



WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

Kundenorientierung im Hotelfach

Die Entwicklung und Validierung eines Situational Judgment Tests

Jan Henning Möller





PSYCHOLOGIE

KUNDENORIENTIERUNG IM HOTELFACH

DIE ENTWICKLUNG UND VALI-
DIERUNG EINES SITUATIONAL
JUDGMENT TESTS

INAUGURALDISSERTATION
ZUR ERLANGUNG DES DOKTORGRADES
DER
PHILOSOPHISCHEN FAKULTÄT DER
WESTFÄLISCHEN WILHELMS-UNIVERSITÄT MÜNSTER

vorgelegt von
JAN HENNING MÖLLER
aus Bönen/Westf.

Juni 2010

Tag der mündlichen Prüfung: 13. Oktober 2010

Dekan der philosophischen Fakultät: Prof. Dr. Christian Pietsch

Gutachter: Prof. Dr. Uwe P. Kanning
Prof. Dr. Stefan Schmukle

Jan Henning Möller

Kundenorientierung im Hotelfach



WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

Wissenschaftliche Schriften der WWU Münster

Reihe VIII

Band 2

Jan Henning Möller

Kundenorientierung im Hotelfach

Die Entwicklung und Validierung eines Situational Judgment Tests



MV WISSENSCHAFT

Wissenschaftliche Schriften der WWU Münster

herausgegeben von der Universitäts- und Landesbibliothek Münster

<http://www.ulb.uni-muenster.de>

Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Dieses Buch steht gleichzeitig in einer elektronischen Version über den Publikations- und Archivierungsserver der WWU Münster zur Verfügung.

<http://www.ulb.uni-muenster.de/wissenschaftliche-schriften>

Jan Henning Möller

„Kundenorientierung im Hotelfach. Die Entwicklung und Validierung eines Situational Judgment Tests“
Wissenschaftliche Schriften der WWU Münster, Reihe VIII, Band 2

© 2010 der vorliegenden Ausgabe:

Die Reihe „Wissenschaftliche Schriften der WWU Münster“ erscheint im Verlagshaus Monsenstein und Vannerdat OHG Münster
www.mv-wissenschaft.com

ISBN 978-3-8405-0026-8 (Druckausgabe)

URN [urn:nbn:de:hbz:6-76429648420](http://nbn-resolving.org/urn:nbn:de:hbz:6-76429648420) (elektronische Version)

© 2010 Jan Henning Möller

Alle Rechte vorbehalten

Satz: Jan Henning Möller

Umschlag: MV-Verlag

Druck und Bindung: MV-Verlag

Für meine Eltern

Danksagung

Als ich mich 2007 dazu entschloss, eine Dissertation zu schreiben, schüttelten viele meiner Freunde und Bekannten nur den Kopf. Dass ich selbst meinen Entschluss *nie* bereut habe, ist einer Vielzahl an Personen zu verdanken:

An erster Stelle möchte ich *Herrn Prof. Dr. Uwe P. Kanning* für die exzellente Betreuung dieser Arbeit danken. Herr Prof. Dr. Kanning hat mir die Chance gegeben, über ein innovatives und spannendes Thema promovieren zu dürfen. Zudem stand er mir – selbst zu manchmal ungewöhnlichen Zeiten oder Anlässen, z. B. zwischen zwei Übungen in einem Assessment Center – immerzu als wertvoller Rat- oder Ideengeber zur Seite. Für die spontane Übernahme der Zweitkorrektur und bereichernde Gespräche habe ich *Herrn Prof. Dr. Stefan Schmukle* zu danken.

Ein riesiger Dank gebührt *Frau Angelina Wendt*, welche mich bei dem gesamten Prozess der Testentwicklung und Datenerhebung(-en) in keinesfalls selbstverständlicher Art und Weise unterstützt hat und in erheblichem Maße zum erfolgreichen Gelingen meiner Forschung beigetragen hat. In Bezug auf die Datenerhebungen habe ich ebenfalls *Frau Verena Bredol* und *Frau Dr. Anke Meizies* zu danken, welche mich in ihrem jeweiligen Verantwortungsbereich haben Daten erheben lassen.

Die Produktion von Videos ist immer mit viel technischem Aufwand verbunden. Deshalb möchte ich an dieser Stelle *Herrn Glaser* und *Herrn Kemper* vom Servicepunkt Film in Münster meinen Dank aussprechen für die professionelle Videoproduktion und flexible, jederzeit hilfsbereite Unterstützung. Videos sind ohne gute und motivierte Schauspieler undenkbar. Hier geht mein Dank an *Frau Sara Bachmann*, *Herrn Christian Gawlista*, *Herrn Philip Klee*, *Herrn Nikolay Kolev*, *Herrn Christian Krämer* und *Herrn Dr. Jens Pöttker*.

In der sechsten oder siebten Klasse fragte mich eine Mitschülerin aus heiterem Himmel, was denn wohl das passende Substantiv zum Verb „promovieren“ wäre. Damals war ich sprachlos, heute wüsste ich die Antwort...

Inhaltsverzeichnis

1. Einleitung	1
2. Theoretischer Hintergrund	3
2.1 Die historische Entwicklung und Bedeutung von SJTs	3
2.2 SJTs als simulationsorientierte Verfahren der Berufseignungsdiagnostik	11
2.3 Die Konstruktion von SJTs.....	17
2.3.1 Die Herkunft der in SJTs abgebildeten Situationen	17
2.3.2 Inhalte und Charakteristiken von SJT-Items	18
2.3.3 Die Wahl der Instruktionsform für die Probanden.....	19
2.3.4 Die Gestaltung der Verhaltensalternativen im SJT	24
2.3.5 Die Festlegung eines Bewertungsschlüssels	26
2.4 Videobasierte SJTs	33
2.4.1 Das Konzept der fidelity.....	33
2.4.2 Videobasierte SJTs im Überblick.....	34
2.5 Der Nutzen von SJTs für die Personalauswahl	46
2.5.1 Kriterienbezogene Validität.....	46
2.5.2 Die Konstruktvalidität von SJTs.....	49
2.5.3 Soziale Validität von SJTs und Testfairness	65
2.5.4 Einflüsse auf die Validität	70
2.6 Kundenorientierung und ihre Bedeutung für ein Unternehmen....	77
2.6.1 Kundenorientierung	79
2.6.2 Kundenzufriedenheit.....	82
2.6.3 Kundenloyalität.....	88
2.6.4 Kundenwert.....	90
3. Die erste Studie: Testentwicklung.....	92
3.1 Durchführung einer Anforderungsanalyse.....	94
3.2 Durchführung von Expertenworkshops.....	102
3.3 Videoproduktion	107
3.4 Diskussion	110
3.5 Ausblick	113

4. Die zweite Studie.....	114
4.1 Beschreibung der eingesetzten Instrumente und Fragebögen.....	114
4.1.1 Das Inventar sozialer Kompetenzen in der Kurzversion	115
4.1.2 Der eingesetzte SJT	118
4.1.3 Demographische Variablen	123
4.2 Ablauf der Datenerhebung und Beschreibung der Stichprobe	123
4.2.1 Ablauf der Datenerhebung	123
4.2.2 Beschreibung der Stichprobe	125
4.3 Deskriptive Statistik.....	126
4.4 Inferenzstatistische Analysen und Hypothesentestung	128
4.4.1 Konfirmatorische Faktorenanalyse	128
4.4.2 Exploration der faktoriellen Struktur des SJT.....	131
4.4.3 Zentrale Statistiken des SJT_kurz.....	139
4.4.4 Itemanalyse: Trennschärfe und Schwierigkeit.....	139
4.4.5 Reliabilitätsanalyse.....	144
4.4.6 Validitätsanalyse	146
4.4.7 Zusammenhänge mit demographischen Variablen	150
4.5 Diskussion	152
4.6 Ausblick.....	159
5. Die dritte Studie	160
5.1 Beschreibung der eingesetzten Instrumente und Fragebögen.....	160
5.1.1 Der SJT: Video - vs. Paper-Pencil	160
5.1.2 Skala zur Messung der sozialen Validität und Akzeptanz des SJT	166
5.1.3 Demographische Variablen.....	168
5.1.4 Die Grundintelligenztest Skala 2-Revision.....	168
5.1.5 Der Wortschatztest	170
5.1.6 Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung	171
5.2 Ablauf der Datenerhebung und Beschreibung der Stichprobe	177
5.2.1 Ablauf der Datenerhebung	177
5.2.2 Beschreibung der Stichprobe	180

5.3 Deskriptive Statistik	181
5.3.1 Fluide Intelligenz	181
5.3.2 Kristalline Intelligenz	182
5.3.3 Soziale Validität und Akzeptanz	183
5.3.4 Zentrale Dimensionen der berufsbezogenen Persönlichkeit	185
5.4 Inferenzstatistische Analysen und Hypothesentestung.....	186
5.4.1 Exploration der faktoriellen Struktur des SJT	186
5.4.2 Zentrale Statistiken des SJT_Hotelfach.....	189
5.4.3 Itemanalyse: Trennschärfe und Schwierigkeit	193
5.4.4 Reliabilitätsanalyse	195
5.4.5 Validitätsanalyse.....	196
5.4.6 Zusammenhänge mit demographischen Variablen	203
5.4.7 Erweiterung des SJT_Hotelfach	204
5.5 Diskussion	207
5.6 Ausblick	219
6. Die vierte Studie.....	220
6.1 Beschreibung der eingesetzten Instrumente und Fragebögen	220
6.1.1 Der eingesetzte SJT	220
6.1.2 Demographische Variablen	221
6.1.3 Das Neo-Fünf-Faktoren-Inventar nach Costa und McCrae	222
6.2 Ablauf der Datenerhebung und Beschreibung der Stichprobe ...	225
6.2.1 Ablauf der Datenerhebung.....	225
6.2.2 Beschreibung der Stichprobe.....	226
6.3 Deskriptive Statistik	228
6.3.1 Die Skala SJT_Hotelfach.....	228
6.3.2 Fünf globale Persönlichkeitsfaktoren	230
6.4 Inferenzstatistische Analysen und Hypothesentestung.....	231
6.4.1 Reliabilitätsanalyse	231
6.4.2 Itemanalyse: Trennschärfe und Itemschwierigkeit.....	232
6.4.3 Validitätsanalyse.....	234
6.4.4 Weitere Ergebnisse	237
6.5 Diskussion	239

7. Generelle Diskussion.....	243
7.1 Zusammenfassung der Studien.....	243
7.2 Implikationen für die weitere Forschung und Praxis.....	253
7.2.1 Implikationen für die weitere Forschung	253
7.2.2 Implikationen für die Praxis.....	256
8. Literaturverzeichnis.....	258
Abbildungsverzeichnis.....	295
Tabellenverzeichnis	297
Anhang A: Erste Studie.....	300
Anhang B: Zweite Studie	314
Anhang C: Dritte Studie.....	324
Anhang D: Vierte Studie.....	355

1. Einleitung

Situational Judgment Tests (SJTs) versetzen den Probanden¹ in realistische, auf einen bestimmten Arbeitsbereich zugeschnittene Situationen (Weekley & Ployhart, 2006). Nach Kanning, Grewe, Hollenberg und Hadouch (2006) entspringen SJTs ähnlich wie das Assessment Center (AC) oder das Situative Interview (SI) dem Konstruktionsprinzip der Simulation. (Komplexe) Simulationen streben dabei eine Erfassung der Leistung und des Verhaltens von Probanden bei der Bewältigung anspruchsvoller und mehrdeutiger beruflicher Entscheidungssituationen an. Sie bilden berufliche Anforderungen nach, um Leistungs- und Verhaltensmaße als Prädiktoren für den Berufserfolg am realen Arbeitsplatz ermitteln zu können (Strauß & Kleinmann, 1995). Dabei können die Situationen, welche ein SJT enthält, entweder im Paper-Pencil-Format oder aber im Video-Format (z. B. Dalessio, 1994; Kanning, 2005) dargestellt werden. An die Präsentation der arbeitsplatzspezifischen Situationen anknüpfend werden den Probanden unterschiedliche Verhaltensalternativen dargeboten. Bei der Auseinandersetzung mit diesen Verhaltensoptionen werden die Probanden aufgefordert „[...]to select or rate the potential responses for likelihood of performing the action or effectiveness of the action“ (Whetzel, McDaniel & Nguyen, 2008; S. 292). Die von den Probanden eingeholten Bewertungen werden mit einem Bewertungsschlüssel, welcher vorab in Zusammenarbeit mit Arbeitsplatzexperten entworfen wurde, abgeglichen und verrechnet (Bergman, Drasgow, Donovan & Henning, 2006).

In den USA erfreuen sich SJTs seit rund 80 Jahren eines regelmäßigen Einsatzes. Dagegen beginnt in Europa erst langsam eine zunehmende Auseinandersetzung mit SJTs (z. B. Salgado, 2001; Behrmann, 2007). Das wachsende Interesse an SJTs sowohl in der Forschung als auch in der Praxis ist nicht zuletzt auf zwei Meta-Analysen von McDaniel, Finnegan, Morgeson, Campion und Braverman (2001) respektive McDaniel, Whetzel, Hartman, Nguyen und Grubb (2007) zurückzuführen, welche übereinstimmend den Nachweis überzeugender kriterienbezogener Validitäten erbracht haben. Ein Zitat von Landry (2007) bringt weitere Vorzüge von

¹ Zugunsten einer besseren Lesbarkeit wird in der vorliegenden Arbeit ausschließlich das grammatikalische Maskulinum verwendet. Gemeint sind jedoch stets beide Geschlechter.

SJTs auf den Punkt: „*They seem to represent psychometric alchemy (adverse impact is down, validity is up), they seem to assess practically important KSAOs [knowledge, skills, abilities, and other characteristics], and assesses like them*” (S. 418).

Im Rahmen der vorliegenden Arbeit soll ein SJT zur Messung der Kundenorientierung im Hotelfach entwickelt und ersten Validierungen unterzogen werden. Der SJT soll bei erfolgreichem Nachweis seiner Nützlichkeit zur Auswahl neuer Auszubildender in diesem Arbeitsbereich verwandt werden. Eine Konstruktion eines SJT, welcher die speziellen Anforderungen und Situationen im Hotelfach erfasst und abbildet, bietet sich aufgrund folgender Konstellationen an: Nach Angaben des Deutschen Hotel- und Gaststättenverbandes DEHOGA waren im Jahre 2008 mehr als 1 Million Menschen in dieser Branche beschäftigt. Einer Statistik des Bundesinstitutes für Berufsbildung BIBB zufolge befanden sich in demselben Jahr rund 36.000 Personen in einer Ausbildung im Hotelfach. Der Kundenorientierung von Mitarbeitern kommt hervorgehobene Bedeutung bei der Bewährung eines Unternehmens im Wettbewerb zu, da sie in beträchtlichem Maße das Ausmaß der Kundenzufriedenheit, der Kundenloyalität und damit des Umsatzes eines Unternehmens beeinflusst (Nerdinger, 2003).

2. Theoretischer Hintergrund

In dem zweiten Kapitel dieser Arbeit erfolgt zunächst eine Darstellung der wichtigsten Forschungsergebnisse in Bezug auf SJTs (Abschnitt 2.1 – Abschnitt 2.5). Zunächst wird die historische Entwicklung von SJTs skizziert (2.1), bevor diese Testgattung mit anderen Verfahren der Berufseignungsdiagnostik verglichen wird (2.2). Im darauf folgenden Abschnitt erfolgt eine Vorstellung der zentralen Schritte bei der Konstruktion von SJTs (2.3). Der vierte Abschnitt dieses Kapitels rückt die Besonderheit videobasierter SJTs in den Vordergrund (2.4). Einen Abschluss erhält der theoretische Überblick über SJTs in Form einer Auseinandersetzung mit dem Nutzen von SJTs für die Personalauswahl (2.5). Der letzte Abschnitt dieses Kapitels widmet sich wiederum dem Punkt, welche Bedeutung der Kundenorientierung von Mitarbeitern für ein Unternehmen zukommt (2.6).

2.1 Die historische Entwicklung und Bedeutung von SJTs

Situational Judgment Tests (SJTs) beziehungsweise Tests, die explizit als SJTs deklariert sind, werden seit rund 60 Jahren in den USA im Rahmen der Berufseignungsdiagnostik eingesetzt. Die Anfänge situativer Aufgaben liegen jedoch zeitlich weiter zurück. So berichtete DuBois (1970) in seiner Übersichtsarbeit zur Geschichte des psychologischen Testens in den USA zum Beispiel, dass bereits 1873 Beschäftigte im öffentlichen Dienst mit Aufgaben konfrontiert wurden, in denen ihnen eine Urteilsbildung in verschiedenen berufsrelevanten Situationen abverlangt wurde.

In den 20ern des zwanzigsten Jahrhunderts erschien mit dem George Washington Social Intelligence Test (GWSIT; z. B. Moss, 1926; Hunt, 1928) ein Test, welchem eine besondere Bedeutung in der geschichtlichen Entwicklung von SJTs zukommt (McDaniel & Nguyen, 2001). Das Konstrukt soziale Intelligenz (social intelligence), welches als Namenspatron für diesen Test fungierte, lässt sich zurückführen auf eine Taxonomie von Thorndike (1920), welcher zwischen drei Facetten der Intelligenz unterschied: Der Fähigkeit, Ideen nachzuvollziehen und selber zu entwickeln (abstract intelligence), der Fähigkeit, Objekteigenschaften zu verstehen und Objekte einsetzen zu können (mechanical intelligence) sowie drittens

der Fähigkeit, andere Personen zu verstehen und mit ihnen umzugehen (social intelligence). Das Konzept der sozialen Intelligenz bietet im Übrigen verschiedene Überschneidungspunkte mit dem zeitgenössischen Konzept der sozialen Kompetenz (vgl. Kanning, 2009; s. Kapitel 4).

Der GWSIT bestand aus einer Vielzahl an Subtests zur Abbildung unterschiedlicher Fähigkeiten, zum Beispiel Judgment in Social Situations, Memory for Names and Faces oder Recognition of Mental States from Facial Expression. Auch wenn in den verschiedenen revidierten Versionen des Tests manche der Subtests ergänzt beziehungsweise eliminiert wurden (Kihlstrom & Cantor, 2000), ist der für diese Arbeit relevante Subtest Judgment in Social Situations Teil aller Versionen des GWSIT. Diesem Subtest wird von Whetzel und McDaniel (2009) bescheinigt, den ersten SJT mit vorgegebenen Verhaltensalternativen im Multiple-Choice Format darzustellen. Mit situativen Items, welche verschiedene Situationen am Arbeitsplatz widerspiegeln, wies dieser Subtest des GWSIT bereits ein weiteres zentrales Merkmal heutiger SJTs auf (vgl. Weekley & Ployhart, 2006). Nach Moss (1926; S. 26) erforderte eine erfolgreiche Bewältigung dieses Subtests „[...]keen judgment, and a deep appreciation of human motives“. Zur Validierung des GWSIT zog Hunt (1928) verschiedene – vom heutigen Standpunkt exotische – Kriterien heran, zum Beispiel die Anzahl extracurricularer Aktivitäten bei Studenten oder Einschätzungen von Führungskräften, wie ihre Mitarbeiter mit anderen Personen umgehen. Da der GWSIT mit Skalen zur Messung der abstract intelligence (s. o.) weitaus höhere Korrelationen erzielte und mehr gemeinsame Varianz teilte als mit den von Hunt (1928) antizipierten sozialen Fähigkeiten, wurde dem GWSIT die Fähigkeit abgesprochen, ein singuläres Konstrukt wie social intelligence zu messen (Thorndike & Stein, 1937).

Ausgehend von den 40ern bis in die 60er des zwanzigsten Jahrhunderts wurde eine Vielzahl an SJTs mit dem primären Ziel entwickelt, Einblicke in die Führungs- und Managementkompetenz der Probanden zu erhalten (s. für einen Überblick McDaniel & Ngyuen, 2001). Im Folgenden sollen verschiedene Tests dieser Ära kurz vorgestellt werden. Ausgeklammert werden solche SJTs, welche dem militärischen Bereich vorbehalten waren (vgl. Northrop, 1989).

Auf Cardall (1942a, 1942b) geht der Practical Judgment Test (PJT) zurück. Wie beim GWSIT liegen die den Probanden zur Verfügung gestell-

ten Verhaltensalternativen im Multiple-Choice-Format vor. Cardall (1942a; S. 1) verknüpfte mit der Entwicklung des PJT die Intention „[...]to measure the element of practical judgment as it operates in everyday business and social situations“. Um den Begriff practical judgment auf theoretischer Ebene zu klären, bat er verschiedene Mitarbeiter aus Personalabteilungen und Psychologen um ihre Vorstellungen zu diesem Begriff und gelangte zu folgendem Konsens (Cardall, 1942a, S. 1):

That practical judgment consists of the recognition of possible alternatives of action and the ability to select the best. It may be observed that in ordinary situations a number of alternatives spring instantly to mind, the element of judgment operating when one course of action is chosen over others. Upon this rationale the present instrument was constructed.

Ferner erhob Cardall (1942a, 1942b) mit dem PJT den Anspruch, ein Konstrukt zu messen, welches auf statistischer Ebene unabhängig wäre von der allgemeinen kognitiven Leistungsfähigkeit oder Statusvariablen wie dem akademischen Hintergrund der Probanden. Validitätskoeffizienten zum Zusammenhang zwischen dem PJT und verschiedenen Persönlichkeitsfaktoren sowie Maßen der Intelligenz wurden von Hogadone und Smith (1947) beziehungsweise Carrington (1949) geliefert.

Der *How Supervise?* (HS) wurde von File (1945) mit dem Ziel entwickelt, das Wissen und die Einsicht von Führungskräften bezüglich sozialer Interaktionen am Arbeitsplatz zu messen. File (1945) sprach in diesem Zusammenhang von einem Generalfaktor der Führungskompetenz. Jedes Item des HS bestand aus einer Verhaltensbeschreibung in Form weniger Sätze und drei dazugehörigen Antwortoptionen für den Probanden: Zustimmung, Ablehnung oder Unentschlossenheit. Rosen (1961) reflektierte und bündelte in seinem Artikel den Forschungsstand zur Validität des HS bis 1960. Als positive Evidenz für die Einsatzberechtigung des HS führte Rosen (1961) exemplarisch die Studien von Canter (1951) und Carter (1952) an. Rosen (1961) listete im Gegenzug auch verschiedene Studien auf, welche negative Evidenz für die psychometrische Güte des HS erbrachten (z. B. Sartain, 1946). Bei einer abschließenden Bewertung und Abwägung der Ergebnisse gelangte Rosen zu der Aussage, dass die überwiegend positive empirische Evidenz einen Einsatz des HS in der Berufs-

eignungsdiagnostik legitimiere. In nachfolgenden Studien (Millard, 1952; File & Remmers, 1971) wurden mehrere signifikante Korrelationen zwischen dem HS und Maßen der allgemeinen kognitiven Leistungsfähigkeit vorgefunden. Auf Basis seiner Ergebnisse schlussfolgerte Millard (1952) dementsprechend, der HS stelle im Grunde einen Test zur Messung der kognitiven Leistungsfähigkeit dar und griff damit – wenn der HS als Vorläufer heutiger SJTs angesehen wird – einer späteren Debatte, welche Konstrukte SJTs vorrangig messen, vor (vgl. Abschnitt 2.5).

Rund 15 Jahre nach der ersten Version des HS folgten Bruce und Learner (1958) mit der Veröffentlichung eines weiteren SJT, dem *Supervisory Practices Test* (SPT). Der SPT sollte dazu genutzt werden, Führungskräfte zu identifizieren, welche effektive Entscheidungen in arbeitsplatzbezogenen Situationen treffen können.

Zur Validierung des SPT zogen Bruce und Learner (1958) unter anderem den HS und mit dem Primary Mental Abilities Test (vgl. Thurstone, 1938) einen Test zur Messung der allgemeinen kognitiven Leistungsfähigkeit heran. In Bezug auf den Gesamtwert des HS ergab sich eine signifikante Korrelation von .56. Für die Subtests des Primary Mental Abilities Tests ergaben sich Korrelationen zwischen .18 (Subtest: Word fluency) und .35 (Subtest: Space). Die Korrelation zwischen dem SPT und dem Gesamtwert des Primary Mental Abilities Test erreichte einen Wert von $r = .27$. Leider geht aus der Studie von Bruce und Learner (1958) nicht hervor, ob die angegebenen Korrelationen zwischen dem SPT und dem Primary Mental Abilities Test signifikant waren.

Zusammengefasst ist für die bisher vorgestellten Verfahren erstens festzuhalten, dass Daten zur Ermittlung von prognostischen Validitäten fehlen. Dies kann darauf zurückgeführt werden, dass die bis dato vorgestellten SJTs vorrangig als Instrumente der Personalentwicklung respektive Potenzialanalyse eingesetzt wurden, um die Eignung der Probanden für Führungsaufgaben (supervisory potential) zu messen (Whetzel & McDaniel, 2009). Zweitens ist weiterhin anzumerken, dass bei vielen Validitätsstudien sehr kleine Stichprobengrößen vorlagen, zum Beispiel dreizehn Probanden bei der oben angeführten Studie von Canter (1951). Zudem sind viele der Kriterien, welche zur Validierung dienten, sehr kritisch in Bezug auf ihre Passung zu heutigen Standards psychologischen Testens,

beispielsweise den Standards for Educational and Psychological Testing der APA (1999) oder der DIN 33430 (2002), zu hinterfragen.

McDaniel et al. (2001) berichteten, dass in den späten 50ern beziehungsweise frühen 60ern des zwanzigsten Jahrhunderts amerikanische Unternehmen SJTs erstmals systematisch in der Personalauswahl eingesetzt haben. Veröffentlichte Studien aus diesem Bereich liegen jedoch nicht vor.

Aufbauend auf einer Studie von Motowidlo, Dunette und Carter (1990) erwuchs ein bis heute anhaltendes Forschungsinteresse an SJTs. Motowidlo et al. (1990) prägten die Unterscheidung zwischen *low fidelity*²-Verfahren wie SJTs auf der einen Seite und Verfahren wie dem AC oder der Arbeitsprobe als *high fidelity*-Verfahren auf der anderen Seite. Um die unterschiedlichen Verfahren gemäß ihrer fidelity einzuordnen, griff das Team um Motowidlo auf ein Kontinuum zurück. Dem oberen Ende des Kontinuums ordneten sie solche Verfahren zu, welche den Probanden in realistische, arbeitsplatzbezogene Situationen hineinversetzen und ihm konkretes Verhalten abverlangen (vgl. Kanning, Grewe, Hollenberg & Hadouch, 2006). Exemplarisch sei ein Rollenspiel aus einem AC genannt, in dem ein Bewerber sich in einem Mitarbeitergespräch wieder findet und Entscheidungen zu treffen oder Lösungen zu finden hat. Das andere Ende des von Motowidlo et al. (1990) aufgestellten Kontinuums wird gebildet durch SJTs, welche Situationen aus dem Berufsalltag abstrakter repräsentieren, zum Beispiel in Form schriftlicher Beschreibungen. Im Gegensatz zu den high fidelity-Verfahren gestatten diese Verfahren nach Ansicht des amerikanischen Forscherteams *nicht*, tatsächliches Verhalten der Probanden zu beobachten. Die Probanden werden viel mehr gebeten zu beschreiben, wie sie sich in der jeweiligen Situation verhalten *würden*³.

Da das Vorgehen von Motowidlo et al. (1990) bis in die heutige Zeit als nahezu idealtypisch für eine wissenschaftliche fundierte Herangehensweise bei der Konstruktion von SJTs angesehen wird (vgl. Ployhart & Erhart,

² Für das englische Wort fidelity existiert nach Sicht des Verfassers kein passendes Äquivalent in der deutschen Sprache. Als Annäherung an den Sinngehalt von fidelity ist der Begriff Wiedergabetreue am besten geeignet. Im weiteren Text wird jedoch weiterhin der englische Begriff benutzt.

³ Im späteren Verlauf dieser Arbeit wird erörtert, dass sich SJTs durchaus sowohl hinsichtlich der Darstellung von Situationen als auch der Art, wie Probanden reagieren können, unterscheiden. Im Rahmen eines historischen Überblicks soll diese verknäppte Darstellung jedoch genügen.

2003; Bledow & Frese, 2009), erfährt diese Studie eine ausführlichere Vorstellung. Im Gegensatz zu den weiter oben beschriebenen Verfahren beabsichtigten Motowidlo et al. (1990) explizit nicht die Erfassung eines spezifischen, singulären Konstrukts, sondern vielmehr die repräsentative Abbildung eines eng umschriebenen Tätigkeitsfeldes. Zudem stützten sie sich in Abgrenzung zu den zuvor genannten zumeist rational-intuitiv entwickelten SJTs auf einen aufwändigen empirischen Konstruktionsprozess, welcher im Folgenden schematisch beschrieben werden soll.

Die Konstruktion des SJT wurde eingeläutet mit einer Sammlung kritischer Ereignisse an einer Stichprobe von Mitarbeitern ohne Führungsaufgaben und Führungskräften aus allen beteiligten Kooperationsunternehmen. Die resultierenden Situationen wurden in einem weiteren Schritt unterschiedlichen Inhaltsbereichen zugeordnet. Zudem wurden verschiedene Führungskräfte aus den partizipierenden Unternehmen gebeten, schriftlich festzuhalten, wie sie sich in den gesammelten Situationen verhalten würden. Nach Sichtung der Ergebnisse ordneten die Autoren jeder Situation fünf bis sieben allgemeine Verhaltensweisen zu. Die Situationen und die dazugehörigen Verhaltensalternativen wurden in einem weiteren Konstruktionsschritt Personen mit langjähriger Berufserfahrung im Managementbereich vorgelegt. Diese sollten einerseits alle Verhaltensweisen hinsichtlich ihrer Effektivität in der jeweiligen Situation einstufen und andererseits je Situation die aus ihrer Sicht beste beziehungsweise schlechteste Verhaltensweise benennen. Auf diesem Wege schieden verschiedene Situationen und Verhaltensweisen aus dem SJT aus. In der Endfassung des SJT befanden sich letztendlich 55 Items.

Bei der Bearbeitung der situativen Aufgaben forderten Motowidlo et al. (1990) ihre Probanden auf, von fünf verschiedenen Verhaltensalternativen sowohl die Verhaltensalternative, die sie am wahrscheinlichsten (*most likely*) in der beschriebenen Situation zeigen würden, als auch die Verhaltensalternative, welche sie am allerwenigsten (*least likely*) zeigen würden, auszuwählen. Als Richtwert, welche Verhaltensalternativen in den Situationen tatsächlich am effektivsten beziehungsweise ineffektivsten sind, dienten die während der Entwicklung des SJT gesammelten Einschätzungen erfahrener Experten. Je nach Situation wurden von den Experten eine oder mehrere Verhaltensalternativen als effektivste respektive als ineffektivste bestimmt.

Für die Bewertung (*Scoring*) der Probandenantworten wählten die Autoren folgenden Schlüssel: Der Proband erhielt 1 Punkt, wenn er eine der Verhaltensalternativen am wahrscheinlichsten zeigen würde, welche auch von den Experten als besonders effektiv eingestuft wurde. Jedoch erhielt der Proband -1 Punkt, wenn er als wahrscheinlichste Verhaltensalternative eine von denjenigen Alternativen gewählt hat, welche Experten als besonders ineffektiv eingestuft haben. 0 Punkte ergaben sich, wenn die Verhaltensalternative, welche der Proband am wahrscheinlichsten zeigen würde, von den Experten weder als besonders effektiv noch als besonders ineffektiv eingeschätzt wurde. Das Scoring der least likely-Antworten der Probanden wurde nach demselben Muster vorgenommen. Mit einem klar definierten Bewertungsschlüssel dieser Art grenzten sich Motowidlo et al. (1990) auf eine weitere Art und Weise deutlich von den oben beschriebenen früheren SJTs ab.

Als Kriterien für die drei Validierungsstudien wählten die Autoren unter anderem Einzelbeurteilungen in verschiedenen Leistungsbereichen des Managements sowie eine Gesamtbeurteilung der Leistung durch Vorgesetzte. Für die Validierung des SJT griffen die Forscher um Motowidlo auf verschiedene Substichproben von Personen im Bewerberpool⁴ der Unternehmen, Mitarbeitern unterschiedlicher Positionen in der Probezeit und Mitarbeitern, welche erst vor kurzem auf Führungspositionen gewechselt waren, zurück. Es ergaben sich Validitätswerte zwischen 20 und 40 zwischen dem SJT und den unterschiedlichen Maßen der Arbeitsleistung. Des Weiteren fanden sich statistisch nicht bedeutsame Korrelationen mit verschiedenen Messungen der Intelligenz. Motowidlo et al. (1990) interpretierten die von ihnen erzielten Ergebnisse als Beweis dafür, dass auch mit low fidelity-Verfahren Validitäten zu erzielen sind, welche einen Einsatz dieser Verfahren zur Personalauswahl gestatten. Ferner billigten sie low fidelity-Verfahren mit der ihrer Ansicht nach kostengünstigeren Entwicklung und ökonomischeren Anwendbarkeit einen Vorteil gegenüber high fidelity-Verfahren zu.

An der Studie des Teams um Motowidlo sind einige Punkte kritisch zu beleuchten. Zunächst ist anzumerken, dass keine prognostischen Validitäten vorliegen, sondern nur konkurrente Validitäten. Gerade vor dem Hin-

⁴ Diese Personengruppe ist nicht mit realen Bewerbern zu verwechseln (s. u.).

tergrund, dass dieser Test als Instrument der Berufseignungsdiagnostik eingesetzt werden sollte, fehlen aufgrund dieses Mangels wichtige Erkenntnisse. Ein zweiter Kritikpunkt sind die teils geringen Stichprobengrößen, welche verallgemeinernde oder repräsentative Aussagen nicht zulassen. Drittens und abschließend ist zu bemerken, dass Probanden aus einzelnen Validitätsstudien aufgrund ihrer Ergebnisse in einem Test zur allgemeinen kognitiven Leistungsfähigkeit und ihrer akademischen Herkunft eingestellt wurden, so dass eine Varianzeinschränkung bezüglich dieser Variablen zu vermuten ist (Whetzel & McDaniel, 2009). Die berichteten statistischen Unabhängigkeiten des SJT, im Sinne von Korrelationen nahe Null, von allgemeiner kognitiver Leistungsfähigkeit und akademischer Leistung sind dementsprechend diskussionswürdig.

