technisch fehlerfreien Präsentation, die online zur Zeit noch nicht als gesichert angesehen werden kann (Reips, 2002), aufgegeben werden. Nur durch die Stand-alone-Version konnte sichergestellt werden, dass

alle Probanden die Filme unter gleichen Bedingungen und in gleicher Qualität betrachten konnten. Dieser Aspekt wurde aufgrund der in Kapitel 2.7.2 dargestellten Überlegungen als wichtig erachtet, dass die Art der Itempräsentation und die Handhabbarkeit des Testmaterials die subjektive Wahrnehmung eines Verfahrens beeinflussen. Weiterhin konnten die von den Probanden produzierten Daten unmittelbar nach der Untersuchung gespeichert und gesichert werden. Bei der für das Schreiben des Programms verwendeten Software handelt es sich um die Version Java 5.0 in Kombination mit der 2.1.1 Version des Java Mediaframeworks, welches zum Einbinden und Abspielen der Videos diente. Als Präsentationsgeräte für die spätere Untersuchung wurden Toshiba Satellite Pro L 20 Notebooks mit einer 15 Zoll Monitor vorgesehen. Die Notebooks waren mit einem Pentium M 740 Prozessor, 512 MB Arbeitsspeicher und einer 60 GB Festplatte (5400 RPM) ausgestattet.

Schritt 6: Expertenrating der Videos im VAK

Aufgrund der in Kapitel 2.8.2 dieser Arbeit dargestellten Besonderheiten von videobasiertem, multimedialem Material wurden trotz der Tatsache, dass die intendierte Manipulation des Beraterverhaltens in den einzelnen Gesprächssituationen durch ein Expertenrating der Drehbücher bestätigt worden war, die fertigen Videos nochmals fünf Experten mit der Bitte um eine Begutachtung präsentiert. Die Berater wurden gebeten die einzelnen Videos im VAK anzusehen und, nachdem sie eine Rangreihe der fünf Videos pro Gesprächsabschnitt gebildet hatten, das ihrer Meinung nach beste Beraterverhalten auszuwählen. Als geeignetes Maß der Übereinstimmung von Expertenratings bei ordinalskalierten Daten und mehr als zwei Ratern wurde nach Wirtz und Caspar (2002) Kendalls Konkordanzkoeffizient W ermittelt. Der ermittelte Koeffizient für $n=5$ Rater für alle 25 Sequenzen ist mit einem Wert von $W=.974$, df 24 als hoch zu bewerten. Die Koeffizienten für die einzelnen Sequenzen sind in Tabelle 4.1 ersichtlich.

Tabelle 4.1: Ergebnisse der Expertenratings für die fünf Abschnitte des VAK

	Kendall-W	Chi-Quadrat	df	Signifikanz
Sequenz 1	1	20	4	.000
Sequenz 2	.904	18.080	4	.001
Sequenz 3	.968	19.360	4	.001
Sequenz 4	1	20	4	.000
Sequenz 5	1	20	4	.000

4.2.3 Beschreibung des VAK

Im folgenden Kapitel soll mit dem VAK das Endresultat der in Kapitel 4.2.2 dargestellten Schritte zur Übertragung eines Rollenspiels in ein video- und computerbasiertes Verfahren beschrieben werden. Hierfür wird zuerst die übergeordnete Struktur des VAK vorgestellt. In einem zweiten Schritt werden die Inhalte des VAK näher beschrieben.

Ebenen des VAK

Das *Videobasierte Analyseverfahren für Kundenberater* besteht aus 11 aufeinander folgenden Bildschirmansichten, die sich in zwei übergeordnete Ebenen unterteilen lassen (siehe Abbildung 4.8).

Die erste Ebene dient der Instruktion der Probanden sowie der Erhebung persönlicher Daten und umfasst insgesamt 5 Bildschirme. Die zweite Ebene beinhaltet 5 aufeinander folgende Ansichtsbildschirme, mittels derer die nach thematischen Abschnitten eines Kundengesprächs geordneten Videosequenzen betrachtet und bewertet werden können, sowie ein der eigentlichen Aufgabenstellung vorangehendes Intro. Die Navigation durch beide Programmebenen erfolgt mit Hilfe einer Hyperlinkverknüpfung, die sich, in Form eines viereckigen Feldes mit der Aufschrift „weiter“, jeweils in der Mitte des unteren Bildschirmrandes befindet. Die beschriebenen Ebenen sind dabei lediglich von konzeptueller Natur und für den Probanden nicht ersichtlich. In der späteren Ergebnisdatei werden zuerst alle Informationen gespeichert und dargestellt, die in Ebene 1 erhoben werden. Hieran schließt sich die Speicherung und Darstellung der Ergebnisse aus den fünf Bearbeitungsbildschirmen an.

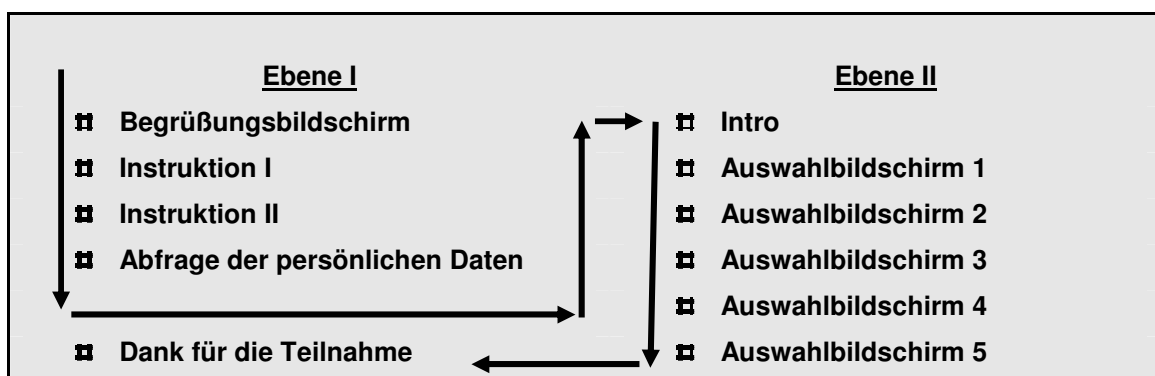


Abbildung 4.8: Die Ebenen des VAK im zeitlichen Ablauf

Inhalte des VAK

Der Abschnitt I beginnt für den Probanden mit der Startseite, die lediglich den Namen des Programms beinhaltet. Durch das Anklicken des Weiterbuttons gelangt der Proband dann auf die erste Instruktionssseite, die das VAK und seinen Aufbau beschreibt. Auf der nach abermaligem Klicken folgenden zweiten Instruktionssseite wird die Aufgabenstellung beschrieben. Für den Fall des VAK wurde in Anlehnung an die in Kapitel 2.8.5 vorgestellte Systematik nach Ployhart und Ehrhart (2003) eine *would do* Instruktion gewählt. Die Probanden wurden gebeten diejenige Verhaltenssequenz auszuwählen, welche ihrem eigenen Verhalten in einer realen Situation am ehesten entspricht (für eine Darstellung siehe Anhang AII). Auf dem vierten Bildschirm des ersten Abschnittes wird der Proband gebeten seine persönlichen Daten anzugeben. Neben dem Namen seines Arbeitgeberinstitutes sowie seinem eigenen Nach- und Vornamen kann der Kandidat hier mittels zweier Kontrollkästchen bestimmen, ob er nur das Videosignal ohne den Ton (Stummfilm), nur das Audiosignal ohne das Bild (Hörspiel) oder beide Signale (vollständige Darbietung) abspielen möchte.

Der Abschnitt II beginnt mit dem Intro, das einen Kunden beim Betreten einer Finanzdienstleistungsfiliale zeigt. Der Kunde orientiert sich kurz in der Schalterhalle und geht dann auf einen Berater am Schalter zu. Dieses Intro wurde eingefügt, um die Probanden mit den handelnden Figuren und dem Szenario vertraut zu machen. Nach dieser „Orientierungsphase“ gelangt der Zuschauer auf den ersten der fünf Auswahlbildschirme, welcher in Abbildung 4.9 dargestellt ist. Hier hat er die Möglichkeit alle fünf Videos zu einem Gesprächsabschnitt zu betrachten. Zur besseren Orientierung verfügt jeder Bildschirm über eine individuelle Kopfzeile, die den Titel des gerade aktuellen Gesprächsabschnittes zeigt. Weiterhin befinden sich auf der linken Seite jedes dieser fünf Bildschirme Startknöpfe zum Abspielen der 5 qualitativ variierten Videosequenzen. Sobald ein Startknopf angeklickt wird, färbt sich sein bisher neutraler Rahmen grün, um dem Probanden eine bessere Übersicht darüber zu geben, welche Videos er bereits betrachtet hat und welche nicht. Nach erfolgter Betrachtung aller Videos kann der Proband das Video, welches dem eigenen Verhalten des Probanden in der betrachteten Situation am ähnlichsten ist, auswählen, indem er ein kleines, rundes Auswahlfeld links neben dem Startknopf anklickt. Das Abspielen und Betrachten der einzelnen Videosequenzen erfolgt in einem großen Feld rechts neben den Startknöpfen. Ein Fortschrittsbalken ermöglicht es dem Probanden den Fortschritt des Videos mitzuverfolgen und gezielt nur einzelne Teile des Videos wiederholt zu betrachten.



Abbildung 4.9: Screenshot des Bearbeitungsbildschirms für die Sequenz 1 des VAK

Die oben beschriebene Hyperlinkverknüpfung in Form des *weiter* Buttons kann nur dann genutzt werden, wenn sowohl alle Videos angeschaut wurden als auch ein Video ausgewählt worden ist. Ist einer der beiden Fälle nicht erfüllt, bekommt der Proband so lange eine Fehlermeldung in Form eines visuellen Hinweises, bis er den Fehler behoben hat.

Bereits während der Bearbeitung generiert das Programm eine Ergebnisdatei im csv Format. Hier werden die persönlichen Daten des Bewerbers, der gewählte Modus (Stummfilm, Hörspiel, Video), die Darbietungsreihenfolge der Videosequenzen in den einzelnen Gesprächsphasen und die Auswahl des Probanden festgehalten. Eine Darstellung aller Bildschirme des VAK findet der interessierte Leser in Anhang D.

Nachdem die einzelnen Schritte der Erstellung der video- und computerbasierten Variante des Rollenspiels beschrieben und das VAK als resultierendes Produkt vorgestellt wurde, soll im folgenden Abschnitt auf die Übertragung des Face-to-Face Rollenspiels in eine Paper-Pencil Form eingegangen werden.

4.2.4 Erstellung der Paper-Pencil Variante

Neben dem angestrebten Vergleich eines klassischen Rollenspiels in Face-to-Face Form mit einer computer- und videobasierten Variante unter Berücksichtigung der jeweiligen kriterienorientierten- und sozialen Validität sollte in der vorliegenden Arbeit auch der Vergleich der klassischen Form mit einer Paper-Pencil Variante unternommen werden. Parallel zur videobasierten Rollenspielversion sollte den Probanden auch für die Paper-Pencil Version die Möglichkeit gegeben werden ihre Fähigkeiten als Kundenberater ähnlich flexibel einzusetzen, wie ihm dies in einem klassischen Rollenspiel möglich gewesen wäre.

Aufgrund dieser Überlegungen wurde eine offene Aufgabenstellung in Aufsatzform gewählt, die den Probanden ein flexibles Bearbeiten der Thematik ermöglichte. Aufgabe der Probanden war es ihr persönliches Vorgehen in einem Kundengespräch mit den bereits für das Face-to-Face Rollenspiel und das VAK beschriebenen inhaltlichen Rahmenbedingungen schriftlich auszuführen. Genau wie bei der Konzeption des VAK wurde die Instruktion in Form einer *would-do* Instruktion formuliert. Konkret sollten die Probanden schriftlich ihr potentiell Verhalten in der beschriebenen Situation in der Ich-Form verfassen und ihr Verhalten so detailliert und ausführlich wie möglich niederschreiben (für eine Darstellung siehe Anhang AIII). Die Bearbeitungszeit für diese Übung wurde auf 30 Minuten festgesetzt und in der Instruktion für die Probanden wurde explizit darauf hingewiesen, dass dieser Zeitraum voll auszunutzen sei.

Als Maßstab zur Bewertung der schriftlichen Leistungen wurde, in Anlehnung an die Beschreibungen des optimalen Beraterverhaltens aus den Drehbüchern, ein Katalog formuliert, der 25 Verhaltensweisen umfasste (siehe Anhang E). Diese setzten sich aus denjenigen fünf Verhaltensweisen aus den fünf Phasen des VAK zusammen, die für ein erfolgreiches Vorgehen in der jeweiligen Phase aus Sicht der Experten als zentral bewertet wurden. Der angelegte Bewertungsschlüssel sah vor für jede beschriebene Verhaltensweise, welche mit den im Katalog beschriebenen Verhaltensweisen übereinstimmte, einen Punkt zu vergeben. Die Zahl der möglichen Punkte lag somit zwischen 0 und 25.

Der endgültige Verhaltenskatalog und der Bewertungsschlüssel wurden vor dem Einsatz in der Hauptuntersuchung im Rahmen einer Voruntersuchung mit einer Stichprobe von $N = 17$ Auszubildenden, die denen der Zielpopulation für die Hauptuntersuchung in Alter und Bildungsstand entsprachen, getestet (ausführlich siehe Kapitel 4.3). Die Analyse zeigte ein zufrieden stellendes Ergebnis. Eine Überarbeitung des Verhaltenskataloges musste deshalb nicht erfolgen.

Nachdem die Auswahl des Face-to-Face Rollenspiels und seine Übertragungen in eine video- und computerbasierte und eine Paper-Pencil Variante beschrieben wurde, soll im folgenden Abschnitt auf die Konstruktion des Fragebogens eingegangen werden, der als Instrument zur Erfassung der sozialen Validität der drei Rollenspielvarianten diene.

4.2.5 Erstellung der Fragebögen zur Erhebung der sozialen Validität

Als Basis für die Konzeption des Fragebogens diene das in Kapitel 2.3.1 beschriebene Konzept der sozialen Validität nach Schuler und Strehle (1983, 1985), Schuler (1993). Dieses Konzept erfuhr im Kontext dieser Arbeit insofern eine Modifikation, als dass nur drei der vier nach Schuler und Strehle (1983, 1985) benannten Aspekte in die Fragebogenkonstruktion eingingen. Der vierte, ursprünglich im Konzept berücksichtigte Aspekt „Urteilkommunikation/ Feedback“ fand keine Verwendung, weil sich die Art des Feedbacks bereits durch die organisatorischen Rahmenbedingungen stark unterschied. Da die Akademie, in Übereinstimmung mit den Personalabteilungen der beteiligten Institute, nur die Leistungen der bisher üblichen Rollenspiele in die spätere Bewertung der Probanden einfließen lassen wollte, wurden auch nur diese Ergebnisse an die Probanden im Rahmen des Verfahrens zurückgemeldet. Die Ergebnisse aus dem Paper-Pencil und dem videobasierten Verfahren wurden den Teilnehmern vom Autor dieser Arbeit individuell nach Absprache zurückgemeldet.

Da die medial variierte Darbietung eines klassischen Rollenspiels ein absolutes Novum im Rahmen des hier beschriebenen *PES* darstellte, sollten mittels des Fragebogens zusätzlich Informationen erhoben werden, die sich auf die erlebte Fairness seitens der Probanden und die Akzeptanz der Probanden für diese Verfahren bezogen. Aus diesem Grund gingen neben den drei bereits genannten Aspekten ergänzend die Gesichtspunkte Fairness und Akzeptanz abgeleitet nach dem in Kapitel 2.3.2 vorgestellten Modell von Gilliland (1993) mit ein.

Sowohl die drei Teilaspekte der sozialen Validität als auch die Aspekte der Fairness und Akzeptanz wurden mit jeweils 4 Fragen abgebildet. Die Generierung der Items orientierte sich für die drei Aspekte der sozialen Validität an den Vorschlägen nach Schuler (1996, S. 184) und für die Aspekte Fairness und Akzeptanz an den *Procedural Justice Rules* des in Abbildung 2.2 dargestellten Modells nach Gilliland (1993) sowie den Arbeiten von Grewe (2005), Hammerschmidt (2004) und Kanning et al. (2006).

Zur Erhebung der Probandenantworten wurde eine siebenstufige Likert-Skala konstruiert, deren Skalenpunkte von 1 (stimme gar nicht zu) bis 7 (stimme voll zu) reichten. Sowohl die Items als auch die Instruktionen für die Probanden waren für alle drei Bedingungen gleich formuliert. Allerdings wurde der zweite Absatz des jeweiligen Fragebogens insofern variiert, als dass hier die jeweilige Rollenspielvariante, auf die der Proband Bezug nehmen sollte, expliziert wurde. Die endgültige Version des Fragebogens zur Erhebung der sozialen Validität und wahrgenommenen Fairness findet der interessierte Leser in Anhang (F).

Nachdem im vorangegangenen Kapitel auf die Erstellung des Materials eingegangen wurde, welches in dieser Untersuchung Verwendung fand, sollen im folgenden Kapitel die Voruntersuchungen beschrieben werden, denen das eingesetzte Material unterzogen wurde

4.3 Empirische Überprüfung der intendierten Materialspezifika

Bevor die erstellten Materialien zur Überprüfung der in Kapitel 3 formulierten Forschungsfragen eingesetzt werden konnten, wurden sie einer Reihe von Vorstudien unterzogen. Die erste hiermit verbundene Zielsetzung bestand darin einen möglichen Optimierungsbedarf in Bezug auf die Realisierung der intendierten Manipulation in der Hauptstudie zu ermitteln. Weiterhin sollten erste Anhaltspunkte dafür gewonnen werden, inwieweit das neu entwickelte VAK zufriedenstellende Ergebnisse im Sinne der in Kapitel 2.2 genannten relevanten Gütekriterien für eignungsdiagnostische Instrumente liefern würde.

Da die Face-to-Face Variante des Rollenspiels aus dem Aufgabenpool der Akademie ausgewählt wurde und somit in ihren inhaltlichen als auch ihren strukturellen Eigenschaften den Standards der Akademie entspricht und eine fortlaufende Evaluation und Verbesserung durch die Akademie gewährleistet ist, wurde sie in die Voruntersuchungen nicht näher mit einbezogen. Die sachgemäße Instruktion durch einen erfahrenen Moderator und die valide Beobachtung und Bewertung durch geschulte Beobachter konnte aufgrund der Standards der Akademie genauso als gesichert gelten, wie die Tatsache, dass das Niveau der Aufgabe für die intendierte Zielgruppe angemessen war. Im Focus der Vorstudien standen deshalb die beiden anderen medialen Variationen, das VAK und die Paper-Pencil Variante sowie der Fragebogen zur Erhebung der wahrgenommenen sozialen Validität.

Insgesamt wurden hierfür drei Studien durchgeführt, von denen zwei im universitären Kontext angesiedelt waren und auf Studenten als Versuchspersonen zurückgriffen. Die dritte Vorstudie, wie die spätere Hauptstudie auch, wurde im Finanzdienstleistungskontext durchgeführt, da für die letzte Kontrolle des Materials auf Bankauszubildende als Probanden zurückgegriffen werden sollte, die in ihrer Alterstruktur und ihrem Bildungsstand den Versuchspersonen in der Hauptstudie entsprachen.

Da die einzelnen Vorstudien teilweise zur Beantwortung mehrerer Fragestellungen dienten, soll zur besseren Übersichtlichkeit für den Leser kurz auf die Struktur eingegangen werden, welche zur Darstellung der Vorgehensweise gewählt wurde. Die Beschreibung der Vorstudien beginnt jeweils mit einer kurzen Darstellung der theoretischen Motivation und den sich daraus ergebenden Fragestellungen. Hieran schließt sich die Beschreibung der Stichprobe, des Materials und des Versuchsablaufes an. Dann folgt, getrennt für die ein-

zelenen Fragestellungen, die Darstellung des Designs und der Ergebnisse. Den Abschluss bildet eine Bewertung der Ergebnisse in Bezug auf die Hauptuntersuchung.

Am Ende des Kapitels werden die Ergebnisse aus allen drei Voruntersuchungen dann zu einem Fazit bezüglich der Eignung des Materials zusammengefasst.

4.3.1 Vorstudie 1: Empirische Prüfung der multimedialen Eigenschaften des VAK

Theoretische Motivation

Die folgende Studie wurde mit dem Ziel durchgeführt das neu entwickelte VAK im Hinblick auf mehrere Fragestellungen zu untersuchen, welche sich aus den in Kapitel 2.8.1 dargestellten Herausforderungen bei der Erstellung von multimedialen Anwendungen ergeben. Konkret sollte geprüft werden, ob sich für das VAK Probleme bei der Bearbeitung zeigen würden, welche sich aus der Kombination von unterschiedlichen Symbolsystemen und angesprochenen Wahrnehmungskanälen ergeben könnten sowie das Problem einer möglichen kognitiven Überlastung durch diese Kombinationen. Weiterhin sollte ein erster Eindruck in Bezug auf die Bewertung der sozialen Validität des VAK durch die Probanden und die empirische Fairness des VAK bei einem Vergleich zwischen männlichen und weiblichen Probanden gewonnen werden.

Fragestellung I

In Kapitel 2.8.1 wurde mit Weidenmann (1996) darauf hingewiesen, dass die optimale Kombination verschiedener Symbolsysteme und Wahrnehmungskanäle in multimedialen Systemen, vor allem bei einem Einsatz von videobasiertem Material, eine gewisse Herausforderung darstellt. Weiterhin wurde in Kapitel 2.8.4 die Übertragung der intendierten inhaltlichen Aspekte schriftlicher Items in adäquate Audio- und Videostimuli multimedialer Items als ein komplizierter Schritt im Produktionsprozess von videobasiertem Material genannt (Olson-Buchanan und Drasgow 2006).

Auch für das VAK stellte sich die Frage, inwieweit die Manipulation des Beraterverhaltens in den Videos, welche durch die Variation von visuellen Hinweisreizen (Mimik, Gestik, Körperhaltung und Handlungen des Bankberaters) und auditiven Hinweisreizen (Sprache und Inhalte, die der Bankberater darstellt) realisiert werden sollte, gleich erfolgreich für beide Kanäle gelungen war. Eine in diesem Sinne erfolgreiche Variation der visuellen und auditiven Hinweisreize in den Videos des VAK würde bedeuten, dass sich für den Betrach-

ter der jeweiligen Videosequenzen ein kongruentes Gesamtbild aus den Einzelinformationen ergibt. Ist die Manipulation auf beiden Kanälen gleich gut gelungen, sollten sich zwei Ergebnisse zeigen. Erstens sollten sich die Leistungen von Probanden, die ein Rating der Videos lediglich über auditive Hinweisreize bzw. visuelle Hinweisreize vornehmen können, in ihrer Qualität nicht unterscheiden. Weiterhin dürften, wenn die Einzelinformationen ein kongruentes Gesamtbild ergeben, die Leistungen von Probanden, die sowohl auditive als auch visuelle Hinweisreize zur Urteilsbildung heranziehen können, nicht schlechter ausfallen als die der beiden anderen Gruppen. Wenn sich die Informationen auf den beiden Kanälen allerdings nicht entsprechen, dann könnte die Leistung schlechter ausfallen, da die inkongruenten Informationen dem Probanden die Lösungsfindung erschweren.

Fragestellung 2

In Kapitel 2.8.1 wurde auf die Besonderheit von Informationsvermittlung mit Hilfe von bewegten Bildern über das pictorale Codierungssystem und auf die Gefahr einer kognitiven Überlastung hingewiesen (Strittmacher, 1994; Singer und Singer, 1983; Thürauf, 2004).

Der zweite Aspekt, der im Rahmen der Voruntersuchung behandelt werden sollte, galt deshalb der Frage, ob die Fülle an Informationen in den einzelnen Videosequenzen in einem angemessenen Rahmen geblieben war. Mit einem angemessenen Rahmen ist in diesem Zusammenhang gemeint, dass die Länge der Sequenzen und damit die Menge an audiovisuellen Informationen, die dem Probanden zur Entscheidungsfindung zur Verfügung gestellt wird, so bemessen ist, dass sie die verarbeitenden Systeme, z.B. das Arbeitsgedächtnis, nicht überfordert. Wäre nämlich letzteres der Fall, könnte neben der tatsächlich interessierenden Eigenschaft, die das VAK zu messen beabsichtigt, auch die Konzentrations- und Merkfähigkeit der Teilnehmer und damit die Häufigkeit der Darbietung der Sequenzen eine wesentliche Auswirkung auf das spätere Abschneiden im VAK haben.

Bei einer angemessenen Länge und Informationsdichte jedoch sollte ein Proband von der Möglichkeit einer wiederholten Darbietung der Videos nicht in dem Sinne profitieren, dass er eine höhere Leistung zeigt als bei einer Bearbeitung des VAK mit einer einmaligen Darbietung des Filmmaterials.

Fragestellung 3

Ein dritter Gesichtspunkt, der im Rahmen der ersten Voruntersuchung von Interesse war, bezog sich auf die Frage, wie Probanden das VAK in Bezug auf ausgewählte Aspekte wie

Fairness, Realitätsnähe und Dauer bewerten. Im diesem Zusammenhang sollte untersucht werden, inwieweit sich signifikante Unterschiede bei der Bewertung durch die Probanden ergeben würden, und zwar in Abhängigkeit davon, ob sie ihre Auswahl auf rein auditives Material (Hörspiel) oder rein visuelles Material (Stummfilm) oder das Betrachten von vollständigen Filmen (Video) stützen konnten. Da es sich bei *Video* um die Variante mit der höchsten *Fidelity* handelte, sollte sich hier eine signifikant bessere Bewertung ergeben als für die anderen beiden Varianten.

Fragestellung 4

Mit den in Kapitel 2.3.3 vorgestellten Studien konnte gezeigt werden, dass sich die Wahrnehmung eines Verfahrens durch die Probanden auf Aspekte wie die Motivation und die Probandenleistung auswirken können. Als vierter Aspekt sollte deshalb untersucht werden, inwiefern sich für die Bedingung *Video*, die in der späteren Hauptuntersuchung Verwendung finden sollte, ein Unterschied in der Bewertung des Verfahrens zwischen weiblichen und männlichen Probanden zeigen würde. Da es sich bei beiden Darstellern in den Videos um männliche Rollenspieler handelt, sollte eine mögliche Auswirkung dieses Umstandes auf die Wahrnehmung der Probanden überprüft werden, um eine eventuelle Benachteiligung weiblicher oder männlicher Probanden auszuschließen.

Stichprobe

Als Versuchspersonen (Vpn) für die vorliegende Untersuchung konnten Studentinnen und Studenten der Westfälischen Wilhelms-Universität Münster (WWU) gewonnen werden. Für die $N = 60$ Studierenden ergab sich folgende Fächerverteilung: Psychologie 29, Masterstudiengänge 11, Lehramt 5, Germanistik 1, Pädagogik 1, Sport 1, BWL 1, VWL 1, Theologie 1, Studium nicht spezifiziert 9. Alle Vpn nahmen freiwillig an der Untersuchung teil und wurden mit Versuchspersonenstunden, die während der psychologischen Ausbildung erworben werden müssen, oder einem Stundenlohn von 10 Euro entlohnt. Die Auswahl der Probanden für die Voruntersuchung fiel auf Studenten, da diese bezüglich des Alters ($M = 23.22$, $SD = 3.22$) als auch des vorherrschenden Bildungsgrades (100% der Versuchspersonen in der Stichprobe verfügten über ein Abitur) den Teilnehmern in der späteren Hauptuntersuchung sehr ähnlich waren. Lediglich bei der Geschlechterverteilung lag die Quote der weiblichen Probanden mit 75% höher als es bei der Stichprobe in der Hauptuntersuchung erwartet wurde.

Material

Um die Frage beantworten zu können, inwieweit bei der Erstellung des Videomaterials eine äquivalente Manipulation der Verhaltensweisen auf auditiver und visueller Ebene realisiert werden konnte, wurde, unter ansonsten gleichen Bedingungen, die Bearbeitung des ursprünglichen Videomaterials mit der Bearbeitung einer Stummfilmvariante (nur visuelle Hinweisreize) und einer Hörspielvariante (nur auditive Hinweisreize) verglichen. Ergänzend hierzu musste allerdings die ursprüngliche Aufgabenstellung dahingehend modifiziert werden, dass die Instruktion gegenüber der Originalversion für alle drei Bedingungen gleich gehalten werden sollte, trotzdem aber für jede Version einzeln betrachtet sinnvoll bleiben musste. So konnte weder für die Stummfilm- noch für die Hörspielvariante mit den ersten drei Bildschirmen des VAK gearbeitet werden, da hier ja auf vollwertige Videos Bezug genommen wird. Aus diesem Grund wurden für alle drei Bedingungen neue Instruktionen verfasst und die Informationen auf den ersten drei Bildschirmen des VAK durch eine papierbasierte Version ersetzt. So wurde beispielsweise in der Hörspielbedingung das Verfahren als AAK (Audiobasiertes Analyseverfahren für Kundenberater) bezeichnet und anstelle von Videosequenz wurde von Audiosequenzen gesprochen. Die tatsächliche Arbeit im VAK begann also erst auf dem vierten Bildschirm mit dem Eintragen der persönlichen Daten.

Auch die Untersuchung der zweiten Frage nach dem Nutzen einer wiederholten Darbietung der Videosequenzen erforderte zusätzliches Material. Da es im VAK nicht möglich ist, nach der Auswahl einer Sequenz aus einem thematischen Abschnitt in diesen nochmals zurückzuspringen, musste die zu bildende Rangreihe nach einmaligem Betrachten der Sequenzen auf einem vorbereiteten Bogen schriftlich notiert werden und nach einem zweiten Durchgang ebenfalls schriftlich festgehalten werden.

Als letztes Element wurde ein neun Items umfassender Fragebogen entworfen, der die Einstellung der Probanden zum absolvierten Verfahren erheben sollte. Der Einsatz des Fragebogens war explorativer Natur und diente dazu einen ersten Eindruck bezüglich der Akzeptanz von Probanden gegenüber dem VAK zu erhalten. Die Generierung der Items erfolgte nicht, wie in der späteren Hauptuntersuchung, streng nach dem Konzept der sozialen Validität nach Schuler und Stehle (1983). In dieser ersten Phase der Materialprüfung standen eher Aspekte wie die Akzeptanz bezüglich der gewählten Rollenspieler, die Realitätsnähe der Szenen und die empfundene Belastung bei der Bearbeitung des Verfahrens im Vordergrund. Die Beantwortung der Fragen erfolgte anhand einer siebenstufigen Ratingskala mit den Ausprägungen von *gar nicht* bis *sehr*. Im Einzelnen lauteten die Fragen:

- Wie gut konntest du dich mit dem Berater identifizieren?
- Wie sympathisch war dir der Berater?
- Für wie realistisch hältst du die Beratungsgespräche?
- Wie schwer fiel dir das Ranking der Ausschnitte?
- Wie deutlich waren für dich die Unterschiede der einzelnen Ausschnitte?
- Hast du die Dauer des Verfahrens als lang empfunden?
- Wie anstrengend fandest du das Verfahren?
- Für wie geeignet hältst du dieses Verfahren?
- Für wie angenehm aus Sicht eines Bewerbers hältst du dieses Verfahren?

Letztendlich bestand das eingesetzte Material aus den schriftlichen Instruktionen, dem VAK mit der oben beschriebenen Modifikation, den Fragebögen, welche die Probanden bei der Bewertung der Sequenzen mittels einer Rangbildung unterstützen sollten und dem Fragebogen zur generellen Bewertung des Verfahrens durch die Probanden. Das Programm wurde den Probanden, wie auch in der späteren Hauptuntersuchung, auf 15 Zoll Notebooks der Marke Toshiba präsentiert. Um eine Gruppentestung zu ermöglichen, wurden die Notebooks zusätzlich mit Kopfhörern (Phillips SBC HL 140) ausgestattet.

Versuchsablauf

Der in Abbildung 4.10 dargestellte Versuchsablauf erfolgte in zwei Schritten. In einem ersten Schritt wurden die Probanden begrüßt, per Zufall auf die drei Bedingungen *Hörspiel*, *Stummfilm* und *Video* aufgeteilt und bekamen die für ihre Bedingung relevante schriftliche Instruktion. Im Rahmen der folgenden Instruktionsphase wurde den Probanden die Möglichkeit gegeben Verständnisfragen an den Versuchsleiter zu stellen.

In einem zweiten Schritt wurden die Vpn gebeten die ersten 5 Videos bzw. Stummfilme oder Hörspiele zum Bereich „Begrüßung des Kunden“ zu starten. Nach dem einmaligen Vorspielen dieser Sequenzen wurden die Vpn nun gebeten eine Rangreihe zu bilden. Das Kriterium zur Bildung der Rangreihe war dabei wie auch in der Originalversion die Frage: „Welches gezeigte bzw. gehörte Verhalten würde ihrem eigenen Verhalten als Bankberater in einer solchen Situation am ehesten entsprechen?“. Nachdem die erste Rangreihe gebildet worden war, hatten die Vpn die Möglichkeit die Sequenzen ein weiteres Mal vorgespielt zu bekommen, um danach gegebenenfalls eine Modifikation ihrer Rangreihe vornehmen zu können. Am Ende des Versuchs wurden die Vpn gebeten eine Bewertung des Verfahrens

anhand des neun Items umfassenden Fragebogens vorzunehmen. Die gesamte Dauer des Verfahrens betrug ca. 2 Stunden.

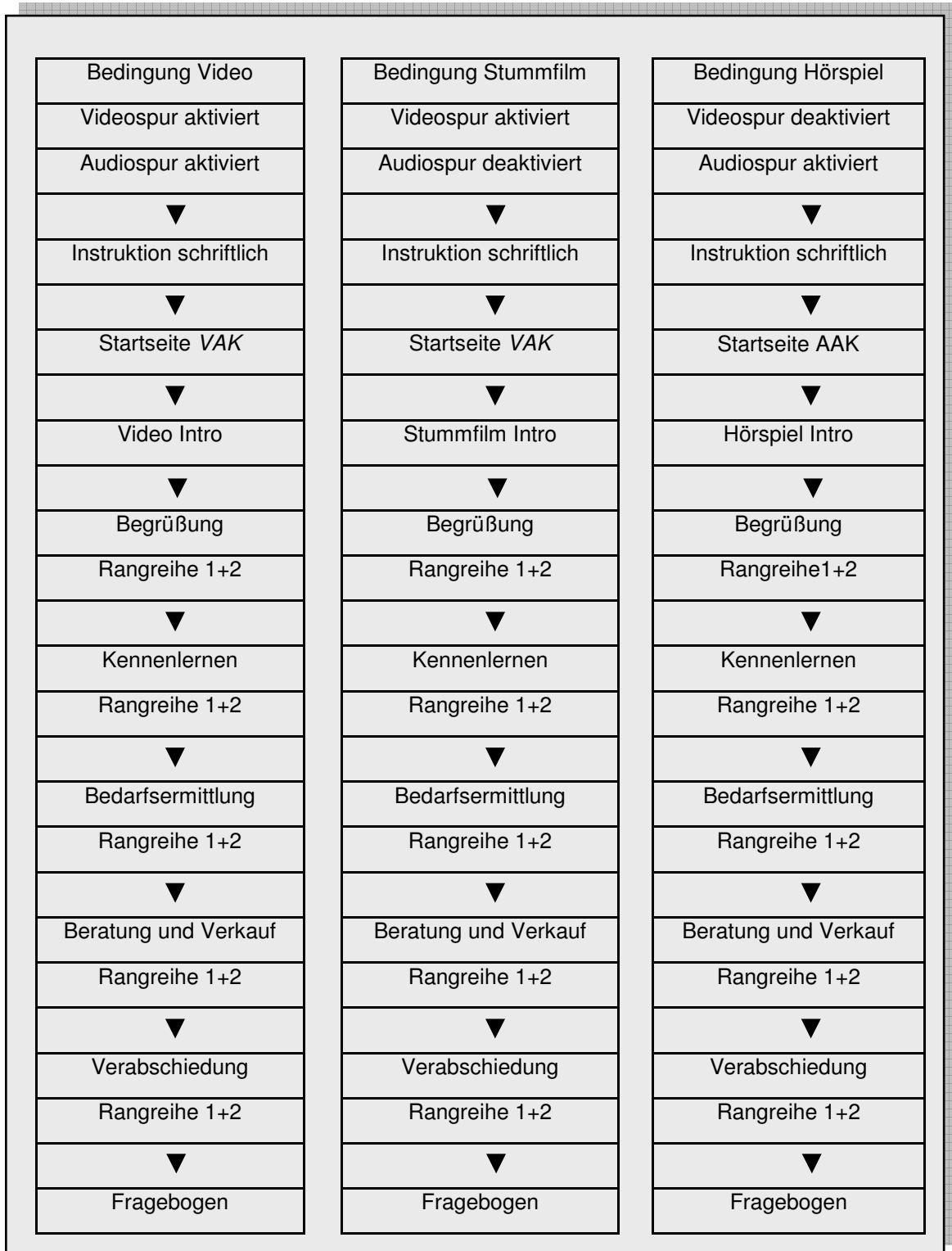


Abbildung 4.10: Graphische Darstellung des Versuchsablaufes

Design Fragestellung 1 und 2

Zur Beantwortung der ersten beiden Fragestellungen wurde eine Häufigkeitsanalyse mit Hilfe von hierarchischen log-linearen Modellen durchgeführt.

Im Folgenden soll zum besseren Verständnis des Lesers zuerst auf die Eigenschaften der betrachteten Häufigkeiten eingegangen werden. In einem zweiten Schritt werden dann die für diese Analyse interessierenden Faktoren vorgestellt, deren Einfluss auf die empirische Verteilung untersucht werden soll.

Als Maß für die Urteilsgüte der Probanden wurde die Übereinstimmung der empirisch erhobenen Rangplätze für die einzelnen Filmsequenzen in den fünf Abschnitten des VAK mit den korrekten und a priori festgelegten Rangplätze für die einzelnen Filmsequenzen erhoben. Konkret wurde ausgezählt, wie oft die Filmsequenz mit dem Rangplatz 1 diesen auch tatsächlich zugewiesen bekam und wie häufig sie den Rangplatz 2, 3, 4 und 5 erhielt. Daraufhin wurde für die Sequenz 2 ausgezählt, wie oft sie den Rangplatz 2 zugewiesen bekam und wie häufig ihr der Rangplatz 1, 3, 4 und 5 zugeteilt wurde. Dieses Vorgehen wiederholte sich für jede der 25 Filmsequenzen.

Da es sich bei den Rangplätzen nicht nur um verschiedene Kategorien, sondern um eine ordinale Abstufung handelt, wird im Ausgangsmodell der log-linearen Analyse nicht von einer Gleichverteilung der erwarteten Häufigkeiten ausgegangen, sondern von einer Bündelung der Beobachtung auf der in Abbildung 4.11 dargestellten Hauptdiagonalen. Praktisch führt dies dazu, dass eine Fehlzuweisung einer Sequenz umso stärker gewichtet wird, je weiter das empirische Urteil von der optimalen Vorgabe abweicht.

Diese Überlegung soll in Abbildung 4.11 für den Fall einer optimalen Zuordnung der Rangfolgen bei einem $N = 15$ Probanden und in Abbildung 4.12 für den Fall einer möglichen empirischen Zuordnung der Rangfolgen bei $N = 15$ Probanden zum besseren Verständnis des Lesers graphisch veranschaulicht werden.

		Zuweisung durch die Probanden				
		1	2	3	4	5
Theoretische optimale Rangfolge	1	15	0	0	0	0
	2	0	15	0	0	0
	3	0	0	15	0	0
	4	0	0	0	15	0
	5	0	0	0	0	15

Abbildung 4.11: Zellenhäufigkeiten bei einer optimalen Zuordnung der Videos aus Sequenz 1

		Zuweisung durch die Probanden				
		1	2	3	4	5
Theoretische optimale Rangfolge	1	9	3	2	1	0
	2	3	9	1	1	1
	3	2	0	9	1	2
	4	1	2	0	9	3
	5	0	1	2	3	9

Abbildung 4.12: Zellenhäufigkeiten bei einer empirischen Zuordnung der Videos aus Sequenz 1

Ausgehend von der oben beschriebenen Datenstruktur wurde nun mittels einer hierarchischen log-linearen Analyse überprüft, inwiefern sich die strukturelle Übereinstimmung der empirisch gefundenen Häufigkeitsverteilung in der Matrix mit einer theoretisch optimalen Häufigkeitsverteilung ändert, wenn verschiedene Faktoren in das Gesamtmodell einbezogen werden.

Bei den betrachteten Faktoren handelt es sich um den Faktor **Gruppe** mit seinen drei Stufen *Video*, *Stummfilm* und *Hörspiel*, um den fünffach gestuften Faktor **Sequenz** mit den Stufen *Begrüßung*, *Kennenlernen*, *Bedarfsermittlung*, *Beratung* und *Verkauf* sowie *Verabschiedung* und um den Faktor **Wiederholung** mit den Stufen *einmaliges Betrachten* und *mehrmaliges Betrachten*. Eine ausführliche Beschreibung findet der interessierte Leser bei (Tabachnik und Fidell, 2007, S. 858 ff.).

Ergebnisse Fragestellung 1 und 2

Vor der Darstellung der eigentlichen Ergebnisse sollen im Folgenden einige Besonderheiten der Analyse, die sich aus dem experimentellen Design ergeben, erläutert werden. Da jede Stufe des Faktors **Optimal** gleich häufig vorgegeben wurde, kann der Haupteffekt dieses Faktors vernachlässigt werden. Gleiches gilt für den Faktor **Empirisch**, da jede Platzierung von jedem Probanden gleich häufig vergeben werden musste. Demzufolge sind auch sämtliche Zweifachinteraktionen, die je einen der Faktoren **Empirisch** oder **Optimal** beinhalten, zu vernachlässigen. So können sich die Gruppen beispielsweise nicht darin unterscheiden, wie häufig der Rangplatz 1 von ihnen vergeben wurde. Inhaltlich von Interesse ist das Ausmaß der Beurteilungsgüte, also der Übereinstimmung der empirischen mit der optimalen Reihenfolge, und wie dieses Maß zwischen den verschiedenen Gruppen, Sequenzen und Wiederholungsmodi variiert. Die Beurteilungsgüte wird statistisch in der Interaktion der Faktoren **Empirisch** und **Optimal** erfasst. Um nun bestimmen zu können, ob sich die Beurteilungsgüte z.B. zwischen den Videosequenzen unterscheidet, wird die Dreifachinteraktion (Empirisch x Optimal x Sequenz) betrachtet. Die folgenden loglinearen Modelle werden also sukzessive überprüft, ob die Aufnahme der jeweiligen Dreifachinteraktion in das Modell zu einer hinreichenden Passung von erwarteter und beobachteter Häufigkeitsverteilung führt. Eine solche ist durch einen nichtsignifikanten χ^2 Wert definiert.

Die berichteten Ergebnisse beziehen sich auf die Daten von N=60 Probanden, da weder fehlende Werte noch sonstige Ausschlussgründe für einzelne Probanden vorlagen.

Als Ausgangsmodell wird die Interaktion zwischen den Faktoren empirisch und optimal betrachtet. Die Analyse ergibt, dass sich die beobachteten Häufigkeiten nicht allein aufgrund dieser Interaktion erklären lassen. ($\chi^2(725) = 1072.805, p < .001$). Die zusätzliche Aufnahme der Dreifachinteraktion (Empirisch x Optimal x Wiederholung) führt nur zu einer sehr geringen Reduktion der Abweichung, die nach wie vor hoch signifikant ist ($\chi^2(700) = 1057.575, p < .001$). Die Beurteilungsgüte unterscheidet sich zwischen den betrachteten Sequenzen also nicht.

Wird das Ausgangsmodell um die Dreifachinteraktion (Empirisch x Optimal x Gruppe) erweitert, zeigt sich ein vergleichbares Bild ($\chi^2(675) = 904.498, p < .001$). Die Beurteilungsgüte unterscheidet sich damit auch zwischen den betrachteten Gruppen nicht.

Die Erweiterung des Ausgangsmodells um die Dreifachinteraktion (Empirisch x Optimal x Sequenz) dagegen ergibt erwartete Häufigkeiten, die nicht mehr signifikant von den beobachteten abweichen ($\chi^2(625) = 584.818, p < .873$). Hieraus ergeben sich zwei Schlussfol-

gerungen. Zum einen scheint die Beurteilungsgüte sich zwischen den betrachteten Sequenzen zu unterscheiden, zum anderen ist die Berücksichtigung weiterer Faktoren und Faktorwechselwirkungen nicht nötig, da die erwarteten Häufigkeiten bereits in ausreichendem Maße erklärt werden.