Ungeachtet der genannten methodischen Einwände bahnte die Studie von Motowidlo et al. (1990) wie bereits oben erwähnt eine Revitalisierung des Forschungsinteresses an SJTs. Zum Abschluss dieses Rückblicks in die Historie von SJTs, wird in der Abbildung 2-1 ein Überblick in Form einer Zeittafel, welche die wichtigsten Entwicklungen aufgreift, geliefert.

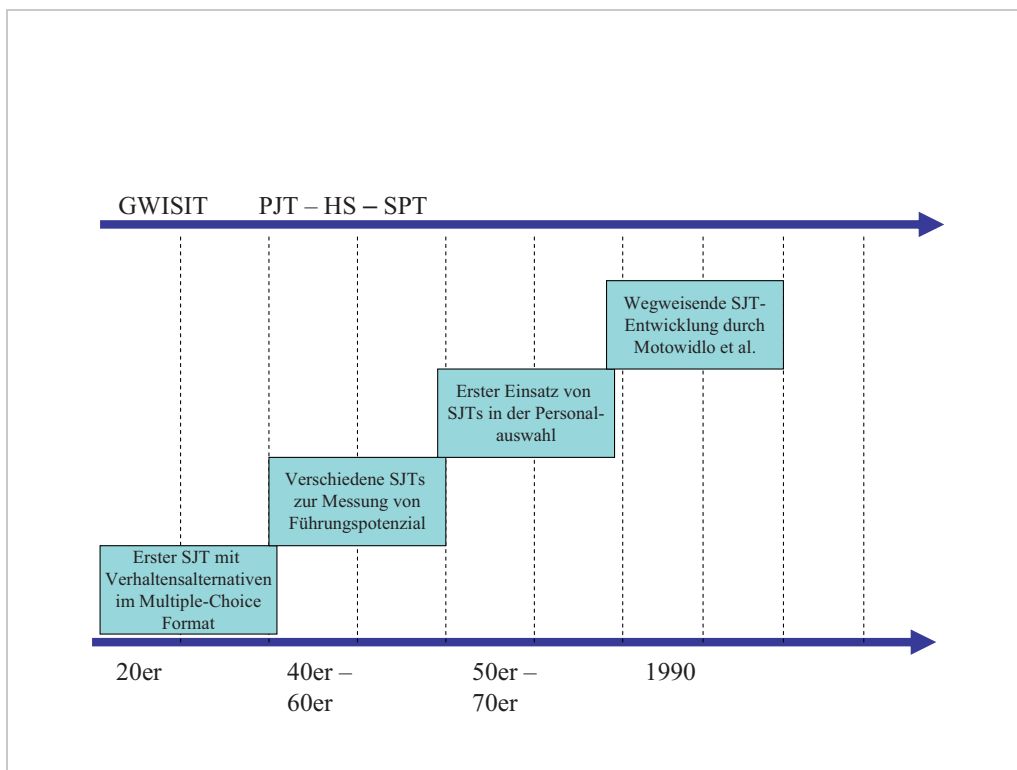


Abbildung 2-1: Zentrale Punkte in der Historie von SJTs

2.2 SJTs als simulationsorientierte Verfahren der Berufseignungsdiagnostik

Bei der Nennung von Verfahren, welche in der Berufseignungsdiagnostik eingesetzt werden, unterschieden Schuler und Höft (2006) drei unterschiedliche Verfahrensarten: Konstruktorientierte Verfahren, biographieorientierte Verfahren und simulationsorientierte Verfahren. Das Ziel dieses Abschnitts ist eine kurze Charakterisierung aller drei genannten Verfahrensarten sowie eine Einordnung von SJTs in die von Schuler und Höft (2006) entworfene Taxonomie.

Konstruktorientierte Verfahren werden zur Erfassung von als zeitlich stabil und konsistent angesehen Merkmalen herangezogen, welche direkt aus vorhergehenden Arbeits- und Anforderungsanalysen abgeleitet wurden (Schuler & Höft, 2006). Dabei werden die Merkmale unabhängig von einem spezifischen beruflichen Kontext oder ohne Zuschnitt auf einen spezifischen Arbeitsplatz mit Hilfe klassisch testtheoretisch konzipierter Verfahren wie Intelligenztests oder Persönlichkeitstests gemessen (Schuler & Höft, 2004).

Verfahren dieser Art beruhen auf dem *sign-Ansatz* (Wernimont & Campbell, 1968), nach welchem die Leistungen von Probanden als Zeichen (signs) oder Indikatoren homogener und theoretisch abgesicherter Prädispositionen ausgelegt werden. So wird beispielsweise das Ergebnis in einem Test zur Erfassung allgemeiner kognitiver Leistungsfähigkeit mit der individuellen Ausprägung der Intelligenz gleichgesetzt.

Zwecks des Einsatzes *biographieorientierter Verfahren* soll der Schluss von vergangenem auf zukünftiges Verhalten vorgenommen werden, denn „[...]the best indicator of future performance is past performance“ (Wernimont & Campbell, 1968; S. 372). Das Verhalten einer Person wird als zeitlich stabil und invariant gedeutet. Jedoch wird anders als bei konstruktorientierten Verfahren ohne den „Umweg über ein Konstrukt“ (Schuler & Höft, 2004; S. 311) auf zukünftiges Verhalten geschlossen.

Schuler und Marcus (2006) nennen als Vertreter dieser Verfahrensart den biographischen Fragebogen, die Sichtung von Bewerbungsunterlagen oder Referenzen sowie biographische Fragen im Interview.

SJTs sind den *simulationsorientierten Verfahren* der Personalauswahl zuzurechnen (vgl. Kapitel 1). Ein Zitat von O`Connell, Hartman, McDan-

iel, Grubb und Lawrence (2007; S. 19) soll diese Zugehörigkeit verdeutlichen: „*Situational judgment tests (SJT) are simulations requiring the respondent to exercise judgment when responding to hypothetical problem situations that occur in work settings*“. Nach Höft und Funke (2006) basieren simulationsorientierte Verfahren auf „[...]der Idee einer direkten „Punkt-zu-Punkt-Übereinstimmung“ zwischen dem Inhalt des diagnostischen Verfahrens und dem zu prognostizierenden Leistungsbereich“ (S. 146). Dies entspricht dem *sample-Ansatz* von Wernimont und Campbell (1968), gemäß welchem im Personalauswahlverfahren eine repräsentative Auswahl realer beruflicher Situationen abgebildet werden sollte, in denen sich die Bewerber behaupten müssen. Bewerber werden mit komplexeren und vielfältigeren Ereignissen konfrontiert als in konstruktorientierten Verfahren (Moser, 2004). Wernimont und Campbell (1968) favorisierten die Nützlichkeit des *sample-Ansatzes* bei der Vorhersage beruflicher Leistung gegenüber dem *sign-Ansatz*, denn „[...]it will be much more fruitful to focus on meaningful samples of behavior, rather than signs of predispositions, as predictors of later performance“ (S. 372). Der Rückschluss von einer Verhaltensstichprobe auf entsprechendes zukünftiges Verhalten sollte umso valider sein, je stärker Prädiktor und Kriterium inhaltlich korrespondieren (Schuler, 2000). Bei der Diskussion des Einsatzes simulationsorientierter Verfahren in Organisationen listete Moser (2004; S. 105 ff.) verschiedene mögliche Probleme bei der Durchführung auf. Exemplarisch seien eine mögliche Überforderung der Bewerber, eine partielle Künstlichkeit und fehlende Realitätsnähe der Situation sowie mögliche Vorerfahrungen der Bewerber mit dem Verfahrensinhalt als Beeinträchtigung der Validität genannt.

Neben SJTs gibt es verschiedene andere Verfahren der Berufseignungsdiagnostik, welche situative Elemente in sich tragen: ACs (z. B. Thornton & Byham, 1982), Arbeitsproben (z. B. Asher & Sciarrino, 1974) sowie situative Interviews (z. B. Latham, Saari, Pursell, & Campion, 1980). Aufgrund ihrer konzeptionellen und methodischen Nähe zu SJTs sollen im weiteren Verlauf dieses Abschnitts das situative Interview und die Arbeitsprobe detaillierter vorgestellt werden.

Das *situative Interview* (SI) wurde als Form des strukturierten Einstellungsinterviews von Latham, Saari, Pursell und Campion (1980) in die Berufseignungsdiagnostik eingeführt. Laut Schuler (2000) entspricht das

SI einer „*mentalen Tätigkeitssimulation*“ beziehungsweise „*Wissensarbeitsprobe*“ (S. 88). Als theoretisches Fundament dienten die Arbeiten von Locke (1968) respektive Locke und Latham (z. B. 1990) zur Zielsetzung (goal-setting), nach denen Ziele und Intentionen als direkte Antezedenten und Prädiktoren realen Verhaltens anzusehen sind (Kleinbeck, 1991). Diesem Ansatz Rechnung tragend werden die Probanden im SI typischerweise gefragt, wie sie sich in der vom Interviewer geschilderten Situation verhalten würden. Die im SI eingesetzten Situationen werden mit Hilfe der CIT (vgl. Flanagan, 1954) gewonnen und enthalten Dilemmata, was die Auftretenswahrscheinlichkeit sozial erwünschter Antworten minimieren soll (Latham & Sue-Chan, 1996; Maurer, Sue-Chan, & Latham, 1999). Allen Probanden werden im Rahmen des SI dieselben Fragen in gleicher Reihenfolge gestellt. Als konzeptionelle Besonderheit des SI wird ein Bewertungsschlüssel für die Antworten der Probanden angesehen (Maurer, 2002), welcher sich an Unternehmenswerten oder der Unternehmenskultur orientieren sollte (Latham & Sue-Chan, 1999). In der Praxis dient der Bewertungsschlüssel zur Konstruktion verhaltensverankerter Einstufungsskalen, mit denen Probandenantworten verglichen und nachfolgend bewertet werden (Schuler, 2000; Schuler & Marcus, 2006).

Aus einer Meta-Analyse von Latham und Sue-Chan (1999) ist eine mittlere korrigierte kriterienbezogene Validität des SI von .47 zu entnehmen. Die beiden Autoren bezogen 20 Studien mit insgesamt 1010 Probanden in ihre Analyse ein. Eine drei Jahre später erfolgende Meta-Analyse von Taylor und Small (2002) beinhaltet die Ergebnisse von 30 Studien und 2299 Probanden zum Zusammenhang zwischen dem SI und Arbeitsleistung. Die beiden Autoren ermittelten eine mittlere korrigierte Validität von .45 und damit einen vergleichbaren Wert wie Latham und Sue-Chan (1999).

Obgleich in Form verschiedener Meta-Analysen der empirische Nachweis für eine kriterienbezogene Validität von SIs erbracht wurde, fehlen bislang Studien, welche statistisch abgesicherte Aussagen über die Konstruktvalidität von SIs erlauben (Conway & Peneno, 1999; Huffcutt, Weekley, Wiesner, Degroot, & Jones, 2001; Van Iddekinge, Raymark, Eidson & Attenweiler, 2004). Damit leidet das SI ähnlich wie zum Beispiel das Assessment Center unter dem sogenannten Validitätsparadoxon (Arthur, Woehr, & Maldegen, 2000).

Das SI weist von den oben genannten simulationsorientierten Verfahren die meisten Gemeinsamkeiten mit SJTs auf. Beide Verfahren streben eine realistische und repräsentative Abbildung eines bestimmten Berufsfeldes an. Dabei setzen sowohl SI als auch SJTs auf die Erfassung von Verhaltenstendenzen respektive -absichten. Die Antworten der Probanden werden in beiden Verfahrensklassen mit vorab festgelegten Bewertungsschlüsseln verglichen, welche zusammen mit Angehörigen von Unternehmen beziehungsweise Personen mit anerkanntem Expertenstatus erstellt wurden. Weiterhin teilen sich SI und SJT die ungeklärte Frage, welche Eigenschaften diese beiden Verfahren eigentlich zu messen vermögen (s. Abschnitt 2.5.2). Hinsichtlich des Antwortformates für die Probanden besteht jedoch auch ein nicht zu verachtender Unterschied zwischen SI und SJT. Während in einem SI Antworten durch die Probanden frei und offen generiert werden, liegen in einem SJT mehrere vom Testautor vorgegebene Antwortmöglichkeiten beziehungsweise Verhaltensalternativen vor.

Eine aktuelle Definition von *Arbeitsproben* stammt von Ployhart, Schneider und Schmitt (2006): „[...]a work sample test is a test in which the applicant performs a selected set of actual tasks that are physically and/or psychologically similar to those performed on the job” (S. 538). Auf Schuler (2000) geht eine inhaltlich entsprechende Auslegung für den deutschsprachigen Raum zurück: „Unter Arbeitsproben werden standardisierte Aufgaben verstanden, die inhaltlich und valide erkennbar äquivalente Stichproben des erfolgsrelevanten beruflichen Verhaltens provozieren” (S. 115). Guion (1998) betonte als ein definierendes Charakteristikum von Arbeitsproben, dass diese einen hohen Grad an fidelity aufweisen. Nach Heneman und Judge (2003) sind Arbeitsproben strikt zu trennen von Praktika oder Probezeiten, da sie ein sehr viel engeres Zeitfenster besitzen und ein enger umgrenztes Tätigkeitsfeld abbilden. Beispielhafte Arbeitsproben wären das Erstellen eines Dokuments im Arbeitsbereich einer Sekretärin oder eines Möbelstücks im Beruf des Schreiners sowie die Bewährung eines Piloten in einem Flugsimulator. Die gewählten Inhalte der Arbeitsproben entspringen Anforderungsanalysen (Ployhart et al., 2006).

Hunter und Hunter (1984) berichteten als ein Ergebnis ihrer Meta-Analyse eine mittlere korrigierte Validität von .54 für Arbeitsproben (vgl.

Schmidt & Hunter, 1998). Als Kriterium fungierte die Leistungsbeurteilung durch Vorgesetzte. Im Rahmen ihrer Studie widmeten sich Hunter und Hunter (1984) ebenfalls einer Reanalyse der Daten von Asher und Scarrino (1974), welche auf Seiten der Arbeitsprobe unterschieden zwischen psychomotorischen und verbalen Arbeitsproben. In diesem Fall ermittelten Hunter und Hunter (1984) eine mittlere korrigierte Validität von .62 für psychomotorische Arbeitsproben und von .45 für verbale Arbeitsproben. Mit der Studie von Schmitt, Gooding, Noe und Kirsch (1984) liegt eine weitere Meta-Analyse zur Validität von Arbeitsproben vor. Schmitt et al. (1984) errechneten unter Bezugnahme auf sieben Studien mit 384 Probanden eine mittlere unkorrigierte Validität von .32.

Den soeben dargestellten Studien wurden verschiedene Einwände entgegengebracht (vgl. Roth, Bobko & McFarland, 2005): Der Arbeit von Asher und Scarrino (1974) wurde beispielsweise angelastet, dass viele der berücksichtigten Tests, welche die beiden Autoren als Arbeitsproben klassifiziert haben, nach heutiger Sicht als SJTs oder Tests zur Messung berufsbezogenen Wissens zu bezeichnen wären. An der Arbeit von Hunter und Hunter (1984) wurde kritisch angemerkt, dass die Autoren keine Angaben über die in die Meta-Analyse einbezogenen Primärstudien machen. In Bezug auf die Meta-Analyse von Schmitt et al. (1984) wurde als Kritikpunkt angeführt, dass die Autoren nur Primärstudien aus zwei amerikanischen Journals betrachtet haben, was in einer geringen Datenmenge resultierte. Aufgrund der beanstandeten Punkte an früheren Meta-Analysen führten Roth, Bobko und McFarland (2005) eine eigene Meta-Analyse durch, in welcher sie eine mittlere korrigierte Validität von .33 ermittelten. Um die oben angesprochenen Mängel und Versäumnisse zu umgehen, beachteten sie bestimmte Entscheidungsregeln bei der Einbeziehung von Primärstudien. Demgemäß bedachten sie zum Beispiel nur Primärstudien mit Bewerbern oder Angestellten und keine Laborstudien. Zudem differenzierten sie zwischen Arbeitsproben auf der einen Seite und SJTs oder Tests zum berufsbezogenen Wissen auf der anderen Seite.

Als Datenbasis konnten Roth et al. (2005) auf 54 Studien mit insgesamt 10469 Probanden zurückgreifen. Es zeigte sich, dass Primärstudien, welche nach 1982 veröffentlicht wurden, im Schnitt geringere korrigierte Validitäten angaben als solche Primärstudien, die vor 1982 in einem Journal erschienen. Ein weiterer für diese Arbeit erwähnenswerter Aspekt ist die

von Roth et al. (2005) berichtete Korrelation in Höhe von .13 zwischen SJTs und Arbeitsproben vor. Allerdings lagen zur Berechnung dieser Korrelation nur drei Studien vor.

SJTs und Arbeitsproben sind in Bezug auf ihren Grad an fidelity zu unterscheiden. Während Arbeitsproben ein hoher Grad an fidelity bescheinigt wird (Guion, 1998), wird SJTs ein niedriger respektive im Falle videobasierter SJTs mittlerer Grad an fidelity (Funke & Schuler, 1998) zugewiesen (vgl. Abschnitt 2.4).

2.3 Die Konstruktion von SJTs

In der relevanten Forschungsliteratur zeichnet sich in den letzten Jahren eine Entwicklung ab, bei welcher die Komponenten von SJTs, zum Beispiel die Art der Situationspräsentation (z. B. Lievens & Sackett, 2006) oder die Form der Instruktion für die Probanden (z. B. McDaniel et al., 2007), systematisch manipuliert und variiert werden, um den Einfluss von Veränderungen in der SJT-Konstruktion auf die Nützlichkeit des SJTs in der Personalauswahl zu beleuchten. In diesem Abschnitt wird aufgezeigt, welche Auswirkungen Veränderungen hinsichtlich der Situationsgestaltung, der Darbietung der Verhaltensalternativen oder der Bewertung der Probandenantworten nach sich ziehen können.

2.3.1 Die Herkunft der in SJTs abgebildeten Situationen

Es sind zwei verschiedene Herangehensweisen bei der Gewinnung von Situationen für SJTs unterscheidbar: Die Durchführung einer *Arbeits- und Anforderungsanalyse*, welche zumeist in Form der Critical Incidents Technique (CIT; Flanagan, 1954) vollzogen wird, oder die Orientierung der Situationsgestaltung an bestehenden *Theorien* beziehungsweise Modellen. Der Einsatz der CIT als Beispiel für eine Arbeits- und Anforderungsanalyse ist nach Weekley, Ployhart und Holtz (2006) die am häufigsten angewandte Methode bei der Generierung von Situationen für einen SJT. Dabei werden im Rahmen der CIT sogenannte *subject matter experts* (SME) befragt. Die SME können sowohl Mitarbeiter ohne Führungsaufgaben als auch Führungskräfte sein. Als leitendes Kriterium für die Auswahl der SME sollte deren Kenntnisstand über denjenigen Arbeitsplatz oder diejenige Klasse an Arbeitsplätzen, für die der SJT entwickelt wird, dienen. Im Rahmen der Befragung werden SME gebeten, sich an selbst oder alternativ von Mitarbeitern respektive Arbeitskollegen erlebte erfolgskritische Ereignisse zu erinnern. Eine umfassendere Darstellung der CIT wird im dritten Kapitel dieser Arbeit geleistet.

McDaniel et al. (2001) vollzogen in ihrer Meta-Analyse einen Vergleich von SJTs, welche einer Arbeits- und Anforderungsanalyse entstammen, und solchen SJTs, die einem alternativen Entwicklungsprozess entspringen sind, hinsichtlich ihrer kriterienbezogenen Validität. Sie stellten fest, dass SJTs, welche auf Basis einer Arbeits- und Anforderungsanalyse ent-

standen sind, eine höhere mittlere Korrelation mit Arbeitsleistung (.38) erreichten als SJTs, deren Inhalte nicht auf eine Arbeits- und Anforderungsanalyse zurückzuführen waren (.29).

Eine zweite mögliche Herangehensweise bei der Entwicklung von Situationen für einen SJT stützt sich auf bereits existierende Modelle oder Theorien. Zumeist zielt die Gestaltung der Situationen dann auf die Messung eines spezifischen Konstrukts wie beispielsweise einer bestimmten Persönlichkeitseigenschaft ab (Weekley et al., 2006). Studien, in denen sich die Konstruktion der SJTs an Theorien oder Modellen orientierte, stammen beispielsweise von Motowidlo et al. (1990; vgl. Abschnitt 2.1), Stevens und Campion (1994, 1999), Weekley und Jones (1997) oder – um ein Beispiel aus dem deutschsprachigen Raum zu nennen – von Schubert, Ortwein, Dumitsch, Schwantes, Wilhelm und Kiessling (2008). Aufgrund der SJTs inhärenten Multidimensionalität (vgl. Abschnitt 2.5) ist ein empirischer Nachweis angenommener theoretischer Strukturen, zum Beispiel mit Hilfe von Faktorenanalysen, als problematisch anzusehen (vgl. McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001; McDaniel & Whetzel, 2005). Zudem mangelt es Situationen, deren Konstruktion vorrangig durch theoretische Strukturen inspiriert wurde, an einer empirischen Verankerung in dem Arbeitsumfeld, welches in dem SJT simuliert werden soll.

2.3.2 Inhalte und Charakteristiken von SJT-Items

Bei der Majorität der entwickelten SJTs dominiert eine inhaltliche Ausrichtung an *interpersonellen oder sozialen Situationen* (z. B. Kanning, 2005; Bledow & Frese, 2009). Seltener kommen SJTs zur Anwendung, in denen *berufsbezogene Problemsituationen* abgebildet werden, welche technischer Natur sind. Weekley et al. (2006) führen als exemplarische Situationen die Behebung eines Computerabsturzes oder die Organisation eines Projektes an. Der Einsatz technisch geprägter Problemsituationen soll vor allem Aussagen über das berufsbezogene prozedurale Wissen der Probanden ermöglichen. Die Gestaltung der Situationen im SJT ist weiterhin unterscheidbar entlang dreier zentraler Charakteristika und Gestaltungsmerkmale (vgl. McDaniel & Nguyen, 2001): *Fidelity*, *Komplexität* und *Verständlichkeit* im Sinne einer inhaltlichen Nachvollziehbarkeit.

In einem SJT werden die Situationen entweder in Text- oder in Videoform präsentiert. Mit der Art der Präsentation ist für McDaniel und Nguyen (2001) die Frage nach der fidelity verknüpft. Die Auslegung des Konzepts der fidelity und die inhaltliche Auseinandersetzung mit videobasierten SJTs erfolgen aufgrund ihrer hervorgehobenen Bedeutung für die vorliegende Arbeit in dem gesonderten Abschnitt 2.4.

Während einige SJTs knappe, prägnante Situationsbeschreibungen enthalten (z. B. der HS von File und Remmers, 1971; s. Abschnitt 2.1), präsentieren andere SJTs den Probanden umfangreiche, komplexe Situationen, in denen sogar situative Verzweigungen möglich sind (z. B. Richman, Olson-Buchanan & Drasgow, 2000). Die Länge und Komplexität von Situationsbeschreibungen steht nach McDaniel und Nguyen (2001) in einem engen Zusammenhang zu der Verständlichkeit respektive Nachvollziehbarkeit von Situationsbeschreibungen aus Sicht der Probanden. Weekley et al. (2006) erwarten aufgrund dieses Zusammenhangs bei SJTs, die von Probanden eine höhere Lesekompetenz verlangen, um Inhalte zu verstehen, die Gefahr, dass nicht für alle Probandengruppen Chancengleichheit bestehe, was wiederum die wahrgenommene Testfairness beeinflussen würde (vgl. Abschnitt 2.5). Weiterhin erhöht sich gemäß McDaniel und Nguyen (2001) mit wachsender Komplexität oder höheren Anforderungen hinsichtlich der Verständlichkeit die kognitive Ladung eines SJT: *„It is a reasonable conjecture that the length, complexity, and comprehensibility of the situation are interrelated and may drive the cognitive loading of the situational stems“* (S. 105). Dementsprechend ist dem Grad an Komplexität und Verständlichkeit von Situationsbeschreibungen ebenfalls ein Einfluss auf die Konstruktvalidität von SJTs zuzuschreiben. Ob der Grad an Komplexität die kriterienbezogene Validität eines SJTs beeinflusst, untersuchten McDaniel et al. (2001) anhand einer Meta-Analyse. Für SJTs mit sehr detaillierter und berufsspezifischer Situationsbeschreibung ergab sich eine leicht niedrigere mittlere Korrelation mit Arbeitsleistung (.33) als für SJTs mit allgemeiner, kurzer Situationsbeschreibung (.35).

2.3.3 Die Wahl der Instruktionsform für die Probanden

Nachdem den Probanden im Rahmen der Testdurchführung der Stimulus präsentiert wurde – zum Beispiel in Form eines Videos in einem videoba-

sierten SJT oder in Form eines Textes in einem SJT im Paper-Pencil-Format – wird ihnen eine Auseinandersetzung mit verschiedenen Verhaltensalternativen abverlangt. Zur Vorbereitung auf diesen Schritt erhalten die Probanden eine Instruktion seitens der Testautoren. Dabei herrschen in SJTs gemeinhin zwei Formen der Instruktion vor (vgl. Ployhart & Ehrhart, 2003; Nguyen, McDaniel & Biderman, 2005; Weekley et al., 2006). Die erste Form der Instruktion enthält eine Aufforderung an die Probanden, aus einer Liste von Verhaltensoptionen die beste (z. B. Reynolds, Winter & Scott, 1999) oder alternativ die beste und die schlechteste (z. B. Weekley & Jones, 1999; Clevenger & Haaland, 2000) je präsentierter Situation zu bestimmen. Als Variation dieser Instruktionsform werden die Probanden in manchen Studien gebeten, die Effektivität aller Verhaltensalternativen einzuschätzen (z. B. Chan & Schmitt, 1997; Pereira & Harvey, 1999; Sacco, Scheu, Ryan, Schmitt, Schmidt & Rogg, 2000). Diese Form der Instruktion soll fortan mit *Instruktion zum Wissensausdruck* (engl.: knowledge instruction) betitelt werden. Eine Instruktion, die beste oder effektivste Verhaltensalternative auszuwählen, korrespondiert mit typischen Instruktionen in Tests zur Erfassung der kognitiven Leistungsfähigkeit oder in Tests zur Messung von berufsbezogenem Wissen. Aufgrund dieser Kongruenz werden von SJTs mit einer Instruktion zum Wissensausdruck hohe Korrelationen mit berufsbezogenem Wissen (z. B. Motowidlo et al., 1990) und allgemeiner kognitiver Leistungsfähigkeit erwartet (McDaniel et al., 2001; Nguyen et al., 2005). Ployhart und Erhart (2003) merkten in diesem Kontext an, dass die Verhaltensweise, welche die Probanden als beste identifizieren, nicht diejenige sein muss, die sie in einer realen Situation tatsächlich an den Tag legen: „[...]the SJT response asks for the most effective behavior in the situation, whereas the same behavior the person actually does on the job may not be the most effective“ (S. 2).

Eine zweite Form der Instruktion fordert die Probanden indessen auf, aus den dargebotenen Verhaltensalternativen diejenigen zu benennen, welche sie einerseits am wahrscheinlichsten (engl.: most likely) und andererseits am unwahrscheinlichsten (engl.: least likely) in der realen Situation zeigen würden (z. B. Motowidlo et al., 1990; Pulakos & Schmitt, 1996). In den weiteren Ausführungen firmiert diese Form der Instruktion unter der Bezeichnung *Instruktion zur Verhaltenstendenz* (engl.: behavio-

ral tendency instruction). Nach Ployhart und Erhart (2003) führt die Instruktion zur Verhaltenstendenz dazu, dass sich die Probanden Verhaltensweisen in Erinnerung rufen, welche sie in vergleichbaren Situationen ihres beruflichen Werdegangs gezeigt haben. Nguyen et al. (2001) ergänzten, dass diese Form der Instruktion Selbstbeschreibungen von Probanden einholt, was dem Vorgehen in Persönlichkeitstests entspreche. Infolgedessen werden hohe Korrelationen zwischen SJTs mit Instruktion zur Verhaltenstendenz und verschiedenen Persönlichkeitsfaktoren prognostiziert.

Die Unterscheidung zwischen einer Instruktion zur Verhaltenstendenz und einer Instruktion zum Wissensausdruck verbanden McDaniel et al. (2007) mit der Differenzierung von maximaler und typischer Leistung (vgl. Cronbach, 1949). Nach Ones und Visweswaran (2007) lassen sich maximale und typische Leistung wie folgt voneinander abgrenzen: *„Typical performance refers to what people will do, and maximal performance refers to what people can do. Thus, typical performance and maximal performance are theoretically distinct”* (S. 294). Die maximale Leistung, wie sie zum Beispiel in ACs oder Arbeitsproben abgefragt wird, hängt nach McDaniel et al. (2007) in stärkerem Maße von der allgemeinen kognitiven Leistungsfähigkeit ab als die typische Leistung. SJTs mit einer Instruktion zum Wissensausdruck sind für McDaniel et al. (2007) ebenfalls Messungen maximaler Leistung. Typische Leistungen assoziierten sie indessen stärker mit nicht-kognitiven Korrelaten wie zum Beispiel Aspekten der Persönlichkeit. Wird ein SJT mit einer Instruktion zur Verhaltenstendenz ausgestattet, misst er ihrer Ansicht nach typische Leistung. Dieser Gedankengang entspricht einer früheren Hypothese von McDaniel und Nguyen (2001), die mutmaßten, dass die in einem SJT gewählte Verhaltensinstruktion einen Einfluss auf seine Konstruktvalidität besäße.

Ployhart und Ehrhart (2003) untersuchten, wie sich sechs Variationen der beiden Instruktionsformen auf die psychometrische Qualität eines SJTs auswirken. Abgesehen von der variablen Instruktion hielten sie die fünf Situationen im SJT als auch die Verhaltensalternativen je Situation konstant. In der ersten Phase ihrer Datenerhebung wendeten die Autoren ein between-subjects-Design an und ließen rund 500 Probanden den SJT einmal absolvieren. Dabei wurden jeweils rund 80 Probanden einer der verschiedenen Instruktionen zugewiesen. Sechs Wochen später startete die zweite Phase der Datenerhebung, in welcher mehr als 200 Probanden

den SJT mit allen sechs möglichen Instruktionen (within-subjects-design) durchliefen. Um Aussagen über die Test-Retest-Reliabilität treffen zu können, waren etwa 150 Probanden an beiden Phasen beteiligt. Auf Grundlage von Berechnungen zur internen Konsistenz in der zweiten Phase der Datenerhebung ließ sich nach Ansicht der Autoren kein systematischer Unterschied zwischen SJTs mit Instruktion zur Verhaltenstendenz und SJTs mit Instruktion zum Wissensausdruck feststellen. Ein anderes Bild ergab sich jedoch bei Analysen der Test-Retest-Reliabilität, bei der sich signifikante Unterschiede je nach benutzter Instruktionsform ergaben. Die höchste Test-Retest-Reliabilität erzielte der SJT mit der Instruktion *rate how likely would do* (.92), während der niedrigste Wert (.20) für den SJT mit der Instruktion *rate how effective* ermittelt wurde. Um Aussagen über die Konstruktvalidität treffen zu können, berechneten Ployhart und Ehrhart (2003) Interkorrelationen zwischen den sechs verschiedenen SJT Versionen aus der zweiten Phase der Datenerhebung. Die durchschnittliche korrigierte Interkorrelation zwischen den drei SJTs mit Instruktionen zur Verhaltenstendenz fiel in diesem Fall deutlich höher aus als die durchschnittliche korrigierte Interkorrelation zwischen den drei SJTs mit Instruktion zum Wissensausdruck. Für die durchschnittliche korrigierte Korrelation zwischen SJTs mit Instruktion zur Verhaltenstendenz und SJTs mit Instruktion zum Wissensausdruck ergab sich indessen ein Wert von .22. Bei der Vorhersage unterschiedlicher Leistungskriterien schnitten SJTs mit Instruktionen zur Verhaltenstendenz in beiden Phasen der Datenerhebung signifikant besser ab als SJTs mit Instruktionen zum Wissensausdruck.

Der Einfluss von unterschiedlichen Instruktionsformen auf die Konstruktvalidität von SJTs wurde weiterführend durch eine Meta-Analyse von McDaniel et al. (2007) untersucht⁵. In Übereinstimmung mit Ployhart und Ehrhart (2003) orientierten sich McDaniel et al. (2007) ebenfalls an einer Differenzierung zwischen SJTs mit einer Instruktion zur Verhaltenstendenz und SJTs mit einer Instruktion zum Wissensausdruck. Sie bezogen Studien mit vier unterschiedlichen Instruktionen zur Verhaltenstendenz

⁵ Interessanterweise erfuhr die Studie von Ployhart und Ehrhart (2003) keine Aufnahme in diese Meta-Analyse, da McDaniel und sein Team nur Studien mit Bewerbern oder Angestellten sowie Vorgesetztenbeurteilungen als Leistungskriterium berücksichtigten.

und Studien mit sieben verschiedenen Instruktionen zum Wissensausdruck in ihre Analyse ein.

Die amerikanischen Forscher stellten fest, dass SJTs mit einer Instruktion zum Wissensausdruck im Mittel signifikant höher mit allgemeiner kognitiver Leistungsfähigkeit (.35) korrelierten als SJTs mit einer Instruktion zur Verhaltenstendenz (.19). Allerdings rieten McDaniel und sein Team zur Skepsis, da sowohl die Breite der Vertrauensintervalle als auch der geringe Anteil an Varianz, welcher durch methodische Artefakte erklärbar ist, für eine substantielle Variabilität der analysierten Einzelkorrelationen spreche. SJTs mit Instruktion zur Verhaltenstendenz korrelierten weiterhin durchschnittlich höher mit den Persönlichkeitsfaktoren Gewissenhaftigkeit (.34 versus .24), Verträglichkeit (.37 versus .19) und Emotionale Stabilität (.35 versus .12). Für die beiden anderen Faktoren aus dem Big 5-Modell ergab sich jedoch ein anderes Bild: SJTs mit Instruktion zum Wissensausdruck korrelierten im Mittel höher mit den Persönlichkeitsfaktoren Extraversion (.14 versus .07) und Offenheit für Erfahrungen (.12 versus .09) als SJTs mit Instruktion zur Verhaltenstendenz. Bezogen die Autoren hingegen nur jene acht Studien ein, in denen derselbe SJT zweimal durchgeführt wurde – einmal mit Instruktion zum Wissensausdruck und ein anderes Mal mit Instruktion zur Verhaltenstendenz –, ergab sich ein klareres Bild: SJTs mit Instruktion zur Verhaltenstendenz korrelierten nun mit allen fünf Persönlichkeitsfaktoren durchschnittlich höher als SJTs mit Instruktion zum Wissensausdruck. Der Meta-Analyse von McDaniel et al. (2007) wurden zwei Punkte entgegengehalten. Zum einen wurde bemängelt, dass sich nur vier der einbezogenen Studien auf andere Daten als von angestellten Mitarbeitern stützten. Zum anderen wurde angeführt, dass die Inhalte der verschiedenen SJTs nicht konstant und damit vergleichbar gehalten wurden, so dass vorgefundene Effekte nicht immer direkt auf die Instruktionsform zurückführen sind (vgl. Arthur & Villado, 2008).