Design Fragestellung 3

Zur Beantwortung der Fragestellung nach Unterschieden in der Bewertung des VAK unter Berücksichtigung der Gruppenzugehörigkeit wurde ein einfaktorielles varianzanalytisches Design gewählt. Als unabhängige Variablen gingen der Faktor **Gruppe** mit den drei Stufen *Video*, *Stummfilm* und *Hörspiel* in die Analyse ein. Bei den betrachteten abhängigen Variablen handelte es sich um die über alle Probanden gemittelten Werte der neun Items Fragebogens.

Die statistischen Voraussetzungen für die Durchführung eines solchen Verfahrens sind nach Bortz (2006), neben einem geforderten Intervallskalenniveau und der Normalverteilung der Variablen, die Homogenität der Varianzen und die Normalverteilung der Residuen. Alle beschriebenen Forderungen an die Datenstruktur wurden im Falle der vorliegenden Analyse erfüllt.

Ergebnisse Fragestellung 3

In die Analyse gingen die Daten von 59 Probanden ein. Ein Proband der Gruppe Video musste ausgeschlossen werden, da er vergessen hatte den Fragebogen auszufüllen. Die Werte für die Items *Schwierigkeit*, *Dauer* und *Anstrengung* wurden zur besseren Übersicht für die Analyse der Daten dahingehend umgepolt, dass auch hier ein hoher Wert eine positive Antwort repräsentiert.

Ein Vergleich der Bewertungen zwischen den Gruppen mittels einer einfaktoriellen Varianzanalyse für unabhängige Stichproben zeigte lediglich einen signifikanten Unterschied in der Bewertung der Dauer des Verfahrens ($F(2,56) = 4.445$; $p < .05$).

Im Anschluss wurde deshalb mittels einer Post Hoc Prüfung nach Scheffè untersucht, zwischen welchen der drei Gruppen ein signifikanter Unterschied bei der Bewertung der Dauer des Verfahrens besteht. Es zeigte sich ein signifikanter Unterschied zwischen der Gruppe *Video* und der Gruppe *Stummfilm* ($p < .05$) nicht aber zwischen den Gruppen *Video* und *Hörspiel* ($p = .099$, ns) und zwischen den Gruppen *Stummfilm* und *Hörspiel* ($p = .801$, ns).

Für den interessierten Leser zeigt Abbildung 4.13 eine graphische Darstellung der Bewertung der 9 Items getrennt nach Gruppen.

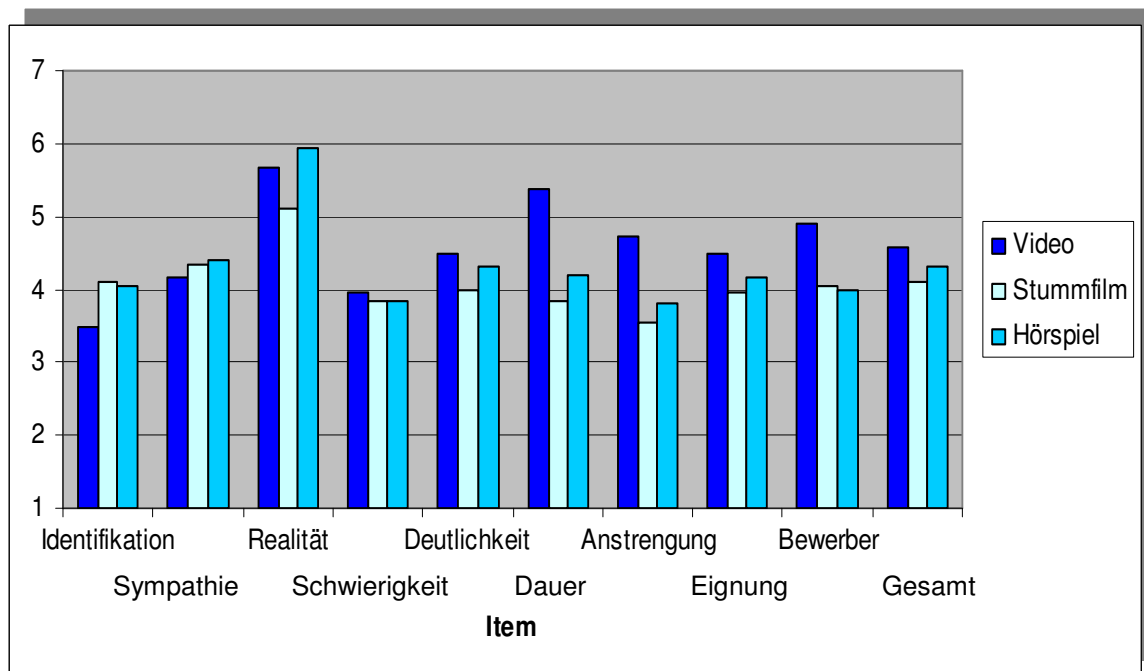


Abbildung 4.13: Mittelwerte der Items unter Berücksichtigung der Gruppenzugehörigkeit

Design Fragestellung 4

Um die Frage beantworten zu können, ob sich ein Unterschied in der Bewertung der Filmsequenzen für die Gruppe Video in Abhängigkeit vom Geschlecht der Versuchspersonen zeigt, wurde eine einfaktorielle Varianzanalyse für unabhängige Stichproben berechnet. Als unabhängige Variable ging der Faktor Geschlecht mit den Stufen männlich und weiblich in die Analyse ein. Bei den betrachteten abhängigen Variablen handelte es sich um die Skalenmittelwerte der oben benannten 9 Items.

Ergebnisse Fragestellung 4

In die Analyse gingen die Daten von 19 Probanden ein, davon waren 10 weiblich und 9 männlich. Ein männlicher Proband musste ausgeschlossen werden, da er vergessen hatte den Fragebogen auszufüllen. Die Werte für die Items *Schwierigkeit*, *Dauer* und *Anstrengung* wurden zur besseren Übersicht für die Analyse der Daten dahingehend umgepolt, dass auch hier ein hoher Wert eine positive Antwort repräsentiert.

Der Vergleich der Stichprobenmittelwerte der männlichen und weiblichen Probanden mittels einer einfaktoriellen Varianzanalyse ergab für keines der untersuchten 9 Items einen

signifikanten Unterschied. Auch die Betrachtung der Gesamtbewertung als Mittelwert über alle Skalen zeigte hier keinen Effekt.

Für den interessierten Leser zeigt Abbildung 4.14 eine graphische Darstellung der Bewertung der 9 Items für die Gruppe Video getrennt nach Geschlecht.

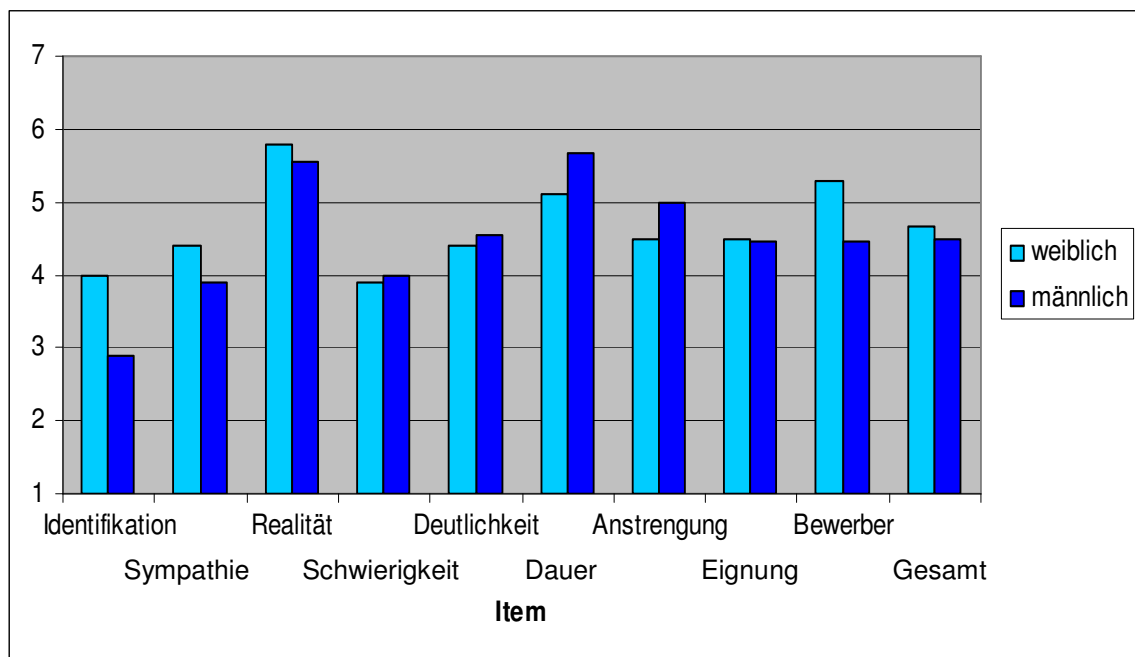


Abbildung 4.14: Mittelwerte der Items unter Berücksichtigung des Geschlechts der Teilnehmer

Diskussion der Ergebnisse in Bezug auf die Hauptuntersuchung

Betrachtet man die Ergebnisse aus der mit Hilfe von hierarchischen log-linearen Modellen durchgeführten Häufigkeitsanalyse, ergibt sich ein positives Fazit. Die erste Fragestellung, ob die Manipulation des Beraterverhaltens in den Videos, in Bezug auf die Variation von visuellen Hinweisreizen (Mimik, Gestik, Körperhaltung und Handlungen) und auditiven Hinweisreizen (Sprache und Inhalte) gleich erfolgreich gelungen ist, kann mit ja beantwortet werden. Gleiches gilt für die Frage, ob diese beiden Gruppen von Hinweisreizen für die jeweiligen Videosequenzen betrachtet ein kongruentes Gesamtbild ergeben. Die Berücksichtigung des Faktors *Gruppe* führte zu keiner signifikanten Verbesserung der Modellpassung. Dies zeigt, dass sich die Leistungen der Probanden in den drei Versuchsbedingungen als gleich gut bezeichnen lassen.

Auch der zweite Aspekt, welcher auf die Angemessenheit der Dauer der einzelnen Videosequenzen fokussierte, kann positiv beantwortet werden. Die Passung des Modells unter Berücksichtigung des Faktors Wiederholung zeigt, dass die Länge der Sequenzen und damit die Menge an audiovisuellen Informationen, die dem Probanden zur Entscheidungsfindung zur Verfügung gestellt wird, so bemessen ist, dass keine Überforderung im Sinne der Fülle des Materials vorliegt. Im gegenteiligen Fall hätten die Probanden von der Möglichkeit einer wiederholten Darbietung der Videos im Sinne einer höheren Leistung profitieren müssen. Da die Aufnahme des Faktors *Wiederholung* aber ebenfalls zu keiner bedeutsamen Modellverbesserung beitrug, kann diese Annahme zurückgewiesen werden.

Beide Ergebnisse bestätigen die Annahmen bezüglich einer gelungenen Manipulation des Beraterverhaltens in den Videosequenzen des VAK unter Berücksichtigung von auditiven und visuellen Stimuli. In Bezug auf den Aspekt der inhaltlichen Gestaltung der Videos wurde deshalb kein weiterer Verbesserungsbedarf gesehen. Auch die Länge der Videos erscheint aufgrund der Ergebnisse angemessen.

Die einzige signifikante Verbesserung der Modellpassung wurde bei der Aufnahme des Faktors Sequenz beobachtet. Dies wurde als Hinweis für die durchaus erwünschte Tatsache betrachtet, dass es sich bei den Sequenzen um Aufgaben unterschiedlicher Schwierigkeit handelt bzw. das VAK eine leicht heterogene Aufgabenstruktur aufweist. Beide Aspekte wurden in die Planung der weiteren Voruntersuchungen aufgenommen.

Die Ergebnisse bezüglich der dritten Fragestellung, inwieweit sich signifikante Unterschiede bei der Einschätzung des Verfahrens durch die Probanden zwischen den drei Gruppen ergeben würden, zeigt bis auf den Aspekt der bewerteten Dauer der Verfahren in den Versuchsbedingungen keine signifikanten Ergebnisse. Eine mögliche Erklärung könnte darin begründet sein, dass ein solch neuartiges Verfahren von den Versuchspersonen generell als interessant und damit positiv wahrgenommen wird, unabhängig von der Art der Präsentationsform. Lediglich die wahrgenommene Dauer der Bedingung Video und der Bedingung Stummfilm zeigte einen signifikanten Unterschied ($p < .05$). Dies erscheint logisch, wenn man bedenkt, dass sich die Gruppe der Stummfilmbedingung ca. 60 Minuten mit dem Betrachten von Filmszenen begnügen musste, ohne irgendwelche inhaltlichen Aspekte angeboten zu bekommen. Die Beurteilungen der Gruppe Video im positiven Bereich der Skala stellen ein zufriedenstellendes Ergebnis dar. Vor allem der hohe Wert für

die Realitätsnähe zeigt, dass das Ziel, das Beratungsgeschehen in einer Geschäftsstelle sehr real abzubilden, erreicht werden konnte.

Als letzter Aspekt wurde untersucht, inwieweit sich ein Unterschied in der Bewertung zwischen weiblichen und männlichen Probanden für die Bedingung Video zeigen würde. Da es sich bei beiden Darstellern in den Videos um männliche Rollenspieler handelt, sollte eine mögliche Auswirkung dieses Umstandes auf die Einschätzung des Verfahrens überprüft werden. Hier konnte für keines der betrachteten Items ein signifikanter Unterschied gefunden werden. Die Bewertung des *VAK* erfolgt für beide Geschlechter gleich gut, was einen Einsatz sowohl bei männlichen als auch bei weiblichen Probanden zulässt, ohne mögliche Unterschiede in der Motivation der Kandidaten, bedingt durch Unterschiede in der Wahrnehmung, zu riskieren.

Nachdem die Untersuchung des im *VAK* eingesetzten Materials unter multimedialen Gesichtspunkten in einem zufriedenstellenden Ergebnis resultierte, wurde eine zweite Studie mit dem Ziel durchgeführt, erste Ergebnisse zur Konstruktvalidität des *VAK* zu erhalten.

4.3.2 Vorstudie 2: Empirische Betrachtung der Konstruktvalidität des VAK

Theoretische Motivation

In Kapitel 2.8.5 der vorliegenden Arbeit wurde anhand empirischer Ergebnisse zur Konstruktvalidität von videobasierten Auswahlverfahren dargelegt, dass sich sowohl kognitive Fähigkeiten als auch spezifische soziale Kompetenzen mit dieser Art von Verfahren messen lassen. In diesem Zusammenhang wurde auch auf die Bedeutung der Instruktionkomponente für die konstruktorientierte Validität hingewiesen. So konnte durch die beschriebenen Studien gezeigt werden, dass *should do* Instruktionen eher geeignet sind, um kognitive Maße zu erheben, während mittels *would do* Instruktionen erhobene Maße eher mit Aspekten der spezifischen sozialen Kompetenz oder Berufserfahrung korrelieren. Nach Ployhart und Erhart (2003) zeigten sich diese Unterschiede selbst bei inhaltlich gleichen Items. Da das VAK konstruiert wurde, um eine spezifische soziale Kompetenz zu erheben, die als *Kundenorientierung* bezeichnet wird, und dementsprechend eine *would do* Instruktion eingesetzt wurde, sollte überprüft werden, inwieweit sich das im VAK gemessene Konstrukt von kognitiven Maßen und Maßen einer berufunspezifischen sozialen Kompetenz abgrenzen lässt.

Fragestellung 1a

Die erste Fragestellung bezieht sich auf den Aspekt, inwieweit ein Zusammenhang zwischen der im VAK erbrachten Leistung von Probanden und ihrer sprachlichen Intelligenz, ihrer generellen berufsbezogenen Kompetenzen und ihrer Konzentrationsfähigkeit besteht. Betrachtet man die als wesentlich erachteten Verhaltenshinweise aus den Experteninterviews, so wird deutlich, dass sich das Konstrukt *Kundenorientierung* aus sehr spezifischen Fähigkeiten zusammensetzt, die es einem Berater ermöglichen in einer Beratungssituation adäquat zu agieren und deren Erwerb in der Regel einer längeren Ausbildung bedarf. Da das VAK versucht, diese spezifischen Fähigkeiten nach dem Ende einer Ausbildung zu erfassen, dürfte allein eine hohe Ausprägung in eher allgemeinen kognitiven oder sozialen Fähigkeitsbereichen nicht ausreichen, um die Übungen erfolgreich zu lösen. Aus diesem Grund wird angenommen, dass die Werte in den Bereichen sprachliche Intelligenz, allgemeine berufsbezogene soziale Kompetenz und Konzentrationsfähigkeit nicht systematisch mit einer hohen Ausprägung in der vom VAK erfassten Fähigkeit einhergehen und im Sinne einer divergenten Validität eher niedrig korrelieren.

Fragestellung 1b

Während mit der Fragestellung unter 1a der Versuch unternommen wird das im VAK gemessene Konstrukt von anderen als eher divergent erachteten Konstrukten abzugrenzen, verfolgt die Fragestellung 1b den Ansatz der Extremgruppenvalidierung. Für die Wahl eines solchen Ansatzes spricht die Überlegung, dass sich eine spezifische soziale Kompetenz erst nach einer gewissen Erfahrung in einem spezifischen Anwendungskontext bilden kann. Hieraus folgt weiterhin, dass der Vergleich von Gruppen unterschiedlicher Expertise und Beratungserfahrung zu signifikanten Unterschieden bei den durch diese Gruppen erzielten Punktescores im VAK führen sollte, und zwar derart, dass der Punktescore umso höher ausfällt, je mehr Expertise bzw. Beratungserfahrung vorhanden ist. Um diesen Ansatz verfolgen zu können, wurden die Scores von drei Gruppen miteinander verglichen. Die erste Gruppe bestand aus $N = 30$ Studierenden, welche nicht über eine Ausbildung im Finanzdienstleistungssektor verfügten. Die zweite Gruppe bestand aus $N = 17$ Auszubildenden zum Bankkaufmann, welche sich im dritten Ausbildungsjahr befanden. Die dritte Gruppe setzte sich aus $N = 5$ Beratungsexperten zusammen, welche über eine mehrjährige Berufserfahrung verfügten und im vergangenen Geschäftsjahr eine Zielerreichungsquote in Bezug auf die jährliche Beratungsleistung von mindestens 100% nachweisen konnten.

Fragestellung 2

Ein zweiter Aspekt galt der Messung der Akzeptanz des Verfahrens, wenn dieses in seiner Originalversion durchgeführt wird. In diesem Rahmen sollte gleichzeitig der Fragebogen zur Erhebung der sozialen Validität und wahrgenommenen Fairness vorgetestet werden, der in der späteren Hauptuntersuchung Verwendung finden sollte.

Stichprobe

Zur Überprüfung der oben genannten Fragen konnten $N = 29$ Studentinnen und Studenten der Westfälischen Wilhelms-Universität Münster (WWU) und ein Teilnehmer ohne Angabe eines Studienganges gewonnen werden. Die fachliche Ausrichtung der 29 Studenten verteilte sich dabei wie folgt: Psychologie 15, Lehramt 5, Magisterstudiengänge 5, Sportwissenschaften 2, Biologie 1 und Informatik 1. Alle Versuchspersonen nahmen freiwillig an der Untersuchung teil und hatten die Wahl sich entweder Versuchspersonenstunden gutschreiben zu lassen oder einen Stundenlohn von 10 Euro zu erhalten. Die Stichprobe setzte sich aus 21 weiblichen und 9 männlichen Probanden im Alter von 20 bis 35 Jahren

($M=24$, $SD= 3,085$) zusammen. Diese Zusammensetzung entsprach sowohl in Bezug auf das Alter als auch in Bezug auf den vorherrschenden Bildungsgrad (100% der Versuchspersonen in der Stichprobe verfügten über ein Abitur) und die Verteilung der Geschlechter (70% weiblich und 30% männlich), der Zusammensetzung der Probanden in der späteren Hauptuntersuchung in zufriedenstellender Weise.

Material

Im zweiten Vortest kamen neben dem *VAK* in seiner Originalform drei Testverfahren zum Einsatz, die im Folgenden kurz vorgestellt werden.

Zur Erhebung der sprachlichen Intelligenz wurde auf drei Untertests des *I-S-T 2000* von Amthauer, Brocke, Liepmann und Beauducel (1999) zurückgegriffen. Diese auch als Verbalskalen bezeichneten Untertest *Satzergänzung*, *Analogien bilden* und *Gemeinsamkeiten finden* erfassen nach Angaben der Testautoren die Fähigkeit zum Umgang mit sprachlichem Material im Rahmen des schlussfolgernden Denkens. Die Ausprägung der verbalen Intelligenz wird hierbei durch das Aufsummieren der Ergebnisse aus den drei Untertests ermittelt, welche jeweils aus 20 Items aufsteigender Schwierigkeit bestehen. Die Bearbeitungszeit der Verbalskalen ist zeitlich begrenzt und dauert inklusive der Instruktionszeit insgesamt ca. 25 -30 Minuten. Der IST-2000 weist eine hohe Durchführungs-, Auswertungs- und Interpretationsobjektivität auf. Die berichteten Reliabilitäten der drei hier verwendeten Untertests werden für die betrachtete Normgruppe mit $\alpha = .51$, $\alpha = .66$ und $\alpha = .66$ (Cronbachs alpha) und $r = .53$, $r = .68$ und $r = .70$ (Split-Half-Reliabilität) angegeben. Der IST-2000 verfügt über eine differenzierte Normierung, die unter anderem Normentabellen für vier Gymnasialgruppen im Alter von 15 bis 25 Jahren beinhaltet, die auf einer Stichprobengröße von $N = 1285$ Probanden beruhen (Amthauer, Brocke, Liepmann und Beauducel, 1999; Wittmann, 2003).

Zur Erhebung der Konzentrationsfähigkeit diente der *KLT* von Düker und Lienert (2001). Das Lösen der Aufgaben aus dem *KLT*, welche aus Rechenaufgaben bestehen, soll die Koordination einer Reihe von Einzeltätigkeiten erfordern, welche als *Auffassen*, *Rechnen*, *Merken*, *Regelabruf* und *Entscheiden* bezeichnet werden. Mit dem Verfahren soll nach Lukesch und Mayerhofer (2001) die Fähigkeit zur Konzentration über einen langen Zeitraum gemessen werden. Die Form für Jugendliche und Erwachsene, die in dieser Untersu-

chung Verwendung fand, besteht aus neun Aufgabenblöcken, die sich aus jeweils 20 Rechenaufgaben zusammensetzen. Die Bearbeitungszeit pro Block beträgt zwei Minuten. Die Objektivität kann in Bezug auf die Durchführung, Auswertung und Interpretation als gewährleistet angesehen werden. Die nach der Zeitfraktionierungsmethode ermittelte Split-half Reliabilität zeigt für die Leistungsmenge zufriedenstellende Koeffizienten von $r = .94$ und $r = .96$ bei einer Stichprobe von $N = 120$ Oberschülern bzw. $N = 120$ Berufsschülern. Auch die berichteten konvergenten und divergenten Validitäten des *KLT* liegen in einem akzeptablen Rahmen. Der *KLT* verfügt über eine differenzierte Normierung (Andree, 2002).

Bei dem dritten in dieser Untersuchung verwendeten Test handelte es sich um das *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP)* von Hossiep und Paschen (1998). Das Instrument wurde für die Persönlichkeitsdiagnostik mit dem Schwerpunkt Personalauswahl Berufsberatung und Training und Coaching entwickelt. Die Konstruktion des Fragebogens erfolgte nach Angaben der Autoren atheoretisch, mit dem Focus auf eine hohe Augenscheinvalidität, weshalb inhaltlichen Erwägungen der Vorrang vor teststatistischen Überlegungen gegeben wurde (Hossiep und Paschen, 1998). Der Test beansprucht 14 berufsbezogene Persönlichkeitsmerkmale zur erfassen, die in vier Bereiche (*Berufliche Orientierung, Arbeitsverhalten, Soziale Kompetenzen und Psychische Konstitution*) zusammengefasst werden. Die Erfassung der 14 Persönlichkeitsmerkmale erfolgt anhand von 210 Aussagen zu Verhaltensweisen in beruflichen Situationen, die vom Probanden, in Form einer Selbsteinschätzung, auf einer sechsstufigen Ratingskala bewertet werden. Die Bearbeitungszeit liegt bei durchschnittlich 45 Minuten. Die Durchführungs- und Auswertungsobjektivität kann nach Kanning (2003) als hoch angesehen werden, auch die Interpretation der Ergebnisse wird durch Beispielfälle erläutert. Die an einer Stichprobe von $N = 5354$ Probanden ermittelte innere Konsistenz nach Cronbachs alpha liegt für die einzelnen Dimensionen zwischen $\alpha = .75$ und $\alpha = .92$. Die Retestreliabilität (8 bis 10 Wochen, $N = 108$) zwischen $r = .77$ und $r = .89$. Die an unterschiedlichen Außenkriterien erhobenen Validität liegen zwischen $r = .00$ und $r = .31$. Eine Befragung nach der Akzeptanz ergab bei $N = 1573$ Teilnehmern ein positives Ergebnis (Hossiep und Paschen, 1998, Kanning 2003). Es liegt eine nach Geschlecht, Alter und beruflichem Status differenzierte Normierung vor (Gesamtstichprobe $N = 5354$).

Versuchsablauf

Alle 30 Probanden absolvierten die Untersuchung im medienpsychologischen Labor der Westfälischen Wilhelms-Universität Münster.

Die Probanden wurden, nach der Begrüßung durch den Versuchsleiter, schriftlich über den Ablauf sowie die Dauer der vor ihnen liegenden Untersuchung hingewiesen, wodurch eine größtmögliche Standardisierung gewährleistet wurde. Nach dieser Informationsphase folgte die Bearbeitung des *KLT*, die ca. 20 Minuten in Anspruch nahm. Nach einer Pause von fünf Minuten folgte dann die Bearbeitung der drei Verbalskalen zur Erfassung der fluiden verbalen Intelligenz aus dem *IST 2000*. Die Anweisungen zur Bearbeitung der Aufgaben erfolgte, wie auch beim *KLT*, durch den Versuchsleiter, der auch auf die Einhaltung der zeitlichen Vorgaben aus den Testmanualen achtete. Die drei Aufgabenblöcke mit je 20 Aufgaben waren in sechs, sieben und acht Minuten zu bearbeiten, so dass zuzüglich der Zeit für die Instruktion durch den Versuchsleiter eine Bearbeitungszeit von ca. 30 Minuten resultierte.

Nach einer weiteren fünfminütigen Pause wurden die Probanden dann gebeten das *Video-basierte Analyseverfahren für Kundenberater* in seiner Originalversion von Anfang bis Ende zu bearbeiten. Die Bearbeitung des *VAK* unterlag keiner zeitlichen Begrenzung und die Probanden erhielten den Hinweis, dass sie die Möglichkeit hätten, sich die Filmsequenzen auch ein weiteres Mal anzuschauen. Die so resultierende Bearbeitungszeit betrug für das *VAK* 60 Minuten. Im Anschluss erfolgte die Beantwortung des Fragebogens zur sozialen Validität, dessen Bearbeitungszeit bei 10 Minuten lag. Hieran schloss sich nach einer abermaligen fünfminütigen Pause die Bearbeitung des *Bochumer Inventars zur berufsbezogenen Persönlichkeitsbeschreibung* an. Ergänzend zu den schriftlichen Hinweisen der Testautoren auf dem Antwortbogen wurden die Probanden von Seiten der Versuchsleiter darum gebeten sich bei der Beantwortung der Fragen auf ihre Erfahrung in Praktika, Studentenjobs oder das Studium zu beziehen und nicht etwa, beeinflusst durch das *VAK*, in der hypothetischen Person eines Kundenberaters zu antworten. Die Bearbeitung des *BIP* dauerte ca. 40 Minuten, so dass für die gesamte Untersuchung ein zeitlicher Aufwand von ca. drei Stunden resultierte.

Der gesamte Ablauf der beschriebenen Verfahren wird in Abbildung 4.15 graphisch veranschaulicht.

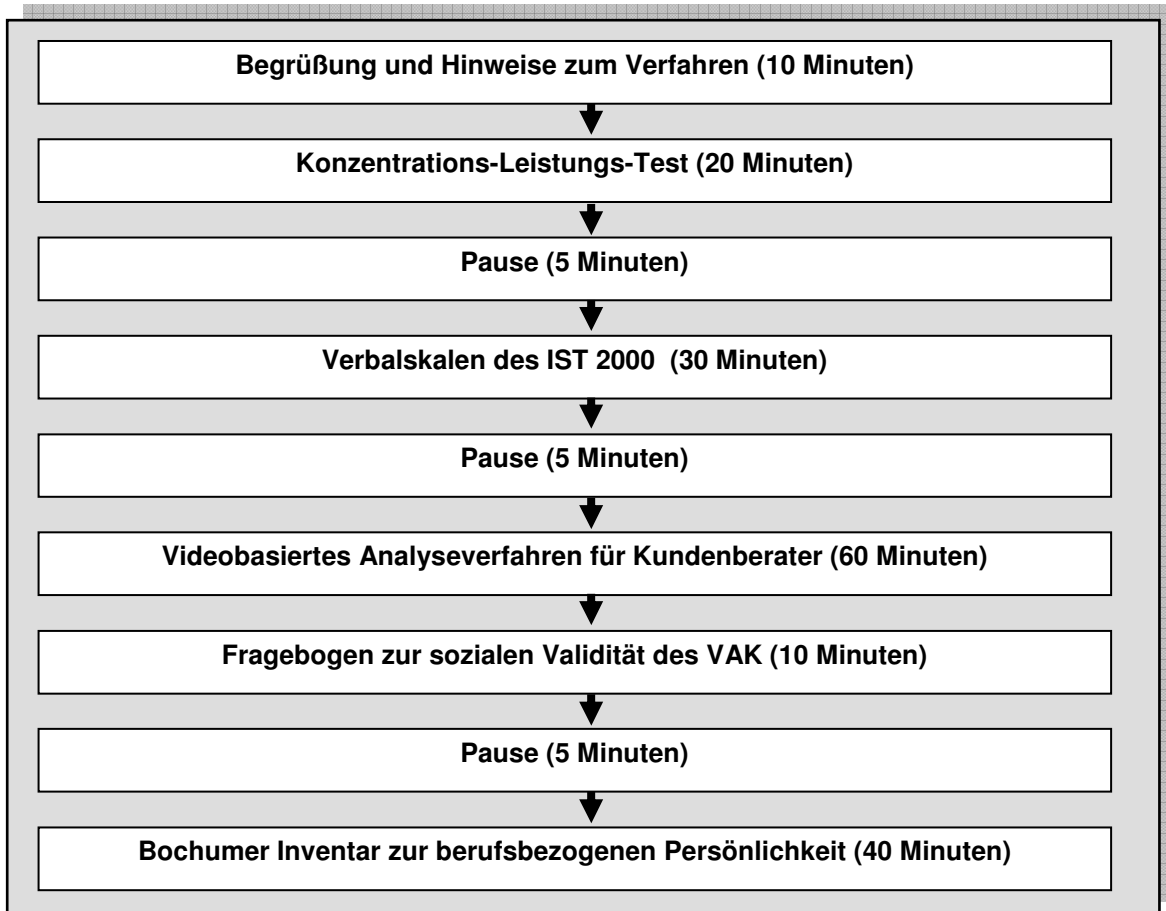


Abbildung 4.15: Graphische Darstellung des Versuchsablaufs (Voruntersuchung 2)

Design Fragestellung 1a

Um die Frage beantworten zu können, inwieweit sich das Abschneiden eines Probanden im VAK durch seine Fähigkeiten in den Bereichen *sprachliche Intelligenz*, *Konzentrationsvermögen* sowie in den Bereichen *Berufliche Orientierung*, *Arbeitsverhalten*, *Soziale Kompetenzen* und *Psychische Konstitution*, gemessen nach dem *BIP*, vorhersagen lässt, wurde eine multiple schrittweise Regression gerechnet. Als Kriterium diente der Gesamtscore des VAK (UV) als Prädiktorvariablen die Ergebnisse aus den 3 Testverfahren (AV).

Alle verwendeten Testverfahren produzieren Daten auf Intervallskalenniveau und die Verteilung der Daten aller drei Tests und der Gesamtleistung im VAK hielten einer Überprüfung auf Normalverteilung nach Kolmogoroff-Smirnoff ($Z = .649$, $p = .793$, ns) stand.

Ergebnisse Fragestellung 1a

In die Analyse der Daten mittels einer multiplen Regression konnten die Werte von $N = 29$ Versuchspersonen einbezogen werden. Ein Proband äußerte, bei der Abgabe der Unterlagen, gegenüber dem Versuchsleiter Vorbehalte in Bezug auf die Qualität des IST 2000 als diagnostisches Instrument. Eine unmittelbare Analyse der Testleistung ergab einen stark unterdurchschnittlichen Wert. Die Daten dieses Probanden wurden von der Betrachtung ausgeschlossen, da eine mangelnde Motivation während der Testung nicht ausgeschlossen werden konnte.

Die Analyse der Daten mittels der multiplen Regression unter Einbezug aller Daten zeigte kein gültiges Modell. Die Überprüfung des ermittelten R^2 von .260 als Maßzahl für den Anteil an Kriteriumsvarianz, der durch die Prädiktoren aufgeklärt wird, mittels eines F-Tests ergab kein signifikantes Ergebnis, $F(7,21) = 1.05$, $p = .425$, *ns*).

Eine komplette Übersicht aller Korrelationskoeffizienten bietet die Tabelle 4.2.

Tabelle 4.2: Korrelationen zwischen dem VAK und dem IST 2000, KLT und BIP

	IST 2000: Verbal	BIP: BO	BIP: AV	BIP: SK	BIP: PK	KLT: GL	KLT: FP
VAK	-.27 (.08)	-.28 (.07)	-.18 (.18)	.09 (.31)	.18 (.18)	.04 (.42)	.14 (.24)

Die Daten in den Zellen zeigen die ermittelten Produkt-Moment-Korrelationen nach Pearson.

Die Daten in Klammern zeigen die Signifikanzen

Design Fragestellung 1b

Um die Frage nach einem signifikanten Unterschied zwischen den Testresultaten von Gruppen mit einem unterschiedlichen Grad an Expertise und Beratungserfahrung beantworten zu können, wurde mit dem Kruskal-Wallis-Test auf ein verteilungsfreies Verfahren zurückgegriffen, da die Voraussetzungen für die Berechnung einer Varianzanalyse verletzt waren. Die Ermittlung von Unterschieden zwischen den drei Gruppen auf Einzel Ebene wurde post hoc mittels Mann-Whitney-Test durchgeführt.

Ergebnisse Fragestellung 1b

Die Betrachtung von Mittelwertsunterschieden in der Gesamtleistung des VAK zwischen den drei nach dem Grad an Expertise unterschiedenen Gruppen mittels Kruskal-Wallis-Test zeigte mit ($\chi^2 = 20.042$, (2), $p < .001$) ein hoch signifikantes Ergebnis.

Auch die Ergebnisse der Einzelvergleiche mittels Mann-Whitney-Test erwiesen sich als signifikant. Sowohl zwischen den Gruppen „Experten“ und „Azubis“ ($U = -3.011$, $p < .01$) als auch zwischen den Gruppen „Experten“ und „Laien“ ($U = -3.449$, $p < .01$) und den Gruppen „Azubis“ und „Laien“ ($U = -3.115$, $p < .01$) zeigten sich signifikante Unterschiede.

Design Fragestellung 2

Die Auswertung der Einschätzungen des Verfahrens durch die Probanden mittels des oben beschriebenen Fragebogens erfolgt deskriptiv. Hierbei werden die einzelnen Items so zu Skalen zusammengefasst, wie es das theoretische Modell nach Schuler und Strehle (1995) vorsieht. Auf eine faktorenanalytische Absicherung der Modellannahmen muss im Rahmen dieser Voruntersuchung verzichtet werden, da die Voraussetzungen, die ein solches Verfahren an die Größe der Stichprobe stellt, in diesem Fall verletzt sind. Für Stichprobengrößen von $N < 60$ kann die Faktorenstruktur nur dann generalisierend interpretiert werden, wenn auf jedem bedeutsamen Faktor mindestens 4 Variablen Ladungen über .60 aufweisen (Bortz, 2006). Diese Vorgaben werden Seitens der Struktur der vorliegenden Daten nicht erfüllt. Auch eine weitere Voraussetzung nach Bühner (2006) konnte nicht als erfüllt angesehen werden. Der Kaiser-Meyer-Olkin-Koeffizient lag für den analysierten Datensatz bei .388 und damit in einem Bereich, der nach Bühner für die Durchführung einer Faktorenanalyse inkompatibel ist, da der Grenzwert bei $> .50$ liegt. Auch die Betrachtung der Anti-Image-Matrix zeigt durchgängig niedrige Werte in der Diagonale. Die MSA-Koeffizienten für lediglich vier der betrachteten Items über .06 und damit nach Bühner (2006) in einem zufriedenstellend hohen Bereich.

Da zwar eine Überprüfung der Passung der in Kapitel 4.2.5 beschriebenen, theoretisch gebildeten Itemgruppen im Rahmen dieser Arbeit als unverzichtbar angesehen wird, um eine sinnvolle Interpretation der Daten zur sozialen Validität in der Hauptstudie zu gewährleisten, der in dieser Vorstudie eingesetzte Fragebogen aber lediglich eine Vorversion des in Kapitel 4.2.5 beschriebenen Instruments darstellt, wird auf eine Analyse der zugrundeliegenden Struktur verzichtet. Die Ergebnisse sollen aus diesem Grund, wie bereits in Studie 1, lediglich als erster Hinweis darauf gewertet werden, wie das VAK, wenn es in seiner Originalversion durchgeführt wird, von den Probanden wahrgenommen wird.

Ergebnisse Fragestellung 2

Wie aus Abbildung 4.16 ersichtlich, liegt die ermittelte Gesamtnote für das VAK bei 4.53 Punkten und damit im positiven, zustimmenden Bereich der siebenstufigen Skala. Eine Betrachtung auf Ebene der fünf einzelnen oben genannten Aspekte zeigt nur für den Bereich der eingeschätzten Fairness ($M=3.87$, $SD=.87$) des Verfahrens einen Wert, der im negativen Bereich der Skala liegt. Die anderen Aspekte wie die Art der *Information* ($M=5.23$, $SD=1.03$), die wahrgenommene eigene *Kontrolle* ($M=4.17$, $SD=.78$) sowie die erlebte *Transparenz* ($M=4.87$, $SD=.84$) und die *Akzeptanz* ($M=4.70$, $SD=.84$) lagen im positiven Skalenbereich.

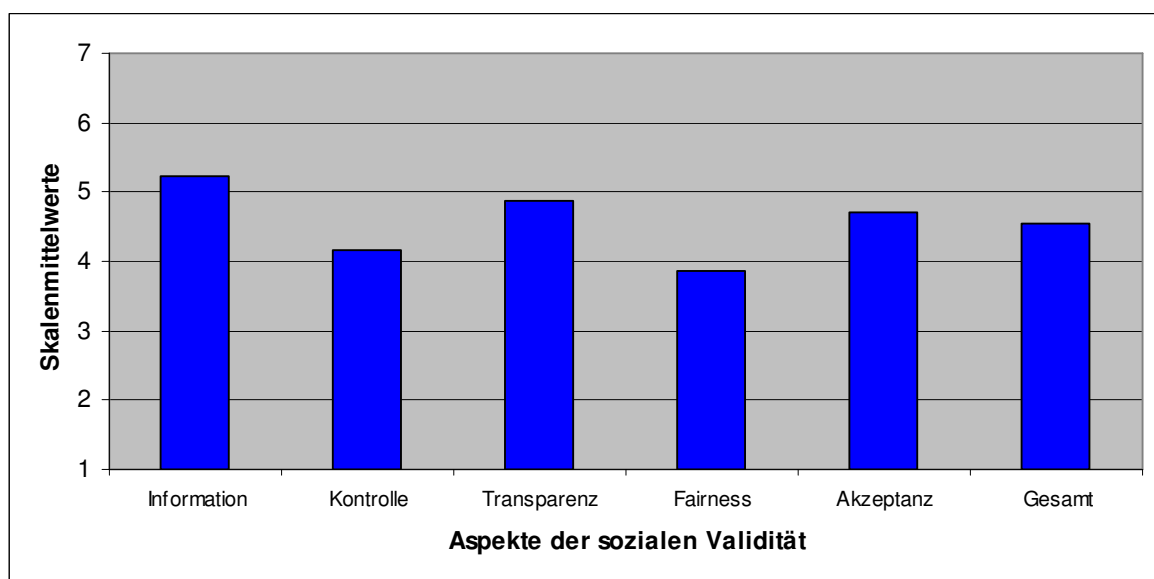


Abbildung 4.16: Skalenmittelwerte der fünf Dimensionen (Studenten)

Diskussion der Ergebnisse in Bezug auf die Hauptuntersuchung

Die Ergebnisse der Datenanalyse in Bezug auf die erste Fragestellung zeigen, dass sich die Leistung, die ein Proband im VAK erzielt, nicht aus seinen Leistungen in den Bereichen *sprachliche Intelligenz*, *Konzentrationsvermögen* und den Aspekten *Berufliche Orientierung*, *Arbeitsverhalten*, *soziale Kompetenzen* und *Psychische Konstitution*, gemessen nach dem *BIP*, vorhersagen lässt. Dieses Ergebnis kann als ein erster Schritt in Richtung einer diskriminanten Validierung gesehen werden und zeigt, dass das VAK ein sehr spezifisches Konstrukt misst. Dieses Ergebnis kann natürlich nur als ein erster Schritt gewertet werden, da weitere Absicherungen in Bezug auf das zu gemessene Konstrukt erfolgen müssen.

Einen solchen ergänzenden Ansatz stellte die Berechnung von Leistungsunterschieden zwischen Probandengruppen mit einem unterschiedlich hohen Grad an Expertise und Beratungserfahrung dar. Durch diese einer Extremgruppenvalidierung vergleichbaren Methode konnte gezeigt werden, dass sich für die Scores im *VAK* sowohl zwischen Beratungsexperten im Finanzdienstleistungssektor und Auszubildenden in diesem Feld als auch zwischen Auszubildenden und absoluten Laien ohne Beratungserfahrung signifikante Unterschiede ergeben. Die Experten zeigten eine signifikant bessere Leistung als die Auszubildenden und diese eine signifikant bessere Leistung als Laien. Dieses Ergebnis kann als ein weiterer erfolgreicher Schritt in Richtung einer Konstruktvalidierung des *VAK* gesehen werden, da sich das gemessene Konstrukt so verhält, wie man es für eine spezifische berufsbezogene Kompetenz erwarten würde, die sich erst nach längerer Tätigkeit in dem jeweiligen Berufsfeld herausbildet.

Die abschließende Bewertung der sozialen Validität des *VAK* kann als zufriedenstellend angesehen werden, da die durchschnittliche Einschätzung über alle fünf Dimensionen im positiven Bereich der Skala lag. Aufgrund der Ergebnisse der zweiten Voruntersuchung mussten somit keine Änderungen an den Materialien vorgenommen werden, welche für die spätere Hauptuntersuchung vorgesehen waren.

4.3.3 Vorstudie 3: Empirische Prüfung der Durchführbarkeit der Verfahren in der Zielpopulation

Ziel der dritten Voruntersuchung war es, Anhaltspunkte für die Durchführbarkeit des VAK und der Paper-Pencil Variante in dem in Kapitel 4.1 beschriebenen Rahmen zu bekommen.

Theoretische Motivation

Die empirische Voruntersuchung der oben genannten Verfahren im Finanzdienstleistungskontext wurde als notwendig angesehen, da der Einsatz eines video- und computerbasierten Verfahrens und eines Paper-Pencil Verfahrens, im Gegensatz zu Face-to-Face Rollenspielen, ein Novum in der Personaldiagnostik sowohl der Akademie als auch der betreuten Institute darstellte. Es lagen deshalb noch keine Erfahrungswerte bezüglich der Durchführbarkeit und der Akzeptanz seitens der Zielpopulation der Probanden im späteren Anwendungskontext vor.

Für das VAK sollten die benötigte Bearbeitungszeit, die Verteilung der Probandenscores, die Aufgabenschwierigkeiten der fünf zu bearbeitenden Phasen und die soziale Validität des Verfahrens aus Sicht von Finanzdienstleistungsauszubildenden bestimmt werden. Weiterhin sollten auch für die Zielpopulation mögliche Unterschiede in der Bewertung des Verfahrens in Abhängigkeit vom Geschlecht der Probanden ermittelt werden (vergleichend siehe Kapitel 4.3.1).

Für die Paper-Pencil Variante sollte geprüft werden, inwieweit die *would do* Instruktion von den Probanden umgesetzt werden konnte. Auch die resultierende Verteilungsform der Probandenscores, verbunden mit der Frage, ob sich durch die schriftliche Form der Rollenspielübung überhaupt eine ausreichende Varianz in Bezug auf die Probandenleistungen erzeugen ließ, war von Interesse. Außerdem sollte der Kriterienkatalog für die Bewertung der Probandenleistung in der Paper-Pencil Version einer empirischen Kontrolle unterzogen werden.

Für den Fragebogen zur sozialen Validität galt es zu ermitteln, inwieweit die Items aus Sicht der Probanden verständlich formuliert waren. Dies erfolgte qualitativ im Rahmen eines abschließenden Gespräches mit den Probanden.

Stichprobe

Für die Voruntersuchung konnten $N = 20$ Auszubildende aus verschiedenen Finanzdienstleistungsinstituten gewonnen werden, welche im Rahmen eines PES angesprochen worden

waren. Da drei Vpn aufgrund von kurzfristiger Erkrankungen zum Erhebungszeitpunkt nicht an der Untersuchung teilnahmen, wurde die Bearbeitung des VAK von $N = 17$ Auszubildenden (weiblich $N = 13$, männlich $N = 4$) durchgeführt. Die Auszubildenden waren zwischen 19 und 23 Jahren alt ($M = 21.11$, $SD = 1.118$). 13 Probanden gaben als höchsten erreichten Bildungsabschluss das Abitur an, drei verfügten über ein Fachabitur und einr über einen Realschulabschluss. Alle waren in Kreditinstituten beschäftigt, die das oben beschriebene *PES* als Instrument der Personalentwicklung einsetzten und von der gleichen Akademie betreut wurden wie die Institute, die sich bereit erklärt hatten, an der späteren Hauptuntersuchung teilzunehmen. Die Probanden stellten sich freiwillig für diese Untersuchung zur Verfügung und die investierte Zeit wurde ihnen als Arbeitszeit angerechnet.

Material

Das für die Untersuchung eingesetzte Material bestand aus dem VAK in seiner Originalform, der Aufgabenstellung für die Paper-Pencil Version sowie dem Fragebogen zur Einschätzung der wahrgenommenen Sozialen Validität für das VAK in seiner in Kapitel 3.1.4 beschriebenen Form. Die Präsentation des VAK erfolgte über Toshiba Satellite Pro Notebooks mit einem 15 Zoll Monitor und Phillips Kopfhörern.

Versuchsablauf

Um die Bearbeitungsbedingungen denen in der späteren Hauptuntersuchung vergleichbar zu machen, wurde den Probanden in einer kurzen mündlichen Einführung erklärt, dass das zu bearbeitende Programm als eine Ergänzung des von ihnen vor einigen Wochen absolvierten *PES* verstanden werden könne. Allerdings wurde, auf Bitte der Personalräte, darauf hingewiesen, dass zwar die Ergebnisse aus dem VAK und der Paper-Pencil Version an die Personalabteilung weitergegeben würden und ein gutes Ergebnis somit sehr wohl zur Kenntnis genommen werde, dass sich aber ein eventuelles negatives Abschneiden nicht nachträglich auf die im *PES* erbrachten Leistungen auswirken würde.

Nachdem alle Probanden das VAK und die Paper-Pencil Version bearbeitet hatten, wurden sie gebeten den Fragebogen zur Einschätzung der sozialen Validität auszufüllen, welcher sich in diesem Fall aber lediglich auf das VAK bezog, da dieses Verfahren im Fokus des Interesses stand. Im Anschluss an die Bearbeitung des Fragebogens wurden die Probanden in Form einer Gruppendiskussion nach ihrer Meinung bezüglich des VAK und der Verständlichkeit der Items des Fragebogens zur sozialen Validität gefragt und konnten Verbesserungsvorschläge machen.

Design zur Analyse des VAK

Die Analyse der Bearbeitungsdauer und der Verteilung der Probandenscores des VAK erfolgt anhand einer Betrachtung der deskriptiven Statistiken.

Für die Berechnung der Aufgabenschwierigkeit des VAK wurde der plausibilitätskorrigierte Schwierigkeitsindex nach Horst (1971) verwendet. Dieser berücksichtigt, dass bei der Bearbeitung eines Verfahrens, wie dem VAK, aufgrund der Stufung der Items von stark unterdurchschnittlich bis stark überdurchschnittlich nicht von einer gleichen Wahrscheinlichkeit bei der Auswahl der Antwortalternativen ausgehen werden kann. Vielmehr muss eine bewusste Strategie bei der Auswahl der Sequenzen, seitens der Probanden, postuliert werden, weshalb die Wahl von qualitativ besseren Antwortalternativen wahrscheinlicher wird (Lienert und Raatz, 1998).

Bei der Betrachtung der sozialen Validität des VAK muss auf eine faktorenanalytische Absicherung der Modellannahmen verzichtet werden, da die Voraussetzungen, die ein solches Verfahren an die Größe der Stichprobe stellt ebenfalls Fall verletzt waren. Die Items werden deshalb nach theoretischen Überlegungen zu den oben beschriebenen fünf Skalen zusammengefasst.

Die Betrachtung von möglichen Unterschieden in der Wahrnehmung des Verfahrens, die durch das Geschlecht der Versuchsteilnehmer bedingt sein könnten, wurde mittels einer einfaktoriellen Varianzanalyse durchgeführt. Als UV ging der zweifach gestufte Faktor **Geschlecht** mit den Stufen *weiblich* und *männlich* in die Analyse ein. Bei den betrachteten abhängigen Variablen handelte es sich um die Skalenwerte der 5 Dimensionen sowie dem Gesamtscore.

Ergebnisse zum VAK

Die benötigte Bearbeitungszeit der Auszubildenden betrug zwischen 45 und 60 Minuten. Alle Auszubildenden konnten das Programm bis zum Ende bearbeiten, es gab keine Abbrecher.

In Bezug auf das Gesamtergebnis konnten keine Hinweise gegen die Normalverteilungsannahme gefunden werden ($Z = .833$, $p = .491$, *ns*). Die Ergebnisse der Probanden variierten zwischen 17 und 25 von 25 möglichen Punkten und lagen im Mittel bei 22.4 Punkten ($SD = 2.03$).

Die mit Plausibilitätskorrektur ermittelten Itemschwierigkeiten der fünf thematischen Sequenzen waren für die Sequenz *Begrüßung des Kunden* $p = .82$, *Kennenlernen des Kunden*

$p = .41$, *Bedarfsermittlung* $p = .12$, *Beratung und Verkauf* $p = .53$ und *Verabschiedung des Kunden* $p = .71$.

Die Auswertung der Fragebögen zur sozialen Validität ergab ein zufriedenstellendes Ergebnis. So vergaben die Probanden für das VAK mit einer Gesamtnote von $M = 5.38$ ($SD = .630$) einen positiven Wert. Diese positive Einschätzung des Verfahrens zeigte sich auch für alle der fünf einzelnen Dimensionen *Information* ($M = 5.78$, $SD = .76$), die wahrgenommene eigene *Kontrolle* ($M = 5.71$, $SD = .62$) sowie die erlebte *Transparenz* ($M = 4.96$, $SD = .76$) die *Fairness* ($M = 4.96$, $SD = .76$) und die *Akzeptanz* ($M = 5.31$, $SD = .90$).

Die Frage nach einem Unterschied in der Wahrnehmung des VAK durch die Probanden bei einem Vergleich von weiblichen und männlichen Teilnehmern konnte anhand der Ergebnisse der varianzanalytischen Betrachtung verneint werden. Es zeigten sich weder für die Gesamteinschätzung ($F(1,15) = .039$, $p = .846$, *ns*) noch für die fünf einzelnen Dimensionen *Information* ($F(1,15) = .009$, $p = .924$, *ns*), die wahrgenommene eigene *Kontrolle* ($F(1,15) = .025$, $p = .877$, *ns*), die erlebte *Transparenz* ($F(1,15) = .138$, $p = .715$, *ns*) sowie *Fairness* ($F(1,15) = .028$, $p = .869$, *ns*) und *Akzeptanz* ($F(1,15) = .181$, $p = .676$, *ns*) signifikante Unterschiede. Eine graphische Darstellung der Einschätzungen der sozialen Validität getrennt nach Geschlecht der Probanden zeigt Abbildung 4.17.

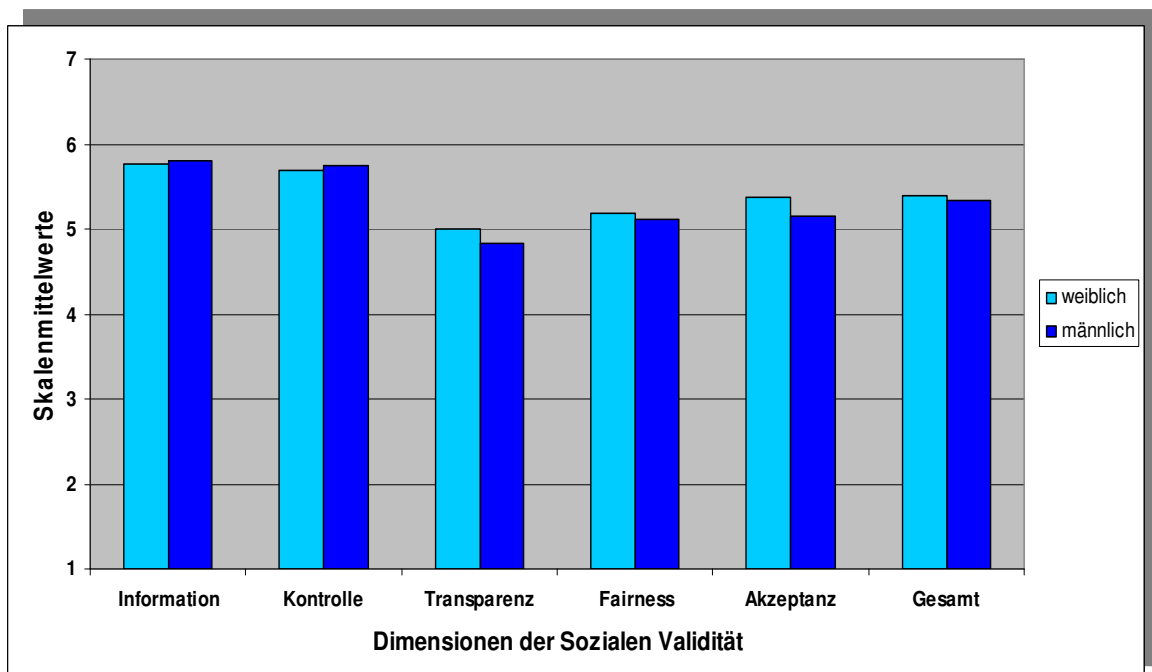


Abbildung 4.17: Skalenmittelwerte der fünf Dimensionen (getrennt nach Geschlecht)

Design zur Analyse der Paper-Pencil Variante

Die empirische Untersuchung bezüglich der Umsetzung der *would do* Instruktion erfolgt qualitativ.

Die Analyse der Verteilung der Probandenscores in der Paper-Pencil Variante erfolgt anhand von deskriptiven Statistiken.

Für die Analyse des Bewertungsschlüssels wurden als Maß der Übereinstimmung von Expertenratings bei ordinalskalierten Daten und mehr als zwei Ratern nach Wirtz und Caspar (2002) Kendalls Konkordanzkoeffizient W ermittelt. Konkret wurden zusätzlich zu der Bewertung der Probandenleistungen von Seiten des Autors drei weitere Experten als Rater hinzugezogen. Diese wurden gebeten aus den insgesamt 17 vorhandenen Aufsätzen vier per Zufall ausgewählte Aufsätze zu bewerten. Die Aufsätze wurden den Bewertern in abgetippter Form präsentiert, so dass eine Beeinflussung der Einschätzung durch Faktoren wie Rechtschreibung, Handschrift oder Interpunktion ausgeschlossen werden konnte.

Ergebnisse zur Paper-Pencil Variante

Die qualitative Analyse bezüglich der Einhaltung der *would do* Instruktion zeigte, dass alle Probanden ihre schriftlichen Ausführungen in der intendierten Form formulierten.

In Bezug auf die resultierenden Probandenscores konnten keine Hinweise gegen die Normalverteilungsannahme gefunden werden ($Z = .823$, $p = .507$, *ns*). Die Ergebnisse der Probanden variierten zwischen 8 und 14 von 25 möglichen Punkten und lagen im Mittel bei 11.24 Punkten ($SD = 1.437$).

In Bezug auf den Verhaltenskatalog und den Bewertungsschlüssel zeigte die ermittelte Übereinstimmung der Expertenurteile das in Tabelle 4.3. dargestellte Ergebnis. Dieses Ergebnis wird als zufrieden stellend angesehen.

Tabelle 4.3 Übereinstimmung der 4 Expertenurteile mittels des Bewertungsschlüssels

N	Kendall-W	Chi-Quadrat	df	Signifikanz
4	.899	10.784	4	.013

Ergebnisse zum Fragebogen zur Erhebung der sozialen Validität

Die qualitative Analyse der Probandenmeinungen bezüglich der Verständlichkeit der Items ergab für drei Items einen Verbesserungsbedarf, da diese nach Meinung von mehreren Probanden missverständlich formuliert waren. Diese Items wurden nach Maßgabe der Probanden umformuliert ohne den inhaltlichen Aspekt zu verändern.

Diskussion der Ergebnisse in Bezug auf die Hauptuntersuchung

Mit der Erstellung des VAK soll der Versuch unternommen werden das klassische Face-to-Face Verfahren *Rollenspiel* in eine adäquate video- und computerbasierte Form zu übertragen. Neben den Ergebnissen aus den beiden Vorstudien im universitären Kontext galt es das VAK an einer Stichprobe zu testen, welche aus der späteren Zielpopulation stammte, um für diese Probandengruppe erste Anhaltspunkt bezüglich der Verteilung der Probandscores, der Itemschwierigkeiten und der Akzeptanz für das Verfahren zu erlangen.

Als positiver Aspekt der Datenanalyse kann die Tatsache angesehen werden, dass die Testscores der Probanden als normal verteilt gelten können. Das VAK hat also wenige Probanden mit einer sehr niedrigen und wenige Probanden mit einer sehr hohen Fähigkeitsausprägung ermittelt, dafür aber viele mit einem mittleren Ausprägungsgrad. Eine solche Verteilung der zu messenden spezifischen sozialen Fähigkeiten in der Gesamtpopulation ist als sehr wahrscheinlich einzustufen.

Vergleicht man die empirisch ermittelten Schwierigkeitsindizes der einzelnen VAK Sequenzen, bezogen auf die Ergebnisse der 17 Vpn des Vortests, mit den Hinweisen zur Auswahl von Aufgaben nach der Schwierigkeit nach Lienert und Raatz (1998), so lässt sich die Streuung von $p = .12$ bis $p = .82$ als sehr zufriedenstellend ansehen. Eine optimale Streuung der Schwierigkeitsindizes für Niveautests sollte hiernach über den ganzen Bereich eines Persönlichkeitsmerkmals streuen, wenigstens aber von $p = .20$ bis $p = .80$ reichen. Die Forderung nach einer so genannten Eisbrecher-Aufgabe, also einer Aufgabe, die von fast allen Probanden richtig gelöst und als einfach empfunden wird, kann mit einer Schwierigkeit von $p = .82$ bei der ersten zu bearbeitenden Sequenz genauso als erfüllt angesehen werden wie die Richtlinie, dass kein Proband der Analytestichprobe ohne Punkte ausgehen sollte und nur wenige Probanden alle Aufgaben richtig lösen können sollten.

Auch die wichtige Frage nach der Akzeptanz der Zielgruppe für das VAK als neuartiges multimediales Verfahren kann positiv beantwortet werden. Die Ergebnisse aus dem Fragebogen zur sozialen Validität zeigten für alle betrachteten Aspekte eine Bewertung im positiven Bereich der Skala. Weiterhin zeigte sich kein Unterschied in der Bewertung der

weiblichen und männlichen Probanden, was ebenfalls für die soziale Validität des VAK spricht. Aufgrund der positiven Ergebnisse wird für das VAK, vor seinem Einsatz in der Hauptuntersuchung kein weiterer Verbesserungsbedarf mehr gesehen.

Mit der Erstellung der Paper-Pencil Variante wurde versucht ein klassisches Rollenspiel in Face-to-Face Form in eine Paper-Pencil Form zu übertragen. Auch für diese Übung galt es einen ersten Eindruck bezüglich der Durchführbarkeit in der Zielpopulation zu gewinnen.

Die Tatsache, dass alle 17 Probanden die *would do* Instruktion richtig umsetzten und die Aufsätze in der Ich-Form verfassten, wird als Beleg dafür gesehen, dass die Instruktion verständlich formuliert ist und zum intendierten Probandenverhalten führt.

Die berichteten deskriptiven Statistiken der Probandenscores zeigten, dass die Bandbreite der möglichen Probandenscores nicht voll ausgenutzt wurde und sich in Bezug auf die erzielten Leistungen eine geringe Streuung ergab. Dieser Aspekt wurde als verbesserungswürdig angesehen und führte zu einer Ergänzung der Probandeninstruktion. Konkret wurde die Instruktion durch die Bitte um eine detaillierte Ausführung und den Hinweis die Bearbeitungszeit von 30 Minuten voll auszunutzen ergänzt. Durch diese Maßnahmen sollten die Probanden angeregt werden mehr Verhaltensweisen zu nennen und ihr Vorgehen genauer zu beschreiben. Beide Aspekte sollten sich in Richtung einer größeren Varianz in den Probandenleistungen auswirken. Positiv wird bewertet, dass die empirische Verteilung der Probandenscores keine Verletzung der Normalverteilungsannahme zeigte.

Auch die Ergebnisse zur Beurteilerübereinstimmung bei der Bewertung der Probandenleistungen mittels des Bewertungsschlüssels lassen ein positives Fazit zu. Aufgrund der hohen Übereinstimmung wurde der Bewertungsschlüssel unverändert übernommen.

Für den Fragebogen zur Erhebung der sozialen Validität galt es zu ermitteln inwieweit die Formulierungen der einzelnen Items für die Probanden aus der Zielgruppe verständlich waren. Aufgrund der Probandenhinweise konnten drei Items identifiziert werden, die aus Probandensicht nicht eindeutig formuliert waren. Diese Items wurden in Kooperation mit den Probanden neu formuliert ohne dass der Inhalt verändert wurde.

Aufgrund der positiven Ergebnisse für das VAK aus allen drei Vorstudien und der erfolgten Änderungen bezüglich der Paper-Pencil Variante aufgrund der dritten Vorstudie wurden die Materialien als geeignet für die Hauptuntersuchung bewertet.

4.4 Hauptuntersuchung

Nachdem die Hauptstudie der vorliegenden Arbeit mit dem Kapitel 4.1 in einen Anwendungskontext eingebettet worden ist und die Erstellung des in der Erhebung verwendeten Materials und seine Testung im Rahmen von Voruntersuchungen in den Kapiteln 4.2 und 4.3 beschrieben wurde, soll im folgenden Kapitel auf die methodische Vorgehensweise bei der Überprüfung der in Kapitel 3 formulierten Hypothesen eingegangen werden. Hierfür wird zuerst die Stichprobe beschrieben und dann der Versuchsablauf dargestellt. Danach wird das der Untersuchung zu Grunde liegende Design bezogen auf die jeweilige Fragestellung erläutert und die unabhängigen sowie abhängigen Variablen benannt.

4.4.1 Stichprobe

Für die Untersuchung der in Kapitel 3 vorgestellten Fragestellung konnten insgesamt $N = 52$ Auszubildende im Alter von 18 bis 29 Jahren ($M = 22.9$, $SD = 1.95$) aus 5 Finanzdienstleistungsinstituten gewonnen werden. Die Geschlechterverteilung lag bei 59,6% weiblichen und 40,4% männlichen Teilnehmern. Alle 5 Institute verwenden das *PES* der Akademie seit vielen Jahren als planmäßiges Instrument der Personalentwicklung, welches am Ende der Ausbildung zum Bankkaufmann eingesetzt wird. Die 52 Auszubildenden wurden darauf hingewiesen, dass die sonst üblichen Übungen im *PES* um zwei neue Übungen ergänzt wurden. Weiterhin wurde auf Bitten der Personalräte betont, dass die in den neuen Übungen erzielten Ergebnisse zwar von der Personalabteilung eingesehen und zur Kenntnis genommen würden, aber keinen Einfluss auf die Entscheidung für oder gegen eine weitere Förderung hätten und somit eine Gleichbehandlung zu den vorangegangenen Jahrgängen gewährleistet sei. Die durch die Bearbeitung des *VAK* zusätzlich anfallende Zeit wurde den Auszubildenden als Arbeitszeit angerechnet.

4.4.2 Material

Das im Rahmen der Untersuchung insgesamt eingesetzte Material bestand zum einen aus Materialien der Akademie und zum anderen aus den in Kapitel 4.2 und 4.3 beschriebenen Verfahren und Fragebögen. Konkret waren dies die Aufgabenstellung für die Face-to-Face

Variante des Rollenspiels, das *VAK* in seiner Originalform, die Aufgabenstellung für die Paper-Pencil Variante und die drei Versionen des Fragebogens zur Erhebung der sozialen Validität. Das *VAK* wurde den Probanden in allen Fällen auf Toshiba Satellite Pro Notebooks mit einem 15 Zoll Monitor und Phillips Kopfhörern dargeboten. Hierdurch konnte eine für alle Probanden einheitliche Darbietung gewährleistet werden.

4.4.3 Versuchsablauf

Um die in Kapitel 3 beschriebenen Fragestellungen beantworten zu können, wurden neben dem bereits etablierten Rollenspiel auch das *VAK* und die Paper-Pencil Übung in das in Kapitel 4.1 beschriebene Personalentwicklungsverfahren integriert. Da der Zeitplan des Verfahrens am ersten Tag nur wenig Spielraum bot, wurde die Paper Pencil Variante als weniger organisations- und zeitaufwändige Übung in den normalen Ablauf des ersten Tages aufgenommen. Das *VAK* folgte für sechs der 11 Gruppen am zweiten Tag des *PES* und wurde bei den fünf übrigen Gruppen aus organisatorischen Gründen erst einige Tage nach dem *PES* durchgeführt. Die Daten aller 52 Teilnehmer wurden in einem Erhebungszeitraum von März bis Juni 2006 gewonnen. Da die Verfahren immer durch einen Trainer der Akademie moderiert wurden, gestalteten sich die Verfahrensbedingungen für die Probanden sowohl in Bezug auf den Ablauf als auch die Dauer der 11 *PES* als durchaus vergleichbar.

Zu Beginn der Untersuchung wurden die Probanden von Seiten des Moderators begrüßt und hatten die Gelegenheit im Rahmen einer kurzen Vorstellungsrunde etwas über ihren persönlichen Hintergrund zu berichten und gleichzeitig die Beobachter kennen zu lernen. Hierauf folgte die Vorstellung der Übungen und des Tagesablaufes, verbunden mit dem Hinweis auf die Erweiterung des *PES* um zwei neue Übungen.

Nach einer Gruppenübung zu Beginn, die von allen Probanden parallel absolviert wurde, durchliefen die einzelnen Teilnehmer die weiteren Übungen in einem individuellen Rhythmus. Dieser beinhaltete die Vorbereitung auf die Übung, die Durchführung und die Aufforderung zu einer schriftlichen Reflexion des eigenen Abschneidens in der Übung. Das Verfahren endete für die Probanden mit einer Abschlussrunde, in welcher diese ihre Eindrücke und Einschätzungen bezüglich der Übungen und der von ihnen erbrachten Leistungen schildern konnten. Mit dem Hinweis auf das *VAK* und die Rückmeldegespräche am nächsten Tag wurden die Probanden dann verabschiedet.

Abbildung 4.18 zeigt den Ablauf des Verfahrens exemplarisch für eine Gruppe mit 4 Probanden für den ersten Tag. Die dieser Studie zugrundeliegende Face-To-Face Variante und die Paper-Pencil Variante werden aus dem gesamten Ablauf farblich hervorgehoben.

Teilnehmer 1	Teilnehmer 2	Teilnehmer 3	Teilnehmer 4
Vorstellungsrunde			
Gruppenübung			
Face-to-Face		Paper Pencil	
	Face-to-Face		Paper Pencil
		Face-to-Face	
			Face-to-Face
Vortrag			
	Vortrag		
		Vortrag	
Paper Pencil	Paper Pencil		Vortrag
Vorbereitung Postkorb	Vorbereitung Postkorb	Vorbereitung Postkorb	Vorbereitung Postkorb
Mittagspause			
			Präsentation Postkorb
		Präsentation Postkorb	
Präsentation Postkorb	Präsentation Postkorb		
			Rollenspiel MG
	Rollenspiel MG	Rollenspiel MG	
Rollenspiel MG			
			Rollenspiel KG
	Rollenspiel KG	Rollenspiel KG	
Rollenspiel KG			
Abschluss			

Abbildung 4.18 Graphische Darstellung des Versuchsablaufes (1. Tag)

Am zweiten Tag folgten die Bearbeitung des *VAK* und das Ausfüllen der Fragebögen zur sozialen Validität. Das Ende der gesamten Erhebung bildete das Feedbackgespräch mit einem Vertreter der Personalabteilung und dem jeweiligen Moderator des PES. Erst im Anschluss an dieses Feedbackgespräch hatten die Probanden die Möglichkeit sich die Ergebnisse aus dem *VAK* mitteilen zu lassen. Die Ergebnisse aus der schriftlichen Parallelform des Rollenspiels wurden den Probanden auf Wunsch nach erfolgter Auswertung schriftlich zugestellt.

Der Ablauf des zweitens Tages wird in Abbildung 4.19 graphisch dargestellt.

Teilnehmer 1	Teilnehmer 2	Teilnehmer 3	Teilnehmer 4
Begrüßung durch den Versuchsleiter			
VAK	VAK	VAK	VAK
Fragebögen zur Sozialen Validität	Fragebögen zur Sozialen Validität	Fragebögen zur Sozialen Validität	Fragebögen zur Sozialen Validität
Abschluss			

Abbildung 4.19 Graphische Darstellung des Versuchsablaufes (2. Tag)

4.4.4 Design

Die Beschreibung des Designs wird zur besseren Übersicht getrennt nach Forschungsfragen vorgenommen.

Design Fragestellung 1

Um die Frage nach Unterschieden in der kriterienorientierten Validität der drei medial variierten Rollenspielvarianten sowie die Frage nach der Höhe der Zusammenhänge zwischen Ergebnissen aus dem Rollenspiel, dem *VAK* und der Paper-Pencil Übung beantwor-

ten zu können, wurde auf ein korrelationsstatistisches Design zurückgegriffen. Für die Bestimmung der Höhe des Zusammenhangs zwischen der IHK-Note und den Leistungen aus den drei Rollenspielvarianten wurde der Produkt-Moment-Korrelationskoeffizient nach Pearson verwendet. Dieses Maß diente auch zur Bestimmung der Zusammenhänge der Leistungen aus den einzelnen Verfahren untereinander.

Prüfung der formalstatistischen Voraussetzungen

Alle eingesetzten Testverfahren produzieren Daten auf Intervallskalenniveau und die Verteilung der Daten aller drei Verfahren, Rollenspiel ($Z = .698$, $p = .714$, ns), Video ($Z = 1.100$, $p = .178$, ns) und Paper Pencil ($Z = 1.039$, $p = .231$, ns) sowie die Verteilung der IHK-Noten ($Z = 1.025$, $p = .244$) hielten einer Überprüfung auf Normalverteilung nach Kolmogoroff-Smirnoff stand.

Design Fragestellung 2

Für die Beantwortung der zweiten Forschungsfrage, welche sich auf Unterschiede in der Bewertung der sozialen Validität der drei Rollenspielvarianten durch die Probanden bezieht, wurde eine einfaktorielle, univariate Varianzanalyse mit Messwiederholung durchgeführt. Als unabhängige Variable ging der dreifach gestufte Faktor **Mediale Variation** mit den Stufen *Rollenspiel*, *VAK* und *Paper-Pencil* in die Analyse ein. Bei den betrachteten abhängigen Variablen handelte es sich um die mittels Fragebogen erhobene Bewertungen der sozialen Validität. Ein solcher Vergleich der drei Bedingungen erscheint allerdings nur dann sinnvoll, wenn sich die a-priori angenommene Struktur des Fragebogens für alle Bedingungen auch empirisch bestätigen lässt oder sich zumindest eine einheitliche empirische Datenstruktur für die drei Bedingungen zeigt. Die Überprüfung der Daten mittels eines konfirmatorischen Designs musste aus formalstatistischen Gründen verworfen werden. Stattdessen wurde eine gestufte Vorgehensweise gewählt, um zu überprüfen ob für jede Bedingung zumindest annäherungsweise die gleichen Iteminformationen eingehen, so dass die Daten vergleichbar sind.

In einem ersten Schritt erfolgt, getrennt für jede der drei medialen Variationen, die Überprüfung der theoretisch angenommenen Struktur des Fragebogens zur sozialen Validität, auf der Ebene der fünf einzelnen Dimensionen mittels einer faktorenanalytischen Betrachtung.

In einem zweiten Schritt wird wiederum getrennt nach den drei Bedingungen die Struktur des gesamten Fragebogens unter Einbezug aller 20 Variablen analysiert. Die Ermittlung der Faktorenstruktur für jede der drei Bedingungen erfolgte mittels einer Hauptachsenanalyse (PAF); als Rotationstechnik wird, ausgehend von korrelierten Faktoren, mit Promax eine oblique Technik gewählt.

Im Anschluss an die separate Analyse der Faktorenstruktur jeder einzelnen Bedingung wird mit dem kompletten Datensatz über alle 3 Bedingungen eine Hauptachsenanalyse mit Promax-Rotation gerechnet. Für die mittels Parallelanalyse identifizierten Faktoren werden dann die Faktorscores bestimmt.

Im Anschluss an diese stufenweise faktorenanalytische Auswertung des Datensatzes erfolgt die Analyse auf vorhandene Mittelwertsunterschiede zwischen den Faktorscores der drei medialen Bedingungen.

Nachdem in den vorangegangenen Abschnitten die methodische Vorgehensweise, welche zur Beantwortung der Forschungsfragen gewählt worden ist, dargelegt wurde, sollen im folgenden Kapitel die Ergebnisse der Auswertungen getrennt nach den Forschungsfragen berichtet werden.

5 Ergebnisse

Im vorliegenden Kapitel werden die Ergebnisse der Hauptuntersuchung getrennt nach den in Kapitel 3 formulierten Forschungsfragen und Hypothesen berichtet. Zuvor werden die deskriptiven Statistiken der drei Rollenspielvarianten und des Kriteriums dargestellt. Die Diskussion der gefundenen Ergebnisse folgt in Kapitel 6.

5.1 Darstellung der deskriptiven Statistiken

5.1.1 Deskriptive Statistiken für die Face-to-Face Variante

Die durchgeführte Analyse der deskriptiven Statistiken für die Probandenleistungen in der Face-to-Face Variante des Rollenspiels ergab die in Tabelle 5.1 dargestellten Kennwerte.

Tabelle 5.1: Darstellung der deskriptiven Statistiken für das Face-to-Face Rollenspiel

	N	Min	Max	M	SD
Gesamtscore Rollenspiel	52	2.15	4.06	3.07	.472

Es zeigt sich, dass der empirische Mittelwert mit $M = 3.07$ ($SD = .472$) sehr gut dem theoretischen Mittelwert der Bewertungsskala in den Beobachtungsbögen der Akademie entspricht. Keiner der Probanden erzielte eine stark unterdurchschnittliche Leistung (Skalenwert 1) oder eine stark überdurchschnittliche Leistung (Skalenwert 5). Die Bandbreite der Skala wurde somit von den Beobachtern nicht voll ausgenutzt.

Eine Überprüfung der Daten auf Normalverteilung nach Kolmogoroff-Smirnoff bestätigt die Normalverteilungsannahme für die Probandenwerte in dieser Übung ($Z = .698$, $p = .714$, *ns*).

5.1.2 Deskriptive Statistiken für das VAK

Die durchgeführte Analyse der deskriptiven Statistiken für die Probandenleistungen aller fünf Abschnitte des VAK sowie des Gesamtscores ergab die in Tabelle 5.2 dargestellten Kennwerte.

Tabelle 5.2: Darstellung der deskriptiven Statistiken für das VAK

	N	Min	Max	M	SD
Sequenz 1	52	3	5	4.71	.605
Sequenz 2	52	2	5	4.38	.867
Sequenz 3	52	3	5	4.02	.828
Sequenz 4	52	3	5	4.56	.539
Sequenz 5	52	2	5	4.31	.961
Gesamtscore	52	15	25	21.98	2.16

Es zeigt sich, dass die Probanden in keiner der fünf Sequenzen des VAK das mit 1 bewertete Video mit dem unzulänglichsten Beraterverhalten gewählt haben und auch nur in Sequenz 2 und Sequenz 5 die Wahl auf ein Video der Kategorie 2 fiel. Die Mittelwerte liegen für alle Sequenzen zwischen 4.02 und 4.71 und damit über dem a priori festgelegten Skalenmittelwert von 3.

Die Überprüfung der Daten auf Normalverteilung nach Kolmogorov-Smirnoff zeigt eine Verletzung der Normalverteilungsannahme für die Sequenz 1 ($Z = 3.401$, $p = .000$), Sequenz 2 ($Z = 2.715$, $p = .000$), Sequenz 3 ($Z = 1.644$, $p = .009$), Sequenz 4 ($Z = 2.675$, $p = .000$), Sequenz 5 ($Z = 2.877$, $p = .000$). Für den Gesamtscore ergibt sich eine Normalverteilung der Items ($Z = 1.100$, $p = .178$, ns).

Eine detaillierte Betrachtung der Verteilungsform jeder der fünf Sequenzen und des Gesamtscores ermöglichen die Abbildungen 5.1 bis 5.6.

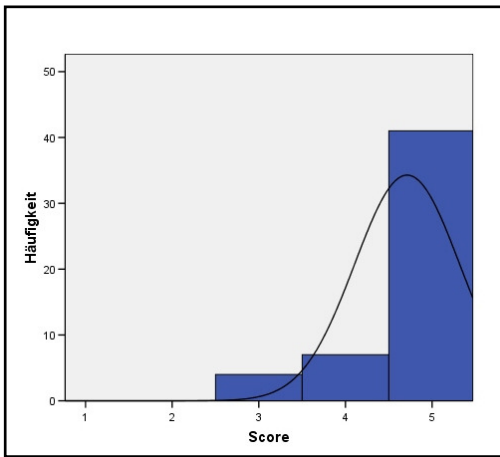


Abbildung 5.1: Scores (Sequenz 1)

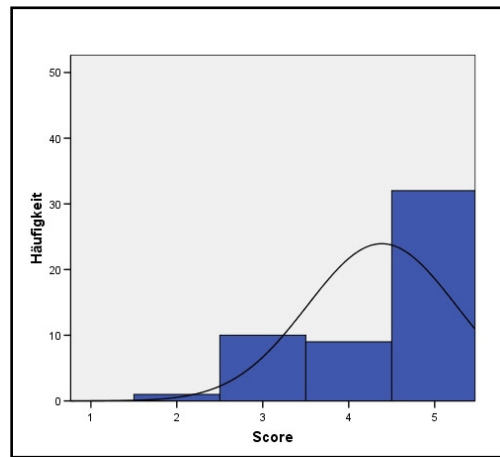


Abbildung 5.2: Scores (Sequenz 2)

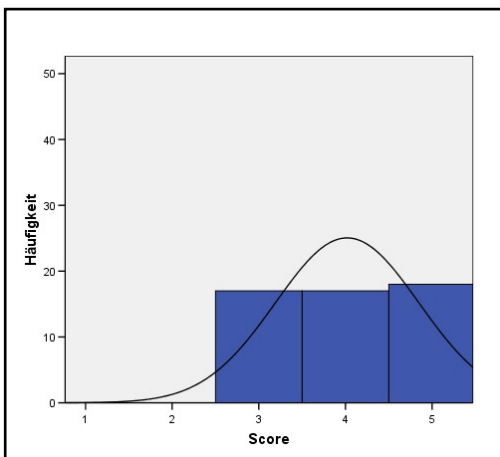


Abbildung 5.3: Scores (Sequenz 3)

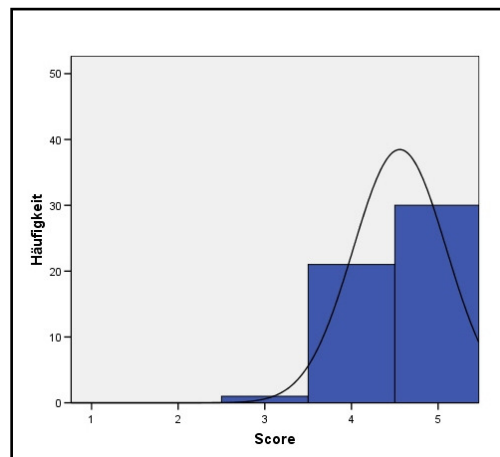


Abbildung 5.4: Scores (Sequenz 4)

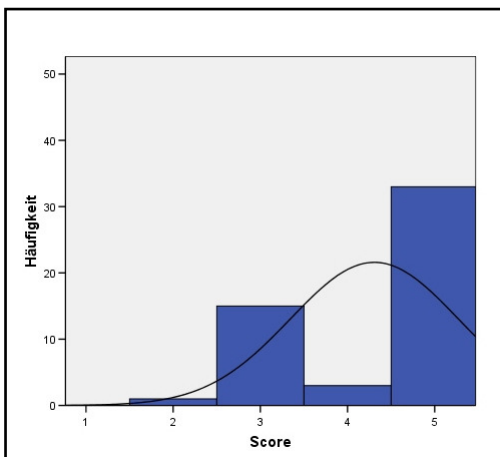


Abbildung 5.5: Scores (Sequenz 5)

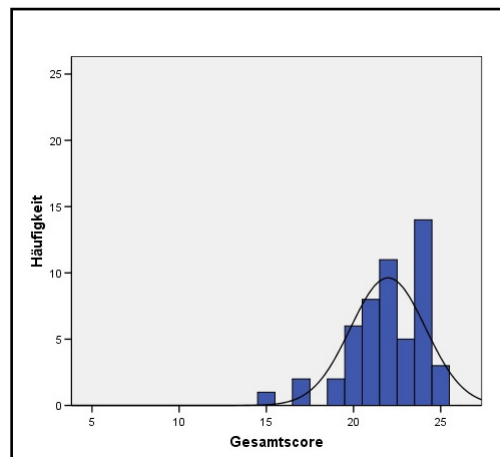


Abbildung 5.6: Scores (gesamt)

Es zeigt sich, dass für alle einzelnen Sequenzen mit Ausnahme der Sequenz 3 die deutliche Mehrheit der Probanden die beste Verhaltensalternative wählte. Zieht man die Einzelleistungen zu einem Gesamtscore zusammen, ergibt sich aber eine Normalverteilung der Probandenleistung für das gesamte virtuelle Kundengespräch. Dieses Ergebnis erscheint insofern nicht ungewöhnlich, als dass es durchaus denkbar ist, dass ein einzelner Proband zwar in der Begrüßungssequenz und in der Kennenlernphase die optimale Lösung wählt, für die Beratungsphase aber nicht, da seine Kompetenz hier weniger stark ausgeprägt ist. Andere Probanden hingegen konnten zwar für die Beratungsphase die beste Sequenz identifizieren, waren aber nicht in der Lage in den anderen Phasen das jeweils optimale Verhalten mit dem eigenen potentiellen Verhalten in diesen Situationen zu verknüpfen. Entscheidend ist die Tatsache, dass der Gesamtscore, welcher zur Beantwortung der Fragestellung als eine AV herangezogen wird, der Normalverteilungsannahme entspricht, da von einer Normalverteilung der spezifischen sozialen Kompetenz *Kundenorientierung* innerhalb der Zielpopulation des VAK ausgegangen wird.

Die Analyse der internen Konsistenz berechnet nach Cronbachs alpha zeigte einen sehr niedrigen Wert von $\alpha = .483$. Ein besonderes Augenmerk galt der für jede Sequenz ermittelten Part-Whole korrigierten Trennschärfe. Es zeigten sich für alle Sequenzen niedrige aber positive Korrelationen. Das empirisch ermittelte Ergebnis der internen Konsistenz für das VAK steht im Einklang mit den in Kapitel 2.7.3 berichteten Befunden aus Studien zu videobasierten SJTs. Da sich einem potentiellen Probanden für jede der fünf Sequenzen eine andere Herausforderung stellt, ist es durchaus möglich, dass die einzelnen Sequenzen untereinander nicht sehr hoch miteinander korrelieren und auch zum Gesamtergebnis nur geringe Korrelationen aufweisen. Entscheidend ist aber, dass sich keine negativen Korrelationen zwischen den Leistungen in den einzelnen Sequenzen bzw. zwischen den Leistungen der einzelnen Sequenzen und dem Gesamtergebnis zeigen. Diese Forderung wird von Seiten des VAK erfüllt.

Die vollständigen Ergebnisse der Reliabilitätsanalyse inklusive der ermittelten Trennschärfen für die einzelnen Sequenzen zeigt Tabelle 5.3.

Tabelle 5.3: Ergebnisse der Reliabilitätsanalyse des VAK

	Skalenmittelwert, wenn Item weggelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item-Skala- Korrelation	R ²	Cronbachs Alpha, wenn Item weggelassen
Sequenz 1	17.27	3.612	.290	.121	.355
Sequenz 2	17.60	3.226	.215	.074	.397
Sequenz 3	17.96	3.253	.237	.112	.378
Sequenz 4	17.42	3.896	.216	.069	.400
Sequenz 5	17.67	2.969	.228	.071	.393

Für die Berechnung der Aufgabenschwierigkeit des VAK wurde der plausibilitätskorrigierte Schwierigkeitsindex nach Horst (1971) verwendet. Sind die Häufigkeiten für die optimale Antwort und die am häufigsten gewählte Alternativantwort ähnlich hoch, so führt dies zu sehr hohen Schwierigkeiten im Sinne niedriger Koeffizienten, auch wenn ein großer Teil der Probanden die optimale Antwort gewählt hat.

Die mit Plausibilitätskorrektur ermittelten Itemschwierigkeiten der fünf thematischen Sequenzen waren für die Sequenz Begrüßung des Kunden $p = .67$, Kennenlernen des Kunden $p = .42$, Bedarfsermittlung $p = .2$, Beratung und Verkauf $p = .17$ und Verabschiedung des Kunden $p = .35$. Dieses Ergebnis bestätigt das positive Fazit bezüglich der Passung der empirisch ermittelten Schwierigkeiten mit den Hinweisen zur Auswahl von Aufgaben nach der Schwierigkeit nach Lienert und Raatz (1998), welches im Rahmen der Voruntersuchung in Kapitel 4.3.3 in Bezug auf die Schwierigkeit der einzelnen Sequenzen des VAK gezogen wurde.

5.1.3 Deskriptive Statistiken für die Paper-Pencil Variante

Die durchgeführte Analyse der deskriptiven Statistiken für die Probandenleistungen in der Paper-Pencil Variante des Rollenspiels ergab die in Tabelle 5.4 dargestellten Kennwerte.

Tabelle 5.4: Darstellung der deskriptiven Statistiken für die Paper-Pencil Variante

	N	Min	Max	M	SD
Gesamtscore	52	6	17	11.65	2.441

Die Betrachtung der durch die Probanden erzielten Minimal- und Maximalwerte zeigt, dass die vorher festgelegte Bandbreite von 0 bis 25 Punkten nicht ausgeschöpft wurde. Die Überprüfung der Daten auf Normalverteilung nach Kolmogoroff-Smirnoff bestätigt die Normalverteilungsannahme ($Z = 1.039$, $p = .231$, ns).

5.1.4 Deskriptive Statistiken für das Kriteriumsmaß

Die durchgeführte Analyse der deskriptiven Statistiken für die Leistungen der Probanden in der mündlichen IHK-Abschlussprüfung, welche als Kriterium in der vorliegenden Arbeit diente, ergab die in Tabelle 5.5 dargestellten Werte.

Tabelle 5.5: Darstellung der deskriptiven Statistiken für das Kriterium

	N	Min	Max	M	SD
Probandenscore	52	58	98	83.69	7.56

Die Betrachtung der durch die Probanden erzielten Minimal- und Maximalwerte zeigt, dass die Bandbreite des Bewertungsbereiches zwischen 50 und 100 Punkten fast vollständig ausgeschöpft wurde. Alle Probanden erreichten den erfolgskritischen Wert von 50 Punkten. Der empirisch ermittelte Mittelwert liegt mit 83.69 Punkten im Bereich einer guten Prüfungsleistung, der nach IHK-Standard bei 82 Punkten beginnt. Eine Überprüfung der Daten auf Normalverteilung nach Kolmogoroff-Smirnoff bestätigt die Normalverteilungsannahme ($Z = 1.025$, $p = .244$, ns).