Zwei Jahre nach der Meta-Analyse von McDaniel et al. (2007) untersuchten Lievens, Sackett und Buyse (2009), wie sich die Variation der Instruktionsform auf einerseits die kriterienbezogene Validität und andererseits die Konstruktvalidität eines SJTs auswirkt, falls dieser im Rahmen einer realen Auswahl-situation (high-stakes selection context) eingesetzt wird. Dazu setzten sie denselben SJT, das heißt der Inhalt wurde gleich

gehalten (s. o.), einmal mit einer Instruktion zum Wissensausdruck und ein anderes Mal mit einer Instruktion zur Verhaltenstendenz ein.

Das Forscherteam stellte dabei einerseits fest, dass die Wahl der Instruktionsform keinen Einfluss auf die Höhe des Testwerts im SJT ausübt. Zudem ergaben sich bei der Korrelation mit zwei Leistungskriterien keine signifikanten Unterschiede je nach gewählter Instruktionsform. Andererseits korrelierte der SJT mit einer Instruktion zum Wissensausdruck in nicht-zufälliger Weise höher (.19) mit Intelligenz als derselbe SJT mit einer Instruktion zur Verhaltenstendenz.

Es bleibt festzuhalten, dass die Wahl der Instruktionsform in einem SJT einen Einfluss auf die Konstruktvalidität des Verfahrens ausüben kann und dass über die Entscheidung für eine bestimmte Instruktionsform – Instruktion zur Verhaltenstendenz oder Instruktion zum Verhaltensausdruck – bis zu einem bestimmten Maße zu steuern ist, mit welchen Kriterien ein SJT höher korreliert und mit welchen niedriger. Für die kriterienbezogene Validität eines SJT kann die Wahl der Instruktion hingegen als unerheblich erachtet werden. Im weiteren Verlauf dieser Arbeit wird dargelegt werden, dass je nach gesetzter Instruktion für die Probanden auch das Ausmaß sozial erwünschter Antworten variiert (s. Abschnitt 2.5).

2.3.4 Die Gestaltung der Verhaltensalternativen im SJT

In den meisten Studien werden SME gebeten, verschiedene Verhaltensalternativen für eine gegebene Situation zu benennen (z. B. Dalessio, 1994). Hinsichtlich der Anzahl von pro Situation eingesetzten Verhaltensalternativen und deren inhaltlicher Komplexität besteht eine große Variabilität zwischen den Studien. Ployhart und Ehrhart (2003) bezeichnen in diesem Zusammenhang eine Zahl von vier bis sechs Verhaltensalternativen als Durchschnitt. Dabei kann die Sammlung von Verhaltensalternativen entweder mit Hilfe derselben SME durchgeführt werden, welche schon an der CIT partizipiert haben, oder es werden bisher nicht involvierte SME herangezogen. Eine Beteiligung von SME bei der Generierung realistischer Verhaltensalternativen wird basierend auf ihrem arbeitsplatzbezogenen Kenntnisstand begrüßt (Weekley et al., 2006). Nichtsdestotrotz existierten auch SJTs, bei denen die Testautoren die Verhaltensalternativen ohne Mitwirkung von SME entwickelt haben (z. B. Weekley & Jones, 1999; Weekley & Jones, 2005). Als Sonderform dieser

alternativen Herangehensweise führten Weekley et al. (2006) die konstruktbasierte Entwicklung von Verhaltensalternativen an. Bezüglich dieses Themas listeten sie in ihrem Überblicksartikel verschiedene Konferenzbeiträge auf, in denen verschiedene konstruktbasierte Vorgehensweisen präsentiert wurden. Exemplarisch berichteten sie von Versuchen, die Konstruktion der Verhaltensalternativen am Big 5 Modell der Persönlichkeit (z. B. Tupes & Christal, 1992) auszurichten oder Verhaltensalternativen von SME a priori in ein theoretisches Rahmenmodell einzuordnen. Systematische Studien zur Überprüfung der Nützlichkeit konstruktbasierter Verhaltensalternativen liegen allerdings bislang nicht vor. In Analogie zur theoriegeleiteten Herleitung der Situationen ist auch an dieser Stelle einerseits die Multidimensionalität von SJTs zu betonen, welche einer empirisch eindeutigen Bestätigung angenommener theoretischer Strukturen im Wege stehen mag, und andererseits der zu befürchtende fehlende Bezug zu realen Verhaltensalternativen am konkreten Arbeitsplatz. Ein Punkt, welcher bei der Gestaltung der Verhaltensalternativen besondere Aufmerksamkeit verdient, ist der, wie leicht es Probanden fällt, besonders effektive oder situativ angemessene Verhaltensweisen als solche zu identifizieren. Dieser Aspekt ist eng mit dem *Problem sozial erwünschter Antworttendenzen* (engl.: Faking) verknüpft (s. Abschnitt 2.4). Um die Gefahr des Fakings zu minimieren, empfehlen Weekley et al. (2006) eine Parallelisierung der Verhaltensalternativen in Bezug auf den Grad ihrer sozialen Erwünschtheit. Um den Grad an sozialer Erwünschtheit, welche eine Verhaltensalternative evoziert, zu bemessen, sind der Literatur zwei Methoden zu entnehmen: Das Ausmaß an sozialer Erwünschtheit wird durch Experten beurteilt (Marlowe & Crowne, 1961) oder Korrelationen zwischen der Beantwortung der Verhaltensalternative und einer Skala zur Messung sozialer Erwünschtheit dienen als Gradmesser (z. B. Jackson, 1994). Da bei einigen SJTs nicht alle Verhaltensalternativen bearbeitet werden beziehungsweise die Beantwortungen der einzelnen Verhaltensalternativen nicht immer statistisch unabhängig sind (s. u.), muss die zweite beschriebene Methode als mit vielen Unwägbarkeiten behaftet angesehen werden.

2.3.5 Die Festlegung eines Bewertungsschlüssels

Während sich bei Tests zur allgemeinen kognitiven Leistungsfähigkeit eine Probandenantwort schnell und eindeutig als richtig oder falsch charakterisieren lässt, gestaltet sich diese Art der Einstufung bei SJTs weitaus diffiziler (Legree, 1995). Verhaltensalternativen, welche den Probanden in den jeweiligen konstituierenden Situationen eines SJT zur Verfügung stehen, müssen vorab, das heißt vor den ersten Testdurchführungen, als richtig/effektiv beziehungsweise falsch/nicht effektiv im Sinne einer erfolgreichen Bewältigung der dargestellten Situation festgelegt werden. Die Festlegung der situationsbezogenen Effektivität von Verhaltensalternativen entspricht der Definition eines Bewertungsschlüssels (scoring key).

In diesem Zusammenhang erwähnten McDaniel und Nguyen (2001) zwei verschiedene Varianten, einen Bewertungsschlüssel festzulegen. Als erste Möglichkeit nannten sie die erneute Einbeziehung von SME oder – alternativ – herausragenden Mitarbeitern. Diese sollten auf individueller Basis oder in Gruppen einvernehmlich festlegen, welche Verhaltensalternativen in der jeweiligen Situation richtig beziehungsweise falsch sind. Erzielen die hinzugezogenen Experten keinen Konsens in Bezug auf die Einstufung bestimmter Verhaltensalternativen, sollten diese nach McDaniel und Nguyen (2001) revidiert werden. Die zweite von McDaniel und Nguyen (2001) aufgelistete Möglichkeit beinhaltet die empirische Durchführung einer Pilotstudie und die anschließende Berechnung von Maßen der zentralen Tendenz, um Verhaltensalternativen als richtig oder falsch einzuordnen.

Im Gegensatz zu McDaniel und Nguyen (2001) nennen Bergman, Drasgow, Donovan und Henning (2006) sogar vier verschiedene Möglichkeiten, um Lösungsschlüssel für die Bewertung der Antworten von Probanden im SJT zu generieren: *Empirisch* (empirical scoring), *theoretisch* (theoretical scoring), *experten-basiert* (expert-based) und eine Mischung mehrerer Lösungsschlüssel (hybrid scoring).

Bei der empirischen Methode werden die Probandenantworten gemäß ihrer statistischen Korrelationen mit einem Kriterium bewertet (vgl. Weekley & Jones, 1997). Zur Anwendung dieser Methode sind vorab verschiedene Bedingungen zu erfüllen. Neben der Wahl eines geeigneten Kriteriums und der notwendigen Kreuzvalidierung der Daten betonen

Drasgow et al. (2006) die Befolgung fester Entscheidungsregeln beim Setzen eines Lösungsschlüssels. Eine Regel sollte thematisieren, welche Höhe eine Korrelation zwischen Verhaltensalternative und Kriterium mindestens erreichen sollte, um Verhaltensalternativen als richtig respektive falsch einstufen zu können. Laut einer weiteren Regel sollte jede Verhaltensalternative von einer Mindestanzahl an Probanden (Basisrate) gewählt werden, um empirisch bewertet werden zu können. Beim Befolgen der empirischen Methode ergibt sich ein dreiteiliger Bewertungsschlüssel: Eine +1 erhalten solche Verhaltensalternativen, welche häufig von solchen Probanden gewählt wurden, die hohe Kriteriumswerte erzielen. Diese Verhaltensalternativen werden als richtig eingestuft. Dagegen werden Verhaltensalternativen mit -1 bewertet, die regelmäßig von Probanden gewählt wurden, welche vergleichsweise niedrige Kriteriumswerte erreicht haben. Folglich werden diese Verhaltensalternativen als falsch angesehen. Verhaltensalternativen, denen keine empirische Beziehung zum Kriterium nachzuweisen ist, erhalten eine 0.

Die theoretische Herangehensweise zur Gewinnung eines Lösungsschlüssels ist eng an den Postulaten bestehender Rahmenmodelle ausgerichtet (s. o.). Verhaltensalternativen, welche theoretische Annahmen widerspiegeln und unterstützen, werden als richtig eingestuft, während Verhaltensalternativen, welche theoretischen Annahmen widersprechen, als falsch bewertet werden. Verhaltensalternativen ohne erkennbaren Bezug zur Theorie erhalten die Funktion von Distraktoren und werden weder als richtig noch falsch gedeutet. Theoriegestützte Bewertungsschlüssel finden sich zum Beispiel in den Studien von Olson-Buchanan et al. (1998) oder Stevens und Campion (1999).

Im Rahmen der *experten-basierten Methode* werden SME gemeinhin aufgefordert, die gegebenen Verhaltensalternativen als richtig oder falsch einzustufen (s. o.). Alternativ erwähnen Drasgow et al. (2006) einen Vergleich zwischen Expertenantworten und Novizenantworten. In diesem Fall werden diejenigen Verhaltensalternativen als richtig bewertet, die ungeachtet der Novizenantworten von besonders vielen Experten gewählt wurden. Wählen hingegen besonders viele Novizen eine Verhaltensalternative, aber nur wenige Experten, wird diese als falsch bewertet.

Die zur Entwicklung des Bewertungsschlüssels hinzugezogenen SME sind zumeist Mitarbeiter ohne Führungsaufgaben (Pulakos & Schmitt,

1996), Führungskräfte (Weekley & Ployhart, 2005), externe Berater (Hunter, 2003) oder Kunden (Weekley & Jones, 1997). Legree (1995) bezeichnete die Festlegung von richtigem Verhalten in Situationen mit sozialem oder interpersonellem Hintergrund als komplex und anspruchsvoll. Nach seiner Ansicht bestehe unter SME oftmals Dissens über die Richtigkeit und situative Adäquatheit bestimmter Verhaltensweisen, so dass er eine größere Gruppe von SME zur Festlegung eines Bewertungsschlüssels empfahl. So bezog Philips (1992, 1993) zum Beispiel nur solche Verhaltensalternativen in die Endfassungen seiner SJT ein, bei denen sich 75 % der involvierten 20 SME hinsichtlich der Einstufung als richtig oder falsch einig waren. Bei der Entwicklung eines Bewertungsschlüssels für einen SJT, welcher bei der amerikanischen Armee eingesetzt werden sollte, griff Legree (1995) bewusst auf Laien zurück. Seine Entscheidung begründete er wie folgt (Legree, 1995; S. 250):

The logic of using nonexperts to develop an expert rating pattern may seem counterintuitive. However, if the ratings of any participant are conceptualized as observed scores composed of common and unique variance, in which the common component corresponds to expert variance and the unique component to nonexpert variance, then expertise is directly proportional to the magnitude of the common component and inversely proportional to the unique component.

Die Frage nach der Anzahl notwendiger SME unberücksichtigt lassend erwähnten Weekley et al. (2006), dass unterschiedliche Expertengruppen womöglich auch unterschiedliche Lösungsschlüssel für ein und denselben SJT entwickeln würden. Vorstellbar wäre zum Beispiel, dass Kunden gänzlich andere Verhaltensalternativen als effektiv und richtig auf Seiten von Verkäufern einschätzen als deren Vorgesetzte. Diesem Punkt vorgreifend setzten Weekley und Jones (1997) zwei Bewertungsschlüssel für einen SJT ein, welcher speziell für Beschäftigte in Pflegeberufen konstruiert wurde. Neben einem empirischen Bewertungsschlüssel kam ein rationaltheoriegeleiteter Schlüssel (s. o.) zum Einsatz, der mit Hilfe von Familienangehörigen der pflegebedürftigen Personen entworfen wurde. Zwischen den beiden Bewertungsschlüsseln fanden die Autoren je nach Stichprobe Korrelationen von .48 respektive .53 vor (s. Abschnitt 2.4).

Bislang blieb die Frage unbeantwortet, ob die Art des Bewertungsschlüssels die Gütekriterien eines SJTs beeinflusst. Bezüglich dieser Frage kamen Weekley et al. (2006) nach der Durchsicht mehrerer Artikel und Konferenzbeiträge aus dem Zeitraum von 1997 bis 2004 zum Vergleich empirischer hergeleiteter Bewertungsschlüssel auf der einen Seite und mit Hilfe von SME konstruierter Bewertungsschlüssel auf der anderen Seite zu dem Schluss, dass SJTs mit beiden Arten von Bewertungsschlüsseln kriterienbezogene Validitäten vergleichbarer Höhe erzielen können. Dennoch spräche die zumeist vorgefundene geringe Interkorrelation zwischen den beiden Arten von Bewertungsschlüsseln dafür, dass unterschiedliche Bewertungsschlüssel unterschiedliche Konstrukte sowie unterschiedliche Aspekte der Kriterien erfassen (vgl. Weekley & Jones, 1997).

Bergman et al. (2006) verglichen in ihrer Studie elf verschiedene Bewertungsschlüssel hinsichtlich ihres Einflusses auf die kriterienbezogene und inkrementelle Validität eines SJT. Als Grundlage diente ein videobasierter SJT zur Messung von Führungskompetenz. Als Leistungskriterien wurden zwei Vorgesetztenbeurteilungen eingeholt, zum einen eine Beurteilung der Führungsqualität der Probanden im speziellen und zum anderen eine Einschätzung der allgemeinen beruflichen Leistung.

Nur für drei der untersuchten Bewertungsschlüssel – experten-basiert, empirisch und hybrid participation – ergaben sich sowohl signifikante Korrelationen moderater Höhe mit den beiden Leistungskriterien als auch statistisch bedeutsame inkrementelle Validitäten bei der Vorhersage der Führungsqualität. Der experten-basierte Bewertungsschlüssel erbrachte einen zusätzlich erklärten Varianzanteil von 4.9 %, der auf empirischem Wege gewonnene Bewertungsschlüssel 3.0 % und der hybride Bewertungsschlüssel 1.7 %. Als weitere Prädiktoren verwendeten Bergman und ihr Team zwei Persönlichkeitsinventare. Aus ihren Ergebnissen folgerten die Forscher, dass „[...] *the validity of an SJT depends in part on its scoring, and that poor choices could lead to the conclusion that the SJT's content is not valid when it may only be the scoring key that is not valid*“ (Bergman et al., 2006; S. 231).

Nach vollzogener Festlegung der Effektivität der verschiedenen Verhaltensalternativen müssen die Probandenantworten eingeholt und mit Hilfe des Bewertungsschlüssels verrechnet werden. Am Ende erhalten die Probanden einen Gesamtwert für ihr Abschneiden im SJT.

Weekley et al. (2006) unterschieden zwei unterschiedliche Formate, wie den Probanden die Verhaltensalternativen vorgegeben werden: Im forced-choice-Format oder in Form einer Likertskala. Beim *forced-choice-Format* werden die Verhaltensalternativen im Multiple-Choice-Profil dargestellt und die Probanden werden dazu aufgefordert, entweder die in ihren Augen effektivste oder diejenige Verhaltensalternative, welche sie in der Situation am wahrscheinlichsten zeigen würden, auszuwählen (s. o.). Bei der forced-choice-Methode liegt eine Abhängigkeit zwischen den Bewertungen der unterschiedlichen Verhaltensalternativen vor (Vasilopoulos, Cucina, Dyomina, Morewitz & Reilly, 2006). Der Einsatz der forced-choice-Methode ist weiterhin verbunden mit dem Wunsch, das Auftreten sozial erwünschter Antworten zu minimieren (Jackson, Wroblewski & Ashton, 2000; S. 373):

Forced-choice questionnaires represent one possible method of minimizing the problem of faking in employment testing by making the task of responding desirably much more difficult. If respondents are motivated to make the best possible impression, being forced to choose between items similar in perceived relevance to the job tends to reduce this type of impression management

Die Erreichung dieses Ziels setzt wie oben bereits besprochen voraus, dass die dargebotenen Verhaltensalternativen in Bezug auf ihren Grad an sozialer Erwünschtheit parallelisiert wurden. Eine sehr einfache und ökonomische Auswertungsform des forced-choice-Format basiert auf der Annahme, dass eine einzige Verhaltensalternative richtig ist, während alle anderen Verhaltensalternativen als falsch gelten (z. B. Stevens & Campion, 1999; Hunter, 2003). Wählen Probanden die als richtig eingestufte Verhaltensalternative, erhalten sie 1 Punkt, während die Auswahl einer anderen und damit nicht richtigen Verhaltensalternative keine Punkte einbringt. Eine komplexere Auswertungsform des forced-choice Formats wurde von Motowidlo et al. (1990) eingeführt und wurde bereits im Abschnitt 2.1 vorgestellt. Nach Weekley et al. (2006) bietet die Auswertung eines SJTs nach Vorgabe von Motowidlo und seinem Team den Vorteil größerer Varianz in Bezug auf die Probandenantworten und damit eine Voraussetzung für höhere Reliabilitäten und Validitäten als bei der ein-

gangs erwähnten einfachen dichotomen Auswertungsform. Einen alternativen Einsatz des forced-choice Formats propagierten Weekley, Ployhart und Harold (2004). Sie ließen ihre Probanden alle vorgegeben Verhaltensalternativen in eine Rangordnung bringen, an deren Spitze die beste und an deren Ende die schlechteste Verhaltensalternative aus Sicht der Probanden eingefügt werden sollte. Um die Rangordnungen der Probanden zu bewerten, wurden diese mit Hilfe von Rangkorrelationen nach Spearman verglichen mit den von SME erstellten Rangordnungen.

Die Bewertung und Auseinandersetzung mit allen Verhaltensalternativen mit Hilfe von Likertskalen – als Gegenvorschlag zum forced-choice Format – erfordert, dass die Probanden jede Verhaltensalternative bezüglich ihrer situativen Effektivität einstufen. Ein Beispiel für diese Art der Herangehensweise findet sich in einer Studie von Chan und Schmitt (2002).

Der Einsatz von Likertskalen bringt nach McDaniel und Nguyen (2001) mehrere Vorteile mit sich. Erstens seien die Bewertungen der einzelnen Verhaltensalternativen als unabhängig voneinander anzusehen, so dass im Gegensatz zum forced-choice-Format keine ipsativen Werte vorlägen. Als zweiten Vorteil führen McDaniel und Nguyen (2001) an, dass bedingt durch die notwendige Einschätzung aller vorgegebenen Verhaltensalternativen eine Vielzahl an Daten und Varianz vorhanden wäre, was wiederum der Reliabilität und Validität des SJTs zuträglich sein könnte. Tatsächlich ermittelte Legree (1994; 1995) für SJTs mit Likertskalen höhere Reliabilitäten – angegeben in Form von Cronbach's Alpha – als für SJTs mit forced-choice Format. Der dritte Vorteil des Einsatzes von Likertskalen, welchen McDaniel und Nguyen (2001) anführen, thematisiert die erstrebte Messung homogener Konstrukte mit SJTs. Nach McDaniel und Nguyen (2001) erweitert die Anwendung des Likertformats anstatt des forced-choice Formats die Möglichkeit, mit einem SJT einzelne Konstrukte zu messen (vgl. Abschnitt 2.5).

In Form der Abbildung 2-2 werden zum Abschluss dieses Abschnitts drei zentrale Konstruktionsschritte für SJTs schematisch zusammengefasst. Die Darstellung orientiert sich an einem Leitfaden, welcher von Ployhart et al. (2006; S. 514-516) für eine Anwendung in der Praxis formuliert wurde, und kann als Heuristik verstanden werden.

Step 1:

Generate situations using the critical incidents technique. Create situations that are fairly common, but where there is no clear course of action.

Step 2:

Edit situations and create response options. Attempt to generate four up to six options for each situation. No option should be obviously correct or incorrect.

Step 3:

Determine the scoring key. Refine and revise items. Delete items that show low agreement. Think carefully about what type of test instructions will be used, how subject-matter experts will be asked to evaluate the options, and how situational judgment scores will be created and combined.

Abbildung 2-2: Leitfaden zur Erstellung von SJTs in der Praxis

2.4 Videobasierte SJTs

In diesem Abschnitt erfolgt ein gesonderter Blick auf verschiedene videobasierte SJTs. Zuvor wird jedoch das Konzept der *fidelity*, welches den Realitätsgrad von Testinhalten thematisiert, besprochen.

2.4.1 Das Konzept der *fidelity*

Jedes Item eines SJT setzt sich wie in Abschnitt 2.3 dargelegt zusammen aus einer Stimuluskomponente, welche den Probanden in eine arbeitsnahe Situation hineinversetzt, und einer Responsekomponente, durch welche der Proband seine Reaktion auf die situativen Erfordernisse – zumeist eine Problemstellung sozialer Färbung (s. u.) – ausdrückt (Motowidlo et al., 1990; Funke & Schuler, 1998). Funke und Schuler (1998) weisen in diesem Zusammenhang daraufhin, dass gerade bei Tests, die arbeitsspezifische Situationen und erfolgsrelevantes Verhalten am Arbeitsplatz abzubilden beabsichtigen, ein hoher Grad an *fidelity* (vgl. Motowidlo et al., 1990; s. Abschnitte 2.1 und 2.3) sowohl hinsichtlich der Stimulus- als auch der Responsekomponente vorliegen sollte.

Auf Seiten des Gestaltungsmerkmals *fidelity* ist zu differenzieren zwischen *psychological* und *physical fidelity*: „*Psychological fidelity refers to the extent that the predictor taps the same constructs required for successful performance on the criterion, and physical fidelity refers to the extent the predictor is physically similar to the criterion*” (Ployhart et al., 2006; S. 377). Der Begriff *physical fidelity* wird demnach herangezogen, um zu bestimmen, inwiefern der Testinhalt eines SJT die Gegebenheiten an einem tatsächlichen Arbeitsplatz widerspiegelt. SJTs, welche Situationen in Videoform beinhalten, wird eine höherer Grad an *physical fidelity* beigemessen als SJTs, welche die Situationen im Paper-Pencil-Format präsentieren (z. B. Motowidlo, Hanson & Crafts, 1997; Lievens & Coetsier, 2002). Die *fidelity* eines SJTs besitzt zudem Bedeutung für die soziale Validität und Augenscheinvalidität eines SJTs (vgl. Abschnitt 2.5).

Bei videobasierten SJTs besteht die Stimuluskomponente aus einer Videoszene, welche je nach Aufwand bei der Konstruktion des SJT eine mehr oder minder realistische Arbeitssituation enthält (Kanning, 2004). Da die Antworten der Probanden im Rahmen videobasierter SJTs zumeist schriftlich oder verbal formuliert erfolgen, schreiben Funke und Schuler

(1998) der Responsekomponente in dieser Art von Tests im Rückgriff auf Motowidlo et al. (1990) einen geringen Grad an fidelity zu (vgl. Ployhart et al, 2006). Hingegen statten sie die Stimuluskomponente in videobasierten Tests mit dem Attribut hoher fidelity aus, so dass sie videobasierte Tests zusammengefasst als „*stimulus-centered*“ (Funke & Schuler, 1998; S. 116) bezeichnen. Die Abbildung 2-3 bietet eine Kontrastierung verschiedener Methoden der Berufseignungsdiagnostik hinsichtlich ihres Grades an (physical) fidelity.

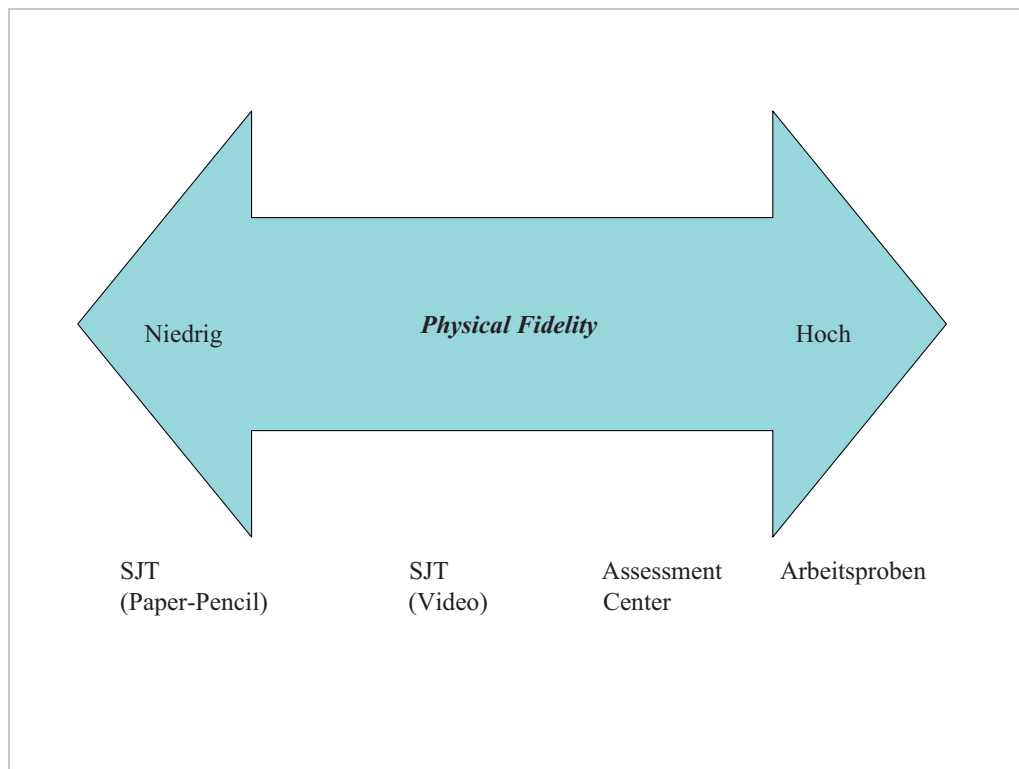


Abbildung 2-3: Einzelne Verfahren der Personalauswahl und ihr Grad an fidelity

Nach der Abbildung 2-3 verfügen Arbeitsproben über den höchsten Grad an physical fidelity (s. Abschnitt 2.2), während SJTs im Paper-Pencil-Format der niedrigste Grad an physical fidelity bescheinigt wird.

2.4.2 Videobasierte SJTs im Überblick

Bereits im Jahre 1949 sah Thorndike – trotz fehlender technischer Möglichkeiten – videobasierte Tests als (zukünftige) Bereicherung bei der Messung von sozialen oder interpersonellen Kompetenzen der Bewerber in einer Personalauswahlsituation an. Laut Thorndike (1949) sprächen

verschiedene Vorteile von videobasierten Tests für ihren Einsatz: Erstens seien sie wie Papier-Bleistift-Tests als Gruppentest durchführbar. Zweitens fügten sie der Auswahl-situation durch ihren anderen Stimulusgehalt – bewegte Bilder – einen höheren Anteil an Realismus hinzu. Als dritten Vorteil nannte er, dass in einem videobasierten Test sowohl die Präsentation des Stimulus als auch die Reaktion des Bewerbers (zeitlich) schneller erfolgen könne als in einem Papier-Bleistift-Test, was den Grad an Realismus noch weiter erhöhe. Weekley und Jones (1997) führten die Argumentation für den Einsatz videobasierter Tests in der Personalauswahl weiter aus. Wie Thorndike (1949) betonten sie, dass in Videos berufsrelevantes Verhalten und arbeitsplatzbezogene Situationen realistischer und detailreicher abgebildet werden können als in Paper-Pencil-Tests. Zudem bezeichneten sie das Hören von Dialogen und Sehen von Bewegungen im Rahmen von Videotests als eine natürlichere und leichter handhabbare Handlung für die meisten Bewerber als das Lesen von Texten in Paper-Pencil-Tests (vgl. Kanning et al., 2006). Als weiteres Argument führten Weekley und Jones (1997) an, dass Bewerber durch videobasierte Tests einen repräsentativen Einblick in ihr mögliches späteres Berufsfeld samt dessen speziellen Anforderungen erhalten. Drasgow, Olson, Keenan, Mober und Mead (1993) bezeichneten zum Beispiel nonverbale Signale von Rollenspielern wie deren Minenspiel als realistische Arbeitsplatzinformationen, die in Papier-Bleistift-Tests nicht adäquat abzubilden sind. Ihr Plädoyer für videobasierte Tests beschlossen Weekley und Jones (1997) mit dem Punkt der Augenscheinvalidität. Videobasierte Tests weisen demnach im Vergleich zu vielen Persönlichkeitstests oder Tests zur Messung der allgemeinen kognitiven Leistungsfähigkeit einen für Bewerber leichter erkennbaren Bezug zum Berufsleben auf (vgl. Abschnitt 2.5). Neben den genannten Vorteilen wiesen Weekley und Jones (1997) auch auf Nachteile von videobasierten Tests hin. Zum einen sprachen sie die hohen finanziellen Entwicklungskosten an und zum anderen nannten sie die Vielzahl an notwendigen technischen Geräten zum Videodreh und zur Präsentation der Videos als Manko gegenüber Papier-Bleistift-Tests. Zur Ergänzung der Vorteile videobasierter Tests soll ein Punkt von Drasgow et al. (1993) ergänzt werden. Sie hoben die Standardisierung der Stimuluspräsentation als weiteren Vorzug videobasierter Tests hervor. Diesen hohen Grad an Standardisierung und damit einhergehend an Ob-

ektivität bezüglich der Durchführung erachteten sie vor allem als Vorteil gegenüber Rollenspielen, in denen eine Standardisierung aufgrund des interaktiven Charakters nicht zwangsläufig gegeben ist. Olson-Buchanan und Drasgow (2006) nahmen weniger eine Abwägung von Vor- und Nachteilen vor als dass sie viel mehr Kriterien benannten, die die Qualität videobasierter Tests beeinträchtigen könnten. Als ein Kriterium führten sie die Befähigung der Rollenspieler, die Situationskripte realitäts- und vor allem arbeitsplatzgetreu umzusetzen, an. Sie empfahlen zur Gewährleistung einer arbeitsplatzgerechten Darstellung eine dauerhafte beratende Anwesenheit von SME beim Videodreh. Obendrein verliehen sie der Forderung nach einer regelmäßigen Revision der Videos Nachdruck, um die in den Videos benutzten Kleidungsstile, Frisuren oder sonstige Accessoires dem jeweiligen Zeitgeist anzupassen.

Die nächsten Seiten geben einen Einblick in verschiedene Studien, in denen videobasierte SJTs für den Einsatz in spezifischen Berufsfeldern konzipiert wurden. Veröffentlicht wurden diese Studien in den 80ern und 90ern des zwanzigsten Jahrhunderts beziehungsweise im ersten Jahrzehnt des einundzwanzigsten Jahrhunderts. Da Studien mit videobasierten SJTs keine Aufnahme in die Meta-Analysen von McDaniel et al. (2001) und McDaniel et al. (2007) erhielten (s. Abschnitt 2.5), erscheint eine differenzierte Auseinandersetzung mit diesen Arbeiten umso bedeutsamer. Dabei beschränkt sich die Darstellung auf Feldstudien mit Bewerberstichproben beziehungsweise Beschäftigten in Unternehmen. Laborstudien wurden zum Beispiel von Chan und Schmitt (1997) oder Richman-Hirsch, Olson-Buchanan und Drasgow (2000) vorgelegt.

Im Jahre 1986 entwickelten Jones und DeCotiis (1986) einen videobasierten Test zur Auswahl kunden- beziehungsweise gastorientierter Mitarbeiter für das Hotel- und Gaststättengewerbe. Da in der vorliegenden Arbeit ebenfalls ein SJT zur Messung der Kundenorientierung im Hotelfach entwickelt werden soll, wird der Studie der beiden amerikanischen Forscher nun verstärkt Aufmerksamkeit gewidmet. Wie bei Motowidlo et al. (1990) erfolgte der erste Konstruktionsschritt in Form der CIT nach Flanagan (1954). Die in den Interviews mit Hotelangestellten gesammelten erfolgskritischen Ereignisse wurden durch eine Gruppe von Managern überprüft und dienten später als Skripte für den Videodreh. Basierend auf den Ergebnissen des ersten Konstruktionsschrittes entstand ein Fragenbo-

gen zur Arbeitsanalyse, welcher weiteren Mitarbeitern der beteiligten Hotels vorgelegt wurde. Die Auswertung der Fragebögen brachte hervor, dass der Umgang mit Gästen nur für bestimmte Berufsgruppen innerhalb des Hotels- und Gaststättengewerbes von hervorgehobener Bedeutung ist. Speziell für diese Berufsgruppen, zum Beispiel Mitarbeiter an der Rezeption oder Mitarbeiter im Bankettbereich, sollte der videobasierte Test entwickelt werden. In die Endfassung des Tests wurden insgesamt 40 Videos aufgenommen. Als Leistungskriterium für einen Pretest mit Hotelmitarbeitern wurde die Beziehungsgestaltung zu Hotelgästen erwählt.

Die Präsentation der einzelnen Videos lief wie folgt ab: Vor der videobasierten Darstellung der einzelnen Situationen leitete eine Erzählstimme aus dem Off das jeweils folgende Szenario inhaltlich ein. Daran schloss sich die Darstellung der Videos an. Für die Beantwortung erschienen den Probanden auf einer Leinwand pro Video vier Verhaltensalternativen, die von derselben Erzählstimme der Reihe nach vorgelesen wurden. Die Aufgabe der Probanden war es, zu entscheiden, wie sich der Mitarbeiter des Hotels in der Situation verhalten sollte. In der Abbildung 2-4 wird eine Situation via eines Skripts dargestellt.

The scene is the front door, where a single door attendant is on duty. A guest approaches the attendant and says, "Hi, I am looking for a good restaurant. Can you recommend me one?"

The door attendant should:

- A. Give only the names of restaurants in the hotel.
- B. Give the names of good restaurants nearby the hotel.
- C. Ask the guest to check with the hotel's front desk.
- D. Tell about the hotel's own restaurants, and then give names of other restaurants.

Abbildung 2-4: Ein Item aus dem Videotest von Jones und DeCotiis (nach Jones & DeCotiis, 1986; S. 71)

In einem Pretest mit rund 100 Mitarbeitern ergaben sich je nach Berufsgruppe der Mitarbeiter kriterienbezogene Validitäten zwischen .16 und .71. Die durchschnittliche um Messfehler korrigierte Validität bezifferten Jones und DeCotiis (1986) mit .55. Da die Testdaten und die Leistungs-

beurteilungsdaten gleichzeitig erhoben wurden, handelt es sich um konkurrente Validitäten. Nach dem Pretest wurde der Test zur Auswahl neuer Mitarbeiter für ein Hotel eingesetzt. Dabei wurden ausschließlich Bewerber getestet.

Ungeklärt bleiben bei der Studie von Jones und DeCotiis (1986) die Frage nach dem Auswertungsschlüssel für die Antworten der Probanden sowie der Verbleib der Ergebnisse für die Bewerberstichprobe, weil diese in dem zitierten Artikel nicht genannt werden. Zudem ist zu vermuten, dass für die einzelnen Berufsgruppen nur sehr kleine Substichproben vorlagen.

Wie Jones und DeCotiis entwickelte Dalessio (1994) einen videobasierten Test für eine spezielle Berufsgruppe, in diesem Fall Versicherungsvertreter. Für die Auswahl der Videos griff Dalessio (1994) aus ökonomischen Erwägungen auf bestehende Videos aus Verkaufstrainings zurück. Die 15 ausgewählten Videos enthielten jeweils eine realistische Berufssituation in Form einer Interaktion zwischen einem Versicherungsagenten und einem Kunden und mündeten in einer offenen Problemstellung, zum Beispiel dem Abschluss einer Versicherungspolice. Dalessio (1994) ordnete aufbauend auf mehreren weiteren Analyseschritten jeder Situation fünf Verhaltensalternativen zu. Als Kriterium wählte Dalessio (1994) keine Leistungsbeurteilung durch Vorgesetzte, sondern den Verbleib der Versicherungsagenten in der Firma (engl.: turnover). Erhoben wurde dieses dichotome Kriterium ein Jahr nach Durchführung des Tests, so dass die im Weiteren aufgeführten Validitäten prognostischer Natur sind.

Die Probanden waren im SJT aufgefordert anzugeben, was sie am wahrscheinlichsten (most likely) beziehungsweise unwahrscheinlichsten (least likely) in der Situation tun würden (vgl. Motowidlo et al., 1990). Nach der Beendigung des Test wurden für jeden Probanden drei Werte berechnet: Ein most likely-Wert, ein least likely-Wert und ein Gesamtwert als Summe aus den most likely- und least likely-Werten.

Bei der Validierung der Daten kam die Methode der Kreuzvalidierung zur Anwendung (vgl. Bortz, 2005). Zu diesem Zweck bildete Dalessio (1994) eine Eichstichprobe und zwei Vergleichsstichproben. Die drei Testwerte (s. o.) der zwei Vergleichsstichproben wurden für die Vorhersage der Kriteriumswerte auf Seiten der Eichstichprobe verwandt. Es ergab sich sowohl für den Gesamtwert als auch für den most likely-Wert

eine signifikante Korrelation von .13. Der least likely-Wert erzielte mit dem Kriterium eine nicht-signifikante Korrelation in Höhe von .09.

Verglichen mit Jones und DeCotiis (1994) verwandte Dalessio (1994) in seiner Studie einen klar definierten scoring key an. Kritisch anzumerken an der Studie von Dalessio (1994) sind die unvollständige empirische Anforderungsanalyse (s. o.) und die unzureichende respektive methodisch ungewöhnliche Validierung des Tests.

Im Gegensatz zu Jones und DeCotiis (1986) oder Dalessio (1994) konstruierten Smiderle, Perry und Cronshaw (1994) für ihre Studie keinen neuen Test, sondern widmeten sich der Evaluation eines bereits existierenden Verfahrens, des aus 65 Items bestehenden Metro Seattle Video Test (MSVT). Mit Hilfe des MSVT sollten Einblicke gewonnen werden in die sozialen Kompetenzen von Bewerbern für Positionen als Busfahrer im Grenzverkehr zwischen den USA und Kanada. Die Evaluation des MSVT wurde an einer Stichprobe von circa 400 kanadischen Busfahrern vorgenommen. Für die Validierung des MSVT wurden mehrere harte Leistungskriterien in Betracht gezogen, zum Beispiel die Anzahl an Beschwerden beziehungsweise an Lob von Fahrgästen oder Vorgesetzten für den Busfahrer, die Anzahl unentschuldigter Fehltag am Arbeitsplatz sowie die Anzahl selbstverschuldeter Unfälle während der Arbeitszeit. Zusätzlich zur kriterienbezogenen Validierung wurde eine Konstruktvalidierung vorgenommen. Zu diesem Zweck wurde zunächst ein strukturiertes Interview durchgeführt. In diesem Interview wurden den Probanden biographieorientierte Fragen gestellt, welche die sozialen Kompetenzen im Umgang mit Fahrgästen in den Vordergrund stellten. Neben dem Interview, welches der Überprüfung der konvergenten Validität dienlich sein sollte, wurde zur Erhebung der diskriminanten Validität ein kognitives Korrelat gemessen.

Auf Seiten der kriterienbezogenen Validierung fand sich für den SJT nur eine signifikant negative Korrelation (-.12) mit der Anzahl an Beschwerden von Fahrgästen. Smiderle und sein Team stellten vor der Validierung die Hypothese auf, dass der MSVT positiv mit dem strukturierten Interview korrelieren würde. Mit einer ermittelten signifikanten Korrelation von .17 konnten sie diese Hypothese bestätigen. Allerdings fügten sie einschränkend hinzu, dass der MSVT nur 1.4 % an Varianz auf Seiten des Interviews erkläre. Für die unterschiedlichen Leistungsbereiche des kog-

nitiven Tests ergaben sich ebenfalls signifikant positive Korrelationen zwischen .24 und .37 mit dem MSVT. Eine Unabhängigkeit des im MSVT gemessenen Konstrukts von kognitiven Fähigkeiten konnte dementsprechend nicht festgestellt werden.

Sowohl bei Smiderle et al. (1994) als auch bei Dalessio (1994) bleibt der Punkt offen, auf welche Art und Weise – schriftlich oder verbal – sich Bewerber im Videotest für Verhaltensalternativen entscheiden mussten. Weitere hinterfragenswerte Aspekte hinsichtlich ihrer Arbeit erwähnen Smiderle et al. (1994) selbst. Zum Beispiel stellen sie Varianzeinschränkungen und Kontaminationen auf Seiten der harten Leistungskriterien heraus, welche zu den niedrigen Korrelationen mit dem MSVT beigetragen haben könnten. Daneben betonen sie das Fehlen prognostischer Validitäten als Manko der Studie. Insgesamt nennen sie den MSVT unter Berücksichtigung der berichteten Ergebnisse als ungeeignet für die Personalauswahl: „[...]at present, the psychometric limitations of the MSVT appear to preclude its operational use“ (Smiderle et al., 1994; S. 18). Was die Studie von Smiderle et al. (1994) weiterhin positiv von den Studien von Jones und DeCotiis sowie Dalessio abhebt, ist eine sehr viele umfangreichere empirische Analyse und Überprüfung der Gütekriterien. Dazu stimmt bei Jones, De Cotiis (1986) und Dalessio (1994) im Gegensatz zu Smiderle und seinem Team der Punkt bedenklich, dass alle drei zu dem Zeitpunkt ihrer Erhebung Angestellte beziehungsweise Teilhaber von Firmen waren, welche von einem validen videobasierten Test finanziell profitiert hätten.

Die von Smiderle et al. (1994) an den Tag gelegte ausführliche empirische Vorgehensweise wurde beibehalten beziehungsweise erweitert durch zwei Studien⁶ mit videobasierten SJTs von Weekley und Jones (1997). Bei beiden Studien bezweckten Weekley und Jones (1997) sowohl die Untersuchung der kriterienbezogenen Validität als auch die der Konstruktvalidität.

Im Blickpunkt der ersten Studie standen Beschäftigte im Einzelhandel. Neben dem videobasierten Test wurde eine neu konstruierte Skala zur Messung der allgemeinen kognitiven Leistungsfähigkeit eingesetzt. Zur Erfassung der bestehenden Arbeitserfahrung wurde eine aus fünf Fragen

⁶ Beide Studien wurden im Rahmen eines einzigen Artikels publiziert.

bestehende Skala entwickelt. Auf Seiten des Kriteriums entschieden sich Weekley und Jones für die Erfassung der Arbeitsleistung (*task performance*; vgl. Borman & Motowidlo, 1993), welche Führungskräfte anhand vierstufiger grafischer Einstufungsskalen beurteilen sollten.

In der ersten Studie wurden die Ergebnisse zweier Datenerhebungen an unterschiedlichen Bewerberstichproben zusammengefasst. Der SJT war jeweils eingebettet in eine Testbatterie bestehend aus der kognitiven Skala und den Fragen zur Arbeitserfahrung. Von den eingestellten Bewerbern wurden Leistungswerte (Kriterium) erhoben. Anhand der Korrelationen zwischen dem videobasierten Test und den Leistungswerten entwickelten Weekley und Jones (1997) einen empirischen Schlüssel (s. Abschnitt 2.3) zum Scoring der Verhaltensalternativen je Situation. Der videobasierte Test erzielte in beiden Stichproben signifikante Korrelationen mit dem Leistungskriterium, der kognitiven Skala sowie mit den Fragen zur Arbeitserfahrung.

Bei ihrer zweiten Studie analysierten Weekley und Jones (1997) die Tätigkeit von Altenpflegekräften in Seniorenresidenzen. Für das Scoring der Verhaltensalternativen verwandten Weekley und Jones (1997) dieses Mal zwei unterschiedliche Herangehensweisen. Aus ihrer ersten Studie übernahmen sie den empirischen Schlüssel zur Gewichtung der Verhaltensalternativen (s. o.). Zusätzlich entwickelten sie auf rationalem Wege einen zweiten Schlüssel. Zu diesem Zweck präsentierten sie Angehörigen von Bewohnern in Altersresidenzen die Videos und baten diese, diejenige Verhaltensalternative zu bestimmen, welche sie sich seitens des Pflegepersonals wünschen würden. Diejenige Verhaltensalternative je Situation, welche von den meisten Angehörigen gewählt wurde, bekam als Punktwert eine 1, die drei anderen Verhaltensalternativen hingegen eine 0. Der videobasierte Test war auch in der zweiten Studie integriert in eine Testbatterie, welche zudem eine neu entwickelte Skala zur Messung kognitiver Fähigkeiten und eine Skala zur Messung von Berufserfahrung enthielt. Das Kriterium war eine Beurteilung der Arbeitsleistung durch einen Vorgesetzten.

Als Probanden für ihre erste Datenerhebung im Rahmen der zweiten Studie konnten Weekley und Jones (1997) rund 400 Personen gewinnen, die zum Messpunkt mindestens sechs Monate in ihrer Position beschäftigt waren. Die Kriteriendaten wurden einen Monat nach Durchführung der

dreiteiligen Testbatterie erhoben. Weekley und Jones (1997) bezeichneten nichtdestotrotz alle an dieser Stichprobe gewonnenen Validitäten als konkurrent. Für den videobasierten Test mit empirischem Bewertungsschlüssel ergaben sich signifikante Korrelationen mit der Leistungsbeurteilung (.35) und der kognitiven Skala (.27). Bezogen auf die Arbeitserfahrung fand sich für den empirisch bewerteten Test keine statistisch bedeutsame Korrelation. Hingegen ermittelten Weekley und Jones (1997) für den videobasierten Test mit rationalem Bewertungsschlüssel signifikante Korrelationen mit dem Leistungskriterium (.14) und mit den beiden Skalen aus der Testbatterie: Kognitive Skala (.32), Skala zur Arbeitserfahrung (.14). Folglich resultierten je nach eingesetztem scoring key Veränderungen der Konstruktvalidität des SJT.

Die zweite von Weekley und Jones (1997) herangezogene Stichprobe bestand aus neu eingestellten Altenpflegekräften. Relevante Leistungsdaten wurden in diesem Fall drei Monate nach Anfang des Beschäftigungsverhältnisses erhoben. Validitäten, welche mit dieser Stichprobe gewonnen wurden, sind laut Weekley und Jones (1997) prognostisch. Wurde der empirische Schlüssel angewandt, ließen sich wie bei der ersten Stichprobe signifikante Korrelationen mit der Arbeitsleistung (.24) und der kognitiven Skala (.15) finden, aber keine signifikante Korrelation mit der Arbeitserfahrung. Eine Anwendung des rationalen Bewertungsschlüssels führte wie bei der ersten Stichprobe zu signifikanten Korrelationen mit der Arbeitsleistung (.33), der kognitiven Skala (.30) und der Arbeitserfahrung (.20).

Als übergreifendes Fazit lassen sich für die beiden Studien von Weekley und Jones (1997) folgende Punkte festhalten: Mit einem videobasierten SJT, welcher auf einer sorgfältigen empirischen Arbeitsanalyse beruht, sind moderate kriterienbezogene Validitäten zu erzielen. Ferner ist die Leistung in dem SJT bis zu einem gewissen Maße als statistisch abhängig von der kognitiven Fähigkeit und Arbeitserfahrung der Probanden zu betrachten.

Aus dem europäischen Raum liegen mit den Arbeiten von Lievens und Coetsier (2002) beziehungsweise von Lievens und Sackett (2006) ebenfalls Studien mit videobasierten SJTs vor. Beide Studien wurden im Gegensatz zu den bislang vorgestellten Studien im Rahmen von realen Auswahlentscheidungen eingesetzt.

Im Jahre 2002 untersuchte Lievens zusammen mit Coetsier (2002) die Nützlichkeit von zwei videobasierten SJTs bei der Auswahl von Medizin- und Zahnmedizinstudenten für fünf flämische Universitäten. Einen SJT widmeten sie der Darstellung typischer Situationen aus der Arzt-Patienten-Interaktion, während der zweite eingesetzte SJT aus mehreren Gesprächsrunden zwischen erfahrenen Medizinern bestand. Die beiden SJTs wurden im Verbund mit zwei Arbeitsproben, einer kognitiven Testbatterie und einem Persönlichkeitsinventar eingesetzt. Die Vorhersage der Abschlussnoten nach einem Studienjahr (Kriterium) durch die videobasierten SJTs fiel ernüchternd aus: Für den SJT zur Arzt-Patienten-Interaktionen ermittelten die Autoren eine nicht-signifikante korrigierte Validität von .12 und für den SJT mit Experten-Diskussionen eine nicht-signifikante korrigierte Validität von .02.

2006 wählte Lievens zusammen mit Sackett eine etwas andere Herangehensweise, indem er primär auf einen Vergleich zweier SJT Formate (Video vs. Paper-Pencil) hinsichtlich ihrer kriterienbezogenen, inkrementellen und sozialen Validität abzielte. Die Entwicklung des videobasierten SJT für diese Studie vollzog sich nach bewährtem Muster (s. Abschnitte 2.1 und 2.3). Für den SJT im Paper-Pencil-Format wurden die verbalen Informationen aus den Videos herangezogen und die Dialoge zwischen Arzt und Patient in eine schriftliche Form gebracht. Neben dem SJT dienten noch zwei weitere Testverfahren der Auswahl zukünftiger Studenten. Mit einer Arbeitsprobe wurde ein zweites simulationsorientiertes Verfahren neben dem SJT in die Testbatterie aufgenommen. Die Testbatterie wurde vervollständigt durch einen Test zur allgemeinen kognitiven Leistungsfähigkeit sowie vier Wissenstests mit den medizinisch relevanten Themengebieten Chemie, Physik, Mathematik und Biologie. Nach Abschluss der Testdurchführung wurden die Studenten auf freiwilliger Basis gebeten, den soeben durchlaufenen Auswahlprozess einzuschätzen. Von diesen Angaben erhofften sich Lievens und Sackett (2006) Einblicke in die Augenscheinvalidität der eingesetzten Testverfahren, insbesondere der SJTs.

Die Kriteriendaten entstammten zwei Quellen. Zum einen lagen durchschnittliche Leistungspunkte (GPA) für die Studenten vor, welche den Test bestanden hatten beziehungsweise nach Bestehen des Tests tatsächlich ein Studium der Medizin oder Zahnmedizin angetreten hatten. Zum

anderen lagen für diese Gruppe an Studenten Noten aus Kursen speziell zur Arzt-Patienten-Kommunikation und Interaktion vor (soziales Kriterium). Beide Datenreihen wurden ein Jahr nach dem Auswahlprozess erhoben.

Die Ergebnisse der kriterienbezogenen Validierung bestätigten vorab formulierte Erwartungen von Levens und Sackett (2006): Der videobasierte SJT erzielte mit einem Wert von .34 eine signifikant höhere Korrelation mit dem sozialen Leistungskriterium als der SJT im Paper-Pencil-Format (.08). Da das soziale Leistungskriterium nicht an allen Studenten erhoben werden konnte, so dass eine Varianzeinschränkung auf Seiten dieses Kriteriums zu erwarten war, korrigierten die beiden Forscher die Validitäten. Auf diesem Wege ergab sich für den videobasierten SJT und das soziale Leistungskriterium eine Korrelation von .35 und für den SJT im Paper-Pencil-Format und das soziale Leistungskriterium eine Korrelation in Höhe von .09. Der SJT im Paper-Pencil-Format brachte weiterhin eine signifikant höhere Korrelation .18 mit dem kognitiven Leistungstest hervor als der videobasierte SJT (.11). Zur Berechnung der inkrementellen Validitäten der beiden SJTs wurden zunächst zwei schrittweise Regressionen auf das soziale Leistungskriterium berechnet. Der SJT wurde hierbei jeweils als dritter und letzter Prädiktor nach dem kognitiven Leistungstest und der Arbeitsprobe in die Regressionsgleichung aufgenommen. Während der videobasierte SJT einen substanziellen Anteil von 11 % an zusätzlicher Varianz aufklären konnte, tendierte der zusätzliche Varianzanteil, welcher durch den SJT im Paper-Pencil-Format erklärt werden konnte, gegen Null (.01). Im Gegensatz zur Vorhersage des sozialen Leistungskriteriums verhalte sowohl die Aufnahme des videobasierten SJT als auch des SJT im Paper-Pencil-Format in die schrittweise Vorhersage der Leistungspunktzahl (GPA) ohne Wirkung. Für eine weitere Hypothese der beiden Forscher fand sich keine empirische Bestätigung: Der videobasierte SJT wurde von den Studenten in Bezug auf seine Augenscheinvalidität nicht signifikant besser als der SJT im Paper-Pencil-Format eingeschätzt. Davon abgesehen wurde der SJT in beiden Ausprägungen jedes Mal positiver bewertet als die anderen beteiligten Prädiktoren.

Zum Abschluss dieses Abschnitts soll ein videobasierter SJT aus dem deutschsprachigen Raum vorgestellt werden (Kanning, 2005). Der SJT

wurde bei Kanning eingebunden in den aus fünf Modulen bestehenden Test Sozialer Kompetenzen (TSK). Folgende fünf Module waren Bestandteil des TSK: Im Modul 1 stand die Messung der allgemeinen kognitiven Leistungsfähigkeit im Vordergrund. Das Modul 2 deckte den Bereich allgemeiner sozialer Kompetenzen ab, welche durch vier übergeordnete Faktoren – Soziale Orientierung, Selbstbehauptung, Selbstkontrolle und Soziale Wahrnehmung – repräsentiert wurden. Mit den Modulen 3 bis 5 wurde zusammengefasst die Messung berufsspezifischer sozialer Kompetenz angestrebt. Das Modul 5 ist einem videobasierten SJT gleichzusetzen. Auf Basis der Sammlung kritischer Ereignisse an einer Expertenstichprobe wurden 50 Situationen ausgewählt und verfilmt.

In Bezug auf die Konstruktvalidität fanden sich – wie von Kanning (2005) erwartet – signifikant positive Korrelationen moderater Höhe zwischen dem Modul 5 (SJT) und den beiden anderen Modulen – 3 und 4 – zur Messung berufsspezifischer Kompetenzen. Aus dem Modul 2 korrelierte alleinig der Faktor Selbstbehauptung signifikant (.13) mit dem Modul 5. Für das Modul 5 ergab sich weiterhin eine Nullkorrelation mit dem Modul 1. Dies spricht dafür, dass die Leistung in dem videobasierten SJT als statistisch unabhängig anzusehen war von der allgemeinen kognitiven Leistungsfähigkeit der Bewerber. Für die weitere Validierung des TSK setzte Kanning mit dem NEO-FFI (Borkenau & Ostendorf, 2008) und dem Interpersonal Competence Questionnaire (Riemann & Allgöwer, 1993) zwei Persönlichkeitsinventare ein. Das Modul 5 erzielte signifikante Korrelationen moderater Höhe mit den Faktoren Offenheit für Erfahrungen (.26) und Soziale Verträglichkeit (.26) aus dem NEO-FFI. Weitere signifikante Korrelationen ergaben zwischen dem Modul 5 und drei Skalen des Interpersonal Competence Questionnaire.

Ähnlich wie bei SJTs insgesamt (vgl. Abschnitt 2.1) ist auch für videobasierte SJTs im Speziellen über die Jahrzehnte ein Zuwachs des Entwicklungsaufwandes und Forschungsumfanges festzustellen. Nichtsdestotrotz existieren zu videobasierten SJTs bislang weitaus weniger Studien als zu SJTs im Paper-Pencil-Format (s. Abschnitt 2.5). Die in diesem Abschnitte vorgestellten Erkenntnisse lassen jedoch auf eine Zunahme in diesem Bereich hoffen.

2.5 Der Nutzen von SJTs für die Personalauswahl

Dieser Abschnitt stellt zentrale Forschungsergebnisse zur kriterienbezogenen Validität, Konstruktvalidität und sozialen Validität von SJTs vor. Zudem werden mögliche Einflüsse auf die Validität eines SJTs angesprochen.

2.5.1 Kriterienbezogene Validität

Als Argument für den Einsatz von SJTs in der Personalauswahl werden zumeist die Ergebnisse zweier Meta-Analysen angeführt, welche die Gültigkeit von SJTs bei der Vorhersage von Arbeitsleistung belegten. Nachfolgend werden zentrale Erkenntnisse der beiden Meta-Analysen berichtet. Ebenso werden aber auch Unzulänglichkeiten und Versäumnisse dieser beiden Studien Erwähnung finden.

McDaniel et al. (2001) führten eine Meta-Analyse durch, um den Stand der Forschungsliteratur zur Beziehung von SJTs auf der einen Seite und der beruflichen Leistung auf der anderen Seite zusammenzufassen⁷. Als Grundlage für ihre Analyse dienten 102 Koeffizienten zur kriterienbezogenen Validität, welche an insgesamt 10640 Probanden ermittelt worden waren.

Studien, denen die relevanten Korrelationseffizienten entnommen wurden, entstammten nach Angabe der Autoren dem Zeitraum von 1987 bis 2000. McDaniel et al. (2001) legten vorab bestimmte Maßstäbe fest, die Studien erfüllen mussten, um in die Meta-Analyse einbezogen zu werden: Erstens mussten die Probanden in den Studien Bewerber oder Angestellte in Unternehmen sein, Laborstudien mit Studenten wurden ausgeschlossen. Zweitens wurden nur solche Studien beachtet, in denen als Leistungskriterien Vorgesetztenbeurteilungen oder Beurteilungen von Personen, welche häufigen und engen Kontakt zu den Beurteilten haben, eingesetzt wurden. Vor der Durchführung der Analysen korrigierten McDaniel et al. (2001) die Varianzschätzungen für die Reliabilitäten der einbezogenen SJTs, welche zumeist in Form von Cronbach's Alpha angegeben waren.

⁷ Die Ergebnisse der Meta-Analyse von McDaniel et al. (2001) zur Konstruktvalidität von SJTs werden in Abschnitt 2.4.2 erörtert.

Neben mehreren seit Jahren oder gar Jahrzehnten erforschten und vielseitig kommerziell genutzten SJTs, zum Beispiel dem HS, bezogen McDaniel et al. (2001) Daten von namentlich nicht näher benannten SJTs ein. Diese entstammten entweder dem Einsatz in amerikanischen Regierungsbehörden (vgl. Legree, 1994; 1995) oder waren von Beratungsunternehmen spezifisch für bestimmte Organisationen entwickelt worden. Es sei angemerkt, dass Studien mit videobasierten SJTs *keine* Aufnahme in die Meta-Analyse erfuhren.

Als zentrales Resultat der Meta-Analyse von McDaniel et al. (2001) ergab sich eine geschätzte mittlere Korrelation von .34 (Populationseffekt) zwischen SJTs und Arbeitsleistung. Bei der Interpretation dieses Ergebnisses merkten McDaniel et al. (2001) an, dass eine große Variabilität in Bezug auf die vorgefundenen Validitäten bestünde. Ferner warf das Team um McDaniel ein, dass der vorgefundene Validitätswert von .34 als konservative Schätzung anzusehen sei, da keine Varianzeinschränkungen korrigiert wurden. Den Anteil an Varianz, welcher durch methodische Artefakte erklärbar sei, bezifferten McDaniel et al. (2001) mit 45 %.

An ihrer eigenen Meta-Analyse wurden von McDaniel et al. (2001) vor allem zwei Punkte als bedenkenswert erachtet: Der erste Punkt bezieht sich auf die geringen Stichprobengrößen in einigen einbezogenen Studien. Darüber hinaus betrifft der zweite Aspekt die Tatsache, dass viele der einbezogenen Studien nie in Zeitschriften mit peer-review veröffentlicht beziehungsweise nur auf Konferenzen vorgestellt wurden.

Sechs Jahre später führten McDaniel et al. (2007) eine weitere Meta-Analyse durch. Aufgrund des gewachsenen Forschungsinteresses an SJTs konnten die Ergebnisse von 24756 Probanden und 118 Korrelationskoeffizienten zwischen SJTs und Arbeitsleistung beachtet werden. Wie bei der oben beschriebenen Meta-Analyse wurden keine Studien mit videobasierten SJTs berücksichtigt. Im Gegensatz zur Meta-Analyse von McDaniel et al. (2001) wurden a priori keine Reliabilitäten auf Seiten der SJTs korrigiert, sondern der Messfehler auf Seiten des Kriteriums, der Beurteilung der Arbeitsleistung. Die Autoren begründeten diesen Schritt wie folgt:

„We declined to make corrections for measurement error in the SJTs because we argued that most situational judgment tests are heterogenous, yet most reported reliabilites for such tests are coefficient alphas. We encourage researchers to consider better ways to estimate the reliability of SJTs” (McDaniel et al., 2007; S. 84).

Es ergab sich mit einem Wert von .26 eine leicht niedrigere mittlere korrigierte kriterienbezogene Validität als in der vorgehenden Meta-Analyse von McDaniel et al. (2001). Durch methodische Artefakte wurde ein Varianzanteil von 38 % erklärt.

Bei der Diskussion ihrer Ergebnisse betonten McDaniel et al. (2007) das Problem des publication bias: Es werden vor allem solche Studien veröffentlicht und damit für Meta-Analysen zugänglich, welche hypothesenkonforme signifikante Ergebnisse hervorbringen. Ferner wiesen sie darauf hin, dass in den meisten Studien ein konkurrentes Design mit Angestellten als Probanden vorlag, so dass Prädiktor- und Kriterienwerte simultan erhoben wurden. Prädiktive Designs mit Bewerbern als Probanden bilden ihrer Meinung nach die Minorität.

Um die kriterienbezogene Validität von SJTs mit denen anderer Verfahren, welche in der Berufseignungsdiagnostik eingesetzt werden, einschätzen zu können, bietet die Abbildung 2-5 einen Überblick über die Validitätswerte. Die jeweiligen Werte wurden – mit der Ausnahme derjenigen von SJTs (.26: s. o.) – einer Meta-Analyse von Schmidt und Hunter (1998) entnommen.

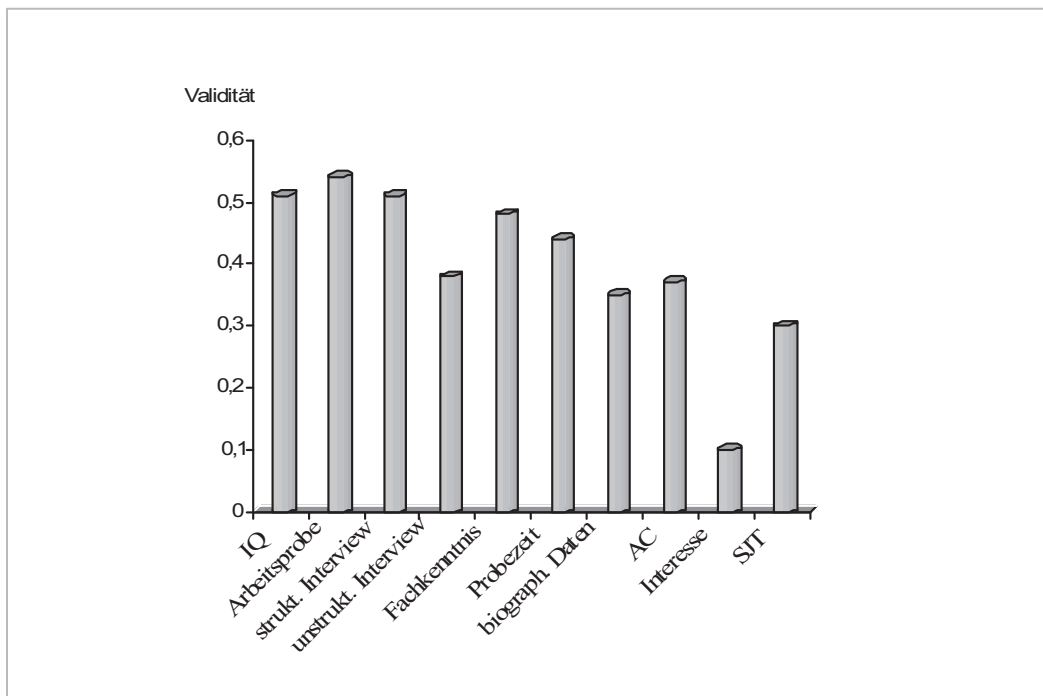


Abbildung 2-5: Validitäten von Verfahren der Personalauswahl im Überblick

2.5.2 Die Konstruktvalidität von SJTs

Die Auseinandersetzung mit Studien und Erkenntnissen zur Konstruktvalidität ist eng verbunden mit der Frage, was SJTs messen (können). Bezüglich dieses Themas stehen sich innerhalb der relevanten Forschungsliteratur fünf verschiedene Positionen gegenüber. Insbesondere vor dem Hintergrund, dass SJTs überzeugende kriterienbezogene Validitäten (vgl. McDaniel et al., 2001; McDaniel et al., 2007; s. Abschnitt 2.5.1) erzielen können, scheint eine Klärung der Konstruktvalidität von SJTs unumgänglich, denn „[...]the validity of a method of measurement is not meaningful outside the context of some construct“ (Schmitt & Chan, 2006; S. 136).

Nachfolgend werden die fünf Positionen samt ihrer Kernaussagen vorgestellt:

SJTs messen ein neuartiges Konstrukt, welches als unabhängig anzusehen ist von der allgemeinen kognitiven Leistungsfähigkeit, der Berufserfahrung und verschiedenen Persönlichkeitsfaktoren. Empirische Evidenz für diese Position liegt in Form verschiedener Studien vor, welche SJTs inkrementelle Validität nachweisen.

Mit SJTs wird nichts Anderes als die *allgemeine kognitive Leistungsfähigkeit* gemessen. Vertreter dieser Position verweisen auf signifikante Korrelationen unterschiedlicher Höhe zwischen SJTs und Skalen zur Messung allgemeiner kognitiver Leistungsfähigkeit.

SJTs erfassen vor allem *berufsbezogenes Wissen*. Als Beweislage führen Anhänger dieser Position Studien mit Korrelationen zwischen SJTs und Messungen des berufsbezogenen Wissens an, in denen berufsbezogenes Wissen über Skalen zur Erfassung der Berufserfahrung operationalisiert wird. Dabei wird Berufserfahrung als Antezedent bei der Aneignung berufsbezogenen Wissens verstanden.

SJTs messen verschiedene Aspekte der *Persönlichkeit*. Auch für die Unterlegung dieser Position werden Korrelationen mit unterschiedlichen Persönlichkeitsfaktoren angeführt.

SJTs sind vorrangig als *Methode* zu betrachten und können derart konstruiert werden, dass mit ihnen eine Vielzahl unterschiedlicher Konstrukte gemessen werden kann.

Im weiteren Verlauf werden die einzelnen Positionen detaillierter exploriert. An diesen Vorgang knüpft eine Zusammenfassung und Integration respektive Synthese der unterschiedlichen Sichtweisen an.