5.2 Ergebnisse zur kriterienorientierten Validität

In die Analyse der Daten aus den drei Verfahren und der IHK-Note konnten die Antworten von allen 52 Probanden aufgenommen werden. Drei fehlende Werte bei der Angabe der mündlichen IHK-Note wurden durch den Gruppenmittelwert ersetzt.

Tabelle 5.6 zeigt die korrelativen Zusammenhänge zwischen der IHK-Note und den Ergebnissen aus den drei medial variierten Übungen.

Tabelle 5.6: Darstellung der korrelativen Zusammenhänge zwischen der IHK-Note und den Ergebnissen aus den drei medial variierten Bedingungen.

	Rollenspiel	VAK	Paper-Pencil
IHK-Note	.226 (.053)	-.048 (.366)	-.048 (.367)

Die Daten in den Zellen zeigen die Höhe des Korrelationskoeffizienten nach Pearson.

Die Daten in Klammern zeigen den p-Wert.

Die Betrachtung der in Tabelle 5.6 dargestellten Zusammenhänge zwischen Probandenergebnissen aus den drei medialen Variationen des Rollenspiels und den Ergebnissen der IHK-Abschlussnote zeigt ein heterogenes Bild. Während sich für die erbrachte Leistung in der Face-to-Face Variante ein leichter positiver Zusammenhang zu der IHK-Note zeigt, ergeben sich für die im VAK und in der Paper-Pencil Variante erbrachten Leistungen negative Zusammenhänge nahe Null. Eine Betrachtung der Signifikanzniveaus zeigt weiterhin, dass für das VAK und die Paper-Pencil Variante nicht von einem signifikanten Zusammenhang positiver oder negativer Art ausgegangen werden kann. Der Zusammenhang zwischen dem Face-to-Face Rollenspiel und der IHK Note zeigt sich als marginal signifikant. In Bezug auf die in Kapitel 3 formulierte Fragestellung lässt sich aufgrund der vorliegenden empirischen Ergebnisse feststellen, dass im Gegensatz zu den berichteten Ergebnissen aus der Forschung zu videobasierten SJTs keine signifikanten Unterschiede zwischen den drei medial variierten Versionen des Rollenspiels in Bezug auf deren kriterienorientierte Validität gefunden werden konnten.

Ergänzend zu diesen Befunden wurde der Zusammenhang zwischen den Probandenleistungen in den einzelnen Übungen mittels eines korrelationsstatistischen Designs ermittelt. Auch hier zeigte sich kein signifikanter Zusammenhang.

Die Korrelationen der Scores in den drei Bedingungen sind in Tabelle 5.7 wiedergegeben.

Tabelle 5.7: Darstellung der korrelativen Zusammenhänge zwischen den Ergebnissen der drei medial variierten Bedingungen

	Rollenspiel	VAK	Paper-Pencil
Rollenspiel	1	-.089 (.531)	.129 (.361)
VAK		1	.152 (.284)
Paper-Pencil			1

Die Daten in den Zellen zeigen die Höhe des Korrelationskoeffizienten nach Pearson.

Die Daten in Klammern zeigen den p-Wert.

Auch eine ergänzende Datenanalyse mittels einer schrittweisen, nach Vorwärts-Technik berechneten Regressionsanalyse, in welche neben dem herkömmlichen Face-to-Face Rollenspiel jeweils eine der beiden anderen Rollenspielvarianten als Prädiktor aufgenommen wurde, ergab kein gültiges Modell. Dieses Ergebnis bestätigt die Befunde zur kriterienorientierten Validität, welche keine oder nur marginal signifikante Zusammenhänge aufzeigen konnten. Eine Erhöhung der Vorhersagegüte durch die Kombination der Face-to-Face Variante mit einer der beiden anderen Versionen kann somit nicht erreicht werden. Mögliche Suppressoreffekte konnten aufgrund der niedrigen Korrelationen der Ergebnisse aus den drei Rollenspielversionen untereinander ebenfalls ausgeschlossen werden

5.3 Ergebnisse zur sozialen Validität

Die Ergebnisdarstellung der Datenanalyse zur Beantwortung der zweiten Forschungsfrage dieser Arbeit erfolgt in zwei Abschnitten. Im ersten Abschnitt wird auf die Ergebnisse der faktorenanalytischen Untersuchung der einzelnen Datensätze und des gesamten Datensatzes eingegangen, die, wie oben beschrieben, in drei Stufen erfolgte. Die Ergebnisse aus der varianzanalytischen Betrachtung der im ersten Analyseschritt ermittelten Faktorscores werden im zweiten Abschnitt berichtet.

5.3.1 Ergebnisse zur Faktorstruktur der fünf Itemgruppen

Die Ergebnisse einer Überprüfung der Datenstruktur der fünf Dimensionen mittels einer Hauptachsenanalyse mit Promax Rotation werden ebenfalls getrennt für die drei Bedingungen dargestellt.

Die faktorenanalytische Betrachtung der Bedingung **Face-to-Face** zeigt für die Dimension *Information* eine Einfaktorenlösung, für die Dimension *Partizipation/Kontrolle* eine Zweifaktorenlösung, für die Dimension *Transparenz* eine Einfaktorenlösung, für die Variablen der Dimension *Fairness* eine Einfaktorenlösung und für die Variablen der Dimension *Akzeptanz* ebenfalls eine Lösung mit einem Faktor.

Für die Bedingung **VAK** zeigt die Faktorenanalysen für die Dimension *Information* eine Zweifaktorenlösung, für die Dimension *Partizipation/Kontrolle* eine Zweifaktorenlösung, für die Dimension *Transparenz* eine Zweifaktorenlösung, für die Variablen der Dimension *Fairness* eine Einfaktorenlösung und für die Variablen der Dimension *Akzeptanz* ebenfalls eine Lösung mit einem Faktor.

Die Analyse der den Items zugrunde liegenden Struktur der Bedingung **Paper-Pencil** zeigt für die Dimension *Information* eine Zweifaktorenlösung, für die Dimension *Partizipation/Kontrolle* eine Zweifaktorenlösung, für die Dimension *Transparenz* eine Einfaktorenlösung, für die Variablen der Dimension *Fairness* eine Einfaktorenlösung und für die Variablen der Dimension *Akzeptanz* ebenfalls eine Struktur, die durch einen Faktor erklärbar ist.

Die Tatsache, dass sich in allen drei Bedingungen für die fünf theoretisch formulierten Dimensionen der sozialen Validität Abweichungen von einer einfaktoriellen Struktur zu

beobachten sind, schließt eine vergleichende Betrachtung der sozialen Validität unter Berücksichtigung der Skalenmittelwerte der fünf Dimensionen aus. Die Ermittlung der tatsächlich zugrunde liegenden Faktorenstruktur für die 3 Bedingungen schließt sich deshalb als zweiter Schritt an.

5.3.2 Ergebnisse zur Faktorstruktur der Fragebögen

Im zweiten Schritt der Datenanalyse wurde eine Hauptachsenanalyse mit Promax Rotation über alle theoretisch gebildeten Dimensionen der sozialen Validität aber weiterhin getrennt nach der medialen Variation durchgeführt. Die Prüfung der Voraussetzungen wird für die drei Bedingungen getrennt dargestellt und erfolgt jeweils vor der Beschreibung der Ergebnisse.

Ergebnisse zur Faktorstruktur der Bedingung Face-to-Face

Für die Bedingung **Face-to-Face** liegt der KMO-Koeffizient mit .793 nach Bühner (2006) in einem mittleren bis guten Bereich. Auch der Bartlett-Test auf Sphärizität zeigt die Eignung der Datenstruktur ($\chi^2 = 689.742$, $df = 190$, $p < .001$). Die Betrachtung der Anti-Image-Matrix zeigt durchgängig hohe Werte in der Diagonale. Die MSA-Koeffizienten für die Items liegen mit Ausnahme der Items 12 ($r = .412$), 17 ($r = .303$) und 19 ($r = .451$) alle über .06 und damit in einem zufriedenstellend hohen Bereich.

Als Extraktionskriterium für die festzulegende Anzahl der Faktoren wurde eine Parallelanalyse mit den Spezifikationen $N_F = 52$ Fälle, $N_V = 20$ Variablen, $N_X = 100$ und 95-Perzent-Perzentil gerechnet. Für die Bedingung **Face-to-Face** zeigte sich nach dieser Analyse eine Einfaktorenlösung. Die Ergebnisse der Parallelanalyse zeigt Tabelle 5.8. Aus Gründen der Übersichtlichkeit werden für alle Bedingungen nur jeweils 5 der 20 in der Analyse beschriebenen Faktoren dargestellt.

Tabelle 5.8: Ergebnisse der Parallelanalyse (Face-to-Face)

Faktor	Raw Data	Means	Percentile (.95)
1	8.765689	1.710827	2.000554
2	1.439456	1.424762	1.646872
3	1.048665	1.213462	1.392785
4	.820390	1.039926	1.212511
5	.738850	.864681	1.017034

In Abbildung 5.7 wird das Ergebnis der Parallelanalyse für die Bedingung **Face-to-Face** graphisch veranschaulicht.

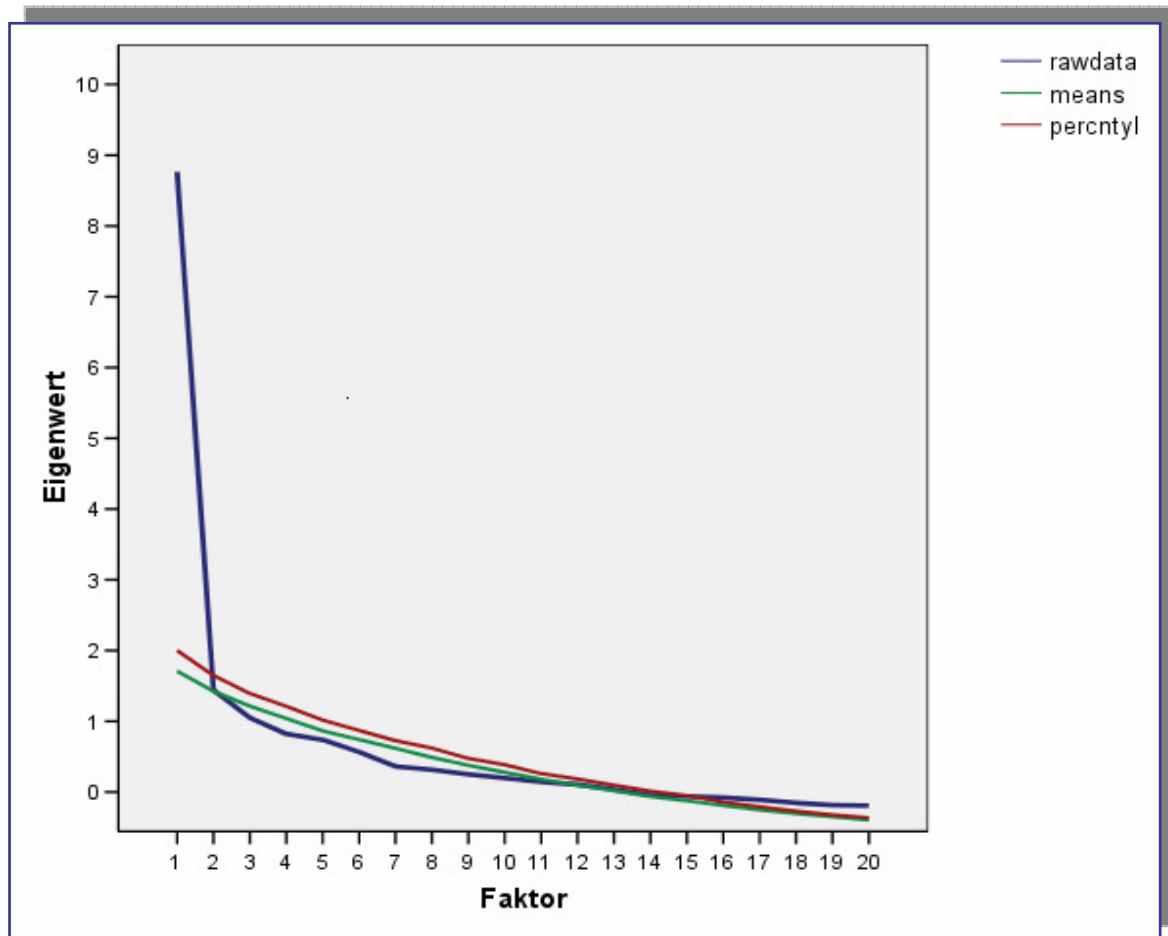


Abbildung 5.7: Ergebnisdarstellung der Parallelanalyse (Face-to-Face Rollenspiel)

Die Strukturmatrix der rotierten Faktorenlösung wird in Tabelle 5.9 berichtet. Da für die beiden anderen Bedingungen eine Zweifaktorenlösung laut Parallelanalyse zugrunde zu legen ist, wird zum Zwecke einer besseren Vergleichbarkeit der drei Bedingungen auch für die Bedingung Rollenspiel eine Lösung mit zwei Faktoren vorgestellt. Dieses Vorgehen ist insofern legitim, als dass die abgebildete Struktur nicht als Grundlage für die Analyse von Unterschieden in den Bedingungen herangezogen wird, sondern nur der vergleichenden Betrachtung der drei Bedingungen dient. Es werden nur diejenigen Variablen dargestellt, die mit mindestens .4, als unterer Grenzwert nach Stevens (2002, S. 394), auf einem der extrahierten Faktoren laden. Variablen, die dieses Kriterium nicht erfüllen, sind in Tabelle 5.9 rot gekennzeichnet.

Tabelle 5.9: Strukturmatrix der extrahierten Faktoren (Face-to-Face)

	Faktor	
	1	2
Item 1		.663
Item 2		.611
Item 3		.641
Item 4	.774	
Item 5	.652	
Item 6	.794	
Item 7	.633	
Item 8	.827	
Item 9	.814	
Item 10	.872	
Item 11		.814
Item 12		
Item 13	.626	
Item 14	.671	
Item 15	.671	
Item 16	.735	
Item 17	.209	
Item 18	.840	
Item 19		
Item 20	.823	

Ergebnisse zur Faktorstruktur der Bedingung VAK

Für die Bedingung **VAK** liegt der KMO-Koeffizient mit .633 nach Bühner (2006) in einem mäßigen, aber noch akzeptablen Bereich. Auch der Bartlett-Test auf Sphärizität zeigt die Eignung der Datenstruktur ($\chi^2 = 444.109$, $df = 190$, $p < .001$). Die Betrachtung der Anti-Image-Matrix zeigt weniger durchgängig hohe Werte in der Diagonale. Die MSA-Koeffizienten für die Items 1 ($r = .434$), 11 ($r = .425$), 12 ($r = .460$), 13 ($r = .345$), 14 ($r = .418$), 16 ($r = .419$) und 17 ($r = .303$) liegen unter dem gewünschten Wert von .6. Alle anderen über .6 und damit in einem zufriedenstellend hohen Bereich.

Für die Bedingung **VAK** zeigte sich eine Lösung mit zwei Faktoren. Die Ergebnisse der Parallelanalyse zeigt Tabelle 5.10.

Tabelle 5.10: Ergebnisse der Parallelanalyse (VAK)

Faktor	Raw Data	Means	Prncntyle (.95)
1	4.890675	1.726435	1.986297
2	2.372072	1.445159	1.665099
3	1.116383	1.219726	1.372176
4	1.016560	1.033966	1.185678
5	.77863	.890532	1.044297

In Abbildung 5.8 wird das Ergebnis der Parallelanalyse für die Bedingung **VAK** graphisch veranschaulicht.

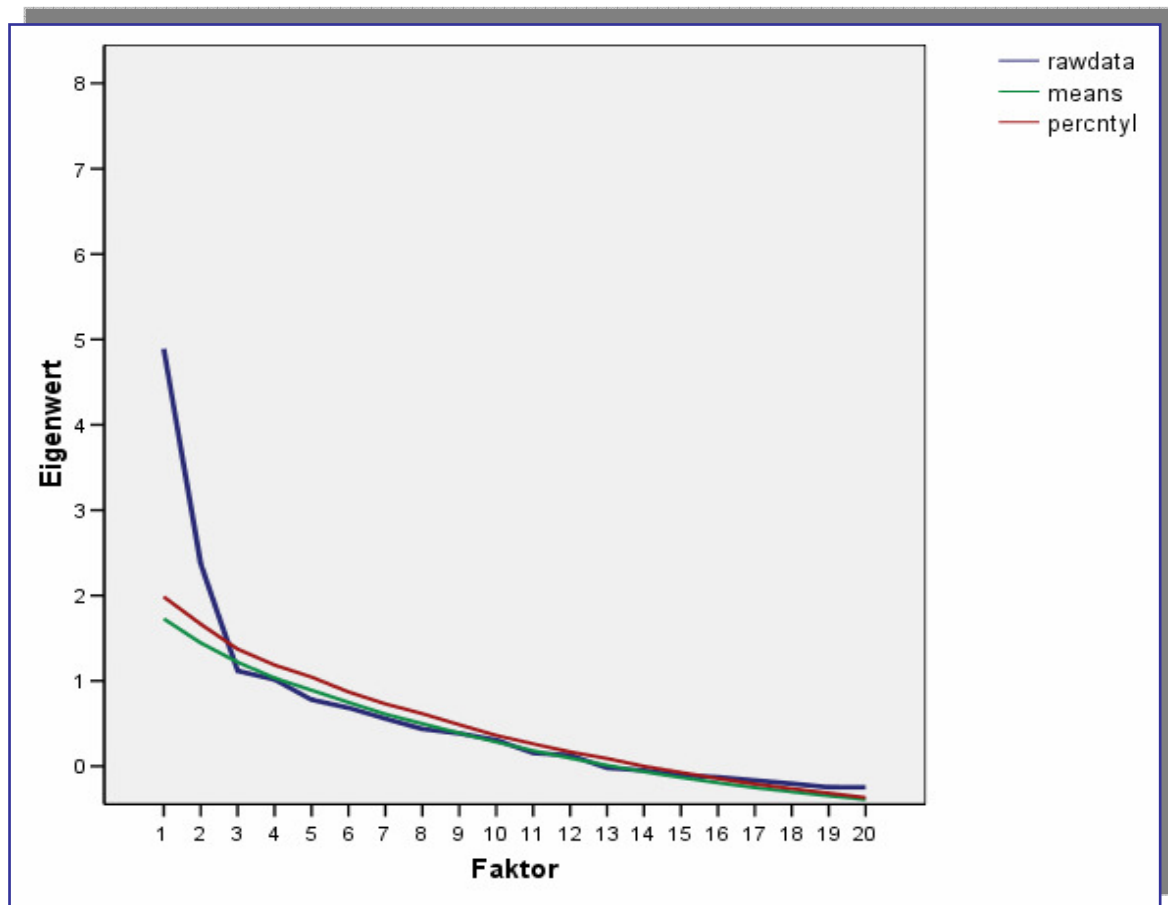


Abbildung 5.8: Ergebnisdarstellung der Parallelanalyse (VAK)

Die Strukturmatrix der rotierten Faktorlösung zeigt Tabelle 5.11. Es werden nur diejenigen Variablen dargestellt, die mit mindestens .4, als unterer Grenzwert nach Stevens (2002, S. 394), auf einem der extrahierten Faktoren laden.

Tabelle 5.11: Strukturmatrix der extrahierten Faktoren (VAK)

	Faktor	
	1	2
Item 1		.806
Item 2	.710	
Item 3		.480
Item 4	.726	
Item 5	.777	
Item 6	.421	
Item 7	.447	
Item 8		.484
Item 9	.627	
Item 10	.826	
Item 11		.814
Item 12		
Item 13		
Item 14	.671	
Item 15		.504
Item 16	.589	
Item 17		
Item 18	.710	
Item 19		
Item 20	.488	

Ergebnisse zur Faktorstruktur der Bedingung Paper-Pencil

Auch für die Bedingung **Paper-Pencil** liegt der KMO-Koeffizient mit .789 nach Bühner (2006) in einem mittleren bis guten Bereich. Der Bartlett-Test auf Sphärizität zeigt ebenfalls die Eignung der Datenstruktur ($\chi^2 = 607.440$, $df = 190$, $p = .000$). Die Betrachtung der Anti-Image-Matrix zeigt durchgängig hohe Werte in der Diagonale. Die MSA-Koeffizienten für die Items liegen mit Ausnahme der Items 12 ($r = .291$), und 19 ($r = .404$) alle über .6 und damit in einem zufrieden stellend hohen Bereich.

Für die Bedingung **Paper-Pencil** zeigte sich ebenfalls eine Zweifaktorenlösung. Die Ergebnisse der Parallelanalyse zeigt Tabelle 5.12. Aus Gründen der Übersichtlichkeit werden nur 5 der 20 beschriebenen Faktoren dargestellt.

Tabelle 5.12: Ergebnisse der Parallelanalyse (Paper-Pencil)

Faktor	Raw Data	Means	Prcentyle (.95)
1	7.695446	1.738113	1.999970
2	1.701836	1.445109	1.658268
3	.950823	1.214979	1.369505
4	.717920	1.028007	1.178824
5	.638935	.889729	.880209

In Abbildung 5.9 wird das Ergebnis der Parallelanalyse für die Bedingung **Paper-Pencil** graphisch veranschaulicht.

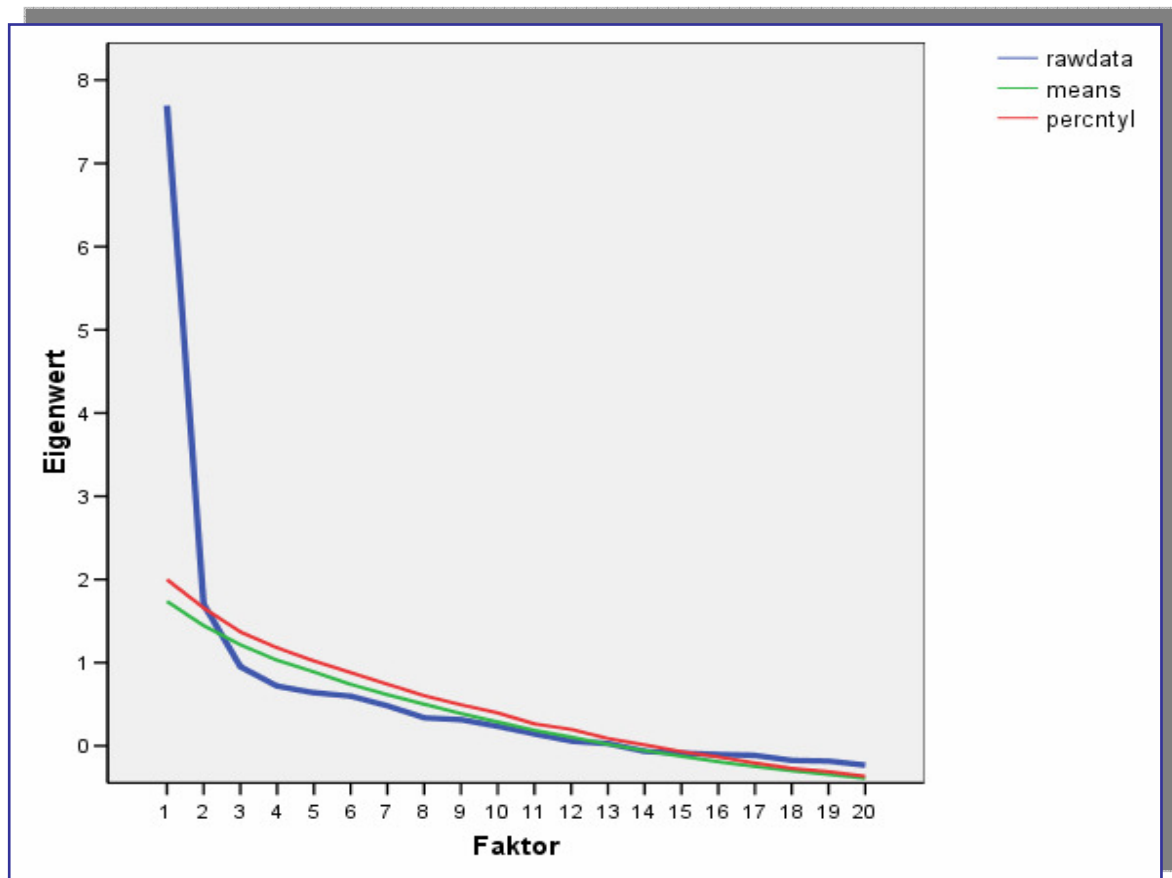


Abbildung 5.9: Ergebnisdarstellung der Parallelanalyse (Paper-Pencil)

Die Strukturmatrix der rotierten Faktorenlösung zeigt Tabelle 5.13. Es werden nur diejenigen Variablen betrachtet, die mit mindestens .4, als unterer Grenzwert nach Stevens (2002, S. 394), auf einem der extrahierten Faktoren laden.

Tabelle 5.13: Strukturmatrix der extrahierten Faktoren (Paper-Pencil)

	Faktor	
	1	2
Item 1		.769
Item 2	.675	
Item 3		.673
Item 4	.808	
Item 5	.810	
Item 6	.848	
Item 7		
Item 8	.812	
Item 9	.431	
Item 10	.924	
Item 11		.877
Item 12		
Item 13	.482	
Item 14	.725	
Item 15	.698	
Item 16	.696	
Item 17		
Item 18	.580	
Item 19		
Item 20	.760	

Fazit

In allen drei Bedingungen ergibt die Analyse der Itemstruktur einen ersten Faktor auf welchem Items aus allen fünf apriori formulierten Aspekten der sozialen Validität laden und einen zweiten Faktor mit wenigen aber hoch ladenden Items. In allen drei Bedingungen weisen die Items 1, 3 und 11 eine hohe Ladung auf dem zweiten Faktor auf und in allen drei Bedingungen können die Items 12 und 19 durch die apriori formulierten Kriterien nicht in die Betrachtung einbezogen werden. Aufgrund der hohen augenscheinlichen Ähnlichkeiten der Strukturen wird es trotz der Tatsache, dass eine Absicherung der Strukturengleichheit mittels einer konfirmatorischen Analyse aus den oben bereits genannten Gründen nicht erfolgen kann, als legitim erachtet, in einem weiteren Schritt die Analyse der Faktorenstruktur über alle Bedingungen vorzunehmen.

5.3.3 Ergebnisse zur Faktorenstruktur der zusammengefassten Fragebögen

In einem dritten Schritt erfolgte die Analyse der Datenstruktur über alle fünf Dimensionen der sozialen Validität und alle drei Bedingungen mittels einer Hauptachsenanalyse mit Promax Rotation.

Für die betrachteten Daten liegt der KMO-Koeffizient mit .89 nach Bühner (2006) in einem guten Bereich. Auch der Bartlett-Test auf Sphärizität zeigt die Eignung der Datenstruktur ($\chi^2 = 1790,745$, $df = 190$, $p < .001$). Die Betrachtung der Anti-Image-Matrix zeigt durchgängig hohe Werte in der Diagonale. Die MSA-Koeffizienten für die Items liegen mit Ausnahme des Items 12 ($r = .528$) alle über .06 und damit in einem zufriedenstellend hohen Bereich.

Auch für den Fall des gesamten Datensatzes wurde als Extraktionskriterium für die festzulegende Anzahl der Faktoren eine Parallelanalyse mit den Spezifikationen $N_F = 156$ Fälle, $N_V = 20$ Variablen, $N_X = 100$ und 95-Prozent-Percentil gerechnet. Es zeigte sich eine Lösung mit zwei Faktoren.

Die Ergebnisse der Parallelanalyse zeigt Tabelle 5.14. Aus Gründen der Übersichtlichkeit werden nur 5 der 20 beschriebenen Faktoren dargestellt.

Tabelle 5.14: Faktorenstruktur über alle Bedingungen laut Parallelanalyse

Faktor	Raw Data	Means	Prcentyle (.95)
1	7.830501	.836958	1.053402
2	1.795360	.700850	.860665
3	.664246	.593065	.721810
4	.460881	.503634	.616508
5	.401136	.424000	.527998

In Abbildung 5.10 wird das Ergebnis der Parallelanalyse des gesamten Datensatzes graphisch veranschaulicht.

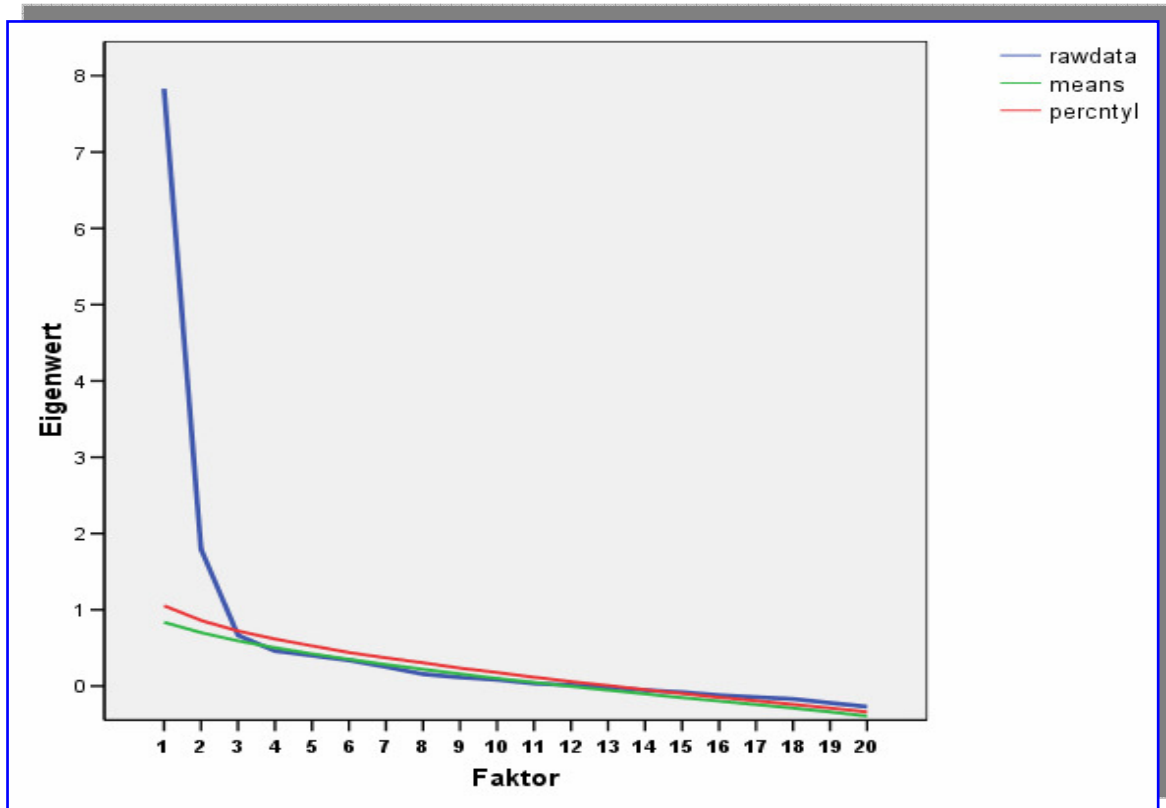


Abbildung 5.10: Graphische Ergebnisdarstellung der Parallelanalyse (alle Bedingungen)

Die Strukturmatrix der rotierten Faktorlösung zeigt Tabelle 5.15. Da für den gesamten Datensatz eine inhaltliche Interpretation erfolgen soll, werden in diesem Fall nach Fürntratt (1969) nur solche Ladungen von Variablen auf einem Faktor berücksichtigt, bei denen der Quotient aus quadrierter Ladung und Kommunalität höher als .5 liegt, d.h. wenn mindestens 50% der aufgeklärten Varianz der Variablen auf den Faktor entfallen, dem diese Variable zugeordnet wurde (siehe hierzu auch Bortz, 2006, S. 552). Darüber hinaus wird als Untergrenze nach Stevens (2002, S. 394) ein Wert von .4 festgelegt.

Die Voraussetzungen für eine generalisierende Interpretation der ermittelten Faktorstruktur, unabhängig von der Größe der Stichprobe, sind nach Guadagnolie und Velicer (1988) gegeben, da auf jedem bedeutsamen Faktor mindestens vier Variablen eine Ladung von über .60 aufweisen (Bortz, 2006, S. 523).

Die so ermittelte Struktur des gesamten Datensatzes zeigt, dass auf dem ersten Faktor 12 Items des Fragebogens eine hohe Ladung aufweisen. Eine inhaltliche Betrachtung zeigt, dass dieser Faktor Items aus jeder der apriori formulierten fünf Itemgruppen zur sozialen Validität bindet. Er wird deshalb mit dem Label *Generelle soziale Validität und Fairness* versehen. Auf dem zweiten Faktor laden zwei Items aus dem Bereich *Information* sowie

jeweils ein Item aus den Bereichen *Transparenz* und *Akzeptanz*. Aus der inhaltlichen Betrachtung der vier Items ergibt sich eine Faktorbeschreibung, welche die Aspekte der Transparenz in der Anforderung an den Probanden und die Verständlichkeit der Aufgabenstellung beinhaltet. Dieser zweite Faktor erhält deshalb das Label *Verständlichkeit und Transparenz*.

Tabelle 5.15: Strukturmatrix der extrahierten Faktoren für den gesamten Datensatz

	Faktor	
	1	2
Item 1		.697
Item 2	.673	
Item 3		.641
Item 4	.850	
Item 5	.811	
Item 6	.749	
Item 7		
Item 8	.731	
Item 9	.574	
Item 10	.919	
Item 11		.874
Item 12		
Item 13	.578	
Item 14	.656	
Item 15		.668
Item 16	.622	
Item 17		
Item 18	.750	
Item 19		
Item 20	.784	

Die gefundene Faktorenstruktur wird durch Studien zur Wahrnehmung der sozialen Validität von videobasiertem Material bei Situational Judgment Tests bestätigt (siehe z.B. Hamerschmidt, 2003; Kanning et al, 2006). Diese Tatsache wird als weiterer Hinweis für die Gültigkeit der in dieser Arbeit ermittelten Struktur gewertet.

5.3.4 Ergebnisse der varianzanalytischen Untersuchung

Im Folgenden wird der Vergleich der wahrgenommenen sozialen Validität für die Bedingungen Paper-Pencil, VAK und Face-to-Face anhand eines Vergleiches der Faktorscores für die zwei gefundenen Faktoren vorgenommen.

Die ermittelten Faktorscores gehen, nachdem sie wieder nach Bedingungen getrennt werden, als abhängige Variablen in die varianzanalytische Betrachtung ein, welche mit dem Ziel durchgeführt wird eventuell vorhandene Mittelwertsunterschiede in der Bewertung der sozialen Validität zwischen den medial variierten Rollenspielvarianten zu ermitteln. Die Voraussetzungen für ein varianzanalytisches Verfahren können als gegeben angesehen werden. Die Daten haben Intervallskalenniveau und zeigen keine Verletzung der Normalverteilungsannahme. Ein Test nach Kolmogorov Smirnov ergab folgende Ergebnisse: Für Faktorscore Paper Pencil 1 ($Z = .792$, $p = .558$, *ns*) und für Faktorscore Paper Pencil 2 ($Z = .613$, $p = .847$, *ns*), für Faktorscore VAK 1 ($Z = .753$, $p = .622$, *ns*) und für Faktorscore VAK 2 ($Z = .883$, $p = .417$, *ns*) sowie für Faktorscore Face-to-Face 1 ($Z = .637$, $p = .811$, *ns*) und Faktorscore Face-to-Face 2 ($Z = .804$, $p = .537$, *ns*).

Die Analyse der Faktorscores erfolgte über zwei einfaktorielle univariate Varianzanalysen mit Messwiederholung mit dem dreifach gestuften Faktor **Mediale Variation**. Die erste ANOVA beinhaltete die Faktorscores Paper Pencil 1, VAK 1 und Face-to-Face 1, die zweite ANOVA bezog die Faktorscores Paper Pencil 2, VAK 2 und Face-to-Face 2 in die Analyse ein.

Ergebnisse für den Faktor *Generelle soziale Validität und Fairness*

Für die Analyse des ersten Faktors *Generelle soziale Validität und Fairness* über die drei Bedingungen ergab der Test auf Sphärizität nach Mauchly keine Verletzung der Voraussetzungen zur Berechnung der Innersubjekteffekte. Mauchly-W = .949, $df = 2$, $p = .273$). Der Test der Innersubjekteffekte ergab ein signifikantes Ergebnis ($F(2,102) = 44,106$, $p < .001$, $\eta^2 = .464$). Ein im Anschluss berechneter Test der Innersubjektkontraste zeigte sowohl einen signifikanten Unterschied zwischen der Bedingung Paper-Pencil und der Bedingung Face-to-Face ($F(1,51) = 66.047$, $p < .001$, $\eta^2 = .564$) als auch einen signifikanten Unterschied zwischen der Bedingung VAK und Face-to-Face ($F(1,51) = 50.972$, $p < .001$, $\eta^2 = .500$). Beide Effekte können nach Cohen (1988) als stark bewertet werden. Zwischen den Bedingungen Paper-Pencil und VAK zeigte sich für den ersten Faktor kein signifikanter Unterschied ($F(1,51) = 2.705$, $p = .106$, *ns*).

Die Abbildung 5.11 zeigt die z-transformierten Mittelwerte und 95% Konfidenzintervalle der Faktorscores für den ersten Faktor getrennt nach den drei Bedingungen.

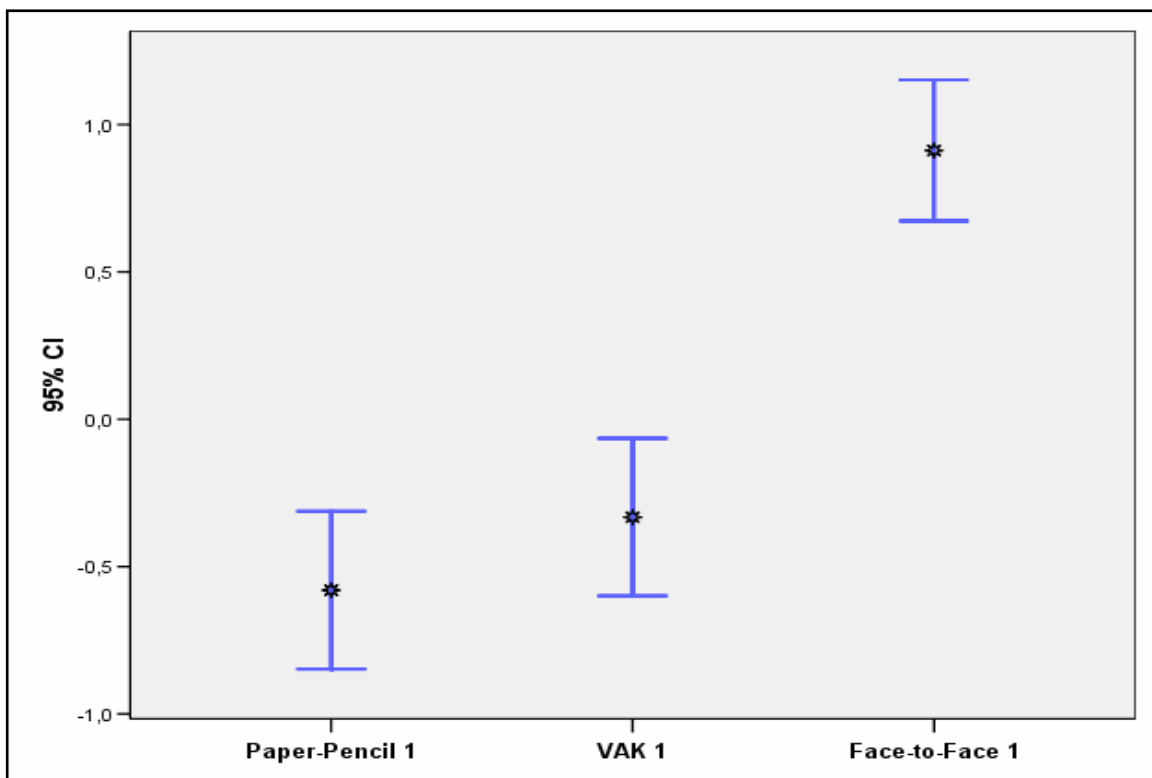


Abbildung 5.11: Mittelwerte und 95% Konfidenzintervalle der Faktorscores (Faktor 1)

Aufgrund der empirisch ermittelten Ergebnisse muss für den ersten Faktor die in Kapitel 3 formulierte Forschungsfrage, ob die Bewertung der sozialen Validität des VAK aus Sicht der Probanden gleich gut oder besser ausfällt, als für die Face-to-Face Variante des Rollenspiels mit nein beantwortet werden. Trotz einer annähernd gleich hohen *Fidelity* beider Verfahrensvarianten und der Tatsache, dass im Falle des VAK eine Bewertung der Probanden erfolgt, die vom Urteil externer Beobachter unabhängig ist, wird das VAK nicht als sozial valider wahrgenommen, sondern signifikant schlechter bewertet. Weiterhin zeigt sich, dass sich die wahrgenommene soziale Validität des VAK für den ersten Faktor nicht signifikant von derjenigen der Paper-Pencil Variante unterscheidet, obwohl es sich bei dem VAK um ein Verfahren mit einer hohen *Fidelity* handelt, während die Paper-Pencil Variante als *Low-Fidelity-Verfahren* zu bezeichnen ist. Ein signifikanter Unterschied zwischen diesen beiden Varianten wäre aufgrund der in Kapitel 2.8.4 berichteten Ergebnisse zu SJTs aber zu erwarten gewesen. Im Gegensatz hierzu ergibt sich für den betrachteten Unterschied zwischen der Face-to-Face Variante und der Paper-Pencil Variante ein Befund, der

in Richtung der in Kapitel 3 formulierten Forschungsfragen weist. Die Bewertung der sozialen Validität unterscheidet sich für diese beiden Varianten signifikant, was aufgrund ihrer unterschiedlichen *Fidelity* erwartet wurde.

Ergebnisse für den Faktor *Verständlichkeit und Transparenz*

Für die Analyse des zweiten Faktors *Verständlichkeit und Transparenz* über die drei Bedingungen ergab der Test auf Sphärizität nach Mauchly keine Verletzung der Voraussetzungen zur Berechnung der Innersubjekteffekte. Mauchly-W = .908, $df = 2$, $p = .089$). Der Test der Innersubjekteffekte ergab ein signifikantes Ergebnis ($F(2,102) = 34,879$, $p < .001$, $\eta^2 = .406$). Ein im Anschluss berechneter Test der Innersubjektkontraste zeigte keinen signifikanten Unterschied zwischen den Faktorscores für die Bedingung Paper-Pencil und die Bedingung Face-to-Face ($F(1,51) = 2.499$, $p = .120$, *ns*). Ein signifikanter Unterschied ergab sich aber für die Faktorscores der Bedingungen VAK und Face-to-Face ($F(1,51) = 56.492$, $p < .001$, $\eta^2 = .526$) und für die Faktorscores der Bedingungen Paper-Pencil und VAK ($F(1,51) = 55.279$, $p < .001$, $\eta^2 = .520$). Beide Effekte können nach Cohen (1988) als stark bewertet werden. Die Abbildung 5.12 zeigt die z-transformierten Mittelwerte und Standardabweichungen der Faktorscores für den zweiten Faktor getrennt nach den drei Bedingungen.