SJTs messen ein neuartiges Konstrukt

Forscher, welche diesen Ansatz unterstützen, proklamieren, dass das Bewältigen realer Situationen, welche in SJTs abgebildet und simuliert werden sollen, ein komplexeres, vielschichtigeres Verhaltensmuster voraussetze als die Lösung eines Items in einem Test zur Erfassung allgemeiner kognitiver Leistungsfähigkeit oder die Bewertung einer Aussage in einem Persönlichkeitsinventar. Sie setzen ihre Argumentation mit der These fort, dass sich das Handeln und Zurechtfinden in realen Problemsituationen grundsätzlich vom Handeln in akademischen Problemsituationen unterscheidet. Im Gegensatz zu akademischen Problemen, die klar strukturiert seien und eine eindeutige Lösung besäßen, stünden zur Lösung realer Probleme kaum Informationen zur Verfügung (Chan, 2000). Es läge keine augenscheinlich richtige Lösung für sie vor, sondern es seien vielmehr mehrere unterschiedlich effektive Lösungen vorstellbar (Hedlund & Sternberg, 2000). Vergleichbar deutete McClelland (1973) an,

dass Tests zur allgemeinen kognitiven Leistungsfähigkeit Probanden mit eindeutigen, genau umrissenen und strukturierten Aufgaben (well-defined) konfrontieren, während reale Situationen mehrdeutig und vielschichtiger (ill-defined) wären.

Fürsprecher dieses Ansatzes stellen nicht in Frage, dass SJTs einzelne Anteile gut erforschter Konstrukte wie Berufserfahrung oder kognitiver Leistungsfähigkeit messen (Weekley & Ployhart, 2005; Schmitt & Chan, 2006). Jedoch akzentuieren sie, dass SJTs zusätzlich ein neuartiges Konstrukt messen und verweisen zur Legitimation ihrer Thesen auf Studien, welche den *Nachweis inkrementeller Validität* erbringen (z. B. Weekley & Jones, 1997, 1999; Clevenger et al., 2001; Weekley & Ployhart, 2005; McDaniel et al., 2007).

In ihrer bereits ausführlich beschriebenen Studie (s. Abschnitt 2.4) ermittelten Weekley und Jones (1997) neben anderen Gütekriterien die inkrementellen Validitäten der zwei eingesetzten videobasierten SJTs. Dabei ergab sich in einer ersten Teilstudie ein Validitätszuwachs von rund 3 % durch die Aufnahme eines SJTs in die Regressionsgleichung mit den Prädiktoren kognitive Leistungsfähigkeit und Berufserfahrung. In einer weiteren Teilstudie, welche vorrangig auf den Vergleich unterschiedlicher Bewertungsschlüssel für Probandenantworten im SJT abzielte, brachte ein SJT mit empirischem Bewertungsschlüssel eine inkrementelle Validität von fast 6 % gegenüber kognitiver Leistungsfähigkeit und Berufserfahrung zustande. Für denselben SJT ergab sich gar ein Validitätszuwachs von rund 10 %, wenn statt des empirischen Bewertungsschlüssels ein rationaler Bewertungsschlüssel eingesetzt wurde.

Zwei Jahre später führten Weekley und Jones (1999) zwei weitere Teilstudien zur Ergründung inkrementeller Validitäten von SJTs durch. Anders als bei der Vorläuferstudie aus dem Jahre 1997 liefen die SJTs jedoch nicht mehr videobasiert, sondern in der Paper-Pencil-Version ab. Als Ergebnis einer ersten Teilstudie konnte hier ein Validitätszuwachs von etwas mehr als 3 % festgehalten werden. Der SJT wurde in diesem Fall als letzter Prädiktor nach kognitiver Leistungsfähigkeit und Berufserfahrung in die Regressionsgleichung aufgenommen. Mit einem Validitätszuwachs von knapp 1 % ergab sich in der zweiten Teilstudie eine geringere inkrementelle Validität als in den anderen Studien der beiden amerikanischen Forscher.

Clevenger, Pereira, Wiechmann, Schmitt und Schmidt Harvey (2001) ermittelten an drei verschiedenen Stichproben die inkrementellen Validitäten dreier SJTs gegenüber weiteren involvierten Prädiktoren. Für jede Stichprobe wurde ein spezifischer SJT konstruiert. Zur Berechnung der inkrementellen Validitäten wurden schrittweise Regressionen berechnet, in die der SJT stets als letzter Prädiktor seine Aufnahme erfuhr. Bei der Vorhersage der Arbeitsleistung erklärte der erste SJT einen signifikanten ($p < .05$) Anteil an zusätzlicher Varianz von rund 3 % gegenüber den anderen beteiligten Prädiktoren – kognitive Leistungsfähigkeit, Gewissenhaftigkeit, Berufserfahrung und berufsbezogenes Wissen. Für den zweiten SJT ergab sich bei Einbeziehung in die schrittweise Regression ein signifikanter ($p < .10$)⁸ zusätzlich aufgeklärter Varianzanteil von rund 2 %. Die anderen beteiligten Prädiktoren waren Gewissenhaftigkeit, Berufserfahrung, berufsbezogenes Wissen sowie zwei berufsspezifische Maße. Im Gegensatz zu den beiden anderen SJTs von Clevenger et al. (2001) konnte für den dritten SJT nicht der Nachweis signifikanter inkrementeller Validität aufgezeigt werden. Ergänzend zum SJT und einer Arbeitsprobe waren Skalen zur Messung von Gewissenhaftigkeit, kognitiver Leistungsfähigkeit und Berufserfahrung Bestandteile der Testbatterie.

Zur Vorhersage der Arbeitsleistung von Managern im Einzelhandel zogen Weekley und Ployhart (2005) neben Skalen zur Messung der kognitiven Leistungsfähigkeit, der Berufserfahrung und der Big 5 einen SJT heran. Zur Anleitung der Probanden im SJT setzten sie eine Instruktion zum Wissensausdruck ein, nach der die Teilnehmer der Studie je Situation die beste und die schlechteste Verhaltensalternative markieren sollten. Um die inkrementelle Validität des SJTs bestimmen zu können, führten Weekley und Ployhart (2005) eine sequenzielle multiple hierarchische Regression durch. Diese Form der Regression lässt später einbezogenen Prädiktoren nur solche Varianzanteile aufklären, welche durch vorhergehende Prädiktoren nicht vorhergesagt werden konnten. Der SJT wurde als finaler Prädiktor in die Regression involviert und erbrachte einen Zuwachs an aufgeklärter Varianz von 2 %. Im Rahmen ihrer Meta-Analyse unterschieden McDaniel et al. (2007) drei verschiedene Szenarien bei der Er-

⁸ Das von Clevenger et al. (2001) an dieser Stelle gewählte Signifikanzniveau erscheint ungewöhnlich.

mittlung inkrementeller Validitäten für SJTs. Sie berechneten einerseits die inkrementelle Validität separat in Bezug auf einen einzigen anderen Prädiktor – entweder die kognitive Leistungsfähigkeit oder die Persönlichkeit operationalisiert über die Big 5 – und andererseits in Bezug auf einen Prädiktor kombiniert aus kognitiver Leistungsfähigkeit und den fünf allgemeinen Persönlichkeitsfaktoren. Je nach im SJT verwandter Instruktion ergaben sich gegenüber der kognitiven Leistungsfähigkeit Zuwächse an erklärter Varianz von 3 % (Instruktion zum Wissensausdruck) bis 5 % (Instruktion zur Verhaltenstendenz). Gegenüber den Faktoren aus dem Fünf-Faktoren-Modell der Persönlichkeit erbrachten SJTs Anteile zusätzlich erklärter Varianz von 6 % (Instruktion zur Verhaltenstendenz) bis 7 % (Instruktion zum Wissensausdruck). Bei der Einbeziehung einer Kombination aus kognitiver Leistungsfähigkeit und den Big 5 konnten SJTs, ungeachtet ihrer Instruktionsform, einen Beitrag von 2 % an zusätzlicher Varianzaufklärung für sich verbuchen.

O'Connell, Hartman, McDaniel, Grubb und Lawrence (2007) untersuchten die inkrementelle Validität eines SJTs bei der differentiellen Vorhersage von task und contextual performance (vgl. Borman & Motowidlo, 1993). Der SJT war eingebettet in eine computergestützte Testbatterie, welcher zudem ein kognitiver Leistungstests und fünf Persönlichkeitskaalen angehörten. Bei der Vorhersage von task performance im Verbund mit kognitiver Leistungsfähigkeit ermöglichte der SJT einen Anstieg der aufgeklärten Varianz um 3 %. Wurden statt kognitiver Leistungsfähigkeit die fünf Persönlichkeitsdimensionen als Prädiktor neben dem SJT eingesetzt, ergab sich durch den SJT ein Anwachsen der erklärten Varianz auf Seiten des Kriteriums um 4 %. Geringer (1 %) fiel der inkrementelle Nutzen des SJTs aus, wenn er neben sowohl der kognitiven Leistungsfähigkeit als auch den fünf Persönlichkeitsdimensionen zur multiplen Regression auf task performance herangezogen wurde. In Bezug auf die Vorhersage von contextual performance konnte für den SJT nur dann eine inkrementelle Validität – 4 % an zusätzlich aufgeklärter Varianz – nachgewiesen werden, wenn er ergänzend zur kognitiven Leistungsfähigkeit Aufnahme in die Regression erhielt. Weder für die fünf Persönlichkeitsdimensionen noch für die Kombination aus kognitiver Leistungsfähigkeit und den Persönlichkeitsdimensionen führte ein Hinzuziehen des SJT zu einem Anstieg an erklärter Varianz. Weitere empirische Bestätigungen, dass SJTs

bedeutsamen inkrementelle Validitäten erzielen können, finden sich unter anderem bei Stevens und Champion (1999), Chan und Schmitt (2002) oder Oswald, Schmitt, Ramsay und Gillespie (2004).

Da das Aufzeigen inkrementeller Validitäten weiterhin die Frage unbeantwortet lässt, was SJTs messen, soll im Nachfolgenden das Konstrukt *praktische Intelligenz* (practical intelligence) vorgestellt werden, welches von vielen Forschern (z. B. Motowidlo et al., 1990; Pulakos, Arad, Donovan & Plamondon, 2000; Sternberg et al., 2000; Chan & Schmitt, 2002) als dasjenige einzigartige Konstrukt bezeichnet wird, welches SJTs messen. Zuvor soll jedoch ein kritischer Blick auf die soeben dargestellten Studien zur inkrementellen Validität von SJTs geworfen werden. Im Rahmen der Diskussion ihrer Studie (s. o.) sprachen O'Connell et al. (2007; S. 28 ff.) selbst zwei wichtige Punkte zur Einordnung der inkrementellen Validitäten von SJTs an: Zum einen widmeten sie dem Punkt des publication bias (s. Abschnitt 2.5.1) Beachtung, indem sie die Vermutung äußerten, dass nur solche Studien publiziert werden, welche SJTs inkrementelle Validität nachweisen können. Zum anderen erwähnte die amerikanische Forschergruppe, dass die inkrementelle Validität eines SJT stark von der Zusammensetzung der Stichprobe und von den Korrelationen des SJTs mit den anderen Prädiktoren abhängt. Dies erschwere stichprobenübergreifende Aussagen über die inkrementelle Validität von SJTs. Ferner ist hinzuzufügen, dass für die angeführten Studien und Meta-Analysen keine Bewerberdaten gewonnen werden konnten – eine Ausnahme bildet teilweise die Studie von Clevenger et al. (2001). Aus Sicht des Autors ist zudem zu differenzieren zwischen statistischer und praktischer Signifikanz. Viele der genannten inkrementellen Validitäten verfehlen trotz gegebener statistischer Signifikanz den Anspruch praktischer Bedeutsamkeit, wenn der finanzielle und personelle Aufwand bei der Entwicklung eines SJT beachtet werden.

Wagner und Sternberg (1985) prägten das Konzept *tacit knowledge* und führten folgende Definition an: „*Such knowledge is typically not directly taught or spoken about.[...]much tacit knowledge is probably disorganized, informal, and relatively inaccessible, making it potentially ill-suited for direct instruction*” (S. 438-439). Sie ließen der Aneignung und Anwendung des tacit knowledge eine entscheidende Rolle bei der Bewältigung von Problemen in realen Situationen zukommen (s. o.). In nachfol-

genden Studien arbeitete Sternberg mit verschiedenen anderen Forschern das Konzept des tacit knowledge weiter aus (z. B. Sternberg, 1997; Sternberg et al., 2000; Sternberg & Horvath, 1999; Sternberg et al., 1995). Dabei assoziierte Sternberg in seinen Arbeiten tacit knowledge mit erfahrungsgebundenem Wissen, welches instinktiv oder intuitiv zur Zielerreichung eingesetzt wird, und schlug eine Brücke zur Experten-Novizen-Forschung (vgl. Chi, Glaser und Farr, 1988). Übertragen auf tacit knowledge folgte Sternberg, dass es sich um kontext-spezifisches Wissen handele, was in der momentanen Situation und unter den gegebenen Umständen zu tun sei. Tacit knowledge sei prozeduraler Natur (vgl. Anderson, 1983) und könne, obgleich es in vielen Situationen eine handlungsleitende Funktion innehat, nicht explizit artikuliert werden. Dass Sternberg tacit knowledge als prozedural und nicht deklarativ ansah, wird weiterhin darin deutlich, dass tacit knowledge in seinen Augen nicht durch direkte Instruktion, zum Beispiel in der Schule oder in einer Personalentwicklungsmaßnahme, erlernbar sei, sondern unbewusst aufgrund von konkreten Erfahrungen aufgebaut würde. Folgerichtig setzte er tacit knowledge nicht mit arbeitsplatzbezogenem Wissen gleich, da diese Art von Wissen sowohl prozedurale als auch deklarative Komponenten enthalte. Im Rahmen späterer Arbeiten band Sternberg (vgl. Sternberg & Heidlund, 2002) tacit knowledge als zentralen Bestandteil in das Konzept der praktischen Intelligenz (practical intelligence) ein. In der relevanten Forschungsliteratur werden die beiden Begriffe unscharf getrennt beziehungsweise synonymisch verwandt (Schmidt & Hunter, 1993). Zur Definition von praktischer Intelligenz soll auf ein Zitat von Sternberg und Heidlund (2002, S. 145) zurückgegriffen werden:

Practical intelligence is defined as the ability that individuals use to find a more optimal fit between themselves and the demands of the environment through adapting to the environment, shaping (or modifying) the environment, or selecting a new environment in the pursuit of personally-valued goals

Schmitt und Chan (2006) reicherten diese Definition an, indem sie praktische Intelligenz als latent treibende Kraft bei der Anpassungsfähigkeit in einer Vielzahl an Situationen und bei der Aneignung kontextuellen Wissens deuteten. Ferner unterstrichen sie in Übereinstimmung mit Sternberg

die Notwendigkeit praktischer Erfahrungen bei der Akquise praktischer Intelligenz.

Zur Messung von tacit knowledge entstanden verschiedene Tests, zum Beispiel das Tacit Knowledge Inventory for Managers (TKIM) von Wagner und Sternberg (1985) oder das Tacit Knowledge Inventory for Military Leaders (Heidlund et al., 1998). Obgleich diese Tests nicht explizit als SJTs benannt wurden, entsprechen sie hinsichtlich ihres Aufbaus und ihrer Ausrichtung gemäß McDaniel und Nguyen (2001) SJTs. Aus rund zwanzig Jahren Forschung liegen Studien mit Tests zur Messung von tacit knowledge vor. Die wichtigsten Erkenntnisse sollen nun genannt und einer kritischen Analyse unterzogen werden.

Für Tests zur Messung von tacit knowledge fand sich nahezu durchgängig ein moderater Zusammenhang mit Berufserfahrung, was sich in vorgefundenen Korrelationen zwischen .20 und .40 ausdrückt. Wagner et al. (1999) führen zum Beispiel eine signifikante Korrelation von .31 ($n = 45^9$) zwischen Berufserfahrung im Vertrieb und tacit knowledge an. In früheren Studien zum tacit knowledge (z. B. Wagner, 1987; Sternberg, Wagner & Okagaki, 1993) wurde übereinstimmend eine Unabhängigkeit von kognitiver Leistungsfähigkeit postuliert und mit nicht-signifikanten Korrelationen geringer Höhe untermauert. Diese Erkenntnis revidierten Sternberg und Heidlund (2002) in ihrer Überblicksarbeit und beziehen sich auf eine Studie von Sternberg et al. (2000), in denen an einer Stichprobe von Armeeführern in Kenia je nach Hierarchieebene signifikant negative oder positive Korrelationen im moderaten Bereich mit verbaler Intelligenz ermittelt wurden. Der Studie von Sternberg et al. (2000) entstammt zudem ein Nachweis inkrementeller Validität: Die zusätzliche Varianzerklärung bei der Vorhersage von Führungserfolg wird je nach untersuchter Hierarchieebene mit 4-6 % beziffert. Bezüglich der Beziehung von tacit knowledge und verschiedenen allgemeinen Persönlichkeitsfaktoren liegen nur wenige Studien vor. Wagner und Sternberg (1990) brachten exemplarisch in einer Studie, in der sie mehrere Persönlichkeitsinventare und -tests einsetzten, keine statistisch bedeutsamen Korrelationen hervor. Verfahren zur Messung von tacit knowledge wurden auch einer kriterienbezogenen Validierung unterzogen. In diesem Zusammenhang ergab

⁹ Die Stichprobengrößen werden in diesem Kontext bewusst angeführt (s.u.).

zum Beispiel eine Studie von Wagner und Sternberg (1985) eine signifikante Korrelation von .46 ($n = 54$) zwischen dem Gehalt von Managern in der Wirtschaft und tacit knowledge beziehungsweise .48 ($n = 22$) zwischen dem Gehalt von Bankangestellten und tacit knowledge.

Zur Relativierung der berichteten (kriterienbezogenen) Validitäten von Tests zur Messung von tacit knowledge sprechen McDaniel et al. (2001; 737 ff.) verschiedene Punkte an. Zunächst betonen sie, dass Tests zur Messung von tacit knowledge nicht auf Arbeitsanalysen beruhen würden, so dass eine Verankerung der Testinhalte im konkreten Arbeitsalltag nicht gegeben sei. Zweitens merken sie an, dass in veröffentlichten Validitätsstudien von Wagner beziehungsweise Wagner und Sternberg (Wagner, 1987; Wagner & Sternberg, 1985, 1991) Stichproben mit einer Varianzeinschränkung hinsichtlich der kognitiven Fähigkeit vorgelegen haben müssen. Nach Meinung von McDaniel und seinem Team waren die in den Stichproben vertretenen Studenten von amerikanischen Eliteuniversitäten und graduierte Psychologen oder Betriebswirte a priori einer Präselektion auf Basis kognitiver Leistungsmaße unterworfen. Mit den geringen Stichprobengrößen, die eine Generalisierbarkeit der erzielten Validitäten ausschließen, bemängeln sie einen dritten Punkt des Vorbehalts gegenüber Tests zur Messung von tacit knowledge. Als vierten Kritikpunkt hinterfragen sie die im Vergleich zu sonstigen Validitätsstudien ungewöhnlichen Leistungskriterien. Der letzte Kritikpunkt, den sie erwähnen, bezieht sich auf die Bewertung der Probandenantworten. Für die Gewichtung der Probandenantworten lag kein einheitlicher, systematischer Schlüssel vor, sondern die Probandenantworten wurden mal nach theoretischen Gesichtspunkten gewichtet, mal mit einem prototypischen Expertenprofil verglichen oder in anderen Fällen nach einfachen Daumenregeln bewertet.

Um die Auseinandersetzung mit tacit knowledge beziehungsweise praktischer Intelligenz abzuschließen, soll verdeutlicht werden, dass zu wenige Studien, insbesondere Meta-Analysen, vorliegen, um den Beitrag dieser Konstrukte bei der Beantwortung der Frage, was SJTs messen, seriös einschätzen zu können. Zudem sind die beiden Konstrukte selbst mit Widersprüchen behaftet. So rücken Sternberg und Heidlund (2002) einerseits den intuitiven, unbewussten Charakter von praktischer Intelligenz in den Vordergrund, nennen aber andererseits die Möglichkeit, praktische Intelligenz gezielt zu instrumentalisieren.

SJTs und allgemeine kognitive Leistungsfähigkeit

Die allgemeine kognitive Leistungsfähigkeit wird gemeinhin als validester Prädiktor bei der Vorhersage von Arbeitsleistung angesehen. Diese Position wird ihr aufgrund einer Meta-Analyse von Schmidt und Hunter (1998) zugebilligt, in der eine korrigierte mittlere Validität von .51 ermittelt wurde (vgl. Abbildung 2-5)¹⁰.

Aus der Meta-Analyse von McDaniel et al. (2001) ist eine korrigierte mittlere Korrelation von .46 zwischen SJTs und der allgemeinen kognitiven Leistungsfähigkeit zu entnehmen. Als Grundlage für diesen Wert dienten 79 Korrelationskoeffizienten, welche an insgesamt 16984 Probanden ermittelt wurden. McDaniel und Nguyen (2001) sowie Schmitt und Chan (2006) merkten diesbezüglich an, dass nur 12 % der Varianz auf methodische Artefakte zurückzuführen seien und dass das Vertrauensintervall für die obige Korrelation einen Bereich von .17 bis .75 umfasse. Daraus folgerten sie, dass in den Studien, welche McDaniel et al. (2001) in ihre Analyse einbezogen haben, sowohl moderate bis hohe Korrelationen, aber auch niedrige Korrelationen und Nullkorrelationen vorgelegen haben müssen. Zwei mögliche Moderatoren werden von McDaniel et al. (2001) selbst angegeben: Erstens ergaben sich für SJTs, welche auf Anforderungsanalysen beruhen, höhere Korrelationen (.50) als für SJTs ohne vorgehende Anforderungsanalyse (.38). Zweitens ergaben sich für SJTs mit weniger komplexen Fragestellungen höhere Korrelationen (.56) im Vergleich zu SJTs mit komplexen, detaillierten Fragestellungen (.47). In einer späteren Metaanalyse ging McDaniel mit seinem Team (McDaniel et al., 2007) unter anderem der Frage nach, ob die Art der Instruktion ein weiterer Moderator der Beziehung zwischen SJTs und der kognitiven Leistungsfähigkeit sei. Es ergab sich für SJTs mit einer Instruktion zur Verhaltenstendenz (behavioral tendency) eine korrigierte mittlere Korrelation von .19, während für SJTs mit einer Instruktion zum Wissensausdruck (knowledge) eine deutlich höhere mittlere korrigierte Korrelation von .35 angegeben wurde. Ungeachtet der Art der Instruktion berichten McDaniel et al. (2007) für alle in die Analyse einbezogenen SJTs (95 Koeffizienten) eine mittlere korrigierte Korrelation von .32 mit kognitiver

¹⁰ Die in der Meta-Analyse von Schmidt und Hunter (1998) berichteten Validitäten werden von einigen Forschern (z. B. Schuler & Höft, 2006) kritisch betrachtet.

Leistungsfähigkeit. Die Grenzen des Vertrauensintervalls für diese Korrelation steckten sie mit .08 und .57 ab, so dass auch hier von einer großen Variabilität bezüglich der einbezogenen Einzelkorrelationen auszugehen ist.

SJTs und Berufserfahrung

Berufserfahrung ist als facettenreiches Konstrukt zu verstehen, bei dem die Art der Operationalisierung einen Einfluss auf die Beziehung zur Arbeitsleistung ausübt (Quinones, Ford & Teachout, 1995; Tesluk & Jacobs, 1998). Frühere Meta-Analysen erbrachten differierende Ergebnisse bezüglich der Beziehung von Berufserfahrung und Arbeitsleistung. So berichten Hunter und Hunter (1984) eine mittlere Korrelation von .18, während der Meta-Analyse von McDaniel, Schmidt und Hunter (1988) eine mittlere korrigierte Korrelation in Höhe von .32 entspringt. Quinones et al. (1995) differenzierten auf Seiten der in ihre Meta-Analyse einbezogenen Studien verschiedene Arten, wie Berufserfahrung gemessen wird. Über alle 44 beachteten Studien hinweg gaben sie eine mittlere korrigierte Korrelation von .27 an. Besonders hohe mittlere korrigierte Korrelationen ergaben sich in Studien, in denen Berufserfahrung auf einem spezifischen Aufgabenniveau (.41) oder anhand der Anzahl ausgeführter Aufgaben (.43) gemessen wurde. Zudem ergaben sich für Berufserfahrung signifikant höhere Korrelationen mit harten Leistungskriterien (.39) als mit weichen Leistungskriterien (.24).

McDaniel und Nguyen (2001) wagten die Prognose, dass sich vor allem dann bedeutungsvolle Korrelationen zwischen SJTs und berufsbezogenem Wissen ergeben sollten, wenn zum einen innerhalb der Stichprobe Varianz bezüglich der Berufserfahrung vorläge und zum anderen die Berufserfahrung und keine andere Variable wie zum Beispiel der akademische Bildungshintergrund die wichtigste Determinante des berufsbezogenen Wissens wäre. Andere Forscher (Schmidt, Hunter & Outerbridge, 1986; McDaniel, Schmidt & Hunter, 1988) betonten, dass Unterschiede bezüglich des berufsbezogenen Wissens stärker ins Gewicht fallen zwischen einer Person mit einjähriger Berufserfahrung und einer Person mit sechsjähriger Erfahrung als zwischen einer Person mit zwölfjähriger Berufserfahrung und einer anderen Person mit sechzehnjähriger Berufserfahrung. Der zur Vorhersage von Arbeitsleistung relevante Anteil an berufsbezo-

genem Wissen werde in den ersten Jahren der Berufstätigkeit erworben und jedes weitere Jahr an Berufserfahrung erweitere das berufsbezogene Wissen nur noch um Nuancen (McDaniel & Nguyen, 2001).

McDaniels und Nguyen (2001) konnten im Rahmen ihrer Meta-Analyse auf 17 Studien zur Beziehung von SJTs und berufsbezogenem Wissen zurückgreifen. Es ergab sich eine mittlere korrigierte Korrelation von $r = .07$. Das Konfidenzintervall für diese Korrelation erstreckte sich von $-.14$ bis $.29$, so dass von McDaniel und Nguyen (2001) auch solche Einzelstudien berücksichtigt wurden, in denen eine hohe Ausprägung in Bezug auf das berufsbezogene Wissen mit niedrigen SJT-Scores einherging.

In der bereits oben erwähnten Studie von Clevenger et al. (2001) mit drei verschiedenen Stichproben wurden die SJTs sowohl mit berufsbezogenem Wissen als auch mit Berufserfahrung korreliert. Je nach Stichprobe ergaben sich folgende Korrelationen zwischen dem SJT und berufsbezogenem Wissen: $r = .13$ (erste Stichprobe), $r = .19$ (zweite Stichprobe) und $.31$ (dritte Stichprobe). Indessen erreichten die Korrelationen zwischen SJT und Berufserfahrung für keine Stichprobe statistische Signifikanz.

Ähnlich wie Clevenger et al. (2001) wählten Weekley und Ployhart (2005) verschiedene Messungen zur Feststellung der Berufserfahrung ihrer Probanden. Sie erhoben die allgemeine Berufserfahrung, die Anzahl der Jahre in der momentanen Position sowie drittens die Trainingserfahrung. Für den von Weekley und Ployhart (2005) eingesetzten SJT ergaben sich signifikante Korrelationen ($p < .05$) mit allen drei eingesetzten Maßen der Berufserfahrung: Allgemeine Berufserfahrung ($.21$), Anzahl der Jahre in der momentanen Position ($.13$) und Trainingserfahrung ($.12$). Anschließende Analysen von Weekley und Ployhart (2005) ergaben weiterhin, dass der Einfluss von allgemeiner Berufserfahrung auf die berufliche Leistung vollständig durch den SJT mediiert wird. Aufgrund dieses Ergebnisses kamen die Autoren zu folgender Schlussfolgerung: „[...]our results suggest this SJT appears to reflect general forms of knowledge“ (Weekley & Ployhart, 2005; S. 99).

SJTs und Persönlichkeit

Ein gesteigertes Interesse am Einsatz von Persönlichkeitsskalen in der Berufseignungsdiagnostik ist einerseits zurückzuführen auf das Aufzeigen zufrieden stellender kriterienbezogener Validitäten in mehreren Über-

blicksarbeiten und Meta-Analysen (Barrick & Mount, 1991; Tett, Jackson & Rothstein, 1991; Salgado, 1997; Motowidlo, Borman & Schmitt, 1997; Beaty, Cleveland & Murphy, 2001) und andererseits erklärbar durch den Nachweis, dass in Persönlichkeitstests eine größere Chancengleichheit für unterschiedliche Stichproben herrscht als in Verfahren zur Messung kognitiver Leistungsfähigkeit (Bobko, Roth & Potosky, 1999). Dabei wird zur Vorhersage der Arbeitsleistung zumeist auf ein fünffaktorielles Modell der Persönlichkeit zurückgegriffen (McCrae & Costa, 1987; Goldberg, 1990; Tupes & Christal, 1992). Die fünf Faktoren dieses Modells werden im deutschsprachigen Raum mit Gewissenhaftigkeit, Verträglichkeit, Extraversion, Emotionale Stabilität und Offenheit für Erfahrungen betitelt. Zur inhaltlichen Erläuterung der fünf Faktoren sei an dieser Stelle beispielhaft auf das Manual zum NEO-FFI (Borkenau & Ostendorf, 2008; vgl. Kapitel 6) verwiesen. Der Einsatz von Persönlichkeitsverfahren in der Berufseignungsdiagnostik wird vor allem in Hinblick auf das Problem des Faking kritisch diskutiert (z. B. Ones & Viswesvaran, 1998; Snell, Sydell & Lueke, 1999). In einigen Laborstudien ergaben sich beispielsweise Mittelwertdifferenzen von bis zu einer Standardabweichung zwischen Probanden, die zu einer besonders positiven Selbstdarstellung aufgefordert wurden und solchen Probanden, welche ehrlich antworten sollten (Viswesvaran & Ones, 1999; McFarland & Ryan, 2000).

Bei einer Meta-Analyse, welche nur eine mögliche Fehlerquelle (sampling error) korrigierte, fanden McDaniel und Nguyen (2001) signifikante mittlere Korrelationen zwischen SJTs und Emotionaler Stabilität ($r = .31$), SJTs und Gewissenhaftigkeit ($r = .26$) sowie SJTs und Verträglichkeit ($r = .25$). Um die Ergebnisse zu relativieren, erwähnten die beiden Autoren zum einen die geringe Anzahl einbezogener Studien und zum anderen die weiten Vertrauensintervalle für die mittleren Korrelationen sowie die Tatsache, dass jeweils nur ein geringer Anteil an Varianz (2.5 bis 4.5 %) durch das berücksichtigte methodische Artefakt erklärbar war.

Interessanterweise sind Gewissenhaftigkeit, Verträglichkeit und Emotionale Stabilität auch diejenigen Persönlichkeitsfaktoren, welche substantielle Beiträge bei der Aufschlüsselung der Konstrukte Kundenorientierung (Frei & McDaniel, 1998), Integrität (Ones, Viswesvaran & Schmidt, 1993) und kontraproduktives Verhalten am Arbeitsplatz (Hough, 1992,

Salgado, 2002) leisten. Ones (1993) bezeichnete den Verbund dieser drei Faktoren als g-Faktor der Persönlichkeit.

Im Vergleich zur oben skizzierten Meta-Analyse von McDaniel und Nguyen (2001) konnten McDaniel et al. (2007) sechs Jahre später eine größere Anzahl an Studien einbeziehen. In der Tabelle 2-1 sind die vorgefundenen Korrelationen zwischen SJTs und den fünf relevanten Persönlichkeitsfaktoren aufgelistet. Je nach eingesetzter Instruktion im SJT, variierte die Höhe der Korrelationen (vgl. Abschnitt 2.3).

Tabelle 2-1: Mittlere Korrelationen zwischen SJTs und Persönlichkeitsfaktoren (nach McDaniel et al., 2007; S. 73)

Persönlichkeitsfaktor	Mittlere Korrelation
Gewissenhaftigkeit	$r = .27$
Verträglichkeit	$r = .25$
Emotionale Stabilität/Neurotizismus	$r = .22$
Extraversion	$r = .14$
Offenheit für Erfahrungen	$r = .13$

Wie der Tabelle 2-1 entnommen werden kann, resultiert für den Faktor Gewissenhaftigkeit die in Relation zu den anderen Faktoren höchste mittlere Korrelation mit SJTs. Gewissenhaftigkeit ist auch diejenige Persönlichkeitsdimension aus dem Verbund der Big 5, welche in vielen Studien (z. B. Mount & Barrick, 1995; Dudley, Orvis, Lebiecki & Cortina, 2007) den bedeutsamen Beitrag zur Vorhersage der Arbeitsleistung liefert.

SJTs als Messmethode

McDaniel und Nguyen (2001) betrachteten Studien, die nachzuweisen versuchen, dass einem SJT ein einziges Konstrukt (z. B. Sternberg et al., 1995) zugrunde läge, als wenig erfolgsversprechend. Sie stimmten vielmehr mit anderen Forschern (z. B. Chan & Schmitt, 1997; Weekley & Jones, 1999; O`Connell et al., 2007) in der Feststellung überein, dass SJTs multidimensionaler Natur wären. Ihre Argumentation setzten sie wie folgt fort: „*Like other measurement methods, such as the employment interview or the job knowledge testing, situational judgment tests can be build to measure a variety of constructs*“ (McDaniel & Nguyen; S. 107). Chan und Schmitt (2005) erwähnten daran anknüpfend, dass einige Konstrukte –

beispielsweise soziale Anpassungsfähigkeit oder kontextbezogenes Wissen – besonders für die Messung mit SJTs geeignet wären.

Der Begriff *Multidimensionalität* ist bezogen auf SJTs unterschiedlich auslegbar. Auf der einen Seite bedeutet dies, dass mit SJTs mehrere Konstrukte erfasst werden können, welche zum Beispiel bei der Vorhersage beruflicher Leistung dienlich sind. Auf der anderen Seite führt die Multidimensionalität dazu, dass Probandenantworten je nach Item im SJT durch unterschiedliche Konstrukte beeinflusst werden können. Die angesprochene Multidimensionalität von SJTs bedingt interne Konsistenzen, welche unter denen anderer Personalauswahlinstrumente liegen (vgl. Smiderle et al., 1994; Kanning, 2005), und Faktorlösungen, die eine eindeutige Interpretation erheblich erschweren beziehungsweise nicht zulassen (z. B. Chan & Schmitt, 2002; McDaniel & Whetzel, 2005). Die Haltung von McDaniel und Nguyen aufgreifend wiesen Kanning et al. (2006) darauf hin, dass die Items eines SJT einem kriterienorientierten Entwicklungsprozess entsprängen und weniger theorie- oder konstruktgeleiteten Überlegungen. Items eines SJT sollten ihrer Meinung nach vorrangig erfolgskritische Situationen aus dem Berufsalltag abbilden. Da der Arbeitsalltag hinsichtlich seiner Anforderungen heterogen und komplex sei, müssten die Items eines SJT dieser Komplexität und Vielfalt gerecht werden. Dementsprechend erfordere die erfolgreiche Bewältigung der Items in einem SJT eine Vielzahl an Fähigkeiten und Kompetenzen, welche auch am konkreten Arbeitsplatz vonnöten seien. Übereinstimmend mit Kanning und seinem Team äußerten Lievens und Sackett (2007), dass SJTs „[...]are commonly designed to sample key domain aspects, rather than to reflect a clearly understood construct“ (S. 1043).