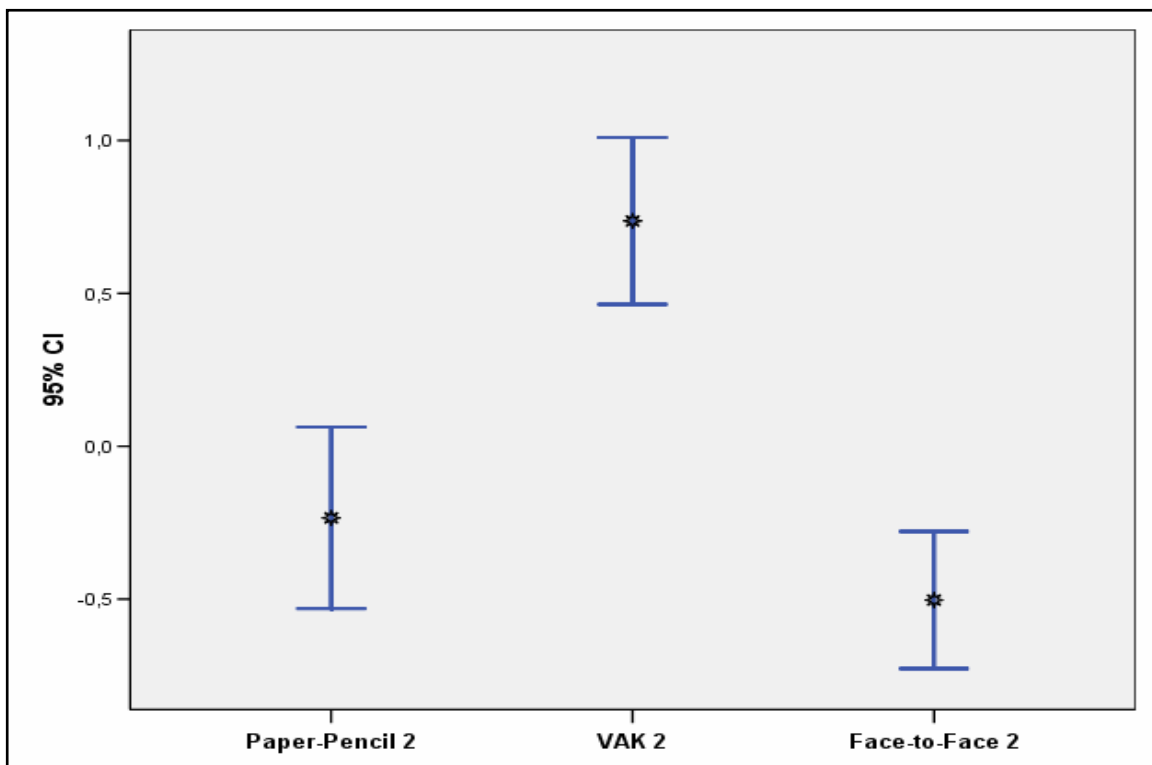


Abbildung 5.12: Mittelwerte und 95% Konfidenzintervalle der Faktorscores (Faktor 2)

In Bezug auf die in Kapitel 3 formulierten Forschungsfragen und Hypothesen ergibt sich für den Fall des zweiten empirisch ermittelten Faktors der sozialen Validität somit ein anderes Bild. Hier wird das *VAK* als videobasiertes Verfahren von den Probanden signifikant besser eingeschätzt als die beiden anderen Verfahrensvarianten. Dieser Befund stimmt mit Ergebnissen aus der Forschung zu *SJT* Verfahren überein, die eine bessere Bewertung der sozialen Validität von videobasierten Verfahren gegenüber herkömmlichen Verfahren aufzeigen. In Bezug auf die Forschungsfrage in Kapitel 3 muss allerdings festgehalten werden, dass für die Betrachtung von Unterschieden zwischen der Face-to-Face und der Paper-Pencil Variante ein signifikant positiveres Urteil zugunsten der Face-to-Face Variante zu erwarten gewesen wäre. Diese Einschätzung konnte empirisch aber nicht bestätigt werden.

6. Diskussion der Ergebnisse

Die Diskussion der in Kapitel 5 berichteten Ergebnisse erfolgt anhand einer inhaltlich dreigeteilten Struktur. Zu Beginn werden in Kapitel 6.1 die deskriptiven Statistiken der drei Rollenspielvarianten und des Kriteriums diskutiert und für das VAK und die Paper-Pencil Variante ein erster Ansatzpunkt zur Optimierung aufgezeigt. In einem zweiten Teil werden mit den Kapiteln 6.2 und 6.3 die Ergebnisse der empirischen Untersuchung in Bezug auf ihre Passung mit den in Kapitel 3 formulierten Forschungsfragen und Hypothesen unter Berücksichtigung des in der vorliegenden Arbeit dargestellten theoretischen Hintergrunds diskutiert. Als ein dritter Aspekt werden speziell für das VAK die Ergebnisse zur Reliabilität (Kapitel 6.4) und zur Schwierigkeit der Items in den fünf Gesprächsphasen (Kapitel 6.5) diskutiert. Da die Entwicklung und empirische Überprüfung dieses Verfahrens im Rahmen der vorliegenden Studie eine zentrale Rolle einnimmt, sollten diese beiden Aspekte für das VAK ergänzend zu den zentralen Fragestellungen bewertet werden.

6.1 Diskussion der deskriptiven Statistiken

Die Betrachtung der deskriptiven Statistiken für alle drei Rollenspielvarianten zeigt, dass sich die empirisch resultierenden Probandenscores in keinem der Verfahren über die gesamte Bandbreite der theoretisch möglichen Scores erstrecken. Die hieraus resultierende Varianzeinschränkung könnte somit als ein erster Erklärungsansatz für die geringen korrelativen Zusammenhänge der drei Verfahrensvarianten untereinander und mit dem Kriterium in Betracht kommen. Bei der Weiterentwicklung bzw. Modifizierung der drei Übungsvarianten sollte daher, aufbauend auf der positiven Tatsache, dass alle drei Verfahrensvarianten normalverteilte Daten lieferten, der Fokus auf mögliche Maßnahmen zur Vergrößerung der Varianz in den resultierenden Probandenscores gelegt werden. Da die Standards und Vorgehensweisen im Falle des Kriteriums und der Face-to-Face Variante von externer Seite bestimmt wurden, sollen nur für das VAK und die Paper-Pencil Variante Vorschläge für eine mögliche Verbesserung formuliert werden.

Ein Ansatz für das VAK könnte in einer Erhöhung der Attraktivität der beiden unteren Qualitätsstufen in den fünf dargestellten Gesprächsphasen bestehen. Diese waren von den Probanden nur in einzelnen Fällen berücksichtigt worden, was sie als Auswahlalternativen

wenig geeignet erscheinen lässt. Eine Erhöhung der Attraktivität und eine damit einhergehende erhöhte Auswahlwahrscheinlichkeit könnte die Varianz in den Probandenantworten erhöhen.

Für den Fall der Paper-Pencil Version könnte eine Unterteilung der Aufgabe in mehrere Phasen dazu führen, dass diese intensiver durchdacht werden als im Falle eines globalen Gespräches und dadurch mehr relevante Verhaltensweisen genannt werden. Hierdurch könnte der Tatsache, dass die Skala in Bezug auf die Probandenleistungen nicht bis zum Maximalscore ausgenutzt wurde, begegnet werden.

6.2 Diskussion der Ergebnisse zur kriterienorientierten Validität

Im Rahmen der ersten Forschungsfrage sollte überprüft werden, inwieweit sich die kriterienorientierte Validität der drei betrachteten Rollenspielvarianten in Bezug auf ein extern erhobenes Kriterium unterscheidet. Ausgehend von Ergebnissen aus Studien zu videobasierten SJTs wurde die Hypothese formuliert, dass sich für die videobasierte, multimediale Variante des Rollenspiels eine höhere kriterienorientierte Validität ergeben sollte als für die klassische Face-to-Face Variante. Für die Paper-Pencil Variante dagegen wurde eine niedrigere kriterienorientierte Validität erwartet. Die empirisch ermittelten Ergebnisse konnten diese Hypothese nur teilweise bestätigen. Hypothesenkonform ist das Ergebnis, dass die kriterienorientierte Validität für die Face-to-Face Variante höher ausfiel als für die Paper-Pencil Variante. Die Tatsache, dass die kriterienorientierte Validität des VAK geringer ausfiel als die der Face-to-Face Variante, widerspricht dagegen der a priori formulierten Hypothese, dass sich für das VAK ein höherer Wert zeigen sollte als für die Face-to-Face Variante. Die positiven Ergebnisse aus Studien zur kriterienorientierten Validität von videobasierten, multimedialen Situational Judgement Tests konnten damit für das Rollenspielverfahren nicht repliziert werden. Im Folgenden sollen für diesen Befund Ansätze für eine mögliche Erklärung aufgezeigt werden.

Ein erster Ansatz ergibt sich aus den in Kapitel 2.5 dargestellten Unterschieden in der *Fidelity* eines herkömmlichen Rollenspiels in Face-to-Face Form und eines herkömmlichen SJTs, welcher in Paper-Pencil Form bearbeitet wird, und den hieraus eventuell resultierenden Unterschieden im kognitiven Aufwand für die Probanden bei der Bearbeitung dieser Verfahrensarten.

Für den Fall des SJT wird dem Probanden mittels einer schriftlichen Stimuluskomponente eine soziale Situation beschrieben. Das Erfassen der sozialen Situation hängt hier von der sprachlichen Kompetenz bzw. Lesefähigkeit des Probanden ab und von seiner Fähigkeit, die komplexen schriftlichen Inhalte mental als soziale Situation zu repräsentieren. Gleiches gilt für die Responsekomponente, welche ebenfalls schriftlich erfolgt. Die Leistung, welche der Proband zeigen kann, hängt damit nicht nur von seiner tatsächlichen sozialen Kompetenz, sondern auch von seinen kognitiven Fähigkeiten und seiner Kreativität ab. Werden dem Probanden die gleichen Inhalte mittels Video dargeboten und entfällt für ihn damit der Schritt des Lesens und „Übersetzens“, sollte sich die Messung der sozialen Kompetenz als weitgehend unabhängig von diesen Faktoren zeigen. Ergebnisse aus empirischen Studien zu videobasierten SJTs bestätigen dieses Bild. So zeigen sich für schriftliche SJTs höhere Übereinstimmungen zu kognitiven Leistungen als zu Maßen der sozialen Kompetenz, während sich für videobasierte Verfahren ein gegenteiliges Bild ergibt. Die Erhöhung der *Fidelity* durch den Einsatz von Videos führt also im Falle eines Kriteriums, welches ebenfalls die soziale Kompetenz erfasst, zu einer Verbesserung der kriterienorientierten Validität (ausführlich hierzu siehe Kapitel 2.8.5).

Betrachtet man die in der vorliegenden Arbeit vorgenommene mediale Variation des Rollenspiels, so wird deutlich, dass ein Video unter diesem Aspekt keine bessere Darstellung oder Übersetzung der Interaktion zweier Individuen sein kann als die direkte Interaktion selbst. Vielmehr könnte für die Probanden das Rezipieren der Videos, verbunden mit der Fragestellung, wie das eigene Verhalten in der gezeigten Situation ausfallen würde, eine Übersetzungsarbeit erfordern, die zwar geringer ausfällt als beim Bearbeiten rein schriftlicher Übungen, aber höher als in der Interaktion selber. Die Face-to-Face Bedingung erfordert vom Probanden den Rollenspieler als tatsächlichen Kunden zu sehen und die umgebenden Beobachter auszublenden, da er nur so seine soziale Kompetenz im Umgang mit einem tatsächlichen Kunden authentisch zeigen kann. Dieser Transfer könnte jedoch einfacher sein als seine soziale Kompetenz durch das Bearbeiten von realitätsnahen Videos zu zeigen. Für den Fall der empirischen Studie der vorliegenden Arbeit ergibt sich damit die folgende Überlegung:

Da sich die Probanden am Ende ihrer Ausbildung befanden und damit schon über eine mindestens zweijährige Erfahrung im Bereich der Kundenberatung verfügten, konnten sie die realitätsfernen Aspekte der Rollenspielsituation aufgrund ihrer Erfahrung und im Rahmen eines routinierten Vorgehens gut auszublenden und ihre Fähigkeiten in der Interaktion optimal zum Ausdruck bringen. Die Bearbeitung des VAK hingegen stellte für alle Probanden

den eine völlig neue Herausforderung dar, da ein solches Verfahren erstmalig in der Organisation zum Einsatz kam. Der kognitive Aufwand, sich auf sein tatsächliches Verhalten zu konzentrieren und seine tatsächlichen sozialen Kompetenzen durch die Auswahl von Videos zu zeigen, könnte deshalb vergleichsweise höher gewesen sein. Dies führte zu einer geringeren Übereinstimmung mit dem ebenfalls Face-to-Face erhobenen Kriteriumsmaß. Ein weiterer Ansatzpunkt für die Erklärung der Ergebnisse betrifft die Art der Verhaltensskripte und -strategien, die durch den jeweiligen Aufgabentypus aktiviert werden. So zeigen Ergebnisse aus empirischen Studien, dass die kriterienorientierte Validität von SJT mit zunehmender Interaktionsmöglichkeit bzw. Handlungsmöglichkeit des Probanden steigt, wenn als Kriterium die Leistung aus Übungen herangezogen wird, die ebenfalls eine interaktive Handlung in einem sozialen Setting verlangen (ausführlich hierzu siehe Kapitel 2.9.2). Das einfache Lesen und schriftliche Bearbeiten einer sozialen Situation scheint andere Skripte und Handlungsstrategien zu aktivieren als das Betrachten und Beantworten derselben mittels eines Videos oder diese als tatsächliche Interaktion mit einem Individuum zu erfahren. Je ähnlicher die *Fidelity* der Stimulus- und Responsekomponente im eignungsdiagnostischen Verfahren derjenigen des Kriteriumsmaßes wird, desto höher scheint der korrelative Zusammenhang auszufallen, da sich die in den jeweiligen Situationen geforderten Strategien und Skripte ähnlicher werden. Für den Fall der schriftlichen SJTs bedeutet die Übersetzung in eine videobasierte Variante genauso wie die Erhöhung der Interaktivität der Responsekomponente damit immer eine Annäherung an die *Fidelity* eines Kriteriumsmaßes, welches die soziale Kompetenz durch eine Interaktionsübungen erhebt. Die kriterienorientierte Validität wird demzufolge erhöht.

Überträgt man diese Überlegung auf das Rollenspiel und den methodischen Rahmen der vorliegenden Arbeit, so entspricht die *Fidelity* des Verfahrens in seiner ursprünglichen Form derjenigen des Kriteriumsmaßes am besten. Die aktivierten Strategien und Verhaltensskripte sollten sich für diese beiden Übungen demnach gut entsprechen. Das gezeigte und bewertete Verhalten in beiden Situationen wäre dann ebenfalls ähnlich und würde zu ähnlichen Ergebnissen führen. Für den Fall des VAK sollte die mittels authentischer Videos generierte hohe *Fidelity* dazu führen, dass die Probanden die Interaktion ähnlich real erleben wie in der Face-to-Face Bedingung. Dies hätte zu einer Aktivierung der gleichen Strategien und Verhaltensskripte führen sollen wie diejenigen, die in der Face-to-Face Variante aktiviert worden waren. Im vorliegenden Fall könnte das Betrachten der dargestellten Lösungsalternativen die Probanden aber dazu verleitet haben eigene Ansätze, die sie in der direkten Interaktion ausführen würden, nochmals zu überdenken. Trotz der in Kapitel 2.8.5

berichteten Unterschiede in den Leistungen von *would do* und *should do* Instruktionen und der Berücksichtigung einer *would do* Instruktion im VAK, die diesem Aspekt Rechnung tragen sollte, scheint sich hier eine leichte Veränderung der gemessenen sozialen Kompetenz im Vergleich zu der im Rollenspiel erhobenen sozialen Kompetenz zu ergeben. Die Übereinstimmung der Leistung in der videobasierten Übung mit der des Kriteriums fällt deshalb geringer aus, als dies für die Face-to-Face Variante beobachtet wurde. Für den Fall der schriftlichen Übung sollten sich die Unterschiede in der *Fidelity* noch deutlicher auswirken. Da in diesem Fall weder eine Interaktion mit einer anderen Person erfolgt noch einzelne Szenen einer Interaktion dargestellt werden, wird von den Probanden keine spontane Reaktion auf das Verhalten einer anderen Person gefordert. Die Bewertung der sozialen Situation hängt lediglich von der schriftlichen Situation ab. Alle weiteren Handlungen werden von den Probanden nach deren anfänglichen Skript erstellt. Die Reaktion des Gegenübers muss nicht ständig in den Lösungsprozess eingebunden werden sondern wird selbstständig formuliert. Folgerichtig sollten sich die aktivierten Handlungsstrategien und Verhaltensskripte von denen in einer Face-to-Face Situation, die zur Bestimmung des Kriteriums dient, unterscheiden. Dies könnte zu den beobachteten Unterschieden in den kriterienorientierten Validitäten der Face-to-Face Übung und der Paper-Pencil Variante geführt haben.

Ein dritter Ansatz zur Erklärung der Ergebnisse ergibt sich aus einer motivationalen Perspektive. Im Rahmen der empirischen Untersuchung der vorliegenden Arbeit wurden die Probanden, auf Bitten der Personalvertretungen, darauf hingewiesen, dass die von ihnen zu absolvierende Multimedia- und Paper-Pencil Variante nicht in das Ergebnis des PES Verfahrens einfließen würden. Während also die Bearbeitung der Face-to-Face Variante und das hieraus resultierende Ergebnis von hoher Relevanz für spätere Personalentscheidungen war, könnte die Bearbeitung der beiden anderen Varianten als weniger relevant für die spätere Karriere empfunden worden sein. Diesem Aspekt wurde im Rahmen der Untersuchung mit dem Hinweis begegnet, dass die Ergebnisse zwar nicht direkt von Bedeutung für Personalentscheidungen seien, aber der Personalabteilung trotzdem zugänglich gemacht würden. Dieser Hinweis könnte aber nicht ausreichend gewesen sein, die Motivation für alle drei Bedingungen gleich zu halten. Aus dieser Überlegung ergibt sich die Schlussfolgerung, dass die Motivation bei der Bearbeitung der Face-to-Face Variante höher ausfiel, da das Ergebnis von Seiten der Probanden als besonders wichtig eingestuft wurde. Weiterhin kann man davon ausgehen, dass die Motivation in der mündlichen IHK Abschlussprüfung ebenfalls sehr hoch ist, da ein positives Abschneiden als hoch relevant für den späteren

Karriereweg angesehen wird. Während also für die Face-to-Face Variante und die IHK Abschlussprüfung die Bearbeitung bei allen Probanden hoch motiviert erfolgte und jeweils die bestmögliche Leistung erbracht wurde, könnten bei der Bearbeitung der beiden anderen Rollenspielvarianten Unterschiede in der Motivation aufgetreten sein. Dies könnte dazu führen, dass Probanden mit einer eigentlich hohen Merkmalsausprägung geringere Werte erzielten, da sie weniger motiviert arbeiteten als Probanden mit einer niedrigen Merkmalsausprägung, die aber hoch motiviert arbeiteten und deshalb bessere Ergebnisse erzielten. Dieser Sachverhalt könnte zu den beobachteten Ergebnissen in Bezug auf die kriterienorientierte Validität geführt haben.

Ein Grund dafür, dass die Ergebnisse aus vergleichenden Studien von videobasierten und herkömmlichen SJT sich von den empirischen Befunden der vorliegenden Arbeit unterscheiden, könnte in der Tatsache liegen, dass die Bedingungen bei der Bearbeitung der medial variierten SJTs gleich gehalten werden konnten und damit die Motivation der Probanden bei der Bearbeitung der Verfahren ausschließlich von der medialen Komponente abhängig war. In einem solchen Fall sollten die multimedialen Verfahren ein höheres Motivationspotential aufweisen und zu besseren Ergebnissen in Bezug auf das jeweils definierte Kriterium führen.

Ein weiterer Erklärungsansatz, der motivationale Aspekte mit einbezieht, ergibt sich aus den Ergebnissen zur wahrgenommenen sozialen Validität der drei Rollenspielvarianten (ausführlich siehe Kapitel 6.2). Die Auswertung der Probandeneinschätzungen zur sozialen Validität ergab, dass die Face-to-Face Variante auf dem Faktor *Generelle soziale Validität und Fairness* signifikant besser bewertet wurde als die beiden anderen Varianten. Weiterhin zeigen Ergebnisse aus anderen empirischen Studien, dass sich die soziale Validität eines Verfahrens auf die Motivation der Probanden und die im Verfahren erbrachten Leistungen auswirken kann (ausführlich siehe Kapitel 2.3.3). Für den Fall der vorliegenden Studie könnten die Probanden bei der Bearbeitung der Face-to-Face Variante motivierter gewesen sein, weil sie diese als sozial valider und fairer empfanden als die beiden anderen Varianten. Hieraus könnte, analog zu den motivationalen Unterschieden aufgrund der unterschiedlich wahrgenommenen Relevanz der Verfahren, der oben beschriebene Unterschied in der kriterienorientierten Validität resultieren.

Für einen letzten Erklärungsansatz sollen die in Kapitel 3 dargestellten Unterschiede in der methodischen Qualität der drei Verfahrensansätze näher betrachtet werden. Es wurde darauf hingewiesen, dass sich für das Rollenspiel Verfälschungen in der Bewertung der Probandenleistungen ergeben können, die auf Fehler in der Beobachtung oder das fehlerhafte

Verhalten eines Rollenspielpartners zurückzuführen sind. Weiterhin wurde dargelegt, dass sich für das VAK solche Fehler nicht ergeben sollten, da es sich um ein computerbasiertes Verfahren handelt, welches damit über eine größtmögliche Durchführungs-, Auswertungs- und Interpretationsobjektivität verfügt. Durch diese Eigenschaften sollte sich die Urteilsgüte erhöhen. Die Paper-Pencil Variante dagegen, deren Ergebnis ebenfalls von der Bewertung durch einen Experten abhängig ist, stellt unter diesem Aspekt keine Verbesserung dar. Diese Überlegungen wurden durch die empirisch gefundenen Ergebnisse der vorliegenden Studie nicht gestützt. Als Erklärung könnte unter methodischen Gesichtspunkten betrachtet die größere Ähnlichkeit der Urteilsfindung in der Paper-Pencil Variante mit der in der IHK-Prüfung gegenüber den beiden anderen Versionen dienen. In beiden Fällen werden die Leistungen der Probanden durch das Urteil von Beobachtern ermittelt. Als Ergebnis resultiert eine höhere Korrelation des Rollenspiels mit dem Kriterium, da sich die methodischen Vorgehensweisen ähnlicher sind. Geht man davon aus, dass sich für das VAK weder eine Passung in der Art der Erhebung noch eine sich in Bezug auf mögliche Fehler ähnelnde Strategie bei der Bewertung ergibt, erklärt sich die empirische gefundene Korrelation mit dem Kriterium.

Da für den Fall der Face-to-Face Variante die kriterienorientierte Validität allerdings ebenfalls wenig zufriedenstellend ausfiel, soll an dieser Stelle ein möglicher Erklärungsansatz für dieses Ergebnis aufgezeigt werden.

Für das Kriteriumsmaß wurde angenommen, dass die resultierenden Einschätzungen bezüglich der Probandenfähigkeiten eine hohe Urteilsgüte aufweisen, da es sich bei den Beurteilern um Experten aus verschiedenen Anwendungsbereichen derselben Fachrichtung handelte. Aufgrund der Tatsache, dass die Probanden den Beurteilern vorher nicht bekannt waren, erschien eine objektive Bewertung ebenfalls als gesichert. Bei kritischer Betrachtung ergeben sich allerdings in Bezug auf die Messung der spezifischen sozialen Kompetenz, welche im Rahmen der Kundenbetreuung erforderlich ist, zwei potentielle Störquellen. Eine erste Störquelle stellt der Aspekt dar, dass sich die Note der mündlichen Prüfung aus der beobachteten sozialen Kompetenz und der fachlichen Kompetenz zusammensetzt. Zwar schreibt das Protokoll eine höhere Gewichtung der sozialen Kompetenz bei der Notenvergabe vor, inwieweit dies von den Beobachtern jedoch eingehalten wird, liegt in der Verantwortung des jeweiligen Komitees. Eine zweite Störquelle besteht darin, dass die Mitglieder des Prüfungskomitees die Noten der schriftlichen IHK-Abschlussprüfung kennen, welche ausschließlich der Bestimmung der fachlichen Kompetenz dient. Die Kenntnis dieser Note könnte sich auf die Bewertung der mündlichen Prüfung auswirken. So könnte

ein Proband, der mit einer mittelmäßigen Leistung seine schriftliche Prüfung abgeschlossen hat, seitens der Prüfer, trotz hoher sozialer Kompetenzen, auch hier nur als mittelmäßig eingeschätzt werden. In beiden Fällen wäre die Bestimmung der sozialen Kompetenz durch den Aspekt des Fachwissens beeinflusst.

Für den Fall der Face-to-Face Variante könnten vor allem in kleineren Institutionen, wie den beteiligten Finanzdienstleistungsunternehmen, die Probanden und Beobachter miteinander bekannt sein. Dadurch besteht seitens der Beobachter bereits eine individuelle Bewertung bezüglich der sozialen Kompetenz und dem Fachwissen der Probanden, welche sich aus dem persönlichen Verhalten des jeweiligen Probanden gegenüber den einzelnen Beobachtern ergibt und zu einer Verzerrung der Beurteilung der Probandenleistung führen kann. Wenn aber sowohl für das Kriteriumsmaß als auch für die Face-to-Face Variante des Rollenspiels die Bewertung der sozialen Kompetenz durch das Vorwissen der Beobachter verfälscht wird, könnte dies die geringe Höhe des korrelativen Zusammenhangs der Probandenleistungen, trotz der Ähnlichkeit ihrer methodischen Bestimmung bewirken.

Aufgrund der wenig zufriedenstellenden Ergebnisse bezüglich der kriterienorientierten Validität der drei betrachteten Rollenspielvarianten wurde ergänzend untersucht, inwieweit sich durch die Kombination der Face-to-Face Variante mit einer der beiden medialen Variationen eine Verbesserung der Kriteriumsvorhersage gegenüber der Face-to-Face Variante als alleinigem Prädiktor ergeben würde. Eine zu diesem Zweck durchgeführte schrittweise, vorwärtsgerichtete Regressionsanalyse zeigte allerdings, dass sich keine Verbesserung der Kriteriumsvorhersage durch den zusätzlichen Einbezug des VAK oder der Paper-Pencil Variante in das Regressionsmodell erzielen ließ.

Im Rahmen einer korrelationsstatistischen Betrachtung zeigte sich weiterhin, dass sich auch zwischen den Leistungen aus den drei Rollenspielvarianten nur sehr geringe Korrelationen um Null finden ließen. Mögliche Suppressoreffekte, welche zu einer Erhöhung der Vorhersagegüte hätten beitragen können, konnten aufgrund der geringen Korrelationen der Leistungen aus den drei Verfahren untereinander und mit dem Kriteriumsmaß ausgeschlossen werden.

Eine Kombination aus den Leistungen medial unterschiedlicher Übungen der gleichen Verfahrenskategorie scheint vor allem dann zu einer Erhöhung der inkrementellen Validität beizutragen, wenn sich das Kriterium ebenfalls aus mehreren Facetten zusammensetzt (siehe z. B. Lievens et al., 2006). Da im Fall der vorliegenden Arbeit als Kriterium aber die Leistungen einer klar umgrenzten, spezifischen Übung gewählt wurden, trägt eine Kombi-

nation der Probandenleistungen aus den drei Rollenspielvarianten nicht zu einer Erhöhung der inkrementellen Validität bei.

6.3 Diskussion der Ergebnisse zur sozialen Validität

Die zweite Forschungsfrage der vorliegenden Arbeit bezog sich auf den Aspekt der sozialen Validität von medial variierten Rollenspielübungen. Konkret sollte untersucht werden, inwieweit sich Ergebnisse zur wahrgenommenen sozialen Validität aus dem Anwendungsbereich der SJTs auch für das Rollenspiel realisieren ließen.

In den dargestellten Studien zur sozialen Validität zeigten sich signifikant bessere Bewertungen für multimediale, videobasierte SJT als für deren herkömmlichen Paper-Pencil Varianten. Für Erstere konnte weiterhin gezeigt werden, dass die Bewertungen bezüglich der sozialen Validität umso positiver ausfielen, je stärker das Gewicht auf den Einsatz von Videos in der Stimulus- und Responsekomponente gelegt wurde (ausführlich siehe Kapitel 2.9.1 und 2.9.2). Als weiterer Gesichtspunkt, der sich auf die Bewertung der sozialen Validität auswirken sollte, wurde auf die höhere Objektivität des VAK als computerbasiertem Verfahren gegenüber dem Rollenspiel hingewiesen.

Ausgehend von diesen Darstellungen wurde für die vorliegende Studie die Hypothese formuliert, dass sich für das VAK eine signifikant höhere Bewertung der wahrgenommenen sozialen Validität zeigen sollte als für das Rollenspiel, während die Paper-Pencil Variante signifikant schlechter abschneiden sollte. Diese Annahmen konnten durch die empirisch gefundenen Ergebnisse ebenso wenig bestätigt werden wie die apriori vorgenommene inhaltliche Gruppierung der Items. Anstelle einer Lösung mit fünf Faktoren, welche die drei Aspekte zur sozialen Validität nach Schuler (1993) und die Aspekte der Akzeptanz und Fairness nach Gilliland (1993) abbilden sollten, zeigte sich eine Zweifaktoren Lösung mit den Faktoren *Generelle soziale Validität und Fairness* sowie *Verständlichkeit und Transparenz*. Da sich für die beiden empirisch gefundenen Faktoren in Bezug auf die formulierte Hypothese heterogene Ergebnisse zeigen, werden diese getrennt nach Faktoren diskutiert.

Für den ersten empirisch bestimmten Faktor *Generelle soziale Validität und Fairness* zeigte sich eine signifikant bessere Bewertung der sozialen Validität für die Face-to-Face Variante als für die beiden anderen Versionen. Zwischen dem VAK und der Paper-Pencil Variante konnte dagegen kein signifikanter Unterschied festgestellt werden. Während der Un-

terschied zwischen der Bewertung der Face-to-Face Variante und der Paper-Pencil Variante erwartet wurde, zeigte sich für das VAK ein Ergebnis, das den apriori formulierten Hypothesen nicht entsprach.

Als ein möglicher Ansatz zur Erklärung der Ergebnisse könnten auch für den Fall der sozialen Validität Unterschiede in der wahrgenommenen Relevanz der Aufgaben für den späteren Werdegang in Frage kommen. Wird eine Übung als weniger relevant in Bezug auf den späteren beruflichen Werdegang angesehen, könnte diese Einschätzung die Bewertung der sozialen Validität und Fairness des Verfahrens beeinflussen. Zieht man weiterhin in Betracht, dass der Faktor *Generelle soziale Validität und Fairness* Items aus allen fünf thematischen Bereichen auf sich vereint, könnte die Bewertung der Verfahren anhand des Fragebogens eher anhand eines globalen Maßstabs erfolgt sein und weniger auf inhaltlicher Basis einzelner Items. Das Verfahren, welches für den eigenen Erfolg als besonders relevant erachtet wird, bekommt damit für einen Großteil der Items auch die höchste Bewertung, unabhängig davon, ob für einzelne inhaltliche Aspekte eventuell ein anderes Verfahren besser bewertet werden müsste. Im Gegensatz hierzu ergibt sich für die als weniger relevant erachteten Verfahren ein globales, weniger positives Urteil: Einzelne inhaltliche Aspekte, die eine bessere Bewertung rechtfertigen würden, werden nicht berücksichtigt.

Ein zweiter Ansatz für eine Erklärung der empirischen Befunde ergibt sich aus dem in Kapitel 2.3.2 dargestellten Zusammenhang von Erfahrungen mit einem diagnostischen Instrument bzw. den Erwartungen an ein solches und der Bewertung der Fairness dieses Instrumentes. So könnte die schlechtere Bewertung des VAK und der Paper-Pencil Version gegenüber der Face-to-Face Variante in der Neuartigkeit dieser beiden Instrumente aus Sicht der Probanden liegen. Die Tatsache, dass die beiden medialen Variationen zum ersten Mal im Rahmen des PES eingesetzt wurden, könnte sich, ähnlich wie der Aspekt der Relevanz, auf die mit dem ersten Faktor vorgenommene globale Einschätzung des Verfahrens ausgewirkt haben. Da die Probanden durch ihre mehrjährige Firmenzugehörigkeit das PES Verfahren und die darin enthaltenen Übungen kannten, wurde das Face-to-Face Rollenspiel als bereits etabliertes Instrument sozial valide und fair eingeschätzt. Die beiden anderen Verfahren, für welche keine Erfahrungswerte im Unternehmen vorlagen und damit für die Probanden keine Möglichkeit gegeben war, sich im Vorfeld des PES Informationen einzuholen, wurden als weniger sozial valide bewertet.

Als ein dritter Erklärungsansatz für die Abweichung der empirischen Ergebnisse von der apriori formulierten Hypothese könnten Unterschiede in der wahrgenommenen Passung des eigenen Handlungsspielraumes in den drei Verfahrensvarianten mit demjenigen in der

Berufsrealität dienen. Dies soll im Folgenden näher erläutert werden. Es wurde dargestellt, dass die Erhöhung der *Fidelity*, im Fall der SJTs, zu einer Verbesserung der wahrgenommenen sozialen Validität führt. Durch den Einsatz von Videos und die Erhöhung der Handlungsspielräume in der Responsekomponente wurden für den Probanden die Realitätsnähe der geforderten Handlungen und die des Settings deutlich verbessert. Für das Rollenspiel gilt dies nicht in gleichem Maße. Es wird zwar das Setting realitätsnäher gestaltet, die vom Probanden zu erbringende Handlung dagegen sinkt in ihrer *Fidelity*. So könnten die Probanden die Tatsache, dass sie im VAK nicht selber mit dem potentiellen Kunden interagieren dürfen, als eine größere Einschränkung der Realität ansehen als die realitätsferne Umgebung und die Abhängigkeit von den Beobachterurteilen in der Face-to-Face Situation. In diesem Fall würden die Probanden bei der Bewertung der Verfahrensfairness den Aspekt einer höheren Objektivität von computerbasierten Verfahren gegenüber einer Face-to-Face Übung sowie mögliche Beurteilerfehler seitens der Beobachter vernachlässigen und den Schwerpunkt auf den Vergleich der Realitätsnähe der eigenen Handlungen in den Verfahren legen. Die Umgebungsfaktoren und andere Charakteristika der drei Rollenspielvarianten treten in den Hintergrund. Für diesen Fall erklärt sich nicht nur die bessere Bewertung der Face-to-Face Variante gegenüber den beiden anderen Variationen, sondern auch der nicht existente Unterschied zwischen dem VAK und der Paper-Pencil Variante. In beiden Fällen können die Probanden ihr Potential nicht in der direkten Interaktion mit einem Kunden zeigen und bewerten diese Varianten deshalb gegenüber dem Face-to-Face Verfahren negativer. Die Bewertung des VAK und der Paper-Pencil Version fällt trotz der anderen Unterschiede in der *Fidelity* gleich aus, da diese für den Schwerpunkt der Bewertung nicht relevant sind. Die wahrgenommene Möglichkeit flexibel und realitätsnah handeln zu können, ist scheinbar eng mit der Einschätzung der Fairness eines Verfahrens verknüpft. Dieser Gedanke wird durch die Darstellungen zu den *procedural justice rules* in Kapitel 2.3.2 unterstützt, welche die *opportunity to perform* als einen Einflussfaktor für die Einschätzung der Verfahrensfairness nennen.

Für den zweiten empirisch gefundenen Faktor *Verständlichkeit und Transparenz* zeigt sich ein anderes Bild in Bezug auf die Bewertung der drei Verfahrensvarianten. Die Probanden bewerteten das VAK signifikant besser als die beiden anderen Rollenspielversionen und urteilten damit hypothesenkonform. Allerdings machten sie keinen Unterschied zwischen der Face-to-Face Variante und der Paper-Pencil Variante, was im Fall der vorliegenden Untersuchung aber erwartet wurde.

Die Tatsache, dass sich für den zweiten resultierenden Faktor eine signifikant positivere Einschätzung für das VAK als für die beiden anderen Rollenspielvarianten ergibt, obwohl dies für den ersten Faktor nicht der Fall war, soll im Folgenden diskutiert werden.

Ein möglicher Erklärungsansatz könnte in der Betrachtung formaler Aspekte liegen. Im VAK werden dem Probanden neben einer Erläuterung des Programms und der Aufgabenstellung auch die einzelnen Gesprächssequenzen, geordnet nach den Phasen eines Kundengesprächs, präsentiert. Die Aufgabe besteht darin, die Sequenzen zu betrachten und zu markieren. Der Proband mag zwar eine gewisse Unsicherheit in Bezug auf die Qualität seiner Auswahl verspüren, er weiß aber zu jeder Zeit genau, welches Verhalten von ihm gefordert ist. Für die beiden anderen Verfahren muss dies nicht in gleichwertigem Maße gelten. Hier bekommt der Proband zwar ebenfalls eine Instruktion, er interagiert aber im Beispiels des Face-to-Face Verfahrens mit einer weiteren Person und muss sich auf diese immer wieder neu und individuell einstellen. Zu einer möglichen Unsicherheit bezüglich der Qualität seiner Lösungsansätze kommt in diesem Fall eine mögliche Unsicherheit in Bezug auf die nächsten Reaktionen oder Aktionen des Kunden. Weiterhin wird hier nicht, wie im VAK, die Struktur klar vorgegeben, sondern muss stattdessen vom Probanden selber gefunden werden. Dieser kann sich nach Ablauf der 15minütigen Übung nicht sicher sein, ob er das von den Beobachtern gewünschte Verhalten gezeigt hat bzw. die von ihm gewählte Struktur des Gesprächs der aus Sicht der Beobachter optimalen Struktur entspricht. Gleiches gilt für die Paper-Pencil Version. Der Proband bekommt auch hier lediglich eine Aufgabenstellung, wird aber danach nicht weiter mittels eines Programms geführt. Er sucht sich seine Struktur individuell und weiß nicht, ob er damit die Erwartungen des Aufgabengestellers trifft. Da sich das VAK positiv im Sinne einer höheren Strukturierung der Aufgabenstellung und Bearbeitung von den beiden anderen Verfahren unterscheidet, ergibt sich für das VAK eine signifikant bessere Bewertung bezüglich der Verständlichkeit und der Transparenz des Verfahrens. Die beiden anderen Verfahren werden schlechter bewertet und unterscheiden sich in der Wahrnehmung der Probanden nicht, da die Probanden in beiden Fällen mit einer wenig strukturierten Aufgabe konfrontiert sind, die mit einer größeren Unsicherheit bezüglich der optimalen Vorgehensweise behaftet ist.

6.4 Diskussion der Schwierigkeiten der fünf Phasen des VAK

Die ermittelten Schwierigkeiten für die fünf Gesprächsphasen lassen ein positives Fazit zu. Sowohl die einzelnen Indizes der jeweiligen Phasen als auch deren Abfolge entsprechen den Hinweisen zur Auswahl von Aufgaben nach ihrer Schwierigkeit nach Lienert und Raatz (1998). Eine optimale Streuung der Schwierigkeitsindizes für Niveautests sollte hiernach über den ganzen Bereich eines Persönlichkeitsmerkmals streuen, wenigstens aber von $P = .20$ bis $P = .80$ reichen. Die Forderung nach einer so genannten Eisbrecher-Aufgabe kann mit einer Schwierigkeit von $P = .67$ bei der ersten zu bearbeitenden Sequenz genauso als erfüllt angesehen werden, wie die Richtlinie, dass kein Proband der Analysestichprobe ohne Punkte ausgehen sollte und nur wenige Probanden alle Aufgaben richtig lösen können sollten. In Bezug auf die Itemschwierigkeiten wird aufgrund der empirischen Ergebnisse kein weiterer Verbesserungsbedarf für das VAK gesehen.

6.5 Diskussion der Reliabilität des VAK

Die Analyse der internen Konsistenz berechnet nach Cronbachs alpha ergab mit $\alpha = .483$ einen niedrigen Wert. Dieses Ergebnis wurde aufgrund der in Kapitel 2.8.5 berichteten Ergebnisse zur Reliabilität von videobasierten Situational Judgement Tests auch für das VAK erwartet. Da das VAK, wie die meisten videobasierten Verfahren, den Probanden mit realitätsnahen und daher multidimensionalen Situationen konfrontiert, die ein multidimensionales Antwortverhalten in Richtung einer erfolgreichen Lösung erfordern, wurden keine hohen Homogenitätswerte erwartet und eine Unterschätzung der wahren Reliabilität in Kauf genommen. Entscheidend für den vorliegenden Fall war lediglich die Forderung, dass sich zwischen den einzelnen Phasen und dem Gesamtergebnis ein positiver Zusammenhang ergeben sollte. Diese Forderung kann aufgrund der empirischen Ergebnisse als erfüllt angesehen werden. Für zukünftige Untersuchungen bezüglich der Zuverlässigkeit des VAK wären Designs wünschenswert, die eine Bestimmung der Retestreliabilität zulassen. Im Fall der vorliegenden Arbeit konnte dies nicht realisiert werden, da die Probanden das PES-Verfahren nur einmal durchliefen und für eine zweite Testung somit nicht mehr zur Verfügung standen.

Nachdem die deskriptiven Statistiken und die empirischen Ergebnisse in Bezug auf die a priori formulierten Hypothesen unter Berücksichtigung der theoretischen Grundlagen sowie Ergebnisse zur Reliabilität und den Schwierigkeiten der Items des *VAK* diskutiert wurden, dient das folgende Kapitel einer kritischen Reflexion der methodischen Vorgehensweise, welche zur Prüfung der Hypothesen gewählt wurde.

7. Diskussion der methodischen Vorgehensweise

Analog zur Darstellung der methodischen Vorgehensweise wird auch deren kritische Reflexion im Folgenden getrennt nach den Aspekten Material, Stichprobe und Versuchsaufbau vorgenommen. Zuvor werden der gewählte Anwendungskontext und das zur Validierung herangezogene Kriterium diskutiert.

7.1 Kritische Reflexion des gewählten Anwendungskontextes

Bei der Wahl des eignungsdiagnostischen Settings, welches als methodischer Rahmen der vorliegenden Studie diente, wurde bewusst eine reale Auswahlssituation einer Erhebung im universitären Kontext vorgezogen. Da sich ein Großteil der berichteten Studien zu videobasierten SJT auf Verfahren bezogen, die in realen Unternehmen mit realen Arbeitnehmern bzw. Bewerbern Anwendung fanden, wurde für den Versuch einer Übertragung der Ergebnisse auf das Rollenspiel ebenfalls ein realer eignungsdiagnostischer Kontext gewählt. Dieses Ziel wurde durch die Kooperation mit einer Akademie im Finanzdienstleistungssektor und dem Einsatz der drei Rollenspielvarianten in den von dieser Institution betreuten PES Verfahren realisiert. Da die Akademie bei der Durchführung dieses Verfahrens den Einsatz geschulter Trainer und Psychologen vorsieht und auch die Anforderungen an Beobachter, Rollenspieler und die eingesetzten Übungen den in Kapitel 2.4 dargestellten Standards entsprechen, wurde das Verfahren unter dem Gesichtspunkt seiner methodischen Qualität als geeignet angesehen. Allerdings ergeben sich bei kritischer Betrachtung zwei Einschränkungen.

Als ein Kritikpunkt wird im Nachhinein die Bedingung der Personalvertreter der beteiligten Institute angesehen, die Probanden im Vorfeld darüber zu informieren, dass die Ergebnisse aus den beiden neuartigen Verfahren nicht unmittelbar in das Gesamtergebnis des PES einfließen würden. Dieser Umstand könnte trotz des Hinweises, dass die Ergebnisse sehr wohl von Seiten der Personalverantwortlichen zur Kenntnis genommen würden, zu den in Kapitel 6.1 und 6.2 beschriebenen Auswirkungen auf die kriterienorientierte und soziale Validität geführt haben.

Ein zweiter Kritikpunkt liegt in der Tatsache, dass die drei Verfahrensvarianten sich in Bezug auf ihren Bekanntheitsgrad im Unternehmen und damit auch bei den Probanden unterscheiden. Die in Kapitel 6.1 und 6.2 diskutierten möglichen Auswirkungen dieses

Umstandes auf die kriterienorientierte und soziale Validität würde vermieden, wenn die drei Verfahrensvarianten nur solchen Teilnehmern zur Bearbeitung vorgelegt würden, welche sich von externer Seite auf eine Stelle im Unternehmen bewerben. Ein solches Setting ist aber für den Fall der vorliegenden Untersuchung nicht realisierbar gewesen, da die Anzahl der freien Stellen in den teilnehmenden Institutionen während des zur Verfügung stehenden Erhebungszeitraumes zu begrenzt war, um den Ansprüchen an eine ausreichend große Stichprobe gerecht werden zu können. Weitere Untersuchungen in Kooperation mit einem großen international agierenden Bankkonzern wären wünschenswert, um die beiden oben genannten Aspekte kontrollieren zu können.