Nach der Darstellung der fünf divergierenden Positionen soll in Form eines Zitats von Schmitt und Chan (2006; S. 148) eine Synthese der Sichtweisen ermöglicht werden.

[...]the answer is “both“, in the sense that although SJTs can be construed as a method that can be used to assess different constructs, the core characteristics of the test content of typical SJTs impose constraints on the range of constructs being assessed and the dominant constructs are probably conceptually distinct from established constructs such as cognitive ability and personality traits.

McDaniel, Whetzel, Hartman, Nguyen und Grubb (2006; S. 194 ff.) fassten die vorliegenden Erkenntnisse zur Konstruktvalidität und kriterienbezogenen Validität (s. Abschnitt 2.5.1) von SJTs in Form eines Modells zusammen. Dabei dachten sie diesem Modell eher die Funktion einer Heuristik zur Orientierung zu als die einer vollständig empirisch abgesicherten Rahmenstruktur. Die verschiedenen (Wirk-)Zusammenhänge, welche ihr Modell aufgreift, sind in der Abbildung 2-6 enthalten.

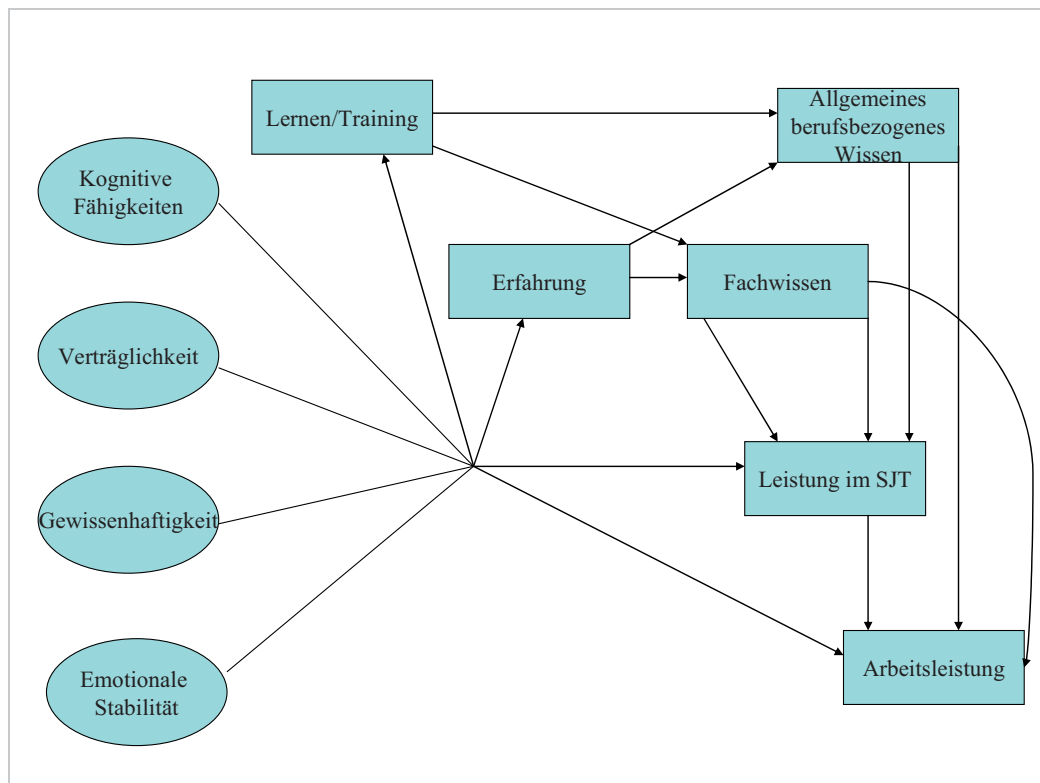


Abbildung 2-6: Heuristisches Modell von McDaniel et al. (2006)

Nach den Prämissen von McDaniel et al. (2006) üben die Persönlichkeitsfaktoren Verträglichkeit, Gewissenhaftigkeit und Emotionale Stabilität sowie die Intelligenz sowohl einen direkten als auch einen indirekten – vermittelt über Variablen wie Lernen/Training und Erfahrung – Effekt auf die individuelle Leistung in einem SJT aus. Eine direkte Einwirkung auf die SJT-Performanz geht weiterhin von Fachkenntnissen und der allgemeinen beruflichen Erfahrung aus. Die (allgemeine) Arbeitsleistung wiederum kann durch die drei genannten Persönlichkeitsfaktoren, Intelligenz, Fachkenntnisse, Berufserfahrung und die Leistung im SJT valide vorhergesagt werden.

2.5.3 Soziale Validität von SJTs und Testfairness

In der relevanten Forschungsliteratur ist eine Vielzahl an Theorien (für einen Überblick: Ryan & Ployhart, 2000) auffindbar, welche die subjektive Wahrnehmung und Einschätzung eines Auswahlprozesses durch die Bewerber thematisieren. Eine intensive Beschäftigung mit der Bewerberseite erscheint umso verständlicher, wenn folgenden Ergebnissen Beachtung geschenkt wird. Die Bewertung eines Personalauswahlprozesses kann unter anderem einen Einfluss auf die erbrachte Leistung der Bewerber ausüben (Ployhart & Harold, 2004; Chan & Schmitt, 2005). Eine negative Bewertung wirkt sich beispielsweise hinderlich auf die Leistungsmotivation des Bewerbers aus, was wiederum in einer geringeren Leistung im Auswahlverfahren münden kann (Arvey, Strickland, Drauden & Martin, 1990). Schmitt und Ryan (1992) konnten weiterhin nachweisen, dass sich die Bewertung eines Auswahlverfahrens durch die Bewerber auf dessen kriterienbezogene Validität auswirkt. In ihrer Studie zeigte sich, dass die kriterienbezogene Validität eines Verfahrens geringer ausfiel, wenn Probanden mit niedriger Motivation und einer negativen Einstellung gegenüber dem Auswahlverfahren beteiligt waren. Darüber hinaus kann das eingesetzte Auswahlverfahren die Außenwirkung einer Organisation beeinflussen (Bauer, Maertz, Dolen & Campion, 1998; Bauer, Truxillo, Paronto, Weekley & Campion, 2004). Während ein positives Image generell als Vorteil für etwa die wahrgenommene Attraktivität oder die Kundenloyalität erachtet wird (Görlich & Schuler, 2006), kann es sich – bezogen auf die Personalauswahl – förderlich auf die Ansprache qualifizierter Bewerber auswirken (Smither, Reilly, Millsap, Pearlman & Stoffey, 1993).

Im weiteren Verlauf dieses Abschnitts werden zwei Modelle – das *Konzept der sozialen Validität* nach Schuler (1990, 1993; Schuler & Stehle, 1983) aus dem deutschsprachigen Raum sowie das in angloamerikanischen Kreisen weit verbreitete *Model of Applicants' Reactions to Employment Selection Systems* von Gilliland (1993) – herausgegriffen und in Bezug auf ihre Bedeutung für SJTs genauer betrachtet. Beide Modelle lassen sich im Übrigen als Leitfäden für die Gestaltung des gesamten Auswahlprozesses, angefangen bei der Vorinformation der Bewerber über

die Datenerhebung bis hin zur Auswahlentscheidung und Rückmeldung der Ergebnisse an den Bewerber, verstehen.

Schuler und Stehle (1983) betrachteten das Konzept der sozialen Validität als Bereicherung „[...]der *technisch-empirischen Validität im testtheoretischen Sinne*“ (S. 35). Durch die Einführung der sozialen Validität als ergänzendes Gütekriterium trugen Schuler und Stehle (1983) dem interpersonalen Charakter einer Test- oder Auswahl-situation stärker Rechnung. Das Konzept der sozialen Validität berücksichtigt in seiner aktuellen Version (Schuler, 1993) vier Komponenten der sozialen Validität: *Information, Partizipation, Transparenz* und *Feedback*. Gemäß diesem Konzept werden Auswahlverfahren dann als sozial valide erlebt und bewertet, wenn sie hohe Ausprägungen auf alle vier Komponenten aufweisen.

Der Kern des Model of Applicants' Reactions to Employment Selection Systems von Gilliland (1993; Schmitt & Gilliland, 1992) enthält die Unterscheidung zwischen zwei Aspekten, welche Probanden bei der Bewertung eines Auswahlprozesses beachten. Demnach nehmen sie sowohl die Fairness des Auswahlprozesses an sich als auch die Fairness des Auswahlresultates respektive des Zustandekommens dieses Resultates wahr. Bei der Bewertung dieser beiden Aspekte dienen nach Gilliland (1993) bestimmte Bewertungsregeln als Grundlage. Gilliland (1993) postulierte, dass die Bewertung negativer ausfällt, wenn Bewerber feststellen, dass Regeln innerhalb eines Auswahlprozesses übertreten oder verletzt anstatt eingehalten werden (vgl. Leventhal, 1980). Abhängig von der Vorerfahrung und Erwartungshaltung der Bewerber fällt eine Gewichtung der Regeln interindividuell anders aus (Ryan & Ployhart, 2000; Cherry, Ordonez & Gilliland, 2003). Gilliland (1993) differenzierte weiterhin zwischen zwei Arten von Bewertungsregeln: Procedural justice rules und distributive justice rules. Procedural justice rules beziehen sich laut Kanning et al. (2006) erstens auf formale Aspekte eines Auswahlprozesses, zum Beispiel den Bezug zum Arbeitsplatz, zweitens auf die Transparenz und den Informationsfluss innerhalb des Auswahlprozesses, zum Beispiel, ob die Bewerber ein Feedback erhalten, und drittens auf die Gestaltung der Interaktion und Kommunikation zwischen Bewerbern und Vertretern der Organisation beziehungsweise den Personen, welche den Auswahl-

prozess gestalten¹¹ Procedural justice rules thematisieren ferner die Konsistenz und Standardisierung in der Durchführung des Auswahlprozesses, zum Beispiel in dem Sinne, ob Bewerber wahrnehmen, dass allen Bewerbern dieselben Aufgaben gestellt werden (Ryan & Ployhart, 1998). Hingegen dominiert unter der Bezeichnung distributive justice rules vor allem das Prinzip der Verteilungsgerechtigkeit (equity): Hat der Bewerber den Eindruck, dass eine bessere Testleistung oder eine höhere Anstrengung im Auswahlverfahren tatsächlich in einem besseren Resultat – optimalerweise in einem Stellenangebot – mündet?

Der Bewertung der Fairness eines Auswahlprozesses wird ein Einfluss auf viele weitere Faktoren wie die Entscheidung eines Bewerbers für oder gegen eine Organisation sowie die Arbeitszufriedenheit, die Selbstwirksamkeit, das Selbstbewusstsein und die Arbeitsleistung der Bewerber zugesprochen (Ployhart & Ryan, 1998; Ryan & Ployhart, 2000; Ployhart & Harold, 2004).

Chan und Schmitt (1997) verglichen in ihrer Laborstudie unter anderem die Augenscheinvalidität¹², welche einem SJT – einerseits videobasiert und andererseits im Paper-Pencil-Format – zugesprochen wird. Es ergab sich aufgrund eines t-Tests für unabhängige Stichproben ein signifikanter Unterschied hinsichtlich der wahrgenommenen Augenscheinvalidität des videobasierten SJT und des SJT im Paper-Pencil-Format. Danach wurden dem SJT im Videoformat durchschnittlich eine höhere Transparenz und ein stärkerer Bezug zum konkreten Arbeitsplatz bescheinigt als seinem Pendant im Paper-Pencil-Format.

Richman-Hirsch, Olson-Buchanan und Drasgow (2000) untersuchten in einer Laborstudie mit Führungskräften Einschätzungen der Validität und Fairness eines SJT. Dabei boten sie, bei gleich bleibendem Inhalt, drei Variationen des SJT auf: Computergestützt-multimedial-interaktiv (neun Videoszenen mit jeweils vier möglichen Verzweigungen), computergestützt-page-turner (computergestützte Darstellung der neun Situationen in schriftlicher Form) und Paper-Pencil. Die Probanden wurden zufällig ei-

¹¹ Inhaltliche Übereinstimmungen mit dem Konzept der sozialen Validität sind offensichtlich.

¹² Das Konzept der sozialen Validität hat noch keinen Einzug in die englischsprachige Literatur gefunden. Hier werden stattdessen Begriffe wie face validity oder applicant reaction zur Umschreibung ähnlicher Phänomene angeführt.

ner der drei SJT-Variationen zugeteilt. Nach Beendigung des SJT wurde die Einschätzung der Probanden hinsichtlich der Augenscheinvalidität des SJT, der wahrgenommenen Fairness, aber auch hinsichtlich des empfundenen Vergnügens, der empfundenen Kurzweile und der empfundenen Aktualität des Verfahrens eingeholt. Wie von den Autoren erwartet wurde der SJT in der computergestützt-multimedial-interaktiven Form signifikant augenscheinvalider, inhaltstvalider, verfahrens- und verteilungsgerechter sowie zufriedener stellender, kürzer und moderner als die beiden anderen SJT-Variationen eingeschätzt. Zwischen den beiden anderen SJT-Varianten ließen sich wider Erwarten keine statistisch bedeutsamen Unterschiede in Bezug auf Einschätzungen und Beurteilungen durch Probanden vorfinden.

Kanning et al. (2006) gingen in zwei Teilstudien der Frage nach, wie experimentelle Veränderungen der fidelity seitens der Stimulus- respektive der Reaktionskomponente die Einschätzung eines SJT durch Probanden und damit einhergehend seine soziale Validität beeinflussen. Die Durchführung des SJT lief in beiden Studien computergestützt ab. Nach Abschluss der Durchführung des SJT wurde dessen soziale Validität durch die Probanden entlang von vier Dimensionen bewertet: *Nützlichkeit* des SJT für die Personalauswahl (usefulness), *positive emotionale Reaktionen* auf das Messinstrument (emotional reaction) in dem Sinne, dass eine Auseinandersetzung mit dem Test Vergnügen bereitet hat, *Transparenz* (transparency), so dass für Probanden erkennbar war, welche Fähigkeiten die Testbearbeitung verlangt, und viertens *Bezug des Tests zum Berufsalltag* (job relatedness)

In der ersten Studie fanden Kanning et al. (2006) heraus, dass SJTs mit interaktivem Situationsablauf über alle vier erhobenen Dimensionen gemittelt signifikant sozial valider als der SJT mit nicht-interaktivem Situationsablauf eingeschätzt wurden. Entgegen der Erwartung von Kanning et al. (2006) konnte jedoch nicht nachgewiesen werden, dass SJTs mit Situationspräsentation in Videoform von den Probanden als sozialer valider eingestuft als SJTs mit Situationspräsentation in Textform. Zudem ließ sich aus den erhobenen Daten ebenfalls nicht ableiten, dass SJTs mit Situationen in Videoform und interaktivem Ablauf eine bessere Bewertung im Sinne höherer sozialer Validität erhalten als zum Beispiel ein SJT mit Situationen in Textform und nicht-interaktivem Ablauf.

Im Rahmen der zweiten Teilstudie von Kanning et al. (2006) schnitt ein SJT mit interaktivem Situationsablauf sowie einer Stimulus- und Reaktionspräsentation in Videoform in Bezug auf die Bewertung der Dimensionen sozialer Validität besser ab als fünf andere SJT-Formen, bei denen die Art der Situationspräsentation und der Interaktivität variiert wurden. Zwischen den fünf anderen SJT-Formen ließen sich jedoch keine statistisch bedeutsamen Unterschiede in Bezug auf die Dimensionen der sozialen Validität feststellen. Bei Sichtung zusätzlich eingeholter Rangordnungen zur Einschätzung der sozialen Validität ergab sich dagegen ein eindeutiges Bild. Hohe Ränge wurden stets von SJTs mit interaktivem Ablauf und Videoeinsatz vereinnahmt, während SJTs ohne Videoeinsatz und nicht-interaktivem Ablauf konstant niedrige Ränge bekleideten.

Da in all den beschriebenen Studien keine Bewerberstichproben getestet wurden, stehen Erkenntnisse zur sozialen Validität von SJTs in Bewerbungssituationen noch aus. Aufgrund der vorhandenen Studien zum Thema SJTs und soziale Validität kann jedoch festgehalten werden, dass SJTs mit einem höheren Grad an fidelity, zum Beispiel videobasierte SJTs oder SJTs mit interaktivem Verlauf, positivere Reaktionen und Einschätzungen auf Seiten der Probanden hervorrufen.

Mit der Frage, ob Probandengruppen verschiedener ethnischer Herkunft oder unterschiedlichen Geschlechts annähernd vergleichbare Testwerte in SJTs erreichen, ergibt sich ein weiterer wichtiger Aspekt bei der Diskussion der Nützlichkeit von SJTs für die Personalauswahl. Im Gegensatz zu den oben dargestellten Konzepten der sozialen Validität oder Gerechtigkeit werden zur Bestimmung der *Testfairness* aber keine subjektiven Einschätzungen der Probanden eingeholt, sondern Mittelwertdifferenzen zwischen verschiedenen Probandengruppen herangezogen.

Eine Zusammenfassung der Studienlage zur Testfairness von SJTs geschah in Gestalt einer Meta-Analyse von Whetzel, McDaniel und Nguyen (2008). Dieses amerikanische Forscherteam stieß zunächst auf die Erkenntnis, dass Probanden weißer Hautfarbe in SJTs durchschnittlich besser abschnitten als Probanden afroamerikanischer ($d = .38$), asiatischer ($d = .29$) sowie hispanoamerikanischer ($d = .24$) Herkunft. Für die aufgetretenen Differenzen konnten sie zwei Moderatoren identifizieren: Die kognitive Ladung des SJT und die eingesetzte Instruktion im SJT. Es zeigte sich, dass die Differenzen zwischen den Probandengruppen anwachsen,

falls der SJT stärker Ressourcen der allgemeinen kognitiven Leistungsfähigkeit beansprucht. Eine höhere kognitive Ladung – im Sinne von Korrelationen mit Maßen der Intelligenz – ist bei SJTs vor allem dann zu verzeichnen, wenn eine Instruktion zum Wissensausdruck eingesetzt wird (vgl. Abschnitt 2.3).

Neben Unterschieden zwischen Probanden unterschiedlicher ethnischer Herkunft gingen Whetzel et al. (2008) der Frage nach, ob sich auch zwischen weiblichen und männlichen Probanden bedeutsame Mittelwertdifferenzen aufzeigen lassen. An dieser Stelle berichten sie, dass weibliche Probanden durchschnittlich höhere Testwerte erreichten als männliche ($d = -.11$). Auch für diesen Befund konnten sie einen Moderator verantwortlich machen. Je höher die SJTs mit den Persönlichkeitsskalen Gewissenhaftigkeit und Verträglichkeit korrelierten, desto besser schnitten weibliche Probanden in ihnen ab.

Die Ausführungen zur Gleichbehandlung von Probanden in SJTs sind zu schließen mit der Aussage, dass Probanden weißer Hautfarbe in den meisten SJTs im Mittel besser abschneiden als Probanden anderer Hautfarbe. Dennoch erreichen diese Unterschiede entweder nicht den Status statistischer Bedeutsamkeit oder liegen aber unter der Differenz von einer Standardabweichung, welche gemeinhin zwischen den Testwerten weißer Probanden und den Testwerten von Probanden anderer Hautfarbe in Tests zur Messung kognitiver Leistungsfähigkeit existiert (z. B. Hunter & Hunter, 1984; Hanson & Ramos, 1996). Für den deutschsprachigen Raum liegen im Übrigen bis dato keine Studien vor, welche sich dem Punkt der Gleichbehandlung von Probanden in SJTs widmen.

2.5.4 Einflüsse auf die Validität

Mit der Tendenz zu sozial erwünschten Antworten (Faking) und dem Coaching werden in diesem Abschnitt zwei potentielle Einflussgrößen auf die Validität eines SJT herausgegriffen.

Unter dem Phänomen des *Faking* ist zu verstehen, dass Probanden, welche einem Testverfahren ausgesetzt sind, bewusst Angaben verzerren, um höhere Testwerte zu erreichen (McFarland & Ryan, 2000; McFarland, Ryan & Alexander, 2002). Vor allem innerhalb von Auswahl- oder Bewerbungssituationen kommt dem Faking hervorgehobene Bedeutung zu: Auf der einen Seite legen Studien (z. B. Ones & Viswesvaran, 1998; Alli-

ger & Dwight, 2000; Salgado, Viswesvaran & Ones, 2001) den Schluss nahe, dass Bewerber ihre Angaben in Persönlichkeitstests nicht in bedeutendem Maße verzerren beziehungsweise dass etwaige Verzerrungen die prognostische Validität dieser Verfahren nicht beeinflussen. Demgegenüber gibt es widersprechende Studien (z. B. Schmitt & Ryan, 1992), welche aufzeigen, dass Faking in Bewerbungssituationen auftritt und ein Absinken der Validität des Gesamtprozesses herbeiführt. In eine ähnliche Richtung zeigen Befunde, welche einen Einfluss des Fakings auf die Frage, ob ein Bewerber eingestellt wird oder nicht, nachweisen (z. B. Rosse, Stechner, Levin & Miller, 2002).

Bei einem diskursiven Vergleich der beiden in Abschnitt 2.3 differenzierten Instruktionsformen in SJTs mutmaßten McDaniel und Nguyen (2001), dass SJTs mit Instruktion zum Wissensausdruck weniger anfällig für sozial erwünschte Antworttendenzen (engl: Faking) wären als SJTs mit Instruktion zur Verhaltenstendenz. Zudem erklärten sie, dass SJTs ungeachtet ihrer Form der Instruktion arbeitsbezogenes Wissen messen, wenn Probanden sich zum Beispiel in einer Bewerbungssituation besonders positiv darstellen wollen: „[...] *when applicants are faking to look good, both instruction sets should be measuring knowledge*[...]“ (McDaniel & Nguyen, S. 106). Bei der Beantwortung der Frage, ob SJTs eine Vulnerabilität gegenüber Faking aufweisen, sind zwei Forschungsströmungen trennbar: Die eine Strömung wird gebildet von Untersuchungen in real-life settings und beinhaltet Bewerber oder Angestellte in Unternehmen (z. B. Vasilopoulos, Reilly & Leaman, 2000; Ployhart, Weekley, Holtz & Kemp, 2003). Eine zweite Strömung untersucht den Einfluss von Faking in Laborstudien, in denen die Probanden systematisch instruiert werden, sich besonders positiv oder im Gegensatz dazu ehrlich darzustellen (z. B. Nguyen, McDaniel & Biderman, 2005; Peeters & Lievens, 2005).

Nguyen, McDaniel und Biderman (2005) überprüften die Frage, ob SJTs je nach gewählter Instruktionsform unterschiedlich anfällig für Faking sind, auf empirischem Wege. Sie ließen Studenten zunächst einen Test zur allgemeinen kognitiven Leistungsfähigkeit durchlaufen. Im Anschluss daran bearbeiteten die Studenten jeweils zweimal eine Testbatterie aus einem SJT und einem Persönlichkeitsinventar zur Messung der Big 5. Dabei erhielten sie bei einem Durchlauf eine Instruktion, ehrlich zu ant-

worten, und bei der anschließenden zweiten Durchführung eine faking good Instruktion bei der Bearbeitung der beiden Tests. Eine Hälfte der studentischen Stichprobe erhielt die Tests zuerst mit Instruktion zur Ehrlichkeit und dann mit faking good Instruktion, während bei der anderen Hälfte die Reihenfolge der Instruktionen umgedreht wurde. Bei der Auseinandersetzung mit dem SJT wurde – unabhängig ob die Testung unter ehrlichen oder faking good Bedingungen erfolgte – stets sowohl eine Instruktion zur Verhaltenstendenz als auch eine Instruktion zum Wissensausdruck angewandt. In diesem Fall blieb die Reihenfolge der Instruktionen konstant, da die Instruktion zum Wissensausdruck immer vorangestellt wurde. Die Reliabilität des SJTs, angegeben als split-half Reliabilität nach der Spearman-Brown Formel, variierte zwischen .76 (faking good und Instruktion zum Wissensausdruck) und .82 (faking good und Instruktion zur Verhaltenstendenz) und liegt deutlich über den Reliabilitäten anderer SJTs (z. B. Smiderle et al., 1994; Kanning, 2005)¹³.

Nguyen et al. (2005) führten verschiedene Varianzanalysen getrennt nach Art der Instruktion im SJT durch, wobei die Art der Versuchsbedingung (faking good versus ehrlich) als Messwiederholungsfaktor fungierte, während die Abfolge der Versuchsbedingungen einen between-subjects Faktor repräsentierte. Bei der Berücksichtigung von SJT-Werten nach einer Instruktion zur Verhaltenstendenz ergab sich ein signifikanter Haupteffekt des Faktors Versuchsbedingung. Unabhängig davon, ob der SJT zuerst unter faking good oder zuerst unter ehrlichen Bedingungen absolviert wurde, waren die SJT-Werte im Mittel unter faking good Bedingungen höher als in der ehrlichen Bedingung. Nguyen et al. (2005) schlussfolgerten auf Basis dieses Resultats, dass sich Probanden in SJTs mit Instruktion zur Verhaltenstendenz bewusst besser darstellen können als sie tatsächlich sind: „[...]participants are able to fake performance in a SJT in which a behavioral tendency response format is used“ (S. 255). Für die SJT-Werte, welche nach einer Instruktion zum Wissensausdruck ermittelt wurden, trifft diese Aussage jedoch nicht zu. An dieser Stelle zeigte sich vielmehr, dass die Reihenfolge der Versuchsbedingungen die Höhe der

¹³ Es ist anzumerken, dass bei den meisten anderen SJTs Reliabilitäten zumeist in Form von Cronbach's Alpha berechnet werden und nicht nach der split-half Methode.

SJT-Werte beeinflusst, da der Mittelwert in der jeweilig zuerst durchlaufenen Bedingung immer höher war als der Mittelwert in der nachfolgenden zweiten Bedingung. Aufgrund dieses Ergebnisses wurden keine weiteren Analysen mit SJT-Werten nach einer Instruktion zum Wissensausdruck vorgenommen. Bei einem Vergleich der Anfälligkeit für Faking zwischen dem SJT und dem Persönlichkeitsinventar anhand von Effektstärken, zeigte sich der SJT robuster gegenüber Faking als drei der fünf einbezogenen Persönlichkeitsfaktoren: Extraversion, Gewissenhaftigkeit und Emotionale Stabilität.

Im Übrigen wurde von Nguyen et al. (2005) berichtet, dass SJT-Werte, welche unter einer Instruktion zum Wissensausdruck ermittelt wurden, höher mit dem Test zur kognitiven Leistungsfähigkeit korrelierten (.56) als SJT-Werte nach vorgehender Instruktion zur Verhaltenstendenz (.38). Ferner konnte nicht nachgewiesen werden, dass SJT-Werte nach Instruktion zur Verhaltenstendenz signifikant höher mit Persönlichkeitsfaktoren korrelieren als SJT-Werte nach Instruktion zum Wissensausdruck. Der SJT wies ungeachtet der enthaltenen Instruktion signifikante Korrelationen mit den Faktoren Gewissenhaftigkeit und Offenheit für Erfahrungen auf, aber nicht mit den drei anderen Faktoren aus dem Big 5-Modell.

Von Peeters und Lievens (2005) stammt eine weitere Studie, in welcher der Einfluss von Faking auf Testergebnisse im SJT untersucht wurde. In diesem Fall kam der SJT in einer realen Bewerbungssituation zum Einsatz. Die beiden Autoren stellten fest, dass Bewerber, welche zum Faking animiert wurden, signifikant höhere SJT-Scores erzielten als Probanden, welche ehrlich antworten sollten. Zudem berichteten sie, dass von den eingestellten Bewerbern, rund drei Viertel der faking-Bedingung entstammten. Bei einem Vergleich der kriterienbezogenen Validitäten wurde dem SJT in der faking-Bedingung eine signifikante niedrigere Korrelation (.09) mit einem Leistungskriterium bescheinigt als dem SJT in der ehrlichen Bedingung (.33). Zudem führte das bewusste Verzerren der Angaben im SJT zu einem Ausbleiben inkrementeller Validität des SJT in Kontrast zu einem zusätzlich aufgeklärten Varianzanteil von mehr als 5 % durch den SJT, bei dem die Probanden ehrliche Antworten abgegeben haben.

Der aktuelle Forschungsstand zum Zusammenhang von Faking und Testwerten wurde von Hooper, Cullen und Sackett (2006) zusammengefasst. Sie konstatierten, dass sich zwischen Testwerten, welche unter fa-

king good Bedingungen auf der einen Seite und ehrlichen Bedingungen auf der anderen Seite erzielt werden, Effektstärken zwischen $d = .08$ und $d = .89$ ergeben. Als Erklärung für diese Variabilität führten sie unterschiedliche Designs, Instruktionen und Konstrukte an. Zudem kamen Hooper et al. (2006) zu dem Schluss, dass SJTs weniger vulnerabel für Faking sind als zum Beispiel Persönlichkeitsskalen (vgl. Mesmer-Magnus & Viswesvaran, 2006; Kuncel & Borneman, 2007).

Zur Kontrolle des Faking in SJTs schlugen Lievens und Peeters (2008) vor, den Probanden Zeit zur Elaboration der Verhaltensalternativen im SJT einzuräumen. Im Rahmen dieses elaborativen Prozesses sollen die Probanden beispielsweise niederschreiben, welche Gründe oder Motivation sie bei der Entscheidung für oder gegen eine Verhaltensalternative anführen. Die beiden belgischen Forscher konnten nachweisen, dass der Effekt von Faking durch die Möglichkeit zur Elaboration deutlich reduziert wurde, falls die im SJT dargestellten Verhaltensalternativen einen gewissen Grad an Bekanntheit für die Probanden besaßen. Weiterhin erhöhte die Einführung des elaborativen Prozesses die wahrgenommene Fairness des SJTs seitens der Probanden.

Es ist zusammengefasst festzuhalten, dass die bestehende Forschungslage zum Faking in SJTs kein einheitliches Urteil zulässt. Durch die Wahl einer Instruktion zum Wissensausdruck kann die Wahrscheinlichkeit des Fakings jedoch vorab verringert werden, da auf diesem Wege eine Motivation seitens der Probanden zu maximaler Performanz (s. o.) besteht (Whetzel & McDaniel, 2009). Levens et al. (2009) sind sogar der Meinung, dass „*the high motivation in a high-stakes selection context seems to wash away possible mean score differences in SJT*“ (S. 1099-1100). Ferner ist aus Sicht des Autors vor dem Hintergrund, dass einige Forscher SJTs die Messung sozialer Anpassungsfähigkeit zuschreiben, zu diskutieren, ob eine positive Selbstdarstellung in SJTs nicht eher als erwünschtes Verhalten zu betrachten ist.

Das Phänomen des *Coaching* rückt dann in den Fokus der Aufmerksamkeit, wenn Probanden oder Bewerber einen Test zweimal durchlaufen und sich in der Zeit zwischen den beiden Testdurchführungen systematisch und gezielt auf die erneute Testung vorbereiten. Dabei unterschieden Messick und Jungeblut (1981) verschiedene Varianten des Coaching, welche von der Beschäftigung mit einzelnen Items bis hin zum Erwerb

allgemeiner Bearbeitungsstrategien reichten. Nach Maurer, Salomon und Troxtel (1998) bedient sich das Coaching sechs unterschiedlicher Prozeduren und Strategien: Reine Praxis – Allgemeine Hinweise zur Testarbeit – Feedback – Lernen am Modell – Tutorium – Theoretische Einführung in die Testkonstrukte. In den meisten Coachingprogrammen werden die genannten Formen variiert und kombiniert (Cullen, Sackett & Lievens, 2006). Mit dem Prozess des Coachings wird das Ziel verbunden, den eigenen Testwert systematisch zu verbessern respektive aufzuwerten (Hurtz & Alliger, 2002). Eine systematische Verzerrung der Testwerte würde ähnlich wie beim Faking die Konstruktvalidität des Tests (Ellingson, Sackett, & Hough, 1999) oder Auswahlentscheidungen und damit verbunden die prognostische Validität beeinflussen (Rosse, Stecher, Miller, & Levin, 1998). Jedoch spricht die vorliegende empirische Evidenz zumindest gegen einen substantiellen Einfluss des Coaching auf die kriterienbezogene Validität unterschiedlicher Verfahren: Bezogen auf Maße der kognitiven Leistungsfähigkeit fanden Allalouf und Ben-Shakkar (1998) in diesem Zusammenhang heraus, dass Coaching entweder zu sehr geringen oder gar keinen Veränderungen hinsichtlich der kriterienbezogenen Validitäten führt. Alliger und Hurtz (2002) berichteten, dass Coaching ebenfalls keine bedeutsamen Veränderungen in Hinblick auf die kriterienbezogene Validität von persönlichkeitsbezogenen Integritätstests verursacht.

Für die Untersuchung des Einflusses von Coaching auf die Validitäten von SJTs liegen bislang kaum Studien vor (vgl. Whetzel & McDaniel, 2009). Cullen et al. (2006) ermittelten beispielsweise in Übereinstimmung mit den oben genannten Studien, dass die kriterienbezogene Validität zweier SJTs nicht durch Coaching beeinflusst wird. Als Design griffen die drei Forscher auf ein Experimental-Kontrollgruppendesign mit Prä- und Posttest zurück, wobei nur die Experimentalgruppe zwischen den beiden Testdurchgängen einen Coachingkurs durchlief. Daneben befragten sie ihre Probanden aus der Experimentalgruppe, wie sehr ihnen das Coaching beim Bearbeiten des SJT geholfen habe. Nach Angaben ihrer Probanden gestaltete sich das Erlernen und Anwenden der Lösungsstrategien vor allem bei demjenigen der beiden SJTs als sehr anspruchsvoll, welcher sehr spezifische Situationen abfragte.

Aus Sicht des Autors treten die Parallelisierung der Verhaltensalternativen gemäß ihrem Grad an sozialer Erwünschtheit sowie der den Probanden nicht bekannte scoring key einem verzerrenden Effekt des Coaching bei SJTs entgegen. Eine weitere Form des Schutzes oder der Vorbeugung wäre nach Whetzel und McDaniel (2009) der Entwurf eines ausreichend großen Itemstammes, aus dem mehrere Parallelformen eines SJTs kreiert werden können. Zudem mutmaßten Cullen et al. (2006; S. 153 ff.) in der Diskussion ihrer Studie, dass das Coaching für solche SJTs, welche mehrere Konstrukte zu messen beanspruchen, mehrere Strategien erfordere und daher einen höheren Anforderungsgrad anbiete.

2.6 Kundenorientierung und ihre Bedeutung für ein Unternehmen

Nach Ryan und Ployhart (2003, S. 377) gilt in Zeiten saturierter und kompetitiv geprägter Märkte folgende strategische Ausrichtung von Unternehmen als gewinnbringend: „[...] *a focus on the customer has become a major component of organizational strategies, regardless of whether the organization is in the service or manufacturing sector*“.

Nerdinger (2003) definierte Kundenorientierung¹⁴ wie folgt: „*Kundenorientierung heißt, die Bedürfnisse und Erwartungen der Kunden zu erkennen und sich zu bemühen, diese zu erfüllen*“ (S. 1). Eine kongruente Sichtweise verfolgten Meyer, Kantsperger und Schaffer (2006; S. 66), welche als Basis der Kundenorientierung „[...] *die Berücksichtigung der Kundenperspektive und Erfüllung der Kundenerwartungen in allen unternehmerischen Aktivitäten, also die durchgängige Verinnerlichung der Kundenperspektive in sämtlichen Leistungs- und Wertschöpfungsprozessen*“ ansahen. Mit der Ausrichtung an Bedürfnissen und Erwartungen von Kunden ist die Kundenorientierung abzugrenzen vom Konzept der Marktorientierung (z. B. Saxe & Weitz, 1982), bei der Mitarbeiter vorrangig nach einem möglichst zeitnahen Geschäftsabschluss respektive Verkauf streben.

Eine Vielzahl von Studien konnte den Nachweis erbringen, dass Kundenorientierung einerseits eine wichtige Voraussetzung zur Erreichung höherer Kundenzufriedenheit darstellt (z. B. Ramsey & Sohi, 1997) und dass sie andererseits vermittelt über die Kundenzufriedenheit eine stärkere Kundenloyalität¹⁵ im Sinne von Treue zum Unternehmen aufbauen kann (Pieters, Bottenschen & Thelen, 1998; Homburg & Bucerius, 2003). Nach Evans und Laskin (1994) nehmen zufriedene Kunden die Angebote eines Unternehmens wiederholt an und empfehlen das Unternehmen über Mund-zu-Mund-Propaganda Freunden oder Bekannten, die aufgrund der Glaubwürdigkeit dieser Form der Propaganda leichter als Neukunden zu gewinnen sind (Anderson, 1998). Zufriedene Kunden sind bereit, mehr für ein und dieselbe Leistung zu investieren und wechseln bei Preiserhöhun-

¹⁴ Als Synonym findet sich in der relevanten Literatur der Begriff Kundennähe.

¹⁵ Statt Kundenloyalität wird in manchen Artikeln auch von Kundenbindung gesprochen.

gen nicht zwangsläufig zu einem preisgünstigeren Konkurrenten (z. B. Reichheld, 1996). Ein weiterer positiver Effekt zufriedener Kunden ist der des erhöhten cross-selling Potentials (Reichheld & Sasser, 1990; Huber, Herrmann & Braunstein, 2009): Zufriedene Kunden tendieren danach stärker dazu, verschiedene Produkte oder Leistungen eines Unternehmens wahrzunehmen und auszuprobieren. Nach Pepels (2003) und Naumann (2005) erweist sich die Gewinnung neuer Kunden als drei- bis siebenmal teurer als die Pflege zufriedener Stammkunden.

In der Abbildung 2-7 wird der Zusammenhang zwischen Kundenorientierung, Kundenzufriedenheit, Kundenloyalität sowie Kundenwert schematisch und vereinfacht dargestellt. Dabei orientiert sich die Darstellung an der sogenannten *Erfolgskette der Kundenorientierung* (Hesskett, Sasser & Schlesinger, 1997). Komplexere Modelle mit weiteren involvierten Variablen und Einflussgrößen finden sich beispielsweise bei Homburg und Bruhn (2008).

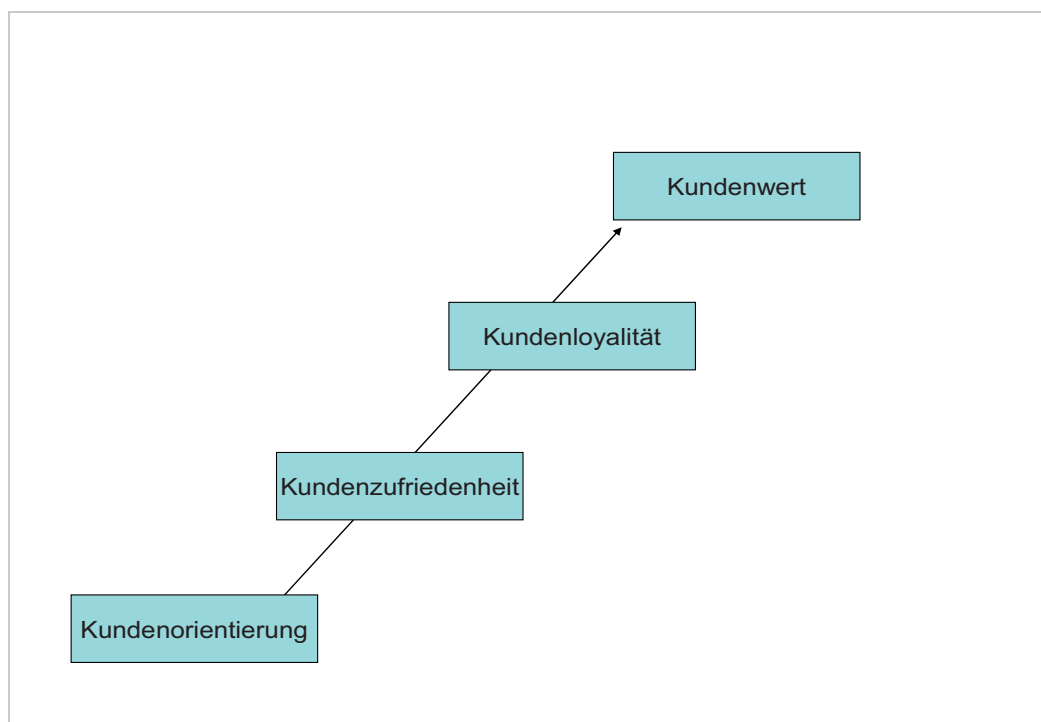


Abbildung 2-7: Erfolgskette der Kundenorientierung

Nach den Prämissen der Erfolgskette (s. Abbildung 2-7) existiert ein linearer Zusammenhang zwischen Kundenzufriedenheit und -loyalität. Dies würde bedeuten, dass sich zufriedene Kunden nahezu immer zu loyalen Kunden entwickeln (Rust & Zahorik, 1993). Nach Ansicht von zum Bei-

spiel Jones und Sasser (1995) sowie Homburg und Giering (2001) ist jedoch auch ein nicht-linearer Zusammenhang zwischen diesen beiden Variablen vorstellbar. So sprachen Homburg und Giering (2001) von einer „*zone of indifference*“ (S. 47), in der sich vor allem Kunden wieder finden, welche mittelmäßig zufrieden mit einer Dienstleistung sind. Zudem identifizierten die beiden deutschen Forscher verschiedene demographische Variablen auf Seiten des Kunden, zum Beispiel sein Alter oder Einkommen, welche die Beziehung zwischen Kundenzufriedenheit und -loyalität moderieren. Speziell in Bezug auf das Hotel- und Gaststättengewerbe konnte festgestellt werden, dass das Image des Hotels einen bedeutsamen Einfluss ausüben kann auf die Beziehung zwischen Kundenzufriedenheit und -loyalität (Ostrowsky, O'Brien & Gordon, 1993; Kandampully & Suhartanto, 2000). Zum Image zählen wiederum gänzlich unterschiedliche, vom Kunden subjektiv gewichtete Facetten wie die Lage eines Hotels oder die Atmosphäre im Hotel.

Mit den in der Abbildung 2-7 genannten Gliedern der Erfolgskette erfolgt nun eine genauere inhaltliche Auseinandersetzung.

2.6.1 Kundenorientierung

In der Literatur zum Themengebiet *Kundenorientierung* wird zwischen einer organisationalen und einer personellen Kundenorientierung differenziert (vgl. Bruhn, 2003; Bruhn, 2006). Unter dem Begriff *organisationale Kundenorientierung* wird die Ausrichtung eines gesamten Unternehmens inklusive seiner Strukturen, seiner Unternehmenskultur und -philosophie an der Erfüllung von Kundenerwartungen verstanden. Unternehmensaktivitäten und Marketingaktivitäten werden an Anforderungen und Bedürfnissen der Kunden angepasst (Grabner-Kräuter & Schwarz-Musch, 2006).

Auf Seiten der Unternehmensstruktur ist unterstützend zum Beispiel die Einrichtung eines Informationssystems, welches Kundendaten erfasst und bündelt (data mining), notwendig (Homburg & Daum, 1997). Zudem wird die Einführung eines kundenorientierten Vergütungssystems als förderlich betont, welches als Bemessungsgrundlage den Umsatz eines Mitarbeiters durch Maße wie die Zufriedenheit oder die Treue der durch den Mitarbeiter betreuten Kunden ablöst. Eine Unternehmenskultur, welche besonders stark der Kundenorientierung verpflichtet ist, existiert in Form der Adhoc-

racy-Kultur, die Werten wie Innovation, Wachstum, Kreativität und der Anpassung an Kundenbedürfnisse eine tragende Rolle zukommen lässt (z. B. Conrad, Brown & Harmon, 1997). Verschiedene Forscher betonten zudem die Wichtigkeit, dass Führungskräfte ihren Mitarbeitern in punkto Kundenorientierung als positive Vorbilder dienen (z. B. Homburg & Werner, 1998; Homburg & Stock, 2004) und einen mitarbeiterorientierten Führungsstil zur Steigerung der Kundenorientierung auf Seiten der Mitarbeiter praktizieren sollten.

Im Gegensatz zum Konzept der organisationalen Kundenorientierung zielt der Begriff *personelle Kundenorientierung* auf das individuelle Verhalten von Mitarbeitern mit Kundenkontakt ab. Die interpersonale Interaktion zwischen Mitarbeiter und Kunde wird als essentieller Grundstein zur Erbringung von Serviceleistungen angesehen (Parasuraman, Zeithaml & Berry, 1985). Ein hoher Grad an Interaktion herrscht insbesondere im Hotelgewerbe vor, so dass hier ein besonders starker Anspruch an die personelle Kundenorientierung gestellt wird (King & Garey, 1997).

Die personelle Kundenorientierung eines Mitarbeiters umfasst nach Nerding (2003) eine fachliche, eine soziale und eine persönlichkeitsbezogene Facette. Aufgrund des Fehlens einer in sich geschlossenen Theorie des kundenorientierten Mitarbeiters beziehungsweise der Tatsache, dass Kundenorientierung je nach beruflichem Kontext anders ausgelegt wird, werden im Folgenden Persönlichkeitseigenschaften benannt, die in mehreren Arbeiten als förderlich für die Kundenorientierung benannt wurden. Eine besonders hohe Ausprägung bezüglich der Kundenorientierung ist vor allem bei solchen Mitarbeitern erforderlich, die in direktem Kundenkontakt stehen, den sogenannten *frontline employees*. Nach Rust, Zahorik und Keiningham (1996) beeinflussen diese Mitarbeiter in entscheidendem Maße die Kundenzufriedenheit: „[...]the personal interaction component of services is often a primary determinant of the customer's overall satisfaction“ (S. 391). In ihrer Funktion als Schnittstelle zwischen Unternehmen und Kunden kommt ihnen im Vergleich zu anderen Mitarbeitern eine hervorgehobene Bedeutung zu (Bettencourt & Brown, 2003). Diese Mitarbeiter sollen durch ihr Verhalten das in ihrem Unternehmen herrschende Konzept von Kundenorientierung leben und zufriedene Kunden hervorbringen (z. B. Schneider, Ehrhart, Mayer & Saltz, 2005) .

Eine Meta-Analyse von Frei und McDaniel (1998) ergab, dass Maße der Kundenorientierung moderat mit den Persönlichkeitsfaktoren Vertrauenswürdigkeit (.43), Gewissenhaftigkeit (.42) und Emotionale Stabilität (.37) korrelieren. Ferner lässt sich der Meta-Analyse entnehmen, dass Maße der Kundenorientierung auf statistischer Ebene unabhängig sind von dem Persönlichkeitsfaktor Offenheit für neue Erfahrungen (.07) sowie der allgemeinen kognitiven Leistungsfähigkeit (-.06).

In demselben Jahr wie Frei und McDaniel (1998) führten Barrick, Mount und Stewart (1998) eine Meta-Analyse durch, um zu überprüfen, welche Faktoren des FFM die Arbeitsleistung in Arbeitsbereichen beeinflussen, in denen die Interaktion zwischen Mitarbeiter und Kunde im Vordergrund steht. Es zeigten sich für Gewissenhaftigkeit (.26), Vertrauenswürdigkeit (.21) und Emotionale Stabilität (.18) moderate Korrelationen mit der Leistungsbeurteilung durch Vorgesetzte. Mit einem hierarchischen Modell von Brown, Mowen, Donovan und Licata (2002) liegt eine theoretische Rahmenstruktur zur Erklärung der persönlichkeitsbezogenen Komponente von Kundenorientierung vor. Brown et al. (2002) bezeichneten Persönlichkeitsfaktoren wie Gewissenhaftigkeit oder Verträglichkeit als basic traits, deren individuelle Ausprägung zeitlich überdauernd und kontextunabhängig beschaffen sei. Hingegen sahen sie Kundenorientierung als surface trait an, welcher kontextabhängig sei und dessen Ausprägung einer Interaktion zwischen einer Situation und bestimmten basic traits entspreche. Als Determinanten der Kundenorientierung ermittelten Brown et al. (2002) die basic traits Emotionale Stabilität, Verträglichkeit und need for activity.

Nerdinger (2003) sah in den drei oben genannten Persönlichkeitsfaktoren die Voraussetzung zum Aufbau einer persönlichen Beziehung zum Kunden. Für die Aufrechterhaltung dieser Beziehung und damit zur dauerhaften Zufriedenstellung des Kunden im Sinne einer langfristigen Loyalität wies er jedoch der intrinsischen Motivation eine hervorstechende Bedeutung zu. Intrinsisch motivierte Mitarbeiter üben eine Tätigkeit aus, weil ihnen diese Spaß bereitet und Befriedigung verschafft, so dass der Arbeit an sich eine stimulierende Funktion zukommt (Deci, 1975; Semmer & Udris, 2004, Gagne & Deci, 2005). Hingegen gehen extrinsisch motivierte Mitarbeiter ihrer Arbeit in der Erwartung positiver Kon-

sequenzen, zum Beispiel einem höheren Einkommen oder der Anerkennung durch Vorgesetzte, nach (Amabile, 1993; Semmer & Udris, 2004).

Andere Eigenschaften, denen studienübergreifend eine positive Beziehung zur Kundenorientierung zugeschrieben wurde, werden im Folgenden kurz aufgeführt. Beispielsweise erbrachten MacKenzie, Podsakoff und Ahearne (1998) mit ihrer Studie die Evidenz, dass zufriedene Mitarbeiter auch kundenorientierter arbeiten. Nach einer Arbeit von Kelley (1992) verhalten sich ferner Mitarbeiter, welche sich ihrem Unternehmen sehr verpflichtet fühlen, kundenorientierter als Mitarbeiter mit geringerem Commitment.

2.6.2 Kundenzufriedenheit

Das Konstrukt *Kundenzufriedenheit* genießt aufgrund seiner Bedeutung für den Unternehmenserfolg und der damit verbundenen Profitabilität (z. B. Andreasen, 1985; Anderson, Fornell & Lehmann, 1994) eine hervorgehobene Bedeutung für Unternehmen.

Homburg und Stock (2001) betrachten Kundenzufriedenheit „[...]als Einstellung, die sich aus dem abwägenden Vergleich zwischen der erwarteten Leistung (Soll) und der tatsächlich wahrgenommenen Leistung (Ist) ergibt“ (S. 791). Dabei bezieht sich die Kundenzufriedenheit nicht auf die Bewertung eines einzelnen episodischen Konsumerlebnisses, sondern ist das Ergebnis einer zusammenfassenden Bewertung aller bisherigen und aktuellen Erfahrungen und Konsumerlebnisse, an der neben einer kognitiven Komponente auch affektive Komponenten beteiligt sind (Boulding; Staelin, Kalra & Zeithaml, 1994; Krause & Dunckel, 2003). Dies trifft insbesondere auch auf die Zufriedenheit mit Serviceleistungen im Hotel- und Gaststättengewerbe zu: „Therefore it is possible to say that satisfaction with a hospitality experience such as a hotel stay [...] is a sum total of satisfactions with the individual elements or attributes of all the products and services that make up the experiences“ (Pizam & Ellis, 1999; S. 329).

Auch wenn noch keine allgemeingültig akzeptierte Faktorenstruktur vorgefunden wurde (Stauss, 1999), dominiert in der Forschung die Vorstellung einer mehrfaktoriellen Struktur der Kundenzufriedenheit (Matzler, Sauerwein & Stark, 2009). Ansätze, welche Kundenzufriedenheit eine mehrfaktorielle Struktur zuschreiben, bieten viele inhaltliche Übereinstimmungen mit der – empirisch nicht abgesicherten – *Zwei-Faktoren-*

Theorie zur Arbeitszufriedenheit von Herzberg (z. B. Herzberg, Mausner & Snyderman, 2004). Nach der Zwei-Faktoren-Theorie werden die Zufriedenheit und Unzufriedenheit von Mitarbeitern als zwei unterschiedliche Konstrukte betrachtet. Hygienefaktoren als eine Gruppe von Faktoren können zwar bei Erfüllung die Entstehung von Unzufriedenheit verhindern, jedoch führt ihre Erfüllung nicht automatisch zur Zufriedenheit, sondern zum neutralen Zustand der Nicht-Unzufriedenheit. Hingegen führt eine Erfüllung der Motivationsfaktoren zu Zufriedenheit am Arbeitsplatz. Ein Fehlen oder zu geringes Vorhandensein der Motivationsfaktoren wiederum führt nicht zwangsläufig zu Unzufriedenheit, sondern zu einem ausgeglichenen Zustand des Nicht-Zufriedenseins (Weinert, 2004).

Bezogen auf den Bereich der Kundenzufriedenheit ist in der aktuellen Forschungsliteratur das Postulat einer Drei-Faktoren-Struktur vorherrschend (z. B. Oliver, 1997). Demnach sind *Basisfaktoren*, *Leistungsfaktoren* oder *Begeisterungsfaktoren* voneinander abzugrenzen (vgl. Matzler et al., 2009).

- Basisfaktoren oder auch Dissatisfiers (Unzufriedenheitsfaktoren) entsprechen den Mindestanforderungen an eine Dienstleistung oder ein Produkt. Sie lösen Unzufriedenheit aus, wenn Kunden sie als nicht vorhanden oder als nicht den eigenen Erwartungen entsprechend wahrnehmen. Exemplarisch sind zu wenige Parkplätze an einem Hotel oder eine zu hohe Lautstärke im Hotelrestaurant zu nennen. Ein Vorhandensein der Basisfaktoren, sogar im Sinne eines Übertreffens von Erwartungen, mündet nicht in Zufriedenheit, sondern in Nicht-Unzufriedenheit. Eine konzeptionelle Nähe zu den oben beschriebenen Hygienefaktoren von Herzberg et al. (1959) wird offensichtlich.
- Leistungsfaktoren¹⁶ führen einerseits direkt zu Kundenzufriedenheit, wenn die Erwartungen von Kunden übertroffen werden, andererseits münden sie aber auch nahtlos in Kundenunzufriedenheit, falls sie die Erwartungen von Kunden nicht erfüllen. Als Beispiel für einen Leistungsfaktor im Hotel- und Gaststättenge-

¹⁶ Für den Begriff Leistungsfaktoren sind verschiedene Synonyme aufzufinden, zum Beispiel *criticals*, *hybrids* oder *dual-threshold factors*.

werbe ist die Kundenorientierung von Hotelmitarbeitern zu nennen (Pizam & Ellis, 1999).

- Begeisterungsfaktoren oder Satisfiers (Zufriedenheitsfaktoren) werden von Kunden im Gegensatz zu Basisfaktoren als nicht-selbstverständlich erachtet, sondern viel mehr als Zugabe oder persönliches Zugeständnis seitens des Unternehmens. Dementsprechend führt ihre Anwesenheit zu Zufriedenheit, während ihr Fehlen keine Unzufriedenheit verursacht, denn „[...]Satisfiers, represent an opportunity to shine, to move ahead of the pack, and to stand out from the crowd“ (Cardotte & Turgeon, 1988; S. 51).

Um festzustellen, welche Basis-, Leistungs- oder Begeisterungsfaktoren für den individuellen Kunden von Bedeutung sind, können verschiedene Methoden angewandt werden (vgl. Matzler, Pachlaner & Siller, 2001), zum Beispiel die CIT (Flanagan, 1954), die Analyse von Kundenbeschwerden oder die sogenannte Importance-Grid Technik (z. B. Matzler & Sauerwein, 2002).

In Anbetracht einer Vielzahl an Theorien zur Entstehung der Kundenzufriedenheit wird im Folgenden nur eine stark begrenzte Auswahl vorgestellt. Begonnen wird mit einer Einführung in die Theorien des intrapersonellen Gleichgewichts beziehungsweise Konsistenztheorien (vgl. Stock, 2001). Mit dem *Diskonfirmations-Modell* (z. B. Olshavsky & Miller, 1972; Prakash, 1984; Stauss, 1999) wird der bekannteste Vertreter der Konsistenztheorien einer längeren Würdigung unterzogen. Neben den Konsistenztheorien existiert mit den Theorien des interpersonellen Gleichgewichts respektive Austauschtheorien eine alternative Sichtweise der Entstehung von Kundenzufriedenheit. Auch für diese Klasse von Theorien wird mit der *Equity-Theorie* ein Repräsentant in den Fokus gerückt.

Nach dem Diskonfirmations-Modell begegnen Kunden Unternehmen und deren Produkten oder Mitarbeitern mit bestimmten Erwartungen und vergleichen in Gestalt eines intrapsychischen Prozesses die vorgefundenen Leistungen eines Unternehmens (IST-Wert) mit der eigenen Erwartungshaltung (SOLL-Wert). Die Erwartungen von Kunden können dabei verschiedenen Quellen entspringen (vgl. Nerdinger, 2003): Persönlichen

Wünschen, vorherigen Konsumerlebnissen, Mund-zu-Mund-Propaganda von Bekannten oder Freunden oder der Werbung von Unternehmen. Georgi (2001) unterschied in diesem Zusammenhang zwischen unterschiedlichen Formen der Erwartungshaltung. Kunden können demzufolge zum einen eine minimale Erwartungshaltung hegen, zum Beispiel die Erwartung, dass ihnen in einem Restaurant sauberes Besteck vorgelegt wird. Zum anderen kann Kunden aber auch eine Idealvorstellung des Konsumerlebnisses vorschweben, so dass sie vom Unternehmen eine kaum zu erbringende Leistung erwarten. Da die Erwartungen von Kunden als variabel und veränderlich erachtet werden, ist es nach Nerdinger (2003) unerlässlich, dass sich Unternehmen kontinuierlich über die Erwartungen ihrer Kunden informieren und die Beschaffenheit ihrer Angebote und Produkte daran ausrichten und anpassen. Zudem sei zu berücksichtigen, dass Kunden unterschiedliche Aspekte oder Facetten von Produkten respektive Dienstleistungen interindividuell unterschiedlich stark bei der Ausbildung ihrer Erwartungen gewichten (z. B. Meister & Meister, 1996).

Gemäß dem Diskonfirmations-Modell sind drei separate Szenarien hinsichtlich des Vergleichs zwischen IST- und SOLL-Wert voneinander abgrenzbar. In einem ersten Szenario kann die wahrgenommene Leistung des Unternehmens die Erwartungen des Kunden nicht erfüllen. In diesem Fall kommt es zu Unzufriedenheit (Diskonfirmation). Als zweites Szenario ist denkbar, dass die wahrgenommene Leistung die Erwartung des Kunden erfüllt, so dass Zufriedenheit (Konfirmation) entsteht. Ein drittes Szenario liegt vor, wenn die Erwartungen des Kunden sogar übertroffen werden. An diesem Punkte besteht innerhalb der Forschung ein Dissens: Eine Gruppe von Forschern argumentiert, dass es unter diesen Umständen zu Euphorie und damit positiver Konfirmation komme (z. B. Oliver, 1993; Homburg, 2001), während eine andere Gruppe erwidert, dass ein unangenehmes Gefühl der inneren Anspannung entstehe, da der Kunde nun das Gefühl haben könnte, das Unternehmen leiste mehr für ihn als er zurückgibt (z. B. Wiendieck, 1994). Mit dem Hinweis auf ein Gefühl der Anspannung wird die Brücke zur Equity-Theorie geschlagen.

Werden die Prämissen der Equity-Theorie, welche ursprünglich der Sozialpsychologie entstammt, auf den Austauschprozess zwischen Kunde und Unternehmen übertragen, so laufen auf Seiten des Kunden zwei intrapsychische Vergleichsprozesse ab (Andreassen, 2000). Zunächst ver-

gleichet der Kunde seinen Aufwand, zum Beispiel den bezahlten Preis, mit dem erhaltenen Nutzen, vorrangig der Qualität einer erhaltenen Dienstleistung. In einem zweiten Schritt vollzieht der Kunde denselben Vergleich aus Sicht des Unternehmens: Was hat das Unternehmen investiert, zum Beispiel in die Entwicklung oder das Marketing eines Produktes, und welchen Betrag erhält es dafür vom Kunden zurück? Im Vergleich zum Diskonfirmations-Modell liegt also nicht mehr ausschließlich das Gleichgewicht innerhalb des Kunden im Fokus, sondern vielmehr das Gleichgewicht im Rahmen einer Beziehung zwischen Kunde und Unternehmen. Aufbauend auf den beiden dargestellten Vergleichsprozessen sind zwei verschiedene Konstellationen konstruierbar: In einem ersten Fall fühlen sich Kunden inäquitabel, da ihr eigenes Verhältnis von Nutzen und Aufwand nicht dem entsprechenden Verhältnis auf Seiten des Unternehmens entspricht (VanYperen & Van de Vliert, 2001). Zum einen können die Kunden hier den Eindruck haben, dass sie deutlich mehr in den Austauschprozess eingebracht haben, zum Beispiel einen zu hohen Preis, als sie Nutzen daraus geschlagen haben. Zum anderen können die Kunden aber auch wahrnehmen, dass das Unternehmen einen Aufwand betrieben hat, welcher den daraus resultierenden Gewinn bei weitem übersteigt. Das in beiden aufgeführten Beispielen aufkeimende Ungleichgewicht bewirkt beim Kunden einen inneren Spannungszustand (vgl. die Theorie der kognitiven Dissonanz nach Festinger, 1957), welcher ihn zur Wiederherstellung des Gleichgewichts antreibt. Eine äquitable Beziehung als zweite mögliche Konstellation im Anschluss an die vom Kunden vorgenommenen Vergleichsprozesse besteht hingegen dann, wenn der Kunde einerseits das Gefühl hat, dass sein Aufwand dem gewonnen Ertrag entspricht, und dass andererseits nach Sicht des Kunden auch das Unternehmen einen Gewinn gutschreiben kann, welcher seinem Aufwand gleichzusetzen ist (Andreassen, 2000).

In einer Auflistung bestehender Verfahren zur Messung von Kundenzufriedenheit unterschied Beutin (2003) zunächst zwischen objektiven und subjektiven Verfahren der Kundenzufriedenheitsmessung. *Objektive Verfahren* erfassen direkt beobachtbare Größen wie die Entwicklung des Marktanteils oder des Gewinns eines Unternehmens. Da diese Maße nach Beutin (2003) kontaminiert und beeinflusst durch Faktoren wie zum Beispiel die aktuelle Konjunktur sind, werden sie als ungeeignet zur Messung

von Kundenzufriedenheit erachtet. Hingegen messen *subjektive Verfahren* die empfundene oder wahrgenommene Zufriedenheit von Kunden. Subjektive Verfahren können aufgrund ihres Messinhalts nach Homburg und Werner (1998) in ereignis- und merkmalsorientierte Verfahren untergliedert werden. Ereignisorientierte Verfahren widmen sich einem einzelnen, als bedeutsam erlebten Konsumereignis, zum Beispiel einem Restaurantbesuch. Für diesen Inhaltsbereich liegen nach Homburg und Werner (1998) vier Klassen von Messinstrumenten vor: Kontaktpunktanalysen, Frequenz-Relevanz-Analysen, die Analyse von Standardereignissen und die CIT. Gemäß Homburg, Schäfer und Schneider (2003) wird durch die Konzentration auf singuläre episodische Ereignisse kein ausreichendes Bild der Kundenzufriedenheit geschaffen. Bei merkmalsorientierten Verfahren wird die Zufriedenheit des Kunden über einen längeren Zeitraum und anhand der Thematisierung einer Vielzahl an Service- oder Produktmerkmalen gemessen (Oderkerken-Schröder, van Birgelen, Lemmink, de Ruyter & Wetzels, 2000). Ein weit verbreiteter Repräsentant merkmalsorientierter Verfahren ist das *SERVQUAL-Verfahren* (z. B. Parasuraman, Zeithaml & Berry, 1988; 1991). Parasuraman, Zeithaml und Berry (1988) charakterisierten ihr Messinstrument wie folgt: „[...] *a concise, multiple-item scale with good reliability and validity that retailers can use to better understand the service expectations and perceptions of consumers, and as a result, improve service*” (S. 30). Gemäß dem Ansatz von Parasuraman et al. (1988, 1991) wird die Zufriedenheit von Kunden mit der Servicequalität eines Unternehmens über fünf Dimensionen gemessen: *Tangibility*, *responsiveness*, *reliability*, *assurance* und *empathy*. In Übereinstimmung mit dem oben dargestellten Diskonfirmationsmodell beruht das SERVQUAL-Verfahren ebenfalls auf einem Vergleich zwischen einem IST-Wert – der durch Kunden wahrgenommenen Ausprägung der genannten fünf Dimensionen – mit einem SOLL-Wert – der von ihnen erwarteten Ausprägung hinsichtlich dieser Dimensionen. Eine umfassende Bewertung und Analyse der aktuellen Forschungsergebnisse zum SERVQUAL-Verfahren wird durch Morrison Coulthard (2004) vorgenommen. Das SERVQUAL-Verfahren erfreut sich im Übrigen einer regelmäßigen Anwendung bei der Messung von Kundenzufriedenheit im Hotel- und Gaststättengewerbe (z. B. Wong Ooi Mei, Dean & White, 1999).

Bei der Messung der Kundenzufriedenheit ist weiterhin zwischen impliziten und expliziten Verfahren zu differenzieren (Beutin, 2003). *Implizite Verfahren* erfassen vorrangig Beschwerden oder Reklamationen von Kunden, da diese als Zeichen von Kundenunzufriedenheit interpretiert werden können. Allerdings zeigt die Mehrzahl von Kunden kein aktives Beschwerdeverhalten, sondern ergeht sich zum Beispiel in negativer Mund-zu-Mund-Propaganda. Aus diesem Grund empfiehlt Beutin (2003), die Kunden mittels *expliziter Verfahren* auf persönlichem, telefonischem oder schriftlichem Wege nach seiner Zufriedenheit zu befragen.

2.6.3 Kundenloyalität

Korrespondierend zu den obigen Ausführungen zur Kundenzufriedenheit werden auch zur Erklärung des Zustandekommens von *Kundenloyalität* mehrere Theorien in der (Marketing-) Literatur diskutiert. Bereits vor 30 Jahren differenzierten Jacoby und Chestnut (1978) zwischen mehr als 53 verschiedenen Ansätzen. Grundsätzlich können jedoch zwei verschiedene Perspektiven voneinander abgegrenzt werden (vgl. Foscht, 2001; S. 40 ff.) – eine verhaltensbezogene und eine stärker psychologisch geprägte Sichtweise –, welche in heutigen Ansätzen zur Auslegung von Kundenloyalität (z. B. Homburg & Giering, 2001) kombiniert werden.

Kundenloyalität (engl.: customer loyalty) kann aus einer *verhaltensbezogenen Perspektive* verstanden werden als „[...] *repeat purchasing of a particular product or service over time*“ (Homburg & Giering, 2001; S. 46). Messgrößen, welcher sich diese Perspektive bedient, entstammen vor allem dem tatsächlich beobachtbaren Verhalten, z. B. der Kaufintensität, der Zeitdauer seit dem letzten Einkauf oder der Kontaktdichte mit dem Unternehmen (Brown, 1952; Day, 1969). Dabei wird das vergangene loyale Verhalten von Kunden als Determinante zukünftigen loyalen Verhaltens angesehen (Foscht, 2001). Nach Day (1969) kann aber zum Beispiel das manifeste Wiederkaufsverhalten zum einen tatsächliche Loyalität indizieren, zum anderen aber auch eine Pseudo-Loyalität (engl.: spurious loyalty), welche dem Zufall oder der Situation entspringt. Zudem fungieren Variablen wie die soziale Norm oder die Motivation des Kunden als wichtige Einflüsse auf den Wiederkauf (Dick & Basu, 1994). Dieser auf die Verhaltensperspektive beschränkte Ansatz liefert weiterhin keine Erklärungsmuster für den Erwerb von Produkten des täglichen Ge-

brauchs oder das Auftreten impulsiver, ungeplanter Käufe (Dabholkar, 1999). Verfechter einer *stärker psychologisch geprägten Sichtweise* lassen die Verhaltensperspektive nicht außer Acht, sehen aber eine positive Einstellung gegenüber dem Anbieter des Produktes oder der Dienstleistung als essentiell für das Zustandekommen einer engen Bindung eines Kunden an das Unternehmen an (Martin, Ponder & Lueg, 2009). Erst aufgrund dieser Einstellung lasse sich demnach unterscheiden zwischen tatsächlich loyalen, bewusst wiederkommenden Kunden und solchen Kunden, welche zufällig oder aus Mangel an Alternativen wiederkaufen (s. o.). Zur Operationalisierung der Kundenloyalität greifen Vertreter dieses Standpunktes auf die Erfragung von nicht-direkt beobachtbaren Größen wie der Wiederkaufabsicht zurück.