7.2 Kritische Reflexion bezüglich der Wahl des Kriteriums

Als Kriterium der vorliegenden Arbeit wurde mit der Probandennote aus der mündlichen IHK-Abschlussprüfung ein externes Kriterium zur Validierung der drei Verfahrensvarianten herangezogen. Für diese Wahl sprachen die in Kapitel 4.1.2 beschriebenen Charakteristika der Prüfungssituation, welche die mündliche IHK-Note aus Sicht des Autors als das beste verfügbare Kriteriumsmaß erscheinen ließ. Als kritisch im Rahmen der vorliegenden Untersuchung könnte allerdings der Umstand betrachtet werden, dass den Prüfern die Noten aus den schriftlichen IHK Prüfungen bekannt waren. So könnte die Leistung in den schriftlichen Prüfungen, die das Fachwissen der Probanden erheben sollen, die Bewertung der sozialen Kompetenz seitens der Prüfer beeinflusst haben. Ein Umstand, der die Eignung der mündlichen Note als Indikator für die spezifische soziale Kompetenz der Probanden mindert. Ein weiterer Kritikpunkt liegt in der großen Ähnlichkeit des Kriteriums mit einer der drei beobachteten Rollenspielvarianten, welche die kriterienorientierte Validität eventuell beeinflusst haben könnte. Die Betrachtung der Kundenzufriedenheit mit den Probanden, das Einholen von Vorgesetztenurteilen oder die Analyse von Zielerreichungsgraden bei den Verkaufs- und Beratungszielen könnten hier sinnvollere Kriterien darstellen um die drei Rollenspielversionen in Bezug auf ihre kriterienorientierte Validität miteinander zu vergleichen. Diese waren aber im Fall der vorliegenden Arbeit nicht verfügbar. Die Messung des Berufserfolges der in dieser Studie teilnehmenden Probanden anhand der oben genannten Kriterien wird erst in zwei bis drei Jahren möglich sein. Dann aber wäre eine Untersuchung des Zusammenhanges zwischen den erbrachten Leistungen in den drei Rollenspielvarianten und den oben genannten externen Kriterien sehr wünschenswert.

7.3 Kritische Reflexion in Bezug auf die Gestaltung der verwendeten Materialien

Im folgenden Kapitel werden das VAK und die Paper-Pencil Variante sowie der Fragebogen zur Erhebung der sozialen Validität und Fairness einer kritischen Betrachtung unterzogen. Die Face-to-Face Variante wird nicht berücksichtigt, da es sich um ein Verfahren der Akademie handelte, dessen Charakteristika von Seiten der Akademie vorgegeben wurden.

Bei der Erstellung des VAK und der Paper-Pencil Variante stand ein Ziel im Vordergrund. Die Probanden sollten bei der Bearbeitung dieser Übungen die Chance haben ihre Kompetenzen in der Kundenberatung ähnlich flexibel und umfassend einzubringen, wie dies in der Face-to-Face Variante der Fall war. Trotz der sorgfältigen Übertragung der Anforderungen in der Face-to-Face Variante in eine videobasierte, multimediale Variante und eine Paper-Pencil Variante, ergeben sich im Hinblick auf die resultierenden empirischen Ergebnisse mögliche Alternativen bei der Materialgestaltung. Diese werden im Folgenden getrennt für beide Varianten vorgestellt.

Im Falle des VAK bestand die Herausforderung darin, die Anforderungen, die an einen Probanden in einer simulierten Face-to-Face Situation zwischen einem Kunden und einem Berater gestellt werden, mittels eines videobasierten, multimedialen Verfahrens realitätsnah abzubilden. Konkret sollten die Probanden durch den Einsatz von Videos, welche das Verhalten eines Beraters in den zentralen Phasen eines Kundengesprächs qualitativ abgestuft darstellten, ein virtuelles Kundengespräch führen, das einem realen Gespräch nahe kommt. Vergleicht man allerdings die Anforderungen an den Probanden in der Face-to-Face Situation mit der im VAK ergibt sich ein wesentlicher Unterschied, der in der Struktur des VAK begründet ist. Während der Proband die Struktur des Gesprächs durch das Programm vorgegeben bekommt muss er sich diese in der Face-to-Face Situation selber erarbeiten. Weiterhin bekommt jeder Proband im VAK, unabhängig davon, wie seine Wahl in der ersten Sequenz ausgefallen ist, alle Videos der zweiten Sequenz gezeigt. Die Präsentation weiterer Items ist also losgelöst von der vorherigen Aktion des Probanden. In der Face-to-Face Situation dagegen ist die Reaktion des Rollenspielers abhängig von der Aktion des Probanden. Die Güte der ersten Aktion beeinflusst hier den weiteren Verlauf des Gesprächs.

Eine Möglichkeit, um auch für die videobasierte Variante vergleichbare Bedingungen wie die in der direkten Interaktion mit einem Kunden zu erreichen, könnte in der Veränderung

dieser festen Struktur liegen. Eine flexiblere Gestaltung würde in diesem Sinne bedeuten, dass die Probanden nach dem ersten Video, in Abhängigkeit von ihrer Auswahl, ein weiteres Video vorgespielt bekommen, welches eine Kundenreaktion zeigt, die für das vorher ausgewählte Verhalten wahrscheinlich ist. Dieser Prozess würde dem in einer Face-to-Face Übung sehr nahe kommen, da auch hier der Rollenspieler individuell auf die Aktionen des Probanden reagiert.

Allerdings sprachen in der Konzeptionsphase dieser Arbeit zwei Gründe gegen eine solche Lösung auch wenn diese unter *Fidelity* Aspekten optimal erscheint. Zum einen führt die Abhängigkeit der Items zu Problemen bei der Bewertung der Probanden, da die Leistungen nicht mehr ohne weiteres vergleichbar sind und die Bewertung der individuellen Lösungswege zu Schwierigkeiten führen kann (vergleichend hierzu siehe die Diskussion zu computerbasierten Problemlöseszenarien in Kapitel 2.6.4) und zum anderen setzt ein solches Verfahren die Produktion einer sehr viel größeren Anzahl an Videosequenzen voraus, was im Kontext dieser Arbeit sowohl den zeitlichen als auch den materiellen Rahmen gesprengt hätte. Mit dem VAK wurde ein Programm geschaffen, welches die Präsentation hochwertiger, realitätsnaher Videos in einer für den Probanden klaren Struktur ermöglicht und gleichzeitig eine objektive Durchführung, Auswertung und Interpretation gewährleistet. Diese Vorteile überwiegen nach Ansicht des Autors der vorliegenden Arbeit den Nachteil einer eingeschränkten Handlungsflexibilität.

Auch für den Fall der Paper-Pencil Version ergibt sich eine mögliche Alternative. Der Proband hat hier die Möglichkeit sein Vorgehen in einem Kundengespräch individuell und flexibel zu beschreiben. Allerdings kann der Proband die Reaktionen des imaginären Kunden selber bestimmen und muss nicht wie in der Face-to-Face Situation auch unerwartete Reaktionen oder Fragen in seinem Lösungsprozess berücksichtigen. Eine Erhöhung der *Fidelity* unter diesem Gesichtspunkt könnte ähnlich wie bei einem computerbasierten Verfahren in der Vorgabe von schriftlichen Fragen und Reaktionen des Kunden liegen, die je nach Antwort des Kunden anhand von vorher definierten Sprungregeln in einem Antwortbogen festgelegt sind. Allerdings würden sich für diese abhängige Form der Itemgestaltung die gleichen Probleme wie für die multimediale Form ergeben. Bei der Erstellung der in dieser Arbeit eingesetzten Paper-Pencil Version wurde weiterhin davon ausgegangen, dass ein Proband, der in seiner schriftlichen Lösung jeweils die fünf wichtigsten Aspekte aus den fünf zentralen Phasen eines Kundengesprächs berücksichtigt, auch in einem realen Gespräch die meisten dieser Aspekte berücksichtigen wird. Dies hätte auch in der Face-to-Face Situation zur Folge, dass er die Struktur des Gesprächs zielführend gestaltet und den

Kunden aktiv durch die Beratung steuert. Dieses Vorgehen sollte zu einem positiven Gesprächsverlauf und zu einem kompetenten Eindruck beim Kunden führen. Ein Proband, der nur sehr wenige der relevanten Aspekte schriftlich berücksichtigt, sollte dies auch in der Beratungssituation tun und wäre damit viel mehr Nachfragen und Reaktionen des Kunden ausgesetzt. Das Gespräch würde negativer verlaufen, da nicht alle relevanten Aspekte angesprochen werden und der Berater hier keinen kompetenten Eindruck auf den Kunden machen würde. Aufgrund dieser Überlegungen wurde die freie Form der oben beschriebenen Alternative vorgezogen.

Als weiterer Aspekt soll der Fragebogen zur Erhebung der wahrgenommenen sozialen Validität und Fairness an dieser Stelle einer kritischen Reflexion unterzogen werden. Da derzeit kein allgemeingültiges Modell etabliert ist oder anerkannte Standards definiert sind und damit auch kein einheitliches Erhebungsinstrument existiert, wurde für diese Arbeit ein eigener Fragebogen entworfen. Bei der Auswahl eines theoretischen Rahmenkonzeptes wurden mit den Modellen von Schuler et al. (1993) und Gilliland (1993) jeweils ein populäres Konzept aus dem europäischen und dem amerikanischen Forschungsraum gewählt. Die Erstellung der Items orientierte sich dabei zum einen an Vorschlägen der Autoren der oben genannten Konzepte (Schuler, 1996, S. 184) und zum anderen an vorhergehenden Arbeiten in diesem Forschungskontext (Grewe 2005, Hammerschmidt, 2003, Kanning et al. 2006). Da die Formulierungen der Items in den betrachteten Studien aufgrund ihres spezifischen Einsatzgebietes variierten und in Bezug auf die empirisch gefundene Faktorenstruktur heterogene Ergebnisse berichtet wurden, konnte keines der Instrumente übernommen werden. Dies machte die Formulierung eigener Items genauso notwendig wie die Prüfung einer Passung der theoretisch angenommenen mit der empirisch gefundenen Struktur.

Als ein möglicher Kritikpunkt kann in diesem Zusammenhang die Vorgehensweise bei der Prüfung der empirischen Struktur des vorliegenden Fragebogens gesehen werden. Zwar wurden die Items im Rahmen von zwei Vorstudien getestet, bevor der Fragebogen in der Hauptuntersuchung zum Einsatz kam, eine faktorenanalytische Überprüfung der Itemstruktur konnte aber aufgrund der geringen Fallzahlen nicht durchgeführt werden. Die Modifikation des Fragebogens basierte damit eher auf Aspekten wie der Verständlichkeit der Items und weniger auf strukturellen Betrachtungen. Dieser Umstand wurde in Kauf genommen, da eine größere Anzahl an Probanden aus der Zielpopulation für die Voruntersuchung nicht verfügbar war und Ergebnisse aus einer Untersuchung an z. B. einfacher verfügbaren Studierenden unter dem Aspekt einer mangelnden Repräsentativität kritisch zu

bewerten gewesen wären. Weiterhin wurde in der vorliegenden Arbeit nicht unmittelbar mit den Itemwerten, sondern mit den Faktorscores gerechnet, was den oben genannten Kritikpunkt ebenfalls relativiert. Nichts desto trotz wäre für weitere Untersuchungen ein faktoranalytisch abgesichertes Instrument wünschenswert, welches die zentralen theoretischen Konzepte der sozialen Validität und Fairness einzeln zu erfassen vermag. Hiermit ließen sich Fragestellungen zu diesen Konzepten sehr viel genauer und detaillierter erfassen, als dies im Falle des vorliegenden Fragebogens möglich wäre.

7.4 Kritische Reflexion der untersuchten Stichprobe

Die zur Beantwortung der in dieser Arbeit formulierten Forschungsfragen untersuchte Stichprobe bestand aus $N = 52$ Auszubildenden, welche bei unterschiedlichen Finanzdienstleistungshäusern eines Vertriebsgebietes beschäftigt waren. Da die Auswahl und Ansprache der Häuser zufällig geschah und die Akademie weiterhin keinen Einfluss darauf hatte, welches Haus zuzugewiesen würde, kann die Stichprobe als repräsentative Auswahl aus der Gesamtpopulation aller Auszubildenden der von der Akademie betreuten Häuser angesehen werden. Als Kritikpunkt ergibt sich allerdings die Stichprobengröße. Aufgrund der relativ geringen Probandenzahl müssten die betrachteten Unterschiede deutlich stärker ausfallen als bei Fallzahlen von mehreren hundert Versuchspersonen, um ein signifikantes Ergebnis zu erzielen. In den meisten Studien zu videobasierten SJTs werden allerdings solche Fallzahlen berichtet. Die Vergleichbarkeit der Ergebnisse solcher Studien mit denen der vorliegenden Arbeit ist deshalb nur bedingt möglich. Da für den Fall des Rollenspiels die zu erwartenden Unterschiede zwischen der herkömmlichen Form und ihrer multimedialen Variante aufgrund der bereits berichteten geringen Unterschiede in der *Fidelity* möglicherweise kleiner sind als für herkömmliche SJTs und deren multimedialer Entsprechung, wäre eine höhere Versuchspersonenzahl wünschenswert gewesen.

Trotz dieser Einschränkung wird die Größe der vorliegenden Stichprobe als ausreichend angesehen, da sich für die Unterschiede in Bezug auf die wahrgenommene soziale Validität ein starker Effekt zeigte. Berücksichtigt man weiterhin, dass Erhebungen im realen Anwendungskontext für die beteiligten Unternehmen einen hohen Kosten- und Organisationsaufwand bedeuten und einige arbeitsrechtliche Aspekte zu berücksichtigen sind, kann die Beteiligung von 52 Auszubildenden aus unterschiedlichen Finanzdienstleistungshäusern als zufriedenstellendes Resultat gewertet werden.

7.5 Kritische Reflexion des Versuchsablaufs

In Bezug auf den in der Hauptstudie der vorliegenden Arbeit realisierten Versuchsablauf ergibt sich ein wesentlicher Kritikpunkt. Bei der Zielsetzung, den Vergleich der drei medialen Variationen eines Rollenspiels in einem realen Personalauswahlprozess durchzuführen, mussten die organisatorischen Rahmenbedingungen der Akademie berücksichtigt werden. Da für das gesamte Verfahren nur ein Tag zur Verfügung stand und die Form des *PES* auf Bitten der Personalvertreter derjenigen der Vorjahre weitestgehend entsprechen sollte, konnten keine alten Übungen durch die neuen Elemente ersetzt werden. Vielmehr mussten das *VAK* und die Paper-Pencil Version in den ohnehin vollen Zeitplan integriert werden. Da die im *PES* gezeigten Videos teilweise ein mustergültiges Vorgehen eines Beraters in einem Kundengespräch zeigten, sollte dieses nicht vor den anderen Übungen durchgeführt werden, da sonst die Leistung in letzteren möglicherweise verfälscht worden wäre. Die lange Bearbeitungszeit sprach weiterhin gegen eine Bearbeitung am Ende des *PES*. Aufgrund dieser Überlegungen wurde lediglich die schriftliche Version in den tatsächlichen *PES*-Ablauf integriert, während das *VAK* am folgenden Tag durchgeführt wurde. Trotz der Tatsache, dass die Bearbeitung des *VAK* noch vor den Feedbackgesprächen durchgeführt wurde und damit in das Gesamtkonzept des Verfahrens integriert werden konnte, stellt dies eine Trennung dar, welche sich auf die Ergebnisse zur kriterienorientierten und sozialen Validität des *VAK* in dem in Kapitel 6 diskutierten Sinne ausgewirkt haben könnte. Eine Betrachtung aller drei Variationen an einem Tag in einem Verfahren ohne die hier vorgenommene Trennung wäre für folgende Untersuchungen somit wünschenswert.

7.6 Kritische Reflexion des methodischen Designs

Im Rahmen der vorliegenden Arbeit wurde ein within-subjects Design gewählt. Ziel war es bei der Bewertung des Verfahrens für alle Probanden über alle drei Bedingungen ein einheitliches Referenzsystem zu gewährleisten und mögliche Einflüsse von Gruppenunterschieden auf die Ergebnisse zu kontrollieren. Diese Vorgehensweise birgt allerdings das Risiko, dass die Probanden bei ihrer Bewertung Aspekte wie die Reihenfolge der Übungen oder die Tatsache einer organisatorischen Trennung, wie im Fall des *VAK* von den anderen Übungen, in ihre Bewertung einfließen lassen. Diese Aspekte würden durch ein between subject Design kontrollierbar. Da für eine vergleichende Betrachtung von medial variierten

Übungen mit identischen Inhalten in Bezug auf deren soziale Validität ein einheitlicher Bewertungsmaßstab sinnvoll erscheint, wurde diesem Aspekt bei der Auswahl des Designs der Vorrang gegeben (vergleichend siehe Kanning et al., 2006). Allerdings könnten Erkenntnisse, denen ein between-subject Design zugrunde liegt, sinnvolle Ergänzungen zu den bisher generierten Ergebnissen liefern.

Nachdem die methodische Vorgehensweise der vorliegenden Arbeit einer kritischen Reflexion unterzogen wurde, soll mit dem folgenden Kapitel eine generelle Diskussion der Ergebnisse erfolgen. Weiterhin werden im Rahmen eines Ausblicks mögliche Implikationen für Forschung und Praxis dargestellt, die sich aus den empirischen Ergebnissen der vorliegenden Arbeit ergeben.

8 Generelle Diskussion und Ausblick

Mit der generellen Diskussion sollen die Ergebnisse aus den einzelnen Abschnitten der gesamten Arbeit zusammengefasst und unter dem übergeordneten Aspekt der Weiterentwicklung und Verbesserung eignungsdiagnostischer Verfahren durch den gezielten Einsatz von Medien betrachtet werden. Eine solche übergeordnete Diskussion setzt voraus, dass die vorliegende Arbeit eine Integration in den bisherigen Entwicklungsprozess der Eignungsdiagnostik unter medialen Aspekten erfährt, da sich sowohl die Motivation der vorliegenden Studie als auch die angelegten Bewertungsmaßstäbe aus ihrer Zugehörigkeit zu einer der medialen Entwicklungsstufe ergeben.

8.1 Integration der Arbeit in den historischen Entwicklungsprozess

In Kapitel 2.7 wurde der mediale Entwicklungsprozess eignungsdiagnostischer Verfahren dargestellt und die zentralen Entwicklungsstufen voneinander abgegrenzt. Als übergeordnete Motivation der Forschung wurde dabei die Weiterentwicklung eignungsdiagnostischer Verfahren durch den Einsatz neuer Technologien beschrieben. Neben dieser generellen Motivation weisen die empirischen Studien der einzelnen Stufen weiterhin sehr spezifische Charakteristika in Bezug auf die ihnen zugrunde liegende Motivation auf. In der Stufe der computerbasierten Auswertung von Paper-Pencil Tests wurde der Focus auf eine Erhöhung der Objektivität und der Verbesserung des diagnostischen Prozesses unter zeitlichen Gesichtspunkten gelegt. In der Stufe der Übertragung von Paper-Pencil Tests in eine Computerversion stand dagegen die Bestimmung der Äquivalenz der Versionen im Vordergrund. Als zentral für die Stufe des computerbasierten adaptiven Testens wurden die Bemühungen einer Optimierung bei der Konstruktion und Durchführung von Tests, die auf probabilistischen Testtheorien basierten, dargestellt, während in der Stufe der computerbasierten Problemlöseszenarien die Abbildung von komplexen Problemwelten zur Bestimmung der Problemlösekompetenz im realen Berufsleben mit Hilfe des Computers optimiert werden sollte. Die vorliegende Arbeit lässt sich in Bezug auf ihre Motivation und die methodische Vorgehensweise der Stufe der multimedialen Diagnostik als der letzten der betrachteten Entwicklungsstufen zuordnen. Als ein zentrales Ziel dieses Forschungsfeldes wurde die Erhöhung der *Fidelity* von Situational Judgement Tests durch den Einsatz multimedialer Komponenten vorgestellt. Die Wiedergabetreue dieser Verfahrensart in Bezug auf die spätere

berufliche Realität wurde mit dem Ziel erhöht eine Verbesserung der Gütekriterien sowie der sozialen Validität und der Verfahrensfairness zu erreichen.

Im Fall der vorliegenden Arbeit sollten erste Antworten auf die Frage gefunden werden, ob sich die positiven Ergebnisse aus Studien zu videobasierten Situational Judgement Tests auch für Face-to-Face Verfahren in der Eignungsdiagnostik realisieren lassen. Zu diesem Zweck wurden zwei medial variierte Versionen eines Rollenspiels und eine klassische Rollenspielübung im Rahmen eines realen Auswahlverfahrens im Finanzdienstleistungssektor darauf hin untersucht, inwieweit sich ihr unterschiedlicher Grad an *Fidelity* auf die jeweilige kriterienorientierte und soziale Validität auswirken würden. Ein besonderes Augenmerk galt hierbei dem VAK als videobasierten, multimedialen Verfahren. Dieses wurde mit dem Ziel entwickelt die *Fidelity* eines herkömmlichen Rollenspiels gleichwertig abzubilden und mit den weiteren Vorteilen eines computerbasierten Verfahrens zu kombinieren. Diese Optimierung stellte den Versuch dar einen Beitrag zur Verbesserung des diagnostischen Prozesses leisten zu können. Neben der vergleichenden Betrachtung der kriterienorientierten und sozialen Validität galt das Interesse dieser Arbeit auch dem Aspekt, inwieweit das VAK in Bezug auf andere relevante Gütekriterien den Anforderungen an ein modernes eignungsdiagnostisches Instrument gerecht werden kann.

8.2 Der Bewertungsmaßstab

Im Folgenden werden für die generelle Diskussion alle relevanten Ergebnisse und Erkenntnisse in Bezug auf die Qualität des VAK, welche sich aus den einzelnen Stufen des gesamten Prozesses dieser Arbeit ergeben haben, zusammengefasst und vor den oben genannten Überlegungen diskutiert. Die Diskussion erfolgt dabei analog der in Kapitel 2.8.5. gewählten Darstellungsweise der Ergebnisse zu videobasierten, multimedialen Verfahren, getrennt nach den einzelnen Gütekriterien. Da für den Fall der vorliegenden Studie nicht alle Gütekriterien in einem direkten Vergleich betrachtet wurden, dienen die Ergebnisse zu Rollenspielübung allgemein sowie zu videobasierten Situational Judgement Tests als Bewertungsmaßstab für diejenigen Werte des VAK, die nicht Bestandteil der Hauptuntersuchung der vorliegenden Arbeit waren.

8.2.1 Bewertung der Objektivität

Da es sich bei dem VAK um ein computerbasiertes Verfahren handelt, welches den Probanden von Anfang bis Ende durch die einzelnen Bearbeitungsebenen führt und die Ermittlung der Probandenleistung am Ende der Übung automatisch erfolgt, kann dem VAK eine hohe Durchführungs- und Auswertungsobjektivität bescheinigt werden. In Bezug auf den Aspekt der Verbesserung des diagnostischen Prozesses folgt hieraus, dass das VAK gegenüber einem herkömmlichen Rollenspiel eine positive Weiterentwicklung darstellt, da die Objektivität bei der Durchführung und Auswertung von herkömmlichen Rollenspielen aufgrund des Einsatzes von Rollenspielern und Beobachtern naturgemäß Schwankungen unterliegen muss. Weiterhin kann das VAK unter dem Aspekt der Objektivität als gleichwertig zu videobasierten SJTs angesehen werden.

8.2.2 Bewertung der Reliabilität

Die Bewertung des VAK unter dem Aspekt der Reliabilität gestaltet sich schwieriger. Die Ermittlung der Reliabilität mittels Cronbachs alpha zeigte, dass das VAK lediglich eine niedrige interne Konsistenz von $\alpha = .483$ aufweisen konnte, was als wenig zufriedenstellend angesehen wird. Im Rahmen der inhaltlichen Diskussion wurde in diesem Zusammenhang allerdings darauf hingewiesen, dass es sich bei dem VAK um ein Verfahren handelt, welches den Probanden mit realitätsnahen und daher multidimensionalen Situationen konfrontiert, die ein multidimensionales Antwortverhalten in Richtung einer erfolgreichen Lösung erfordern. Weiterhin wurde bemerkt, dass diese Eigenschaften keine hohen Homogenitätswerte erwarten lassen, weshalb die Berechnung der internen Konsistenz eine Unterschätzung der wahren Zuverlässigkeit darstellen dürfte. Für das herkömmliche Rollenspiel wurden in Kapitel 2.4.2 Reliabilitätswerte von .57 bis .80 berichtet. Diese Werte lassen sich mit den Werten des VAK aber nicht sinnvoll vergleichen, da unterschiedliche Methodiken zur Bestimmung gewählt wurden. Für die Rollenspielübung allgemein wurde die mittels Beobachterübereinstimmung ermittelte Reliabilität berichtet und für das VAK wurde die interne Konsistenz berechnet. Ein Vergleich kann erst dann sinnvoll gezogen werden, wenn für beide Verfahrensarten z. B. die Retestrelabilität bestimmt würde. Die Frage, inwieweit das VAK in Bezug auf das Gütekriterium der Reliabilität eine Verbesserung gegenüber einem herkömmlichen Rollenspiel darstellt, kann deshalb im Rahmen dieser Ar-

beit nicht beantwortet werden. Vergleicht man die Ergebnisse zur internen Konsistenz des VAK mit den in Kapitel 2.8.5 berichteten Ergebnissen zu videobasierten SJT lässt sich ein positives Fazit ziehen. In der Spannweite der berichteten Ergebnisse, welche von $\alpha = .29$ bis $\alpha = .65$ reichte, liegt das VAK mit $\alpha = .483$ in einem mittleren Bereich.

8.2.3 Bewertung der Inhaltsvalidität

Im Rahmen der Darstellung und Bewertung der Inhaltsvalidität von videobasierten SJT Verfahren wurde darauf hingewiesen, dass bereits aufgrund der einzelnen Stufen der Itemgenerierung, die bei der Erstellung eines videobasierten Verfahrens durchlaufen werden, von einer hohen inhaltlichen Validität solcher Verfahren ausgegangen werden kann (ausführlich siehe Kapitel 2.8.3).

Auch die Item- und die Responsekomponenten des VAK wurden auf der Basis von Experteninterviews und Expertenrankings erstellt, weshalb die Inhaltsvalidität dieses Instruments als hoch bewertet wird. Bei einer vergleichenden Betrachtung des VAK mit einem herkömmlichen Rollenspielverfahren in Face-to-Face Form stellt das VAK unter dem Aspekt der Inhaltsvalidität weder eine Verbesserung noch einen Rückschritt gegenüber dem herkömmlichen Verfahren dar, da diese über ebenfalls über eine hohe Inhaltsvalidität verfügen (ausführlich siehe Kapitel 2.4.2).

8.2.4 Bewertung der Konstruktvalidität

Auf eine vergleichende Bewertung des VAK mit herkömmlichen Rollenspielverfahren vor dem Hintergrund der Konstruktvalidität wird an dieser Stelle verzichtet, da die inhaltliche und kriterienorientierte Validität für diese Verfahrensart eine vorrangige Bedeutung haben. Weiterhin gestalten sich die Methoden im Kontext der Konstruktvalidierung derart heterogen, dass ein Vergleich schwierig erscheint. Trotz der oben genannten Aspekte wurden erste Schritte in Richtung einer Konstruktvalidierung des VAK unternommen, deren Ergebnisse positiv bewertet werden können. Dies sei im Folgenden kurz erläutert. Das VAK wurde entworfen, um eine spezifische soziale Kompetenz von Beratern im Finanzdienstleistungssektor bestimmen zu können. Die Produktion und der Einsatz eines videobasierten Verfahrens erscheinen aber nur dann sinnvoll, wenn die zu messende Fähigkeit nicht durch andere Konstrukte erklärbar ist, welche anhand kostengünstigerer Verfahren erhoben wer-

den können. Im Rahmen der vorliegenden Arbeit konnte gezeigt werden, dass sich die Probandenergebnisse im *VAK* nicht durch kognitive oder generelle soziale Fähigkeiten vorhersagen lassen. Weiterhin konnte gezeigt werden, dass die Ergebnisse des *VAK* bei Probandengruppen mit langjährigen Berufserfahrung und einem hohen Grad an Beratungsexpertise signifikant höher ausfielen als bei Gruppen, deren Probanden über wenig Berufserfahrung und einem niedrigen Expertisegrad verfügten. Diese Ergebnisse werden als erster Hinweis darauf interpretiert, dass das *VAK* eine sehr spezifische Kompetenz misst, die durch Leistungen aus anderen Testverfahren nicht erklärbar ist und einen Zusammenhang zu Berufserfahrung und Beratungsexpertise aufweist. Diese Resultate stehen nicht nur im Einklang mit Ergebnissen zu videobasierten Situational Judgement Tests, sondern zeichnen das *VAK* als Instrument aus, welches ergänzend und gewinnbringend zu herkömmlichen Tests eingesetzt werden kann, da es zusätzliche Informationen über die Kompetenzen eines Probanden liefert. Ein Aspekt, der als eine Weiterentwicklung des diagnostischen Prozesses gewertet wird.

8.2.5 Bewertung der kriterienorientierten Validität

Aufgrund der empirischen Ergebnisse aus der Hauptuntersuchung der vorliegenden Arbeit lässt sich in Bezug auf die kriterienorientierte Validität derzeit kein positives Fazit ziehen. Die Werte für das *VAK* bewegten sich in einem ungenügenden Bereich. Zieht man weiterhin in Betracht, dass sich in metaanalytischen Betrachtungen von herkömmlichen Face-to-Face Rollenspielübungen eine mittlere kriterienorientierte Validität von .30 ergibt (siehe Kapitel 2.3.2), so kann aufgrund der derzeitigen Datenlage nicht von einer Verbesserung des diagnostischen Prozesses durch das *VAK* gesprochen werden. Allerdings sollte berücksichtigt werden, dass es sich hierbei um den ersten Versuch einer Validierung des *VAK* handelte und sich in Bezug auf das gewählte Außenkriterium Potential zur Optimierung ergibt. Eine abschließende Bewertung der kriterienorientierten Validität des *VAK* kann deshalb an dieser Stelle nicht vorgenommen werden. Allerdings lassen die Ergebnisse zur kriterienorientierten Validität von videobasierten Situational Judgement Tests darauf hoffen, dass sich bei der Wahl eines im realen Arbeitskontext erhobenen Außenkriteriums, wie z. B. dem Vorgesetztenurteil, eine Verbesserung der kriterienorientierten Validität des *VAK* ergeben könnte. Ob diese höher ausfällt, als dies bei herkömmlichen Rollenspielen der Fall ist, lässt sich derzeit aber nicht vorhersagen.

8.2.6 Bewertung der sozialen Validität

In Bezug auf die von den Probanden wahrgenommene soziale Validität und Fairness ergibt sich ein heterogenes Bild für den direkten Vergleich des VAK mit der herkömmlichen Rollenspielvariante in Face-to-Face Form. Auf dem ersten resultierenden Faktor *Generelle soziale Validität und Fairness* wurde die Face-to-Face Variante besser bewertet als das VAK, während sich für den zweiten Faktor *Verständlichkeit und Transparenz* ein umgekehrtes Bild ergab. Die wahrgenommene soziale Validität wurde in der vorliegenden Arbeit, aus den in Kapitel 2.3.3 dargestellten Gründen, als ein wichtiges Kriterium für die Eignung eines personaldiagnostischen Instrumentes dargestellt. Ein neu konzipiertes Instrument sollte deshalb vor allem unter diesem Gesichtspunkt eine Verbesserung darstellen. Die Tatsache, dass sich das VAK lediglich auf dem Faktor *Verständlichkeit und Transparenz* positiv von der herkömmlichen Rollenspielvariante abheben konnte, lässt deshalb derzeit kein generelles positives Fazit in Bezug auf eine Verbesserung der sozialen Validität zu. Allerdings sollten auch hier die in der Diskussion der empirischen Ergebnisse dargestellten methodischen Einschränkungen und die Tatsache berücksichtigt werden, dass es sich um die erste Untersuchung des VAK unter diesem Aspekt handelt. Unabhängig von der vergleichenden Untersuchung haben die Ergebnisse aus den Voruntersuchungen der vorliegenden Arbeit nämlich gezeigt, dass die Mittelwerte für alle fünf Skalen zur sozialen Validität und wahrgenommenen Fairness im positiven Bereich der Skala lagen. Dieses Ergebnis zeigt, dass das VAK als multimediales Verfahren von Seiten der Probanden Zustimmung erfährt und fügt sich passend in die Berichte zu anderen multimedialen videobasierten Verfahren ein.

8.2.7 Zusammenfassende Bewertung des VAK

Fasst man die obigen Ergebnisse zusammen, ergibt sich für das VAK als personaldiagnostisches Instrument ein heterogenes Bild. In Bezug auf die Objektivität und Inhaltsvalidität entspricht das VAK den Kriterien, die ein modernes eignungsdiagnostisches Verfahren erfüllen sollte. Im Fall der Objektivität stellt das VAK sogar eine Verbesserung gegenüber einer herkömmlichen Rollenspielübung dar. Für die Reliabilität und Konstruktvalidität sind die resultierenden Ergebnisse nicht vergleichend mit Ergebnissen zu herkömmlichen Rollenspielen interpretierbar. Hier gilt es anhand weiterer Untersuchungen vergleichbare und

brauchbare Koeffizienten zu ermitteln. Die ersten Resultate bezüglich einer Konstruktvalidierung des *VAK* lassen allerdings auf weitere positive Ergebnisse hoffen. Die Bewertung der kriterienorientierten Validität des *VAK* lässt dagegen derzeit kein positives Fazit im Hinblick auf eine Weiterentwicklung und Verbesserung zu. Im direkten Vergleich mit einem herkömmlichen Rollenspiel schnitt das *VAK* hier schlechter ab. Diese Aspekte gilt es in weiteren Studien zu untersuchen, da sich für die Interpretation der Ergebnisse die in Kapitel 6 dargestellten Einschränkungen ergeben. In Bezug auf die soziale Validität stellt das *VAK* im direkten Vergleich mit einem herkömmlichen Rollenspiel derzeit nur unter dem Aspekt der Verständlichkeit und Transparenz eine Verbesserung dar. Die generell zustimmende Bewertung der Probanden ist als erfreulich zu bewerten.

Auf der Basis der ersten Ergebnisse lässt sich damit zusammenfassen, dass das *VAK* bereits jetzt einige positive Eigenschaften aufweist. In Bezug auf die kriterienorientierte und soziale Validität gilt es allerdings noch weitere Ergebnisse zu generieren, da sich für diese Bereiche entweder noch Verbesserungspotential feststellen lässt bzw. derzeit keine einheitlichen Aussagen möglich sind.

8.3 Ausblick

Die Heterogenität der resultierenden Ergebnisse und Erkenntnisse aus den einzelnen Abschnitten der vorliegenden Arbeit lassen eine abschließende Bewertung des VAK derzeit nicht zu. Auch die Frage, inwieweit sich durch den Einsatz von multimedialen Elementen eine Verbesserung von herkömmlichen Rollenspielverfahren in Face-to-Face Form erzielen lässt, kann derzeit nicht beantwortet werden. Vielmehr ergeben sich weitere Vorschläge in Bezug auf das methodische Vorgehen und mögliche Fragestellungen. Neben den spezifischen Implikationen für weitere Forschungsarbeiten, welche in den Kapiteln 6 und 7 dargestellt wurden, ergeben sich aus der generellen Betrachtung des Anwendungsgebiets der multimedialen, videobasierten Eignungsdiagnostik, aus Sicht des Autors der vorliegenden Arbeit, weitere Implikationen für die Forschung und Praxis. Hiervon sollen einige abschließend vorgestellt werden.

8.3.1 Implikationen für die Forschung

Ein erster Aspekt, der in Bezug auf multimediale, videobasierte Verfahren aus Sicht des Autors derzeit nicht zufriedenstellend gelöst ist, betrifft die Bestimmung der Reliabilität. Sowohl die berichteten Ergebnisse zu videobasierten SJTs als auch die Reliabilitätsbestimmung für das VAK mittels Cronbachs alpha stellen eine Unterschätzung der wahren Zuverlässigkeit videobasierter eignungsdiagnostischer Verfahren dar. Aus diesem Grund wären für zukünftige Untersuchungen Desings wünschenswert, welche eine Bestimmung der Retestreliabilität ermöglichen.

Ein weiterer als wichtig erachteter Aspekt betrifft die Bestimmung der sozialen Validität und wahrgenommenen Fairness von videobasierten Personalauswahl- und Personalentwicklungsverfahren. Im Rahmen der vorliegenden Arbeit wurde dargestellt, dass es derzeit kein allgemein anerkanntes, global verfügbares Instrument zur Bestimmung dieser wichtigen Charakteristika gibt. Vielmehr existieren verschiedene theoretische Modelle, aus denen wiederum die unterschiedlichsten Erhebungsinstrumente resultieren. Weiterhin ergeben sich für Instrumente, welche nach dem gleichen theoretischen Modell und für einen ähnlichen Anwendungskontext entwickelt wurden, individuelle Unterschiede in Bezug auf die resultierende Faktorstruktur. Auch für den Fall des VAK wurde die aufgrund inhaltlicher Überlegungen a priori angenommene Struktur der Items empirisch nicht bestätigt. Ein

Vergleich von unterschiedlichen Verfahren anhand von allgemein etablierten Kennwerten, wie es im Fall der Reliabilität oder der Validität möglich ist, gestaltet sich damit derzeit als schwierig. Auch die Frage, welche Aspekte der sozialen Validität und Fairness hauptsächlich für eine Verbesserung von Aspekten wie der Leistungen oder Motivation der Probanden verantwortlich sind, lässt sich derzeit nicht klar beantworten. Die Bemühungen der Forschung sollten dahingehend gebündelt werden, für die Wahrnehmung und Bewertung multimedialer Verfahren durch die Probanden ein einheitliches theoretisches Modell zu etablieren. Ein solches Modell könnte dann als Basis für die Entwicklung von Erhebungsinstrumenten dienen, deren Ergebnisse in Bezug auf die resultierende Faktorstruktur stabil sind.

Ein dritter Gesichtspunkt ergibt sich aus dem Entwicklungsprozess von multimedialen, videobasierten Verfahren. Im Rahmen der vorliegenden Arbeit wurde dargestellt, dass es sich bei der Generierung von videobasierten Item- und Responsekomponenten um einen aufwendigen Prozess handelt. Als besondere Herausforderung wurde die Aufgabe des Konstrukteurs beschrieben, die schriftlich formulierten Inhalte der Aufgabenstellung und Antwortmöglichkeiten in den Videosequenzen adäquat abzubilden. Im Idealfall sollten alle Informationen, die für die Lösungsfindung relevant sind, auf den unterschiedlichen Wahrnehmungskanälen und Symbolebenen gleichwertig abgebildet werden. Für den Fall des VAK wurde dieser Aspekt berücksichtigt und im Rahmen einer Voruntersuchung empirisch getestet (ausführlich siehe Kapitel 4.3.1). Neben einer empirischen Prüfung des Zusammenspiels der einzelnen Kanäle und Symbolebenen ergibt sich für die Forscher und Entwickler im Kontext multimedialer, eignungsdiagnostischer Verfahren ein weiteres interessantes Forschungsfeld. Die Frage, welche Kanäle der Informationsverarbeitung für die Lösung von Problemen in multidimensionalen sozialen Situationen von besonderer Bedeutung sind, erscheint genauso interessant wie die empirische Untersuchung einer optimalen Gestaltung der einzelnen Kanäle für die Darstellung solcher Situationen. Für den Fall des VAK zeigten sich keine signifikanten Unterschiede zwischen den Probandenleistungen einer Hörspiel-, Stummfilm- oder Videogruppe. Auch die Bewertung der Verfahren variierte nur in Bezug auf die Verfahrenslänge, welche für die Stummfilmvariante schlechter ausfiel. Hieraus könnte die Überlegung resultieren Rollenspielverfahren in einer weniger aufwendigen Audioversion zu produzieren. Weitere Untersuchungen in dieser Richtung könnten helfen multimediale Verfahren noch effizienter zu gestalten, da die multimedialen Instrumente noch sorgfältiger produziert und die als besonders relevant erscheinenden Kanäle noch gezielter eingesetzt werden könnten.

Ein abschließender Aspekt bezieht sich auf den speziellen Fall des multimedialen Rollenspiels. Während für den Bereich der SJT Verfahren bereits eine recht hohe Zahl an empirischen Studien zu multimedialen, videobasierten Verfahren existiert, wurde mit der vorliegenden Arbeit ein erster Versuch unternommen, die Rollenspielübung in eine vollständig computer- und videobasierte Version zu übertragen. Hierbei musste in Bezug auf Aspekte wie die Instruktion, den Bewertungsschlüssel und des zur Validierung herangezogenen Außenkriteriums Entscheidung getroffen werden, die sich aus den bisherigen Ergebnissen zu videobasierten SJT Verfahren oder aus den Rahmenbedingungen des Erhebungskontextes ergaben. Es wäre wünschenswert, wenn sich auch für den Fall der multimedialen, videobasierten Face-to-Face Verfahren ein Forschungsfeld etablieren würde, welches dem Anwender Informationen bezüglich der Auswirkung von unterschiedlichen Instruktionsformen, Bewertungsschlüsseln oder Vorschlägen zu sinnvollen Außenkriterien bieten kann. Ein erster Schritt zur Beantwortung des letztgenannten Aspekts wird mit der Validierung des VAK an weiteren Außenkriterien unternommen, welche in den nächsten ein bis zwei Jahren verfügbar sein werden. Weitere Studien mit neuen Verfahren sollten folgen.

8.3.2 Implikationen für die Praxis

In der Einleitung der vorliegenden Arbeit wurde die Notwendigkeit einer permanenten Weiterentwicklung und Verbesserung der in der Personalauswahl und Personalentwicklung eingesetzten Instrumente in Bezug auf deren Gütekriterien und soziale Aspekte dargestellt. Welche Rolle könnten aber die Anwender solcher Verfahren in der Praxis bei diesem Prozess leisten? Nach Meinung des Autors gibt es zwei zentrale Ansatzpunkte. Ein erster Ansatzpunkt liegt in der reflektierten Nutzung bereits etablierter Instrumente. Ein Verfahren ist nicht per se gut, nur weil es multimedial gestaltet ist. Die empirischen Ergebnisse aus Studien zu videobasierten Verfahren zeigen, dass die aus einem Verfahren resultierenden Ergebnisse in Abhängigkeit von einzelnen Komponenten wie der Instruktion, dem Bewertungsschlüssel oder dem verwendeten Außenkriterium zu sehen sind. Inwieweit ein Verfahren nützliche Ergebnisse liefert, hängt deshalb nicht zuletzt davon ab, ob der Anwender sich über die Wirkung dieser Komponenten im Klaren ist. Als Beispiel sollte, je nachdem ob bei einer Betrachtung eher kognitive Fähigkeiten oder soziale Kompetenzen im Vordergrund stehen, die Wahl auf eine *should do* oder *would do* Instruktion fallen. Da aus der reflektierten Nutzung in der Regel bessere Ergebnisse resultieren und diese wiederum zu

einer höheren Akzeptanz für die eingesetzten multimedialen Verfahren in den Unternehmen führen sollten, könnte das Interesse der Praxis an weiteren Verfahren dieser Art gestärkt werden. Nur durch ein solches Interesse entstehen Kooperationen mit der Forschung, welche die Entwicklung und Testung weiterer Verfahren ermöglichen und mit neuen Ergebnissen das Forschungs- und Anwendungsfeld bereichern können.

Der zweite Gesichtspunkt bezieht sich auf die Rolle der Praxis als Kooperationspartner bei der Entwicklung von multimedialen, videobasierten Verfahren. Die Entwicklung von videobasierten Verfahren ist nach wie vor mit einem hohen zeitlichen Aufwand und nicht unerheblichen Kosten verbunden. Weiterhin erscheint es zur Bestimmung der Gütekriterien und anderer zentraler Kriterien sinnvoll videobasierte Verfahren in Kooperation mit Experten aus den jeweiligen Arbeitsbereichen zu konzipieren und sie im realen Anwendungskontext zu produzieren und später zu testen. Ohne die Bereitschaft seitens der Konzerne ihre Expertise in den jeweiligen Anwendungsbereichen zur Verfügung zu stellen und neue Verfahren zu Testzwecken in bestehende Auswahlprozesse zu integrieren, fällt die Entwicklung neuer Verfahren extrem schwer. Zudem kann gerade für videobasierte Verfahren das Erstellen der Videos an realen Arbeitsplätzen von großer Bedeutung für die resultierende *Fidelity* sein. Solche real wirkenden Settings sind ohne die Mithilfe der Konzerne nur schwer zu realisieren, da eine gute Nachbildung in Studios sehr aufwendig ist.

Die Kooperationsbereitschaft der Konzerne ist aber nicht nur in der Entwicklungsphase von großer Bedeutung. Nach der Einführung eines neuen Verfahrens stellt eine sorgfältige Evaluation der resultierenden Ergebnisse einen wichtigen Faktor für die Qualität des diagnostischen Prozesses dar. Auch für das VAK hat sich gezeigt, dass derzeit nicht für alle Gütekriterien ein positives, abschließendes Fazit gezogen werden kann und weitere Untersuchungen folgen müssen. Die Zusammenarbeit mit der Forschung garantiert hier ein methodisches Vorgehen nach wissenschaftlichen Standards. Eine konsequente Zusammenarbeit von Forschung und Praxis bei der Konzeption und Weiterentwicklung von multimedialen eignungsdiagnostischen Verfahren lässt somit beide Seiten profitieren.

Die vorliegende Arbeit zeigt, wie viele andere empirische Studien, dass sich die Praxis dieser Tatsache bewusst ist. Setzt man allerdings die Vielzahl der allein in Deutschland ansässigen Unternehmen ins Verhältnis zu den in diesem Forschungsbereich publizierten Studien, ergibt sich ein großes Potential für weitere Kooperationen. Diese vielen potentiellen Kooperationsmöglichkeiten lassen auf eine positive Entwicklung hoffen.

9. Zusammenfassung

Im Rahmen der vorliegenden Arbeit wurde, ausgehend von einer Beschreibung der derzeitigen wirtschaftlichen Entwicklung, auf die hohen Anforderungen hingewiesen, denen HR-Manager und die von ihnen verwendeten Personalauswahl- und Personalentwicklungsverfahren gerecht werden müssen, um im Kampf um die besten Mitarbeiter bestehen zu können. Die Anforderungen, welche an moderne Verfahren in diesem Kontext gestellt werden, wurden dabei als vielschichtig beschrieben. So müssen diese, neben einer hohen Objektivität, Zuverlässigkeit und Vorhersagegüte, auch in Bezug auf die wahrgenommene soziale Validität und ihre Fairness höchsten Ansprüchen genügen und sollen weiterhin über eine hohe Wiedergabetreue der realen Situation am Arbeitsplatz verfügen. Als ein möglicher Ansatz, um den oben genannten Forderungen an moderne Auswahlverfahren gerecht zu werden, wurde die Übertragung von Erkenntnissen aus dem Bereich der multimedialen Trainings auf herkömmliche eignungsdiagnostische Anwendungen vorgeschlagen. Mit der vorliegenden Arbeit wurde, motiviert durch die positiven Ergebnissen zu videobasierten SJT Verfahren, ein erster Versuch unternommen, eine Rollenspielübung in multimedialer, videobasierter Form abzubilden. Anhand des Vergleichs eines zu diesem Zweck konstruierten multimedialen Verfahrens (VAK) mit einer herkömmlichen Rollenspielübung in Face-to-Face Form und einer Paper-Pencil Variante, welche inhaltlich gleich gestaltet waren, sollte die Frage beantwortet werden, inwieweit sich auch für das Rollenspiel eine Verbesserung der kriterienorientierten sowie der wahrgenommenen sozialen Validität und Fairness durch den Einsatz von Multimedia realisieren lässt.

Die empirische Untersuchung erfolgte anhand einer Stichprobe von $N = 52$ Auszubildenden im Finanzdienstleistungssektor, die im Rahmen eines realen Personalentwicklungsverfahrens alle drei Übungen durchliefen und diese am Ende bewerteten. Als Außenkriterium zur Bestimmung der kriterienorientierten Validität diente die mündliche IHK-Note. Die Bewertung der wahrgenommenen sozialen Validität und Fairness erfolgte mittels eines zu diesem Zweck entworfenen 20 Items umfassenden Fragebogens.

Die resultierenden Ergebnisse aus der empirischen Studie zeigten, dass nicht automatisch von einer Verbesserung der Rollenspielübung durch den Einsatz von Multimedia ausgegangen werden kann. Konkret ergab sich in Bezug auf die kriterienorientierte Validität keine Verbesserung gegenüber der herkömmlichen Rollenspielvariante oder der Paper-Pencil Variante. Für die soziale Validität zeigte sich eine Bevorzugung der herkömmlichen

Rollenspielübung auf dem Faktor *Generelle soziale Validität und Fairness* und eine Bevorzugung des VAK auf dem Faktor *Transparenz und Information*. Eine abschließende Beantwortung der Frage, inwieweit der Einsatz von Multimedia auch für ein Face-to-Face Verfahren wie das Rollenspiel zu einer Verbesserung der Gütekriterien und anderer wichtiger Charakteristika führen kann, lassen diese ersten Ergebnisse allerdings nicht zu. Weitere Untersuchungen des VAK und anderer Verfahren mit gleicher Intention sind nötig, um hier verlässliche Aussagen treffen zu können.

Literaturverzeichnis

- Andree, S. (2002). Beschreibung des Konzentrations-Leistungs-Test (KLT). In E. Brähler, H. Holling, D. Leutner & F. Petermann (Hrsg.) *Brickenkamp Handbuch psychologischer und pädagogischer Tests* (3., vollständig überarb. und erw. Aufl.). (S. 286-288). Göttingen: Hogrefe.
- Amelang, M. & Schidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention* (4. Aufl.). Berlin: Springer.
- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention* (3. Aufl.). Berlin: Springer.
- American Psychological Association (APA). (1999). *Guidelines for computer-based tests and interpretations*. Washington, DC: APA.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A., (1999). *Intelligenz-Struktur-Test 2000*. Göttingen: Hogrefe.
- Arthur, Jr., W., Day, E.A., McNelly, T.L. & Edens, P.S. (2003). A meta-analysis of the criterion-related validity of assessmentcenter dimensions. *Personnel Psychology*, *56* (1), 125-153.
- Arvey, R.D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, *43*, 695–716.
- Baddeley, A. (1992). Working Memory. *Science*, *255*, 556-559.
- Ballstaedt, S. (1990). Integrative Verarbeitung bei audiovisuellen Medien. In K. Böhme-Dürr, J. Emig & M. Seel (Hrsg.). *Wissensveränderung durch Medien: Theoretische Grundlagen und empirische Analysen*. (S. 185-196). München: Saur.
- Bartram, D. & Hambleton, R.K. (Hrsg.)(2006). *Computer-Based Testing and the Internet: Issues and Advances*. Chichester: Wiley & Sons.
- Bauer, T.N., Truxillo, D.M., Paronto, M.E., Weekley, J.A., & Champion, M.A. (2004). Applicant reactions to different selection technology: Face-to-face, interactive voice response, and computer-assisted telephone screening interviews. *International Journal of Selection and Assessment*, *12*, 135–148.
- Beller, H. (2005). Aspekte der Filmmontage. Eine Art Einführung. In H. Beller (Hrsg.), *Handbuch der Filmmontage*. (S. 9-32). München: TR-Verlagsunion.
- Bergman, M.E., Drasgow, F., Donovan, M.A., Henning, J.B., Juraska, S.E. (2006). Scoring Situational Judgement Tests: Once You get the Data, your troubles begin. *International Journal of selection and assessment*, *14*(3), 223-235.

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinees ability. In F.M. Lord & M. R. Novick (Hrsg.), *Statistical theories of mental test scores* (S. 395-479). Reading, Mass.: Addison-Wesley.
- Booth, J. (1999). Computerdiagnostik. In R.S. Jäger & F. Petermann (Hrsg.) *Psychologische Diagnostik: Ein Lehrbuch (4. Aufl.)* (S. 186-197). Weinheim: Beltz.
- Bordwell, D. & Thomson, K. (2004). *Film Art. An Introduction*. Boston: Mcgraw-Hill Publ. Comp.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler (6. Auflage)*. Heidelberg: Springer.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion (2., aktualisierte Auflage)*. München: Pearson.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15*, 233–254.
- Chan, D., & Schmitt, N. (2004). An agenda for future research on applicants reactions to selection procedures: A construct oriented approach. *International Journal of Selection and Assessment, 12*, 9–23.
- Chan, D., Schmitt, N., DeShon, R.P., Clause, C.S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology, 82*, 300–310.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. New York: Erlbaum.
- Dalessio, A.T. (1994). Predicting insurance agent turnover using a video based situational judgement test. *Journal of Business and Psychology, 9*, 23-32.
- Dawes, R.M., Faust, D. & Mehl, P.E. (1989). Clinical versus actuarial judgement. *Science, 243*, 1668-1674.
- Dörner, D. Kreuzig, H.W., Reither, F. & Stäudel, T. (1983). *Lohausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Drasgow, F., Olson-Buchanan, J.B., & Moberg, P.J. (1999). Development of an interactive video assessment: Trials and tribulations. In F. Drasgow (Ed.), *Innovations in computerized assessment* (pp. 177–196). Mahwah: NJ: Erlbaum.
- Düker, H. & Lienert, G. (2001). *KLT-R- Konzentrations-Leistungs-Test*. Göttingen: Hogrefe.
- Etzel, S. (1999). *Multimediale, computergestützte diagnostische Verfahren: Neue Perspektiven für die Managementdiagnostik*. Aachen: Shaker.

- Fennekels, G.P. (1987). *Validität des Assessmentcenters bei Führungskräfteauswahl und –entwicklung*. Philosophische Dissertation, Bonn.
- Fisseni, H.-J. (2004) *Lehrbuch der psychologischen Diagnostik (3. Aufl.)*. Göttingen, Hogrefe.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327–358.
- Franke, G. (1997). Über die Möglichkeiten der computergestützten Darbietung beim revidierten Freiburger Persönlichkeitsinventar – Zwei experimentelle Studien. *Zeitschrift für Experimentelle Psychologie*, XLIV, 332-356.
- Fricke, R. (1990). Videotests: „True-to-life“-Testsituationen durch interaktives Video. In W. Sarges (Hrsg.), *Management-Diagnostik* (S. 463-466). Göttingen. Hogrefe.
- Funke, J. (1991). Solving complex problems: exploration and control of complex systems. In R. J. Sternberg & P.A. Frensch (Hrsg.), *Complex problem solving: principles and mechanisms* (S. 185-222). Hillsdale: LEA.
- Funke, J. (1998). Computer-based testing and training with scenarios from complex problemsolving research: advantages and disadvantages. *International journal of Selection and Assessment*, 6, 90 -97.
- Funke, U. (1995). Using Complex Problem Solving Tasks in Personnel Selection and Training. In P.A. Frensch & J. Funke (Hrsg.), *Complex problem solving. The european perspective* (S. 219-240). Hillsdale: LEA.
- Funke, U. & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection and Assessment*, 6(2), 115-123.
- Gagalick, T. (1988). *Kontinuität und Diskontinuität im Film. Die frühen Filme Jean Luc Godards*. Münster: MAKs Publikationen.
- Gilliland, S.W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *The Academy of Management Review*, 18, 694–734.
- Gilliland, S.W. & Hale, J. (2005). How can justice be used to improve employee selection practice? In J. Greenberg & A. Jason (Eds.), *Handbook of organizational justice* (pp. 411– 438). Mahwah, NJ: Erlbaum.
- Gregory, R. J. (1996). *Psychological Testing: History, Principles and Applications. Second Edition*. Boston: Allyn and Bacon.
- Grewe, K. (2005). *Multimediale interaktive Items in der Personalauswahl aus Sicht der Bewerber*. Unveröffentlichte Diplomarbeit, Westfälische Wilhelms-Universität Münster.

- Haack, J. (2002). Interaktivität als Kennzeichen von Multimedia und Hypermedia. In L.J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia und Internet - 3. vollst. überarb. Auflage* (S. 127-136). Weinheim: Beltz/PVU.
- Hageböck, J. (1994). *Computergestützte Diagnostik in der Psychologie. Die Entwicklung eines computergestützten Diagnosesystems für die Einzelfallhilfe in der Schulpsychologie*. Göttingen, Hogrefe
- Hammerschmidt, J. (2003). *Einfluss der Multimedialisierung auf Reliabilität und soziale Validität eines Personalauswahlverfahrens*. Unveröffentlichte Diplomarbeit. Westfälische Wilhelms-Universität Münster.
- Havighurst, L.C., Fields, L.E. & Fields, C.L. (2003). *High Versus low fidelity simulations: Does the type of Format Affect candidates Performance or Percpetion?* Paper presented at the 27th Annual IPMAAC Conference on Personnel Assessment, Baltimore.
- Höft, S. & Funke, U. (2006). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie (2. Aufl.)*. (S. 145-188). Göttingen: Hogrefe.
- Hossiep, R. & Paschen, M. (1998). *BIP Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung*. Göttingen: Hogrefe.
- Howard, A., & Choi, M. (2000). How do you assess a manager's decision-making abilities? The use of situational inventories. *International Journal of Selection and Assessment*, 8, 85-88.
- Jäger, R.S. & Petermann (1999). *Psychologische Diagnostik (4. Aufl.)*. Weinheim: PVU.
- Jones, C. & DeCotiis, T.A. (1986). Video-based selection of hospitality employees. *The Cornell Hotel and Restaurant Administration Quarterly*, 67-73.
- Juraska, S.E. & Drasgow, F. (2001). *Faking Situational Judgement: A test of the conflict resolution skills assessment*. Paper presented at the 16th annual meeting of the society for industrial and Organizational Psychology, San Diego, CA.
- Kanning, U. P. (2003). Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP) von R. Hossiep und M. Paschen (1998). In E. Brähler, H. Holling, D. Leutner & F. Petermann (Hrsg.) *Brickenkamp Handbuch psychologischer und pädagogischer Tests* (3., vollständig überarb. und erw. Aufl.) (S. 527-530). Göttingen: Hogrefe.
- Kanning, U.P. (2004). *Standards der Personaldiagnostik*. Göttingen: Hogrefe.
- Kanning, U.P. (2005). Computergestützte Personalauswahl von Polizeibeamten. *Zeitschrift für Arbeits- und Organisationspsychologie*, 49, 140-146.

- Kanning, U.P., Grewe, K., Hollenberg, S. & Hadouch, M. (2006). From the subjects' point of view: Reactions to different types of situational judgement items. *European Journal of Psychological Assessment*, 22, 168–176.
- Kersting, M. (1999). *Diagnostik und Personalauswahl mit computergestützten Problemlösenszenarien?* Göttingen: Hogrefe.
- Kersting, M. (2001). Zur Konstrukt- und Kriteriumsvalidität von Problemlösenszenarien anhand der Vorhersage von Vorgesetztenurteilen über die berufliche Bewährung. *Diagnostica*, 47, 67-76.
- Kirbach, C., Montel, C., Oenning, S. Wottawa, H. (2004) *Recruiting und Assessment im Internet. Werkzeuge für eine optimierte Personalauswahl und Potenzialerkennung.* Göttingen: Vandenhoeck + Ruprecht
- Kisser, R. (1999). Adaptive Strategien. In R.S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik 4. Aufl.* (S. 161-169). Weinheim: Beltz.
- Kleinmann, M. (1997). *Assessment Center: Stand der Forschung – Konsequenzen für die Praxis.* Göttingen: Hogrefe.
- Kleinmann, M. (2003). *Assessment Center.* Göttingen: Hogrefe.
- Kleinmann, M. & Strauß, B. (1998). Validity and Application of Computersimulated Scenarios in Personnel Assessment. *International Journal of Selection and Assessment*, 6 (2), 97-106.
- Klinck, D. (2002). *Computergestützte Diagnostik. Beeinflusst das Medium der Testbearbeitung die Testcharakteristika, die Testfairness oder das Erleben der Testsituation?* Göttingen: Hogrefe.
- Klingler, D.E., Miller, D., Johnson, J. & Williams, T. (1977). Process evaluation of an on-line computer-assisted unit for intake assessment of mental health patients. *Behaviour Research Methods and Instrumentation*, 9, 110-116.
- Konradt, U. & Sarges, W. (Hrsg.) (2003). *E-Recruiting und E-Assessment.* Göttingen: Hogrefe
- Kravitz, D.A., Stinson, V., & Chavez, T.L. (1996). Evaluation of tests used for making selection and promotion decisions. *International Journal of Selection and Assessment*, 4, 24–35.
- Krone, H.W. & Hock, M. (2007). *Psychologische Diagnostik, Grundlagen und Anwendungsfelder.* Stuttgart: Kohlhammer.
- Kubinger, K. D. (2006). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens.* Göttingen: Hogrefe.

- Laguna, K. & Babcock, R. (1997). Computer anxiety in young and older adults: Implications for human-computer interactions in older populations. *Computers in Human Behaviour, 13*, 317-326.
- Horst, P. (1971). *Messung und Vorhersage*. Weinheim: Beltz.
- Lienert, G.A. & Raatz, U. (1998). *Testaufbau und Testanalyse (6. Aufl.)*. Weinheim: PVU.
- Lievens, F., Buyse, T. & Sackett, P.R. (2005). The operational validity of a video-based situational judgement test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90* (3), 442-452.
- Lievens, R. & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment, 10*, 245-257.
- Lievens, F., DeCorte, W., & Brysse, K. (2003). Applicants perceptions of selection procedures: The role of selection information, belief in tests, and comparative anxiety. *International Journal of Selection and Assessment, 11*, 67-77.
- Lievens, F. & Sackett, P.R. (2006). Video-based versus written situational judgement tests: A comparison in terms of predictive validity. *Journal of applied psychology, 91*(5), 1181-1188.
- Lukesch, H. & Mayrhofer, S. (2001). *KLT-R- Konzentrations-Leistungs-Test – Revidierte Fassung*. Göttingen: Hogrefe
- Macan, T.H., Avedon, M.J., Paese, M., Smith, D.E. (1994). The effects of applicants reactions to cognitive ability tests and an assessment center. *Personnel Psychology, 47*, 715-738.
- Mead, A.D. & Drasgow, F. (1993). Equivalenz of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*, 449-458.
- McDaniel, M.A., & Nguyen, N.T. (2001). Situational judgment tests: A review of practice and constructs assessment. *International Journal of Selection and Assessment, 9*, 103-113.
- McDaniel, M.A., Hartman, N.S. & Grubb III, W.L. (2003). *Situational Judgement Tests, Knowledge, Behavioral Tendency and Validity: A meta-Analysis*. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando.
- McDaniel, M.A., Morgeson, F.P., Finnemagan, E.B., Campion, M.A., & Braverman, E.P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730-740.

- McHenry, J.J. & Schmitt, N. (1994). Multimedia testing. In M.G. Rumsey & C.B. Walker (Hrsg.), *Personnel selection and classification* (S. 193-232). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Meehl, P.E. (1954). *Clinical versus statistical prediction. A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Meehl, P.E. (1956). Wanted – a good cookbook. *American Psychologists*, *11*, 263-272.
- Miles, E. W. & King Jr., W.C. (1998). Gender and administration mode effects when pencil-and-paper personality tests are computerized. *Educational and Psychological Measurement*, *58* (1), 68-76.
- Millstein, S.G. (1987). Acceptability and reliability of sensitive information collected via computer interview. *Educational and Psychological Measurement*, *47*, 523 – 533.
- Monaco, J. (1995). *Film verstehen. Kunst, Technik, Sprache, Geschichte und Theorie des Films und er Medien. Mit einer Einführung in Multimedia*. Reinbeck bei Hamburg: Rowohlt.
- Motowidlo, S.J., Dunnette, M.D., & Carter, G.W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, *75*, 640–647.
- Olson-Buchanan, J.B. (2001). Computer-based advances in assessment. In F. Drasgow & N. Schmitt (Hrsg.), *Measuring and analyzing Behaviour in organizations* (S. 44-87). San Francisco, CA: Jossey-Bass.
- Olson-Buchanan, J.B. & Drasgow, F. (2006). Multimedia Situational Judgement Tests: The medium Creates the message. In J.A. Weekley & R. E. Ployhart (Hrsg.), *Situational Judgement Tests: Theory, Measurement and Application* (S. 253-278). Mahwah: LEA.
- Olson-Buchanan, J.B., Drasgow, F., Moberg, P.J., Mead, A.D., Keenan, P.A., & Donovan, M.A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology*, *51*, 1–24.
- Oswald, F.L., Schmitt, N., Kim, B.H., Ramsay, L.J., & Gillespie, M.A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, *89*, 187–207.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- Ployhart, R.E., & Harold, C.M. (2004). The applicant attribution reaction theory (AART): An integrative theory of applicant attributional processing. *International Journal of Selection and Assessment*, *12*, 84–98.
- Ployhart, R.E., & Erhart, M.G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment*, *11*, 1–16.

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut.
- Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, 49(4), 243-256.
- Richman-Hirsch, W.L., Olson-Buchanan, J.B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, 85, 880-887.
- Rost, J. (2004). *Lehrbuch Testtheorie, Testkonstruktion (2. vollst. überarb. und erw. Aufl.)*. Bern: Huber
- Ryan, A.M., & Ployhart, R.E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, 26, 565-606.
- Rynes, S.L. (1993). When recruitment fails to attract: Individual expectations meet organizational realities in recruitment. In H. Schuler, J.L. Farr & M. Smith, (Hrsg.), *Personnel Selection and Assessment: Individual and Organizational Perspectives*. (S. 27-40). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rynes, S.L. Bretz, R. D. & Gerhart, B. (1991). The importance of recruitment in job choice: A different way of looking. *Personnel Psychology*, 44 (3), 487-521.
- Schaumburg, H. & Issing, L.J. (2004). Interaktives Lernen mit Multimedia. In R. Mangold, P. Vorderer & G. Bente (Hrsg.), *Lehrbuch der Medienpsychologie (S. 717-742)*. Göttingen: Hogrefe.
- Scheurer, H. & Jäger, R.S. (1999). Experimentelle Psychologie. In R.S. Jäger & F. Petermann (HRSG.), *Psychologische Diagnostik (4. Aufl.)* (S. 22-27). Weinheim: PVU.
- Schiff, W. Arnone, W. & Cross, S. (1994). Driving assessment with computer-video scenarios: More is sometimes better. *Behaviour Research Methods, Instruments & Computers*, 26(2), 192-194.
- Schmidt, F.L. & Hunter, J.E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, 124 (2), 262-274.
- Schmitt, M.J., & Ryan, A.M. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology*, 77, 629-637.
- Schuler, H. (1993). Sozial validity of selection situations: A concept and some empirical results. In H. Schuler, J.L. Farr & M. Smith (Hrsg.), *Personnel Selection and Assessment: Individual and Organizational Perspectives*. (S. 11-26). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Schuler, H., Diemand, A. & Moser, K. (1993). Filmszenen. Entwicklung und Konstruktvalidierung eines neuen eignungsdiagnostischen Verfahrens. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 3-9.
- Schuler, H. & Höft, S. (2006). Konstruktorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie (2. Aufl.)* (S. 101-144). Göttingen: Hogrefe.
- Schuler, H. & Marcus, B. (2006). Biografieorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie (2. Aufl.)* (S. 189-230). Göttingen: Hogrefe.
- Schuler, H. & Stehle, W. (1983). Neuere Entwicklung des Assessment-Center- Ansatzes beurteilt unter dem Aspekt der sozialen Validität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 27, 33-44.
- Schuler, H. & Stehle, W. (1985). Soziale Validität eignungsdiagnostischer Verfahren: Anforderungen für die Zukunft. In H. Schuler & W. Stehle (Hrsg.), *Organisationspsychologie und Unternehmenspraxis: Perspektiven der Kooperation* (S. 133-138). Stuttgart: Verlag für Angewandte Psychologie.
- Schwenkmezger, P. & Hank, P. (1993). Papier-Bleistift- versus computerunterstützte Darbietung von State-Trait-Fragebögen: eine Äquivalenzprüfung. *Diagnostica*, 39, 189-210.
- Shotland, A. Alliger, G.M. & Sales, T. (1998). Face validity in the context of personnel selection: a multimedia approach. *International Journal of Selection and Assessment*. 6(2), 124-131
- Singer, J.L. & Singer, D.G. (1983). Psychologists look at television: Cognitive, developmental, personality and social implications. *American Psychologist*, 38, 826-834.
- Smiderle, D., Perry, B.A., & Cronshaw, S.F. (1994). Evaluation of video-based assessment in transit operator selection. *Journal of Business and Psychology*, 9, 3-22.
- Steiner, D.D., & Gilliland, S.W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, 81, 134-141.
- Strauß, B. Hasselmann, D. & Hasselmann, G. (1992). Validitätsaspekte computergestützter Szenarien in der Managementdiagnostik. In A. Gebert & U. Winterfeld, (Hrsg.), *Arbeits-, Betriebs- und Organisationspsychologie vor Ort. Bericht über die 34. Fachtagung der Sektion Arbeits-, Betriebs- und Organisationspsychologie im Berufsverband Deutscher Psychologen e.V.* Bad Lauterberg: DPV.
- Strittmatter, P. (1994). Wissenserwerb mit Bildern bei Film und Fernsehen. In B. Weidenmann (Hrsg.). *Wissenserwerb mit Bildern: Instruktionale Bilder in Printmedien, Film/Video und Computerprogrammen* (S. 177-194). Bern: Hans Huber.

- Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics (Fifth Edition)*. Boston: Pearson and Allyn and Bacon.
- Thornton III, G.C. (1993) The effect of selection practices on applicants' perceptions of organisational characteristics. In H. Schuler, J.L. Farr & M. Smith (Hrsg.), *Personnel Selection and Assessment: Individual and Organizational Perspectives*. (S. 57-69). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Thornton, G. C. III, Gaugler, B. B., Rosenthal, D. & Bentson, C. (1992). Die prädiktive Validität des Assessment Centers – eine Metaanalyse. In H. Schuler & W. Stehle (Hrsg.), *Assessment Center als Methode der Personalentwicklung* (2. Aufl., S. 36-77). Göttingen: Hogrefe.
- Thürauf, D. (2004). *Die Bedeutung von Konzentrationsfähigkeit und Ängstlichkeit für die Beantwortung multimedialer Items zur Messung sozialer Kompetenzen*. Unveröffentlichte Diplomarbeit. Westfälische Wilhelms-Universität Münster.
- Truxillo, D.M., Steiner, D.D., & Gilliland, S.W. (2004). The importance of organizational justice in personnel selection: Defining when selection fairness really matters. *International Journal of Selection and Assessment*, 12, 39–53.
- Volz-Sidiropoulou, E. (2004). Computerbasierte Psychodiagnostik. In Fisseni, H.J. (Hrsg.), *Lehrbuch der Psychologischen Diagnostik. Mit Hinweisen zur Intervention* (3. Aufl.). (S. 278-308). Göttingen: Hogrefe.
- Wagner-Menghin, M.M. (2003). Computerdiagnostik. In K.D. Kubinger & R.S. Jäger (Hrsg.) *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 68-81). Weinheim: Beltz.
- Walsh, W. B., Betz, N.E. (2001). *Tests and Assessment*. (4. Aufl.). New Jersey: Prentice Hall.
- Weekley, J.A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology*, 50, 25–49.
- Weekley, J.A., & Ployhart, R.E. (2005). Situational judgment: Antecedents and relationships with performance. *Human Performance*, 18, 81–104.
- Weidenmann, B. (1996). Instruktionsmedien. In F. E. Weiner (Hrsg.), *Psychologie des Lernens und der Instruktion* (Enzyklopädie der Psychologie, Band D1/2. (S. 319-368). Göttingen: Hogrefe.
- Weidenmann, B. (2002). Multicodierung und Multimodalität im Lernprozess. In L.J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia und Internet* (3. vollst. überarb. Auflage, S. 45-62). Weinheim: Beltz/PVU.
- Wiechmann, D., & Ryan, A.M. (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection and Assessment*, 11, 215–229.

- Wild, B. (1989). Neue Erkenntnisse zur Effizienz des „Tailord“-adaptiven Testens. In K.D. Kubinger (Hrsg.). *Moderne Testtheorie* (S. 179-186). Weinheim. Beltz.
- Wirtz, M. & Caspar, F. (2002) *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe
- Wittmann, A. J. (2002). *IST 2000*. In E. Brähler, H. Holling, D. Leutner & F. Petermann (Hrsg.) *Brickenkamp Handbuch psychologischer und pädagogischer Tests* (3., vollständig überarb. und erw. Aufl.) (S. 164-167). Göttingen: Hogrefe.

Rollenspiel-Übung **Kundengespräch**

Aufgabenstellung für den Berater

Versetzen Sie sich bitte in folgende Situation:

Sie sind Kundenberater der Sparkassenorganisation und arbeiten in einer Geschäftsstelle Ihres Institutes, wo Sie ein Beratungsbüro zur Verfügung haben. In wenigen Minuten wird eine Ihnen unbekannte Person die Geschäftsstelle betreten und sich mit der Bitte um eine Beratung an Sie wenden. Bei der Person handelt es sich um einen potentiellen Neukunden (bisher kein Sparkassenkunde), der Sie am Serviceschalter in der Kundenhalle anspricht.

Kurzinformation zum Kunden

Er hat ein Studium in einer in Süddeutschland liegenden Stadt absolviert und beginnt in zwei Wochen sein erstes Arbeitsverhältnis bei einem Unternehmen in Ihrem Geschäftsgebiet.

Ihre Aufgabe:

**Führen Sie ein erstes Kontaktgespräch mit dem Kunden!
Sie haben für das Gespräch 15 Minuten Zeit!**

Ihre Vorbereitungszeit beträgt 15 Minuten!

Sehr geehrter Teilnehmer / sehr geehrte Teilnehmerin!

Herzlich willkommen zum Videobasierten Analyseverfahren für Kundenberater!

Das VAK besteht aus 25 kurzen Videoclips, die alle einer typischen Beratungssituation aus dem Sparkassenalltag entnommen sind.

**Es werden Ihnen immer 5 dieser Videoclips präsentiert, die zu einer thematischen Sequenz zusammengefasst sind.
Es gibt insgesamt 5 solcher Sequenzen.**

Die einzelnen Sequenzen lauten:

Sequenz 1: Begrüßung des Kunden

Sequenz 2: Gegenseitiges Kennenlernen

Sequenz 3: Bedarfsermittlung

Sequenz 4: Beratung und Verkauf

Sequenz 5: Verabschiedung des Kunden

Weiter

Wir möchten Sie nun bitten aus diesen 5 Sequenzen, die Ihnen nacheinander präsentiert werden, immer denjenigen Videoclip auszuwählen, der Ihrem persönlichen Verhalten als Kundenberater in einer solchen Situation am besten entspricht.

Stellen Sie sich vor Ihrer Auswahl bitte immer die Frage:

„Wie würde ich mich als Kundenberater in einer solchen Situation verhalten?“

Sie können immer nur eine Sequenz auswählen!

Sie dürfen sich die Sequenzen so oft anschauen, wie sie zu Ihrer Entscheidungsfindung brauchen. Es gibt kein zeitliches Limit!

Das Abspielen der Filme erfolgt über das Anklicken des jeweiligen Filmfensters.

Die Auswahl erfolgt über das Anklicken der runden Box links neben dem jeweiligen Filmfenster.

Weiter

Schriftliche Übung **Kundengespräch**

Aufgabenstellung für den Berater

Versetzen Sie sich bitte in folgende Situation:

Sie sind Kundenberater der Sparkassenorganisation und arbeiten in einer Geschäftsstelle Ihres Institutes, wo Sie ein Beratungsbüro zur Verfügung haben. In wenigen Minuten wird eine Ihnen unbekannte Person die Geschäftsstelle betreten und sich mit der Bitte um eine Beratung an Sie wenden. Bei der Person handelt es sich um einen potentiellen Neukunden (bisher kein Sparkassenkunde), der Sie am Serviceschalter in der Kundenhalle anspricht.

Kurzinformation zum Kunden

Er hat ein Studium in einer in Süddeutschland liegenden Stadt absolviert und beginnt in zwei Wochen sein erstes Arbeitsverhältnis bei einem Unternehmen in Ihrem Geschäftsgebiet..

Ihre Aufgabe:

Erstellen Sie bitte in schriftlicher Form den Ablauf für ein Ihrer Meinung nach optimales und vollständiges Erstgespräch von der Begrüßung bis zur Verabschiedung des Kunden, wie Sie persönlich es in der oben beschriebenen Situation führen würden! Folgen Sie in Ihren Ausführungen bitte dem Ihrer Meinung nach typischen Verlauf eines solchen Gespräches. Gehen Sie hierbei bitte ausführlich und detailliert auf folgende Aspekte ein:

- Ihre Handlungen und ihr Verhalten gegenüber dem Kunden (z.B. die Hand geben).
- Fragen die Sie dem Kunden stellen würden.
- Informationen bzw. Hinweise auf Produkte und Serviceangebote, die sie dem Kunden geben würden.

Es handelt sich hierbei um eine rein schriftliche Aufgabe! Sie werden im Anschluss an die Konzepterstellung **kein** Beratungsgespräch führen. **Nur Ihr schriftliches Konzept wird bewertet!**

Achten Sie bitte darauf, dass Sie schreiben, wie Sie persönlich sich in diesem Gespräch verhalten würden. **Schreiben Sie in der Ich-Form!**

Bearbeitungszeit: 30 Minuten! Nutzen Sie diese Zeit bitte voll aus!



WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

PSYCHOLOGISCHES INSTITUT I
Psychologische Diagnostik und Medienpsychologie

Fragebogen

Sehr geehrte(r) Interviewteilnehmer(in)!

Als Ergänzung zu dem 45minütigen Interview zum Thema: „Welche Verhaltensweisen machen ein Beratungsgespräch im Finanzdienstleistungssektor erfolgreich?“, möchte ich Sie nun bitten, den folgenden Fragebogen zu bearbeiten.

Hierbei interessiert mich, wie bereits im Interview, Ihre Perspektive als Kundenberater.

Die Antworten auf die Fragen sollten dabei Ihre ganz persönliche Meinung widerspiegeln. Es gibt weder richtige noch falsche Antworten.

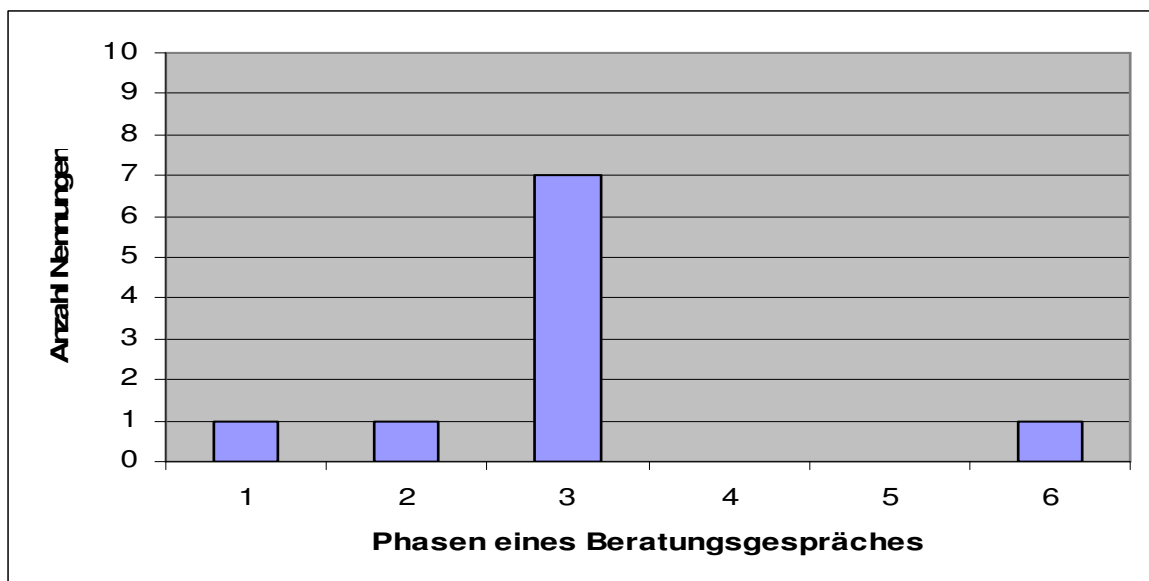
Vielen Dank für Ihre Teilnahme!

1) Welche der folgenden fünf Phasen eines Kundengesprächs halten Sie für die wichtigste?

Bitte kennzeichnen Sie die Ihrer Meinung nach wichtigste Phase, indem Sie eine der vorgegebenen Antwortmöglichkeiten ankreuzen.

1. Begrüßung am Schalter und Überleitung in das Beratungszimmer
2. Gegenseitiges Kennen lernen
3. Bedarfserhebung
4. Beratung, Produktverkauf & Crossselling
5. Terminvereinbarung & Verabschiedung
6. Alle Phasen in einem Kundengespräch sind gleich wichtig

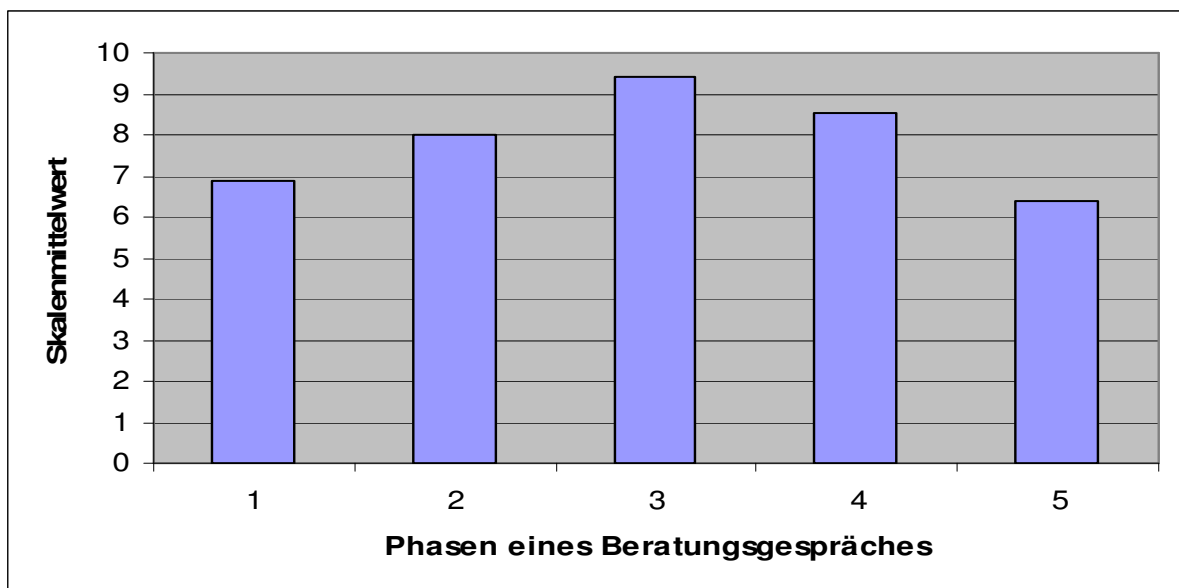
Ergebnis aus der Befragung von N=10 Experten:



2) Wie würden Sie die Wichtigkeit der einzelnen Phasen eines Kundengesprächs auf einer Skala von 1-10 bewerten. Bitte geben Sie zu jeder der aufgeführten Phasen eine Bewertung ab, indem Sie die entsprechende Zahl auf der jeweiligen Linie notieren.

1. Begrüßung am Schalter und Überleitung in das Beratungszimmer _____
2. Gegenseitiges Kennen lernen _____
3. Bedarfserhebung _____
4. Beratung, Produktverkauf & Crossselling _____
5. Terminvereinbarung & Verabschiedung _____

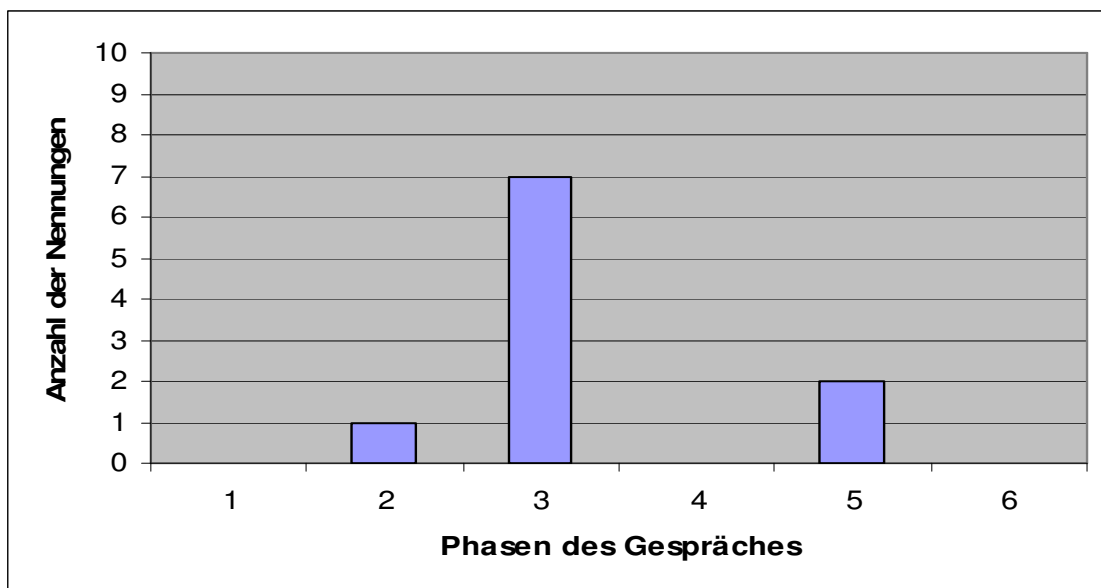
Ergebnis aus der Befragung von N=10 Experten:



3) Welche der folgenden Phasen eines Kundengesprächs fordert Sie als Kundenberater am meisten? Bitte kennzeichnen Sie die Ihrer Meinung nach anspruchsvollste Phase indem Sie eine der vorgegebenen Antwortmöglichkeiten ankreuzen!

1. Begrüßung am Schalter und Überleitung in das Beratungszimmer _____
2. Gegenseitiges Kennen lernen _____
3. Bedarfserhebung _____
4. Beratung, Produktverkauf & Crossselling _____
5. Terminvereinbarung & Verabschiedung _____
6. Alle Phasen in einem Kundengespräch sind gleich anspruchsvoll _____

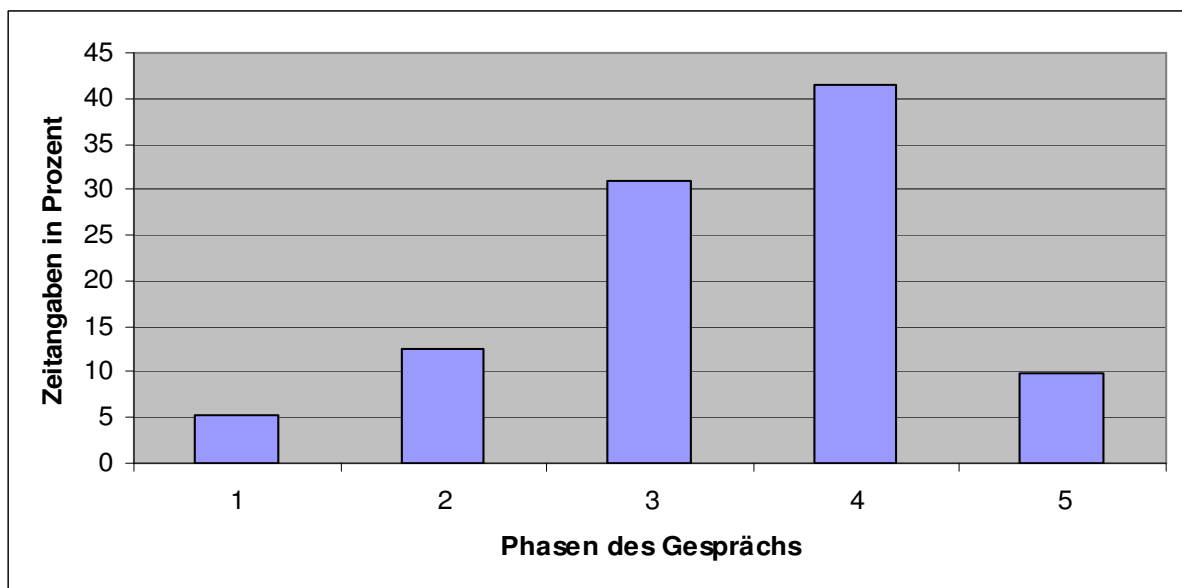
Ergebnis aus der Befragung von N=10 Experten:



4) Wie hoch ist Ihrer Meinung nach der prozentuale Zeitanteil, der im Folgenden genannten Phasen, gemessen am gesamten Zeitaufwand eines Kundengesprächs? Bitte tragen Sie Ihre Einschätzung als Prozentzahl hinter der jeweiligen Phase ein!

1. Begrüßung am Schalter und Überleitung in das Beratungszimmer _____
2. Gegenseitiges Kennen lernen _____
3. Bedarfserhebung _____
4. Beratung, Produktverkauf & Crossselling _____
5. Terminvereinbarung & Verabschiedung _____

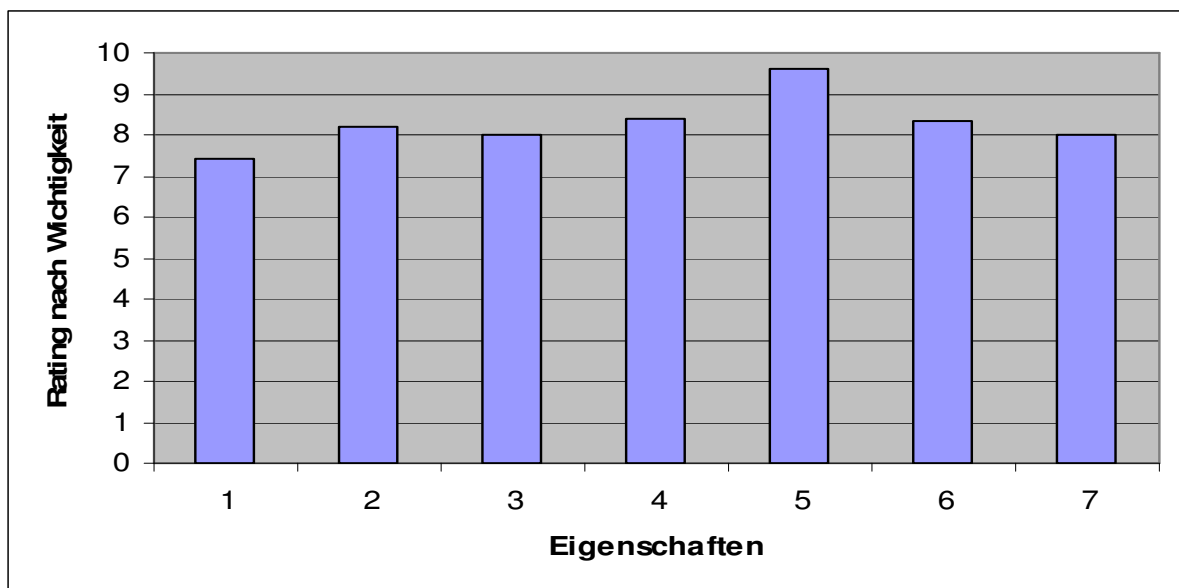
Ergebnis aus der Befragung von N=10 Experten:



5) Wie wichtig auf einer Skala von 1-10 sind die im Folgenden genannten einzelnen Fähigkeiten, um ein gutes Kundengespräch führen zu können? Bitte geben Sie zu jeder der aufgeführten Fähigkeiten eine Bewertung ab, indem Sie die entsprechende Zahl auf der jeweiligen Linie notieren.

1. Unternehmerisches Denken _____
2. Durchsetzungsvermögen / Verhandlungsgeschick _____
3. Sprachlicher Ausdruck / Rhetorik _____
4. Problemlösevermögen / Planung _____
5. Kontaktfähigkeit / soziale Kompetenz _____
6. Verkäuferische Kompetenz _____
7. Eigeninitiative _____

Ergebnis aus der Befragung von N=10 Experten:



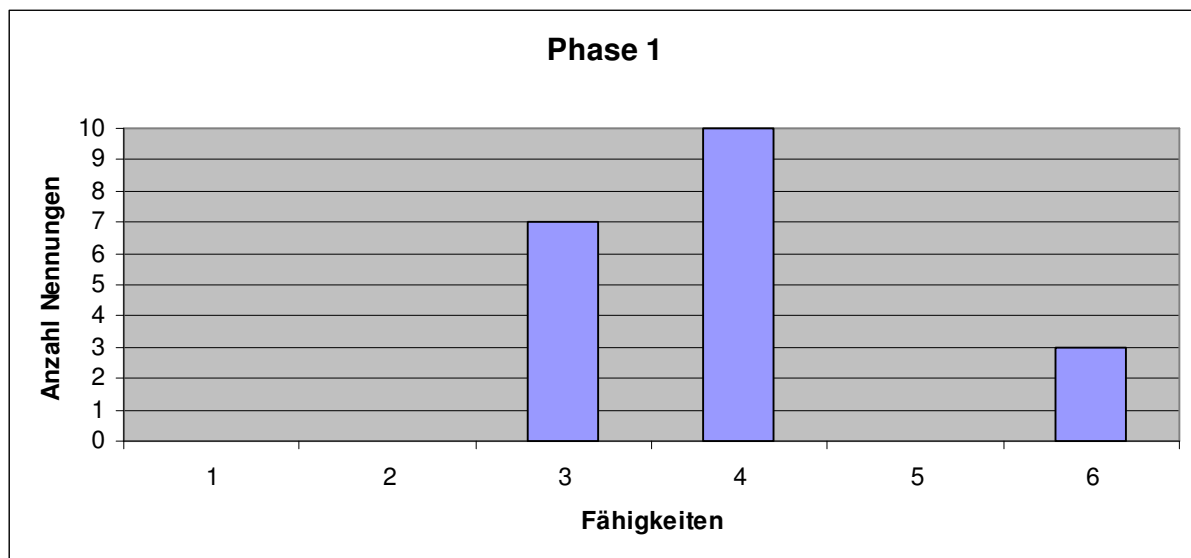
6) In welcher der folgenden fünf Phasen eines Kundengesprächs lassen sich die unten genannten Fähigkeiten Ihrer Meinung nach am besten beobachten?

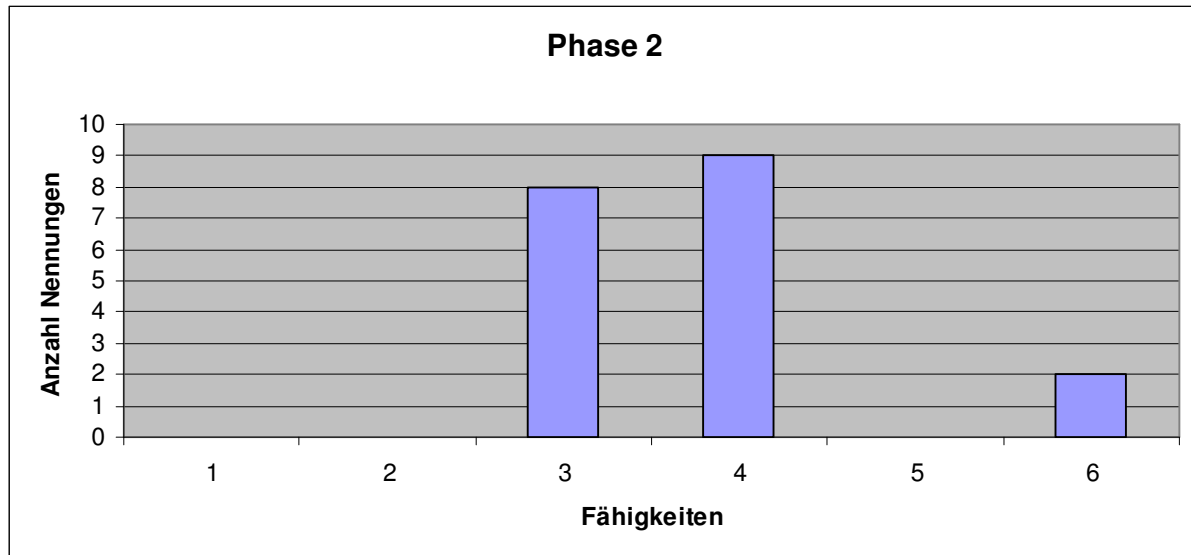
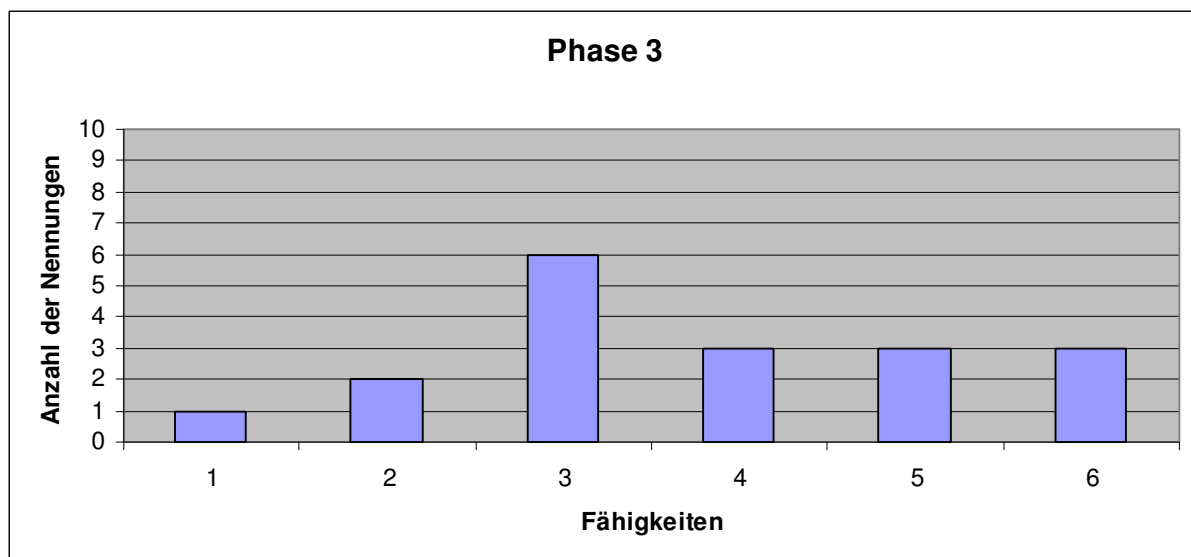
1. Begrüßung am Schalter und Überleitung in das Beratungszimmer _____
2. Gegenseitiges Kennen lernen _____
3. Bedarfserhebung _____
4. Beratung und Produktverkauf & Crossselling _____
5. Terminvereinbarung & Verabschiedung _____

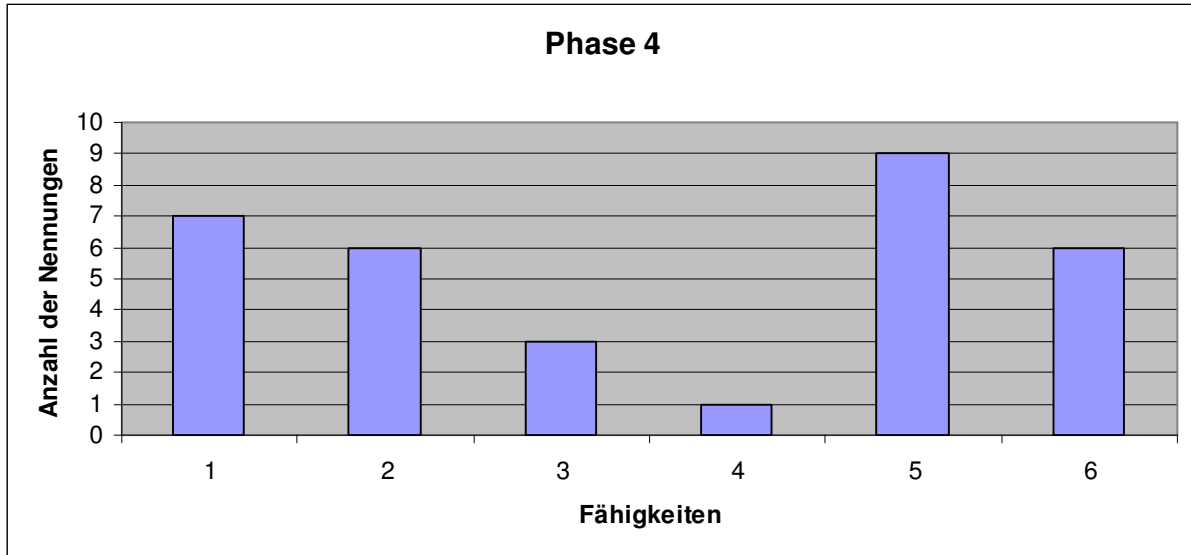
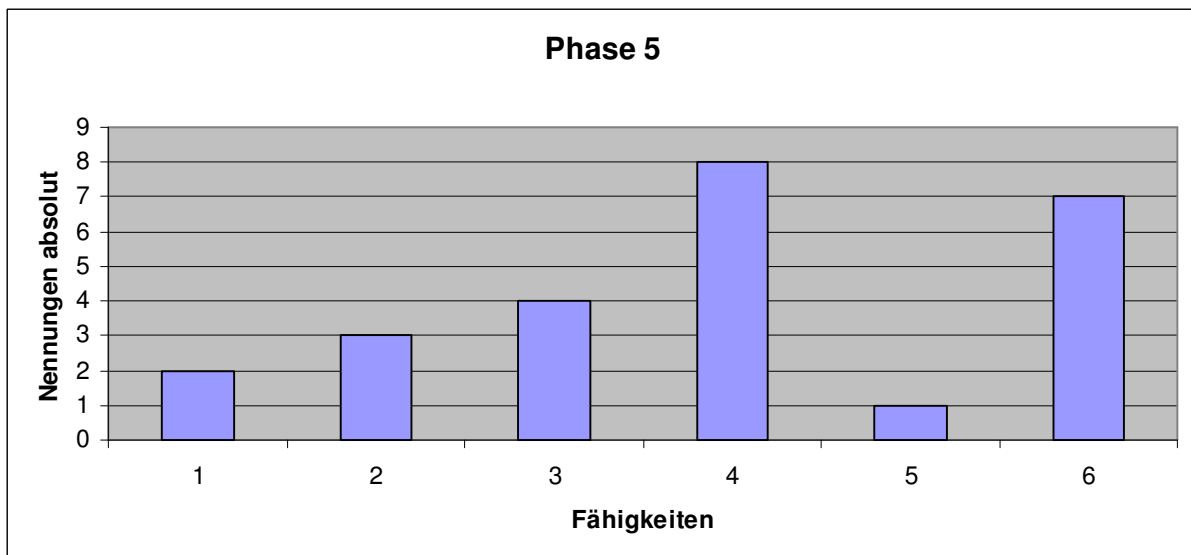
1. Unternehmerisches Denken
2. Durchsetzungsvermögen / Verhandlungsgeschick
3. Sprachlicher Ausdruck / Rhetorik
4. Kontaktfähigkeit / soziale Kompetenz
5. Verkäuferische Kompetenz
6. Eigeninitiative

Ordnen Sie bei Ihrer Antwort bitte die Fähigkeiten den einzelnen Phasen zu, indem Sie die Nummer der jeweiligen Fähigkeit hinter die genannte Phase schreiben. Sie haben hierbei die Möglichkeit, mehrere Fähigkeiten pro Phase zu nennen.

Ergebnis aus der Befragung von N=10 Experten:



Ergebnis aus der Befragung von N=10 Experten:**Ergebnis aus der Befragung von N=10 Experten:**

Ergebnis aus der Befragung von N=10 Experten:**Ergebnis aus der Befragung von N=10 Experten:**

Das „Continuity-Editing-System“ oder „Hollywood-Continuity-System“

Methode des Schnitts, die sich in den 20er Jahren überwiegend in den USA und speziell in Hollywood entwickelt hat. Sie beherrscht noch heute die Filmsprache und kann inzwischen als nahezu global verbreiteter, Standard bestimmter filmischer Konventionen angesehen werden. Durch dieses System, das sowohl die Position verschiedener am Set beteiligter Kameras zueinander als auch die Abfolge von Kameraeinstellungen und Regeln zu Kamerabewegungen festlegt, sollen glatte Übergänge zwischen einzelnen Einstellungen ermöglicht werden indem der Schnitt so gut wie unsichtbar gemacht wird (Monaco, 1996, S. 548).

Eine solche typische Abfolge beginnt mit dem *Master shot* oder *Establishing Shot*, der dem Zuschauer einen Überblick über die gesamte Szene vermitteln soll. Dieser *Establishing Shot* wird von Einstellung abgelöst, z.B. dem *cut-in*, die stärker auf die handelnden Personen fokussieren, indem der Bildausschnitt immer kleiner gewählt wird, bis nur noch jeweils eine Person im Bild zu sehen ist. Die Kamera zeigt nun immer abwechselnd die beiden handelnden Personen (*Shot/Reverse Shot*). Die Abfolge endet nach einiger Zeit mit dem so genannten *Re-Establishing Shot*, um den szenischen Bezugsrahmen für den Zuschauer zu erneuern (Beller, 2005, S.16).

Die 180 Grad Regel

Schlüsselement des *Continuity-Editing-Systems*. Die 180 Grad Regel besagt, dass sich alle am Dreh beteiligten Kameras auf ein und derselben Seite einer imaginären Linie befinden müssen, die den Set in der Mitte durchtrennt. Diese Linie wird als *Handlungsachse* bezeichnet. Durch die Tatsache, dass sich alle Kameras nur auf einer Seite dieser Linie befinden dürfen, wird dem Zuschauer die räumliche Orientierung stark erleichtert. Ein irritierender Perspektivwechsel wird durch die *180 Grad Regel* vermieden (Beller, 2005, S. 15 ff; Galalick, 1988, S. 68ff).

Die 30 Grad Regel

Neben der grundlegenden *180 Grad Regel* gibt es einige weitere Konventionen, die das Rezipieren von filmischem Material für den Zuschauer angenehmer machen und die Aufmerksamkeit von technischen Details, wie dem Schnitt, ablenken. Die erste ist die so genannte *30 Grad Regel*. Sie sorgt für einen harmonischen Ablauf der Filmsequenz. Hier wird festgelegt, dass sich die Positionen der Kamera, von zwei aufeinander folgenden Kameraeinstellungen, um mindestens 30 Grad unterscheiden müssen. Hierdurch wird der Effekt vermieden, dass ein Objekt oder eine Person in der Handlung zu springen scheint, weil die neue Perspektive der alten vermeintlich gleicht, die Kamera sich aber an einem neuen Ort befindet (Gallick, 1988, S. 78ff.).

Die „Match on Action“ Technik

Mit der *„Match on Action“ Technik*, als zweiter Konvention, welche vorsieht einen Schnitt immer in die Mitte einer Handlung zu legen, wird die Aufmerksamkeit des Betrachters vom Schnitt abgelenkt, so dass die Handlung in den Focus der Aufmerksamkeit fällt. (Beller, 2005, Seite 15ff; Bordwell und Thompson, 2004, S. 310 ff)

Die „Eyeline Match“ Technik

Auch die *„Eyeline Match“ Technik* trägt dazu bei die Aufmerksamkeit des Betrachters auf die Handlung und die Personen zu lenken. Bei dieser Technik folgt die Kameraeinstellung dem Blick einer handelnden Person, um nach dem Schnitt dann die Person oder das Objekt zu zeigen, welches Ziel des vorherigen Blickes gewesen ist. (Beller, 2005, Seite 15ff; Bordwell und Thompson, 2004, S. 310 ff)

Ziel der oben beschriebenen Konventionen und Regeln ist es dem Betrachter die Inhalte in einem kontinuierlichen und klaren Erzählstil zu präsentieren und es ihm dadurch einfacher zu machen der Handlung zu folgen bzw. diese so real wie möglich erscheinen zu lassen.

Die Darstellung der verwendeten technischen Komponenten erfolgt zur besseren Verständlichkeit des Lesers am Beispiel der konkreten Vorgehensweise bei der Erstellung des VAK.

Das Filmen der Sequenzen erfolgte mit 3 professionellen Kameras (Sony Camcorder DSR 370-P, Sony Kamerakopf DXC D30, Sony Kamerakopf DXC D30 mit DVcam Andockrecorder DSR-1P), die alle über einen externen Timecode - Eingang verfügten. Durch den gleichzeitigen Einsatz von 3 Kameras, deren Timecodesignal gleichgeschaltet war musste nicht, wie sonst beim professionellen Filmen üblich, die Kamera für jede neue Einstellung bewegt werden. Stattdessen filmten alle 3 Kameras alle Sequenzen von Anfang bis Ende und die interessierende Einstellung wurde beim späteren Schneiden dann einfach der jeweiligen Kamera entnommen. Durch die parallel geschalteten Timecodes ließen sich die Einstellungen der 3 Kameras später problemlos aneinander fügen, da für die Übergänge immer nur nach der Passung der jeweiligen Nummer im Timecode geschaut werden musste.

Der Ton wurde mit Hilfe von Ansteckmikrofonen (Lavaliermikrofone mit Kugelcharakteristik und einem Taschensender) für jeden der Rollenspieler getrennt aufgenommen, da die Tonabnahme unmittelbar an der handelnden Person eine qualitativ hochwertigere Tonspur gewährleistet als der Gebrauch eines an der Kamera selbst befindlichen Mikrofons. In der Praxis wird der dazugehörige Empfänger an derjenigen Kamera befestigt, die den Shot auf die Person mit dem Mikro aufzeichnete. Die zusammengehörigen Bild und Toninformationen befinden sich somit auf einem Band. Bei dem verwendeten Sende/Empfangssystem handelte es sich jeweils um Geräte des Typs Sennheiser Evolution Wireless ew 112p G2. Bei den Aufnahmen im großen Schalterraum wurde ergänzend ein Richtmikrofon (Sennheiser ME 66 mit Speiseadapter K6) eingesetzt um auch hier den Ton so authentisch wie möglich abbilden zu können.

Um kontrastreiche und freundlich wirkende Bilder gewährleisten zu können wurde für die Beleuchtung des Schalterraums auf drei 800W Strahler (Arrilite 800) zurückgegriffen. Für die Dreharbeiten in dem kleineren Beratungszimmer kamen drei 300W Strahler (Sachtler Reporter 300H, mit Diffuserfolie) zum Einsatz.

Videobasiertes Analyseverfahren für Kundenberater

Weiter

Sehr geehrter Teilnehmer / sehr geehrte Teilnehmerin!

Herzlich willkommen zum Videobasierten Analyseverfahren für Kundenberater!

Das VAK besteht aus 25 kurzen Videoclips, die alle einer typischen Beratungssituation aus dem Sparkassenalltag entnommen sind.

Es werden Ihnen immer 5 dieser Videoclips präsentiert, die zu einer thematischen Sequenz zusammengefasst sind.

Es gibt insgesamt 5 solcher Sequenzen.

Die einzelnen Sequenzen lauten:

Sequenz 1: Begrüßung des Kunden

Sequenz 2: Gegenseitiges Kennenlernen

Sequenz 3: Bedarfsermittlung

Sequenz 4: Beratung und Verkauf

Sequenz 5: Verabschiedung des Kunden

Weiter

Wir möchten Sie nun bitten aus diesen 5 Sequenzen, die Ihnen nacheinander präsentiert werden, immer denjenigen Videoclip auszuwählen, der Ihrem persönlichen Verhalten als Kundenberater in einer solchen Situation am besten entspricht.

Stellen Sie sich vor Ihrer Auswahl bitte immer die Frage:

„Wie würde ich mich als Kundenberater in einer solchen Situation verhalten?“

Sie können immer nur eine Sequenz auswählen!

Sie dürfen sich die Sequenzen so oft anschauen, wie sie zu Ihrer Entscheidungsfindung brauchen. Es gibt kein zeitliches Limit!

Das Abspielen der Filme erfolgt über das Anklicken des jeweiligen Filmfensters.

Die Auswahl erfolgt über das Anklicken der runden Box links neben dem jeweiligen Filmfenster.

Weiter

Bevor sie mit der Bearbeitung der einzelnen Sequenzen beginnen bitten wir Sie noch folgende Angaben zu machen:

Name Ihrer Sparkasse:

Vorname:

Nachname:

Video anzeigen:

Audio abspielen:

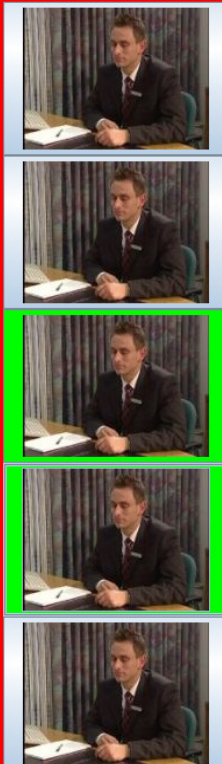
Weiter

Intro



Weiter

Sequenz 1: Begrüßung des Kunden



Weiter

Sequenz 2: Gegenseitiges Kennenlernen

Weiter

Sequenz 3: Bedarfsermittlung

Weiter

Sequenz 4: Beratung und Verkauf



00:02:16.11 / 00:02:34.24

Weiter

Sequenz 5: Verabschiedung des Kunden



00:01:24.75 / 00:01:25.88

Weiter

Vielen Dank für Ihre Teilnahme!

Neustart

Schliessen

Bewertungsschlüssel zur Paper-Pencil Variante des Rollenspiels

Für die im Folgenden genannten Verhaltensweisen bekommt der Proband jeweils 1 Punkt wenn er diese Verhaltensweise in seinem Vorgehen beschreibt! Die in Klammern gesetzten Hinweise sind erklärende Beispiele für mögliche richtige bzw. falsche Ausführungen der Teilnehmer! Jede dieser Verhaltensweisen kann nur einmal bepunktet werden! Die Maximalpunktzahl beträgt demnach 25 Punkte!

Die vorgegebene Gesprächsstruktur muss vom Probanden nicht eingehalten worden sein! Sie dient nur einer erleichterten Bewertung durch den Gutachter!

Begrüßung des Kunden:

1. Ich begrüße den Kunden mit Handschlag – oder - Ich begrüße den Kunden freundlich.
2. Ich frage den Kunden **aktiv** nach seinen Wünschen (*was kann ich für Sie tun? und nicht: kann ich etwas für sie tun?*).
3. Ich weise den Kunden auf die Vorteile einer Beratung in den Beratungszimmern hin.
4. Ich begleite den Kunden in das Beratungszimmer und biete Ihm einen Platz.
5. Ich biete dem Kunden etwas zu trinken an.

Gegenseitiges Kennenlernen:

1. Ich stelle mich persönlich vor und weise auf meine Rolle im Unternehmen hin (*Mein Name ist XY und ich bin hier als Berater der Ansprechpartner für alle Anliegen unserer Kunden*).
2. Ich stelle offene Fragen und bin aufgeschlossen und freundlich – oder - Hinweis auf small talk.
3. Ich erkundige mich nach dem bisherigen Werdegang des Kunden (Ausbildung, Beruf, Familie, Wohnsituation).
4. Ich bestärke den Kunden positiv in seiner jetzigen Situation, hier alternativ: Ich Sorge für eine offene und freundliche Atmosphäre. (Gratulation nach abgeschlossenem Studium oder zum neuen Job, etc.).
5. Ich frage den Kunden nach seinen aktuellen Zielen und Plänen (Familie, Erwerb von Eigentum).

Bedarfsermittlung:

1. Ich weise den Kunden auf den Nutzen einer zielgerichteten und umfassenden Beratung hin – oder – Hinweis auf das S-Finanzkonzept.
2. Ich frage den Kunden welche Produkte er bisher bei seiner alten Bank genutzt hat (Online-Banking, Kreditkarte, Sparverträge).
3. Ich empfehle dem Kunden den Umzugsservice oder ich eröffne ein Girokonto.
4. Ich frage den Kunden ob er sich schon Gedanken darüber gemacht hat, was er mit dem Geld, das ihm jetzt zur Verfügung steht, anfangen will (z.B. Geldrücklage, Ansparpläne, Bausparen oder Vermögenswirksame Leistungen).
5. Ich frage den Kunden wie seine bisherige Absicherung gewesen ist und weise ihn auf die Wichtigkeit einer zusätzlichen Beratung in diesem Bereich hin (Versicherungen).

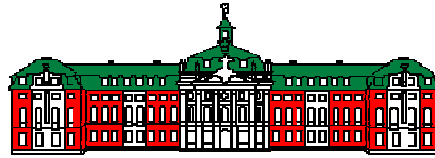
Beratung und Verkauf:

1. Ich frage den Kunden nach seinen Anlagezielen.
2. Ich erkläre dem Kunden den Zusammenhang zwischen Rendite, Risiko und Verfügbarkeit indem ich diesen Zusammenhang visualisiere (Goldenes Dreieck).
3. Ich setze Verkaufshilfen ein (Prospekte von DEKA, LBS).
4. Ich erkläre die Passung von Produktmerkmalen und Kundenwünschen – oder – ich weise auf die Vorteile hin, die das Produkt für den Kunden mit sich bringt.
5. Ich empfehle **ein** konkretes Produkt.

Verabschiedung des Kunden:

1. Ich fasse das Gespräch zusammen.
2. Ich gebe dem Kunden die Möglichkeit selber noch Fragen zu stellen bzw. Unklarheiten anzusprechen.
3. Ich stelle mich dem Kunden als Ansprechpartner auch außerhalb der Öffnungszeiten zur Verfügung.
4. Ich vereinbare einen konkreten Termin mit dem Kunden.
5. Ich verabschiede den Kunden mit Handschlag und wünsche ihm alles Gute oder - und bedanke mich für das Gespräch.

Name: _____



Liebe Untersuchungsteilnehmerin, lieber Untersuchungsteilnehmer,

ich möchte mich an dieser Stelle herzlich für Ihre Bereitschaft bedanken, an dieser Untersuchung teilzunehmen.

Im Folgenden bitte ich Sie um die Bewertung von einigen Aussagen, die sich alle auf die Übung „**Rollenspiel**“ beziehen, die Sie während des Verfahrens bearbeitet haben.

Bitte geben Sie bei jeder der nachfolgenden Aussagen an, in wie weit sie dieser in Bezug auf das VAK zustimmen.

Hierzu kreuzen Sie bitte für jede Aussage diejenige Zahl von 1 (= stimme gar nicht zu) bis 7 (= stimme voll zu) auf der jeweils folgenden Zahlenreihe an, die Ihre Bewertung der Aussage am besten wiedergibt.

Bitte kreuzen Sie immer nur eine Zahl an und machen Sie Ihre Kreuze nicht zwischen den Zahlen! Ihre Bearbeitungszeit ist nicht begrenzt. Bitte treffen Sie Ihre Entscheidung jedoch möglichst spontan.

Im Folgenden finden Sie ein Beispiel dafür, wie sie die Aussagen mittels der Zahlenreihe bewerten können:

(Dieses Beispiel bezieht sich noch nicht auf das Rollenspiel!)

	Stimme gar nicht zu		Stimme voll zu				
	↓		↓				
Ich treibe gerne Sport!	1	2	3	4	5	6	7 <input checked="" type="checkbox"/>

Die befragte Person in diesem Beispiel treibt sehr gerne Sport und hat daher die 7 angekreuzt, da Sie der Aussage voll zustimmt!

Eine Person, die überhaupt nicht gerne Sport treibt, würde die 1 markieren, da sie der Aussage gar nicht zustimmt und eine Person, die weder gerne noch ungerne Sport treibt würde das Kreuz bei der 4 setzen.

Für weitere Fragen zur Bearbeitung steht Ihnen der Untersuchungsleiter gerne zur Verfügung!

Wenn Sie keine weiteren Fragen zur Bearbeitung mehr haben blättern Sie jetzt bitte um und beginnen mit der Bearbeitung der Aussagen.

Ihr Untersuchungsleiter!

	Stimme gar nicht zu ↓						stimme voll zu ↓
1. Ich fühlte mich vor der Bearbeitung der Aufgaben hinreichend informiert über das, was in dem Verfahren auf mich zukam.	1	2	3	4	5	6	7
2. In dem Verfahren konnte ich meine Stärken gut zum Ausdruck bringen.	1	2	3	4	5	6	7
3. Während der gesamten Bearbeitungszeit war mir bewusst welches Verhalten von mir erwartet wurde.	1	2	3	4	5	6	7
4. Das Verfahren ist dazu in der Lage, meine Fähigkeiten als Berater zutreffend zu erfassen.	1	2	3	4	5	6	7
5. Ich würde dieses Verfahren zur Personalauswahl einsetzen.	1	2	3	4	5	6	7
6. Die Bearbeitung des Verfahrens gibt mir einen realistischen Einblick in die spätere Beratertätigkeit.	1	2	3	4	5	6	7
7. Mein Abschneiden in dem Verfahren hängt vor allem von mir selbst ab.	1	2	3	4	5	6	7
8. Das Verfahren hat einen engen Bezug zum wahren Berufsleben und den dort benötigten Fähigkeiten.	1	2	3	4	5	6	7
9. Das Abschneiden im Verfahren hängt entscheidend von den Fähigkeiten des jeweiligen Teilnehmers ab.	1	2	3	4	5	6	7
10. Das Verfahren eignet sich gut zur Auswahl von Kundenberatern.	1	2	3	4	5	6	7
11. Ich fühlte mich während der Durchführung des Verfahrens hinreichend über die an mich gestellten Anforderungen informiert.	1	2	3	4	5	6	7
12. Äußere Faktoren entscheiden darüber, wie gut meine Leistung in dem Verfahren ist.	1	2	3	4	5	6	7
13. Dieses Verfahren ist gerecht in der Beurteilung seiner Teilnehmer.	1	2	3	4	5	6	7

	Stimme gar nicht zu ↓							stimme voll zu ↓
14. Mir ist klar, was mit diesem Verfahren gemessen werden soll.	1	2	3	4	5	6	7	
15. Die Art der Aufgabendarstellung war motivierend.	1	2	3	4	5	6	7	
16. Das Verfahren bezieht sich auf realistische Verhaltensweisen aus dem Alltag eines Bankberaters.	1	2	3	4	5	6	7	
17. Mein Abschneiden in dem Verfahren ist unabhängig von der Bewertung durch andere Personen.	1	2	3	4	5	6	7	
18. Das Verfahren erhebt relevante und wichtige Informationen über den Bewerber.	1	2	3	4	5	6	7	
19. Das Ergebnis des Verfahrens ist durch den Teilnehmer leicht zu verfälschen.	1	2	3	4	5	6	7	
20. Das eingesetzte Medium ist zur Erfassung von Beratungskompetenz gut geeignet.	1	2	3	4	5	6	7	

Abschließend noch ein paar Fragen zu Ihrer Person:

Geschlecht: _____

Alter: _____

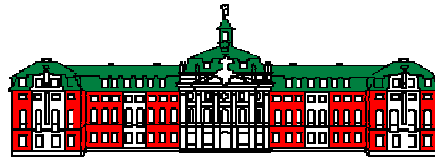
Höchster erreichter
Bildungsabschluss: _____

Beruf/ Studium: _____

Alle Daten werden selbstverständlich vertraulich behandelt!

Vielen Dank!

Name: _____



Liebe Untersuchungsteilnehmerin, lieber Untersuchungsteilnehmer,

ich möchte mich an dieser Stelle herzlich für Ihre Bereitschaft bedanken, an dieser Untersuchung teilzunehmen.

Im Folgenden bitte ich Sie um die Bewertung von einigen Aussagen, die sich alle auf das „**Videobasierte Auswahlverfahren für Kundenberater**“ (VAK) beziehen, das Sie während des Verfahrens bearbeitet haben.

Bitte geben Sie bei jeder der nachfolgenden Aussagen an, in wie weit sie dieser in Bezug auf das VAK zustimmen.

Hierzu kreuzen Sie bitte für jede Aussage diejenige Zahl von 1 (= stimme gar nicht zu) bis 7 (= stimme voll zu) auf der jeweils folgenden Zahlenreihe an, die Ihre Bewertung der Aussage am besten wiedergibt.

Bitte kreuzen Sie immer nur eine Zahl an und machen Sie Ihre Kreuze nicht zwischen den Zahlen! Ihre Bearbeitungszeit ist nicht begrenzt. Bitte treffen Sie Ihre Entscheidung jedoch möglichst spontan.

Im Folgenden finden Sie ein Beispiel dafür, wie sie die Aussagen mittels der Zahlenreihe bewerten können:

(Dieses Beispiel bezieht sich noch nicht auf das VAK!)

	Stimme gar nicht zu	Stimme voll zu					
	↓	↓					
Ich treibe gerne Sport!	1	2	3	4	5	6	7
							✗

Die befragte Person in diesem Beispiel treibt sehr gerne Sport und hat daher die 7 angekreuzt, da sie der Aussage voll zustimmt!

Eine Person, die überhaupt nicht gerne Sport treibt, würde die 1 markieren, da sie der Aussage gar nicht zustimmt und eine Person, die weder gerne noch ungerne Sport treibt würde das Kreuz bei der 4 setzen.

Für weitere Fragen zur Bearbeitung steht Ihnen der Untersuchungsleiter gerne zur Verfügung!

Wenn Sie keine weiteren Fragen zur Bearbeitung mehr haben blättern Sie jetzt bitte um und beginnen mit der Bearbeitung der Aussagen.

Ihr Untersuchungsleiter!

	Stimme gar nicht zu ↓						stimme voll zu ↓
1. Ich fühlte mich vor der Bearbeitung der Aufgaben hinreichend informiert über das, was in dem Verfahren auf mich zukam.	1	2	3	4	5	6	7
2. In dem Verfahren konnte ich meine Stärken gut zum Ausdruck bringen.	1	2	3	4	5	6	7
3. Während der gesamten Bearbeitungszeit war mir bewusst welches Verhalten von mir erwartet wurde.	1	2	3	4	5	6	7
4. Das Verfahren ist dazu in der Lage, meine Fähigkeiten als Berater zutreffend zu erfassen.	1	2	3	4	5	6	7
5. Ich würde dieses Verfahren zur Personalauswahl einsetzen.	1	2	3	4	5	6	7
6. Die Bearbeitung des Verfahrens gibt mir einen realistischen Einblick in die spätere Beratertätigkeit.	1	2	3	4	5	6	7
7. Mein Abschneiden in dem Verfahren hängt vor allem von mir selbst ab.	1	2	3	4	5	6	7
8. Das Verfahren hat einen engen Bezug zum wahren Berufsleben und den dort benötigten Fähigkeiten.	1	2	3	4	5	6	7
9. Das Abschneiden im Verfahren hängt entscheidend von den Fähigkeiten des jeweiligen Teilnehmers ab.	1	2	3	4	5	6	7
10. Das Verfahren eignet sich gut zur Auswahl von Kundenberatern.	1	2	3	4	5	6	7
11. Ich fühlte mich während der Durchführung des Verfahrens hinreichend über die an mich gestellten Anforderungen informiert.	1	2	3	4	5	6	7
12. Äußere Faktoren entscheiden darüber, wie gut meine Leistung in dem Verfahren ist.	1	2	3	4	5	6	7
13. Dieses Verfahren ist gerecht in der Beurteilung seiner Teilnehmer.	1	2	3	4	5	6	7

	Stimme gar nicht zu ↓						stimme voll zu ↓
14. Mir ist klar, was mit diesem Verfahren gemessen werden soll.	1	2	3	4	5	6	7
15. Die Art der Aufgabendarstellung war motivierend.	1	2	3	4	5	6	7
16. Das Verfahren bezieht sich auf realistische Verhaltensweisen aus dem Alltag eines Bankberaters.	1	2	3	4	5	6	7
17. Mein Abschneiden in dem Verfahren ist unabhängig von der Bewertung durch andere Personen.	1	2	3	4	5	6	7
18. Das Verfahren erhebt relevante und wichtige Informationen über den Bewerber.	1	2	3	4	5	6	7
19. Das Ergebnis des Verfahrens ist durch den Teilnehmer leicht zu verfälschen.	1	2	3	4	5	6	7
20. Das eingesetzte Medium ist zur Erfassung von Beratungskompetenz gut geeignet.	1	2	3	4	5	6	7

Abschließend noch ein paar Fragen zu Ihrer Person:

Geschlecht: _____

Alter: _____

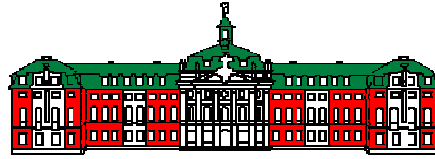
Höchster erreichter
Bildungsabschluss: _____

Beruf/ Studium: _____

Alle Daten werden selbstverständlich vertraulich behandelt!

Vielen Dank!

Name: _____



Liebe Untersuchungsteilnehmerin, lieber Untersuchungsteilnehmer,

ich möchte mich an dieser Stelle herzlich für Ihre Bereitschaft bedanken, an dieser Untersuchung teilzunehmen.

Im Folgenden bitte ich Sie um die Bewertung von einigen Aussagen, die sich alle auf die „**schriftliche Übung für Kundenberater**“ beziehen, die Sie während des Verfahrens bearbeitet haben.

Bitte geben Sie bei jeder der nachfolgenden Aussagen an, in wie weit sie dieser in Bezug auf das VAK zustimmen.

Hierzu kreuzen Sie bitte für jede Aussage diejenige Zahl von 1 (= stimme gar nicht zu) bis 7 (= stimme voll zu) auf der jeweils folgenden Zahlenreihe an, die Ihre Bewertung der Aussage am besten wiedergibt.

Bitte kreuzen Sie immer nur eine Zahl an und machen Sie Ihre Kreuze nicht zwischen den Zahlen! Ihre Bearbeitungszeit ist nicht begrenzt. Bitte treffen Sie Ihre Entscheidung jedoch möglichst spontan.

Im Folgenden finden Sie ein Beispiel dafür, wie sie die Aussagen mittels der Zahlenreihe bewerten können:

(Dieses Beispiel bezieht sich noch nicht auf die schriftliche Übung!)

	Stimme gar nicht zu	Stimme voll zu
	↓	↓
Ich treibe gerne Sport!	1	2
	3	4
	5	6
		7 X

Die befragte Person in diesem Beispiel treibt sehr gerne Sport und hat daher die 7 angekreuzt, da Sie der Aussage voll zustimmt!

Eine Person, die überhaupt nicht gerne Sport treibt, würde die 1 markieren, da sie der Aussage gar nicht zustimmt und eine Person, die weder gerne noch ungerne Sport treibt würde das Kreuz bei der 4 setzen.

Für weitere Fragen zur Bearbeitung steht Ihnen der Untersuchungsleiter gerne zur Verfügung!

Wenn Sie keine weiteren Fragen zur Bearbeitung mehr haben blättern Sie jetzt bitte um und beginnen mit der Bearbeitung der Aussagen.

Ihr Untersuchungsleiter!

	Stimme gar nicht zu ↓						stimme voll zu ↓
1. Ich fühlte mich vor der Bearbeitung der Aufgaben hinreichend informiert über das, was in dem Verfahren auf mich zukam.	1	2	3	4	5	6	7
2. In dem Verfahren konnte ich meine Stärken gut zum Ausdruck bringen.	1	2	3	4	5	6	7
3. Während der gesamten Bearbeitungszeit war mir bewusst welches Verhalten von mir erwartet wurde.	1	2	3	4	5	6	7
4. Das Verfahren ist dazu in der Lage, meine Fähigkeiten als Berater zutreffend zu erfassen.	1	2	3	4	5	6	7
5. Ich würde dieses Verfahren zur Personalauswahl einsetzen.	1	2	3	4	5	6	7
6. Die Bearbeitung des Verfahrens gibt mir einen realistischen Einblick in die spätere Beratertätigkeit.	1	2	3	4	5	6	7
7. Mein Abschneiden in dem Verfahren hängt vor allem von mir selbst ab.	1	2	3	4	5	6	7
8. Das Verfahren hat einen engen Bezug zum wahren Berufsleben und den dort benötigten Fähigkeiten.	1	2	3	4	5	6	7
9. Das Abschneiden im Verfahren hängt entscheidend von den Fähigkeiten des jeweiligen Teilnehmers ab.	1	2	3	4	5	6	7
10. Das Verfahren eignet sich gut zur Auswahl von Kundenberatern.	1	2	3	4	5	6	7
11. Ich fühlte mich während der Durchführung des Verfahrens hinreichend über die an mich gestellten Anforderungen informiert.	1	2	3	4	5	6	7
12. Äußere Faktoren entscheiden darüber, wie gut meine Leistung in dem Verfahren ist.	1	2	3	4	5	6	7
13. Dieses Verfahren ist gerecht in der Beurteilung seiner Teilnehmer.	1	2	3	4	5	6	7

	Stimme gar nicht zu ↓						stimme voll zu ↓
14. Mir ist klar, was mit diesem Verfahren gemessen werden soll.	1	2	3	4	5	6	7
15. Die Art der Aufgabendarstellung war motivierend.	1	2	3	4	5	6	7
16. Das Verfahren bezieht sich auf realistische Verhaltensweisen aus dem Alltag eines Bankberaters.	1	2	3	4	5	6	7
17. Mein Abschneiden in dem Verfahren ist unabhängig von der Bewertung durch andere Personen.	1	2	3	4	5	6	7
18. Das Verfahren erhebt relevante und wichtige Informationen über den Bewerber.	1	2	3	4	5	6	7
19. Das Ergebnis des Verfahrens ist durch den Teilnehmer leicht zu verfälschen.	1	2	3	4	5	6	7
20. Das eingesetzte Medium ist zur Erfassung von Beratungskompetenz gut geeignet.	1	2	3	4	5	6	7

Abschließend noch ein paar Fragen zu Ihrer Person:

Geschlecht: _____

Alter: _____

Höchster erreichter
Bildungsabschluss: _____

Beruf/ Studium: _____

Alle Daten werden selbstverständlich vertraulich behandelt!

Vielen Dank!