Zusammengefasst lässt sich Kundenloyalität als zweidimensionales Konstrukt verstehen, in welchem sich eine einstellungsbezogene Präferenz (nicht-beobachtbar) gegenüber bestimmten Produkten oder Dienstleistungen von Unternehmen ausdrückt, was mit einer verstärkten Wiederinanspruchnahme (beobachtbar) eben jener einhergeht: „[...] *a loyal customer is a customer who repurchases from the same service provider whenever possible, and who continues to recommend or maintains a positive attitude towards the service provider*“ (Kandampully & Suhartanto, 2000; S. 346). An positiven Effekten für ein Unternehmen ist anzuführen, dass die durch den Kunden geäußerte Loyalität zum Unternehmen zu einer stärkeren Toleranz gegenüber Servicefehlern und Prozessen wie Preiserhöhungen seitens des Unternehmens führen kann (Zeithaml, 1988; Müller & Riesenbeck, 1991; Hess, Ganesan & Klein, 2003). Loyale Kunden hegen außerdem zumeist stärkeres Interesse an einer langjährigen Verbindung zu Unternehmen und daran, dass beide Seiten – sowohl Kunde als auch Unternehmen – von der Beziehung profitieren (Gilliland & Bello, 2002). Ferner sprechen Reichheld und Kenny (1990) an, dass Mitarbeiter eines Unternehmens das Vorhandensein loyaler Kunden auf ihre eigene Aktivität zurückführen und darauf aufbauend ihre eigene Zufriedenheit und Treue gegenüber dem Unternehmen steigern können. Allerdings birgt eine Fokussierung auf loyale Kunden auch Risiken in sich, da womöglich andere Kunden vernachlässigt werden (Fournier, Dobscha & Mick, 1998) oder ein nicht-profitabler Betrag in den loyalen Kundentamm investiert wird. Zudem ist aus Sicht eines Unternehmens zu beach-

ten, dass nicht alle Kunden sämtliche Serviceleistungen aus einer Hand erhalten wollen (Day, 2000) und dass in bestimmten Branchen Kunden eine loyale Bindung an mehrere Unternehmen gleichzeitig knüpfen können (Jacoby & Chestnut, 1978).

Um die Loyalität der Kunden langfristig aufrechtzuerhalten und zu belohnen, initiiert eine Vielzahl von Unternehmen sogenannte *loyalty programs* respektive *frequent-guest programs* im Hotel- und Gaststättengewerbe, an denen sich zum Beispiel in den USA mehr als 90 % der Kunden beteiligen (Berman, 2006). Ein bekanntes und weit verbreitetes loyalty program in Deutschland stellt zum Beispiel PAYBACK® dar.

In Anlehnung an Mimouni-Chaabane und Volle (2010; S. 32 ff.) haben diese Programme für den Kunden einen wahrgenommenen utilitaristischen Nutzen – der Kunde spart Geld und Zeit für die (Kauf-)Entscheidung –, einen hedonistischen Nutzen – der Kunde empfindet Freude beim Sammeln von Boni – und einen symbolischen Nutzen – der Kunde fühlt sich distinguiert und einer selektierten Gruppe angehörig. Auf der anderen Seite investieren Unternehmen jährlich hohe Beträge, um diese Programme aufrechterhalten zu können. So nannte Bond (1995) einen Betrag zwischen 60 und 80 Millionen US-Dollar, den amerikanische Hotels aufzubringen haben. Mit der Nennung des finanziellen Aufwands für Unternehmen wird der Bogen geschlagen zum finalen Glied der in Abbildung 2-7 aufbereiteten Erfolgskette der Kundenorientierung – dem Kundenwert.

2.6.4 Kundenwert

Der *Kundenwert* repräsentiert „[...]eine ökonomische Größe der Unternehmenssteuerung“ (Bruhn, 2009; S. 39) und stellt damit eine vor allem für den betriebswirtschaftlichen Kontext bedeutsame Variable dar. Aus diesem Grund beschränkt sich die Beschäftigung mit dieser Variable im Rahmen der vorliegenden Arbeit auf eine knappe Ausführung.

Zur Bestimmung des (finanziellen) Werts, welcher ein Kunde für ein Unternehmen besitzt, sind zwei Klassen von Modellen – eindimensionale und mehrdimensionale Kundenwertmodelle – unterscheidbar (vgl. Eggert, 2006; S. 45 ff.). *Eindimensionale Kundenwertmodelle* stützen sich entweder allein auf monetäre Kriterien, zum Beispiel Kundenerfolgsrechnungen und customer lifetime-Analysen, oder beziehen alternativ nicht-

monetäre Kriterien, beispielsweise Kundenzufriedenheitsanalysen und Kaufhäufigkeitsanalysen, in die Berechnung ein. Das dominante Modell innerhalb dieser Klasse ist die sogenannte ABC-Analyse (vgl. Krafft & Marzian, 1997), welche nach Höhe des Umsatzes geordnete Kundenkategorien und -segmente schafft. Gemäß Eggert (2006) kann auf Seiten der *mehrdimensionalen Kundenwertmodelle* differenziert werden zwischen Kunden-Portfolioanalysen, welche sich an Kundenstämmen orientieren, und ganzheitlichen Kundenwertmodellen. Ein exemplarisches Kundenwertmodell geht auf Cornelsen (2000) zurück, welcher die Dimensionen Umsatz, Referenzwert, Informationswert sowie cross-selling-Wert zu einem einheitlichen Kundenwert verrechnete.

Durch die in diesem Abschnitt vorgenommene Darstellung der Erfolgskette der Kundenorientierung sollte einerseits erläutert werden, dass die Variable Kundenorientierung nicht losgelöst von Wirkzusammenhängen mit anderen Variablen zu betrachten ist. Andererseits sollte betont werden, welche wichtige Bedeutung der Kundenorientierung von Mitarbeitern für die Konkurrenz- und letztendlich Überlebensfähigkeit eines Unternehmens zukommt.

Die soeben erfolgte Auseinandersetzung mit der Erfolgskette der Kundenorientierung beschließt die theoretische Einführung dieser Arbeit. Im nachfolgenden Kapitel wird die Entwicklung eines videobasierten SJTs zur Messung von Kundenorientierung im Hotelfach beschrieben. Daran schließen sich weitere Studien an, welche Einblicke verschaffen sollen in die Nützlichkeit dieses videobasierten SJTs für die Personalauswahl.

3. Die erste Studie: Testentwicklung

In diesem Kapitel wird die methodische Vorgehensweise bei der Entwicklung eines videobasierten SJT zur Messung von Kundenorientierung im Hotelfach beschrieben. Die Entwicklung des SJT erfolgte in Kooperation mit einem in Münster ansässigen, zu einer internationalen Kette gehörenden Hotel. An seinem Münsteraner Standort beschäftigt das Unternehmen rund 160 Mitarbeiter. Im weiteren Verlauf soll das kooperierende Hotel mit *Hotel Münster* benannt werden. Von dem SJT erhofft sich das Hotel Münster eine Bereicherung seiner bisherigen Verfahren – Sichtung der Bewerbungsunterlagen, strukturiertes Einstellungsinterview und Arbeitsprobe – zur Auswahl neuer Auszubildender für eine Beschäftigung im Hotelfach. Die Entwicklung des SJT gestaltete sich als mehrstufiger Prozess, dessen Teilschritte im Folgenden nacheinander vorgestellt werden.

Um den zugrunde liegenden Arbeitsbereich samt seiner speziellen Ansprüche an Bewerber zu explorieren und im SJT adäquat abbilden zu können, wurde zu Beginn eine *Anforderungsanalyse* durchgeführt (Abschnitt 3.1). Im Rahmen der Anforderungsanalyse wurden *Interviews* mit verschiedenen SME des Hotel Münster durchgeführt. In diesen Interviews wurden sowohl typische berufsbezogene Situationen als auch Verhaltensweisen zur Bewältigung der Situationen erhoben. Die auf diese Weise ermittelten Verhaltensweisen dienten in einem weiteren Schritt als Grundlage zur Ableitung der für eine Beschäftigung im Hotelfach notwendigen Anforderungen (Abschnitt 3.2). An die Anforderungsanalyse schloss sich die nächste Entwicklungsstufe des SJT in Form von *Workshops mit Führungskräften* des Hotels Münster an. Während der Expertenworkshops wurden die zu verfilmenden Situationen samt dazugehöriger Verhaltensalternativen festgelegt. Zudem wurde im Rahmen der Workshops ein *Bewertungsschlüssel* für den SJT formuliert. Ihren Abschluss fand die Entwicklung des SJT in Form der *Produktion von Videoclips*, welche vor Ort in den Räumlichkeiten des Kooperationspartners gedreht wurden (Abschnitt 3.3). In der Abbildung 3-1 werden die verschiedenen Entwicklungsstufen des SJT schematisch aufgezeigt.

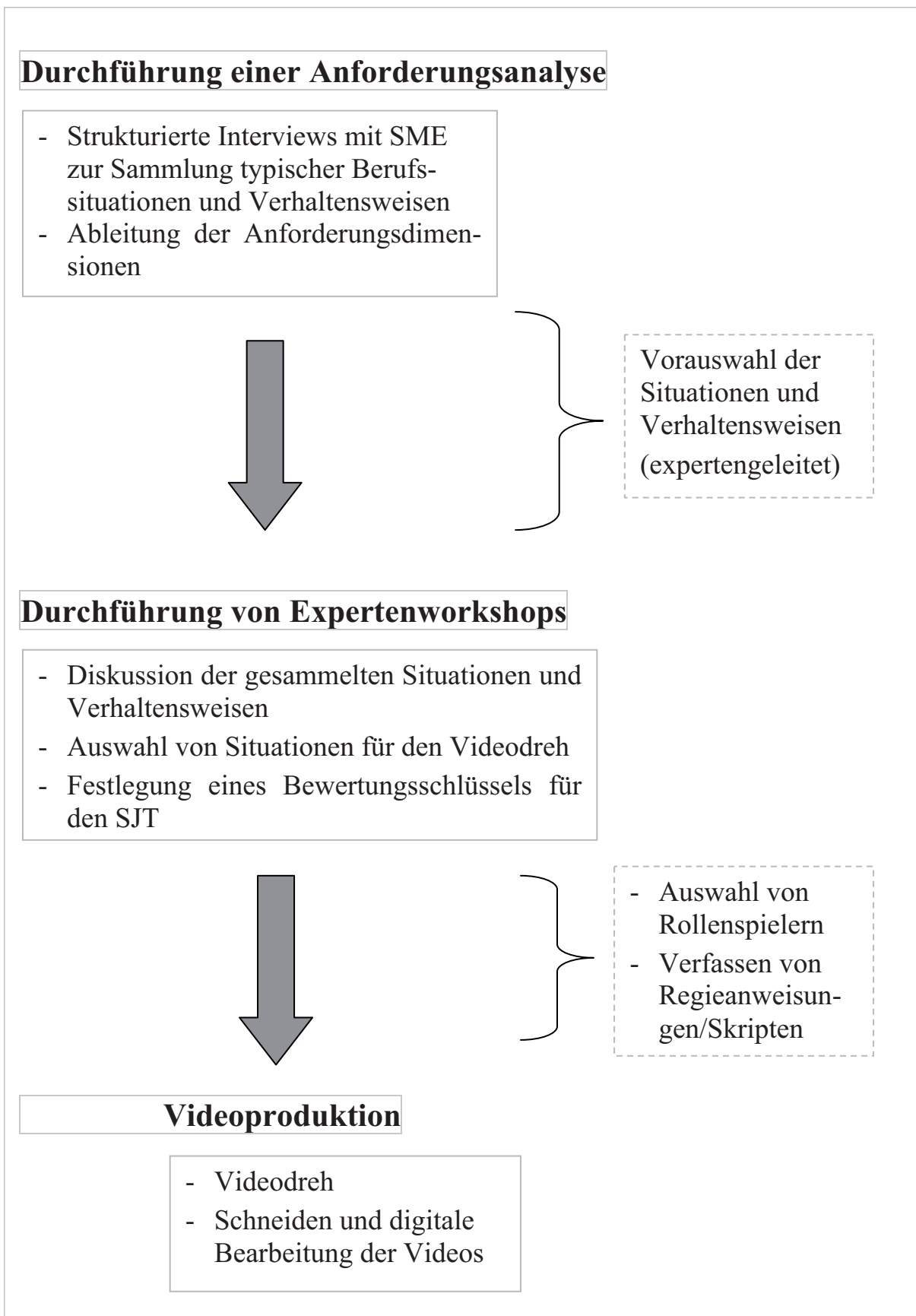


Abbildung 3-1: Methodisches Vorgehen bei der Konstruktion des SJT

3.1 Durchführung einer Anforderungsanalyse

Eine Meta-Analyse von McDaniel et al. (2001) ergab, dass SJTs¹⁷, welche auf einer Anforderungsanalyse beruhen, eine signifikant höhere mittlere kriterienbezogene Validität erreichen als SJTs, welche ohne vorhergehende Durchführung einer Anforderungsanalyse entstanden sind (vgl. Abschnitt 2.3). Aus diesem Grund wurde auch für den dieser Arbeit zugrunde liegenden SJT die Verwirklichung einer Anforderungsanalyse als unerlässlich erachtet.

Zur Anwendung in dieser Arbeit kam die *Methode der kritischen Ereignisse*, welche ursprünglich auf die Studien von Flanagan (Critical Incidents Technique, CIT, vgl. Flanagan, 1954) zurückzuführen ist. Flanagan (1954) beabsichtigte mit der CIT folgenden Einsatzzweck: „*The critical incident technique consists of a set of procedures for collecting direct observations of human behavior in such a way as to facilitate their potential usefulness in solving practical problems and developing broad psychological principles*” (S. 327). Zunächst werden einige allgemeine Ausführungen zu dieser Methodik angeführt, bevor das konkrete Vorgehen in dieser Studie präsentiert wird.

In der heutigen organisationspsychologischen Tätigkeit wird die CIT zumeist zur Sammlung sogenannter *erfolgskritischer Situationen* im Dialog mit SME eingesetzt (Schuler & Höft, 2004). Dieser Prozess des Zusammentragens von Situationen beinhaltet das Erfragen von Verhaltensweisen, welche für die effektive Ausübung der Tätigkeit Bedeutsamkeit besitzen (Zemke & Kramlinger, 1982). Das American Institute for Research (1998) ergänzt zu diesem Punkt, dass auf systematische Art und Weise Verhaltensweisen identifiziert werden können, welche zum Erfolg oder Misserfolg einer Person oder Organisation in spezifischen Situationen beitragen. Die auf diesem Wege gesammelten Informationen dienen dann zum Beispiel als Grundlage für Personalentwicklungsmaßnahmen oder zur Erstellung von Anforderungsprofilen in der Personalauswahl (vgl. Kanning, Pöttker & Klinge, 2008). Schuler und Höft (2004) zählen die CIT zu den Formen der Anforderungsanalyse, welche auf der Verhaltensebene eines Individuums ansetzen, um daraus Fertigkeiten abzuleiten,

¹⁷ Es sei nochmals daran erinnert, dass in die Meta-Analyse von McDaniel et al. (2001) keine videobasierten SJTs einbezogen wurden.

welche zum Beispiel durch Simulationen wie SJTs gemessen werden können. Laut Stitt-Gohdes, Lambrecht und Redmann (2000) ist die CIT jedoch nicht für die Anwendung in allen Arbeitsumfeldern angebracht, sondern vor allem für die Analyse solcher Tätigkeiten empfehlenswert, bei denen den Handelnden stets mehrere Handlungsalternativen offen stehen, wie zum Beispiel Angestellten im Hotelfach. Auf der Ebene der Datenanalyse dominiert bei der CIT, ähnlich wie bei anderen qualitativen Verfahren, die Inhaltsanalyse (Bitner et al., 1990). Eine Bestimmung psychometrischer Gütekriterien im Sinne von Reliabilität oder Validität ist nach Schuler (2006) eher die Ausnahme als die Regel (s. Abschnitt 3.4).

Bezogen auf die Mitarbeiter-Kunden-Interaktion müssen von SME beschriebene Situationen gemäß Bitner, Bommers und Tetreault. (1990; S. 73) folgenden Anforderungen gerecht werden, um als erfolgskritisch im Sinne der CIT anerkannt zu werden: *„An incident was required to meet the four criteria of (1) involving employee-customer interaction, (2) being very satisfying or dissatisfying from the customer's point of view, (3) being a discrete episode, and (4) having sufficient detail to be visualized by the interviewer.“* Diese Voraussetzungen waren auch für die vorliegende Arbeit prägend. Ihre Umsetzung wird nachfolgend weiter ausgearbeitet.

Da Auszubildende im Hotel Münster in den drei Jahren ihrer Ausbildung ähnlich einem Traineeprogramm für Hochschulabsolventen verschiedene Stationen innerhalb des Hotels durchlaufen – Rezeption, Service (Restaurant und Bar) sowie Bankett –, wurden für die Interviews Vertreter aller relevanten Ausbildungsstationen ausgewählt. Die Selektion der Gesprächspartner und Festsetzung der Gesprächstermine wurde durch die Personalchefin des Hotel Münster vorgenommen. Seitens des Autors bestanden jedoch zwei Vorgaben: Erstens sollten Gesprächspartner mit einem heterogenen Maß an Berufserfahrung im Bereich Hotelfach gefunden werden, um den Arbeitsbereich aus diversen Perspektiven und vor dem Hintergrund unterschiedlicher Expertise zu beleuchten. Zweitens sollten nur solche Mitarbeiter als Gesprächspartner gewonnen werden, welche in direktem Kundenkontakt stehen, die sogenannten *frontline employees* (z. B. Nerdinger, 2003). Diese Restriktion schloss zum Beispiel Mitarbeiter der Küche – ausgenommen Köche in leitender Position – aus.

Im Oktober und November 2007 führte der Autor zwölf Interviews mit SME des Hotel Münster durch. Alle Gespräche fanden in Räumlichkeiten

des Hotel Münster statt. Die Tabelle 3-1 gibt einen Überblick über relevante Charakteristika der zwölf beteiligten Personen, welche dem Autor für ein Gespräch zur Verfügung standen.

Tabelle 3-1: Charakteristika der befragten SME

	Geschlecht	Arbeitsbereich im Hotel	Berufserfahrung in Jahren (Hotelfach) ¹⁸
1	Männlich	Service (Restaurant)	20
2	Männlich	Service (Restaurant)	10
3	Weiblich	Bankett	18
4	Weiblich	Auszubildende (1. Jahr)	2 Monate
5	Weiblich	Bankett	7
6	Weiblich	Rezeption	17
7	Weiblich	Auszubildende (3. Jahr)	2
8	Männlich	Küche	42
9	Weiblich	Rezeption	6
10	Männlich	Rezeption	5
11	Männlich	Service (Bar)	20
12	Weiblich	Service (Restaurant)	10

Von den zwölf Gesprächspartnern hatten fünf (41.67 %) zum Zeitpunkt der Interviews Führungsaufgaben inne. Im Schnitt verfügten die Gesprächspartner über eine Berufserfahrung von rund 14 Jahren. Die durchschnittliche Gesprächsdauer belief sich auf 65 Minuten, wobei das längste Interview anderthalb Stunden andauerte und das kürzeste Gespräch 45 Minuten.

Zu Beginn eines jeden Gesprächs wurden die SME über den Zweck des Gesprächs sowie die Anonymität ihrer Daten aufgeklärt. Dabei wurde insbesondere darauf geachtet, dass die SME darüber informiert werden, dass ihre Angaben nicht an Vorgesetzte oder andere Entscheidungsträger weitergeleitet werden und keinesfalls als irgendeine Form der Leistungsbeurteilung gedacht sind (vgl. Moser, 2004).

¹⁸ Einige SME konnten ihre Berufserfahrung nicht exakt in Jahren bestimmen, so dass manche Werte als Schätzungen zu betrachten sind.

Die Einleitung der Interviews orientierte sich an einer von Bitner, Booms und Mohr (1994, S. 97) vorgeschlagenen Instruktion, welche in der Abbildung 3-2 einzusehen ist.

Put yourself in the shoes of customers of your firm. In other words, try to see your firm through your customers' eyes. Think of a recent time when a customer of your firm had a particularly satisfying (dissatisfying) interaction with yourself or a fellow employee. Describe the situation and exactly what happened. They were then asked the following questions:

1. When did the incident happen?
2. What specific circumstances led up to this situation?
3. Exactly what did you or your fellow employee say or do?
4. What resulted that made you feel the interaction was satisfying (dissatisfying) from the customer's point of view?
5. What should you or your fellow employee have said or done? (for dissatisfying incident only)

*Abbildung 3-2: Vorlage zur Instruktion der Interviewpartner
(nach Bitner et al., 1994; S. 97)*

Vom Autor wurde folgende standardisierte Instruktion eingesetzt: „Bitte versetzen Sie sich für dieses Gespräch in die Lage Ihrer Hotelgäste. Erinnern Sie sich nun bitte an eine berufsbezogene Situation aus den letzten zwölf Monaten, in der eine Begegnung zwischen Ihnen und einem oder mehreren Hotelgästen aus Kundensicht entweder zufrieden stellend oder aber nicht zufrieden stellend verlaufen ist“. Eine standardisierte Instruktion dieser Art sollte die Gesprächspartner auf solche Situationen lenken, welche aus Kundensicht entweder in einem zufrieden stellenden oder aber nicht zufrieden stellenden Ergebnis mündeten. Über diese spezifische Ausrichtung und leichte inhaltliche Abkehr von der in der Abbildung 3-2 dargestellten Instruktion nach Bitner et al. (1994) sollte das Thema Kundenorientierung im Hotelfach stärker in den Vordergrund gerückt werden, da wie in der theoretischen Einleitung der vorliegenden Arbeit beschrieben, kundenorientiertes Verhalten als entscheidende Stellgröße bei der Erreichung von Kundenzufriedenheit angesehen wird (s. Abschnitt 2.6). Die von Bitner et al. (1994) genannten Anschlussfragen nach den Vorbe-

dingungen der Situation, dem konkreten Verhalten des Mitarbeiters in der Situation und den daraus erwachsenden Konsequenzen für den Gast, welche dem sogenannten A-B-C-Konzept der CIT entsprechen, wurden im Verlaufe der Situationsbeschreibung gestellt. Ein weiterer Punkt, welchem beim Ergründen berufstypischer Situationen erhöhte Beachtung geschenkt wurde, war der, dass der Autor den Gesprächspartner während seiner Situationsbeschreibung dazu aufforderte, möglichst konkrete von Hotelgästen oder von ihm selbst in der Situation genannte Ausdrücke oder gar komplette Sätze anzuführen (s. Abschnitt 3.4). Obendrein sollte der Gesprächspartner schildern, wie Hotelgäste in seinem konkreten Arbeitsbereich begrüßt, verabschiedet oder nach weiteren Wünschen befragt werden. Auf diesem Wege sollte ein späteres Verfassen der Skripte und Drehbücher für den Videodreh sowie deren Bezug zum konkreten Arbeitsplatz erleichtert werden (s. u.). Gelangte der Gesprächspartner zu einem Punkt, an dem er keine Situationen mehr nennen konnte, welche er selbst mit Hotelgästen erlebt hatte, bat ihn der Autor um die Beschreibung von Situationen, an denen Arbeitskollegen beteiligt waren. Nach Beendigung einer Situationsbeschreibung gab der Autor das Gesagte in eigenen Worten wieder, um zu überprüfen, ob die dargestellten Situationen richtig erfasst worden sind. Wurde die Zusammenfassung des Autors vom Gesprächspartner als passend eingestuft, begann der nächste Teil des Gesprächs. In diesem forderte der Autor den Gesprächspartner auf, Verhaltensweisen zu generieren, welche aus seiner Sicht in der soeben beschriebenen Situation effektiv im Sinne einer erfolgreichen Situationsbewältigung wären. Zusätzlich fragte der Autor den Gesprächspartner nach Verhaltensweisen, welche nach dessen Meinung ineffektiv oder sogar hinderlich im Sinne einer erfolgreichen Situationsbewältigung wären. Vorab wurde der Gesprächspartner vom Autor dazu angehalten, lediglich realistische und authentische Verhaltensweisen anzugeben. Nannte der Gesprächspartner stark verallgemeinernde Verhaltensweisen wie zum Beispiel „freundlich sein“ als positive Verhaltensweisen, bat ihn der Autor um eine spezifischere Formulierung und Auslegung. Aufgrund der oben beschriebenen Konzentration auf Situationen, welche zufriedene oder unzufriedene Kunden hinterließen, sind die gesammelten Verhaltensalternativen nicht nur als situativ effektiv respektive ineffektiv, sondern auch als situationsspezifisch kundenorientiert beziehungsweise nicht kundenorien-

tiert zu titulieren. Alle Antworten und Angaben der Gesprächspartner wurden vom Interviewer in ein formales Protokollblatt übertragen. Nach abgeschlossener Durchführung der zwölf Interviews lagen dem Autor 82 Situationen vor. Von den zusammengetragenen Situationen mündeten 39 Situationen (47.56 %) in einem für den Hotelgast nicht zufrieden stellenden Ergebnis, während der Ausgang von 43 Situationen (52.44 %) einen zufriedenen Hotelgast hinterließ. In der Abbildung 3-3 befindet sich sowohl ein Beispiel für eine aus Sicht des Kunden zufrieden stellend verlaufene Situation als auch für eine nicht zufrieden stellend verlaufende Situation.

A. Eine den Hotelgast nicht zufrieden stellende Situation

Im Rahmen einer saisonalen Wildgerichtaktion werden spezielle Speisen angeboten. Ein Hotelgast beschwert sich über sein Gericht (Fasanenbrust), dass es ihm nicht saftig genug, sondern viel zu trocken sei. Der angesprochene Hotelmitarbeiter bietet dem Gast ein neues Gericht auf Kosten des Hauses an, was jener jedoch ablehnt.

B. Eine den Hotelgast zufrieden stellende Situation

Beim Sonntagsbrunch rempelt ein Gast unabsichtlich einen Hotelmitarbeiter an. Daraufhin vergießt dieser ein Getränk auf das Jackett des Gastes. Der Mitarbeiter entschuldigt sich prompt und sagt dem Gast ad hoc eine Übernahme der Reinigungskosten zu. Der Hotelgast akzeptiert die Entschuldigung und bedankt sich für die rasche Abwicklung des Problems.

Abbildung 3-3: Beispiele für Situationen, welche bei Hotelgästen Zufriedenheit beziehungsweise Unzufriedenheit ausgelöst haben

Zur Ableitung der relevanten Anforderungsdimensionen wurden die in Gesprächen mit SME protokollierten Verhaltensweisen losgelöst von ihrem konkreten situativen Kontext und ihrer Valenz (situativ effektiv versus situativ ineffektiv) analysiert. Bei diesem Schritt der Anforderungsanalyse unterstützen den Autor zwei weitere mit der CIT vertraute Diplom-Psychologen. Die einzelnen Verhaltensweisen wurden durch die Gruppe zu Clustern thematischer Ähnlichkeit gruppiert. Damit eine Gruppierung Akzeptanz erhielt, mussten alle drei beteiligten Personen der Zuordnung zustimmen. Falls kein Konsens erreicht werden konnte, ordnete

die Gruppe die Verhaltensweisen bis zum Erreichen einer Übereinstimmung neu. In einem letzten Schritt wies die Gruppe den gebildeten Clustern von Verhaltensweisen inhaltliche Label zu.

Folgende fünf Anforderungsdimensionen wurden abgeleitet: *Serviceorientierung*, *Beschwerdemanagement*, *Durchsetzungsfähigkeit*, *Handlungsfähigkeit* und *Gewissenhaftigkeit*. Die Definitionen und die inhaltlich prägenden Verhaltensweisen sind Bestandteil des Anhangs A.

In einem hierarchischen Modell der Kundenorientierung treten die fünf genannten Anforderungsdimensionen als Faktoren 1. Ordnung auf. Den fünf Anforderungsdimensionen übergeordnet wird wiederum ein Faktor 2. Ordnung, welcher mit *Kundenorientierung* benannt werden soll. Einen Eindruck des aufgestellten Modells verschafft die Abbildung 3-4.

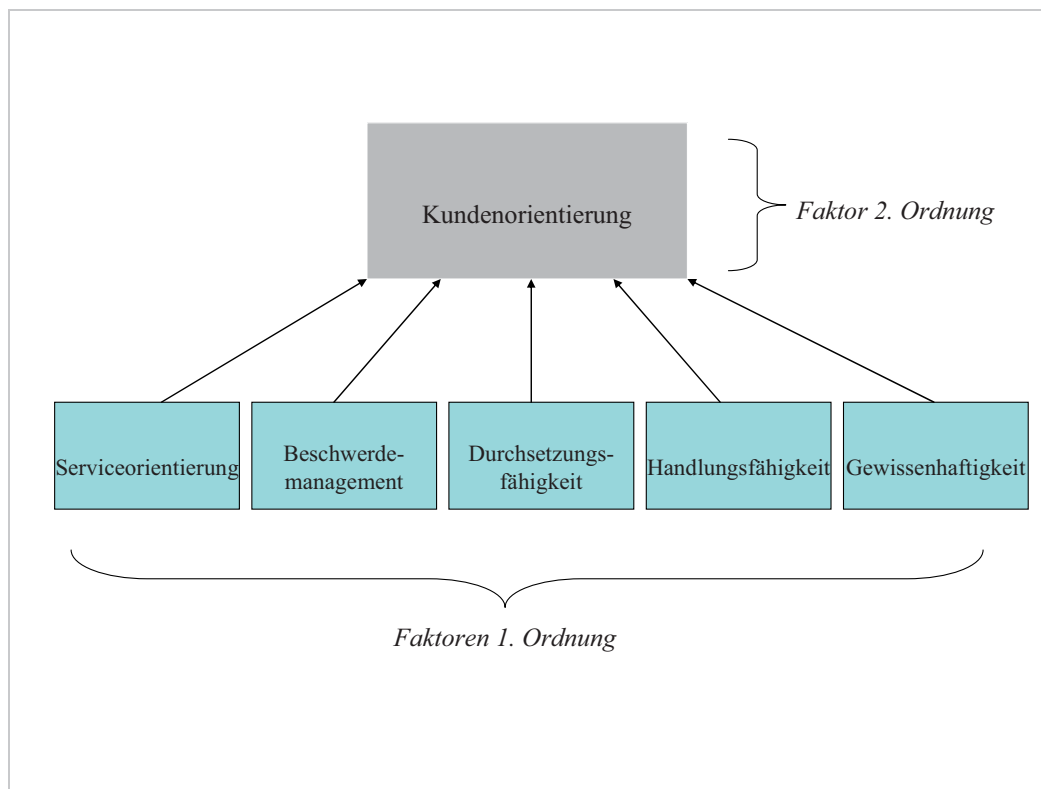


Abbildung 3-4: Hierarchisches Modell der Kundenorientierung

Bevor mit der Durchführung der Expertenworkshops die zweite Entwicklungsstufe des SJT eingeläutet wurde, fanden zwei die Workshops vorbereitende Zwischenschritte statt.

Im Rahmen des ersten Zwischenschrittes nahm der Autor eine *Sichtung und Vorauswahl* der in den Interviews mit SME gesammelten Situationen

vor. Dieser Prozess sollte verschiedenen Zwecken dienlich sein: Zunächst wurden geschlechtsspezifische Situationen aus der weiteren Entwicklung des videobasierten SJT ausgeschlossen. Als Beispiele für geschlechtsspezifische Situationen sind zwei Situationen verkürzt in der Abbildung 3-5 enthalten.

Ein männlicher Gast versucht über den ganzen Abend hinweg, eine Mitarbeiterin an der Bar in ein Gespräch zu verwickeln. Nach ihrem Dienstschluss bitte er sie, mit ihm auf sein Hotelzimmer zu gehen.
Beim Abendbuffet beschwert sich ein Gast bei einem männlichen Mitarbeiter, er habe den Eindruck, der Mitarbeiter starre seiner Ehefrau unentwegt ins Dekolleté.

Abbildung 3-5: Beispiele für geschlechtsspezifische Situationen

Neben geschlechtsspezifischen Situationen schied man zu ungewöhnliche, nur äußerst selten auftretende – das heißt maximal einmal pro Jahr – Situationen aus. Um die Situationen derart beurteilen zu können, schätzten die Personalchefin des Hauses und eine Abteilungsleiterin die verbleibenden Situationen – unabhängig voneinander – hinsichtlich ihrer Auftrenshäufigkeit und Repräsentativität für den Arbeitsalltag ein. Herrschte Dissens zwischen der Personalchefin und der Abteilungsleiterin in Bezug auf den Realitätsgrad einer Situation, wurde aufgrund der Menge vorhandener Situationen gegen den Verbleib der fragwürdigen Situation in der weiteren Testentwicklung entschieden. Des Weiteren identifizierte der Autor inhaltlich redundante Situationen und fasste diese zu einer einzigen situativen Kategorie zusammen. Beispielshalber schilderten die SME wiederholt Situationen, in denen sich Gäste über die mangelhafte Qualität des servierten Essens oder die unzureichende Sauberkeit ihres Hotelzimmers beschwert haben. Nach Einsatz der drei genannten Selektionskriterien (Geschlechtsspezifität, Realismus und Redundanz) bildeten 41 Situationen die Grundlage für die Workshops mit SME. Zur Vorbereitung der Expertenworkshops wurden diese 41 Situationen weiterhin den verschiedenen Ausbildungsstationen (s. o.) im Hotel zugeordnet. Dementsprechend fasste der Autor zum Beispiel alle Situationen, welche sich an der Hotelrezeption ereignet haben, zusammen. In einem zweiten Zwischenschritt ordnete der Autor jeder der 41 Situationen jeweils vier der in den

Interviews gesammelten Verhaltensweisen zu. Die Zuweisung von vier bis sechs Verhaltensalternativen zu einer Situation wird in der Forschungsliteratur empfohlen (Ployhart & Ehrhart, 2003). Von diesen vier Verhaltensalternativen wurde erwartet, dass sie ein unterschiedlich ausgeprägtes Spektrum an Kundenorientierung abbilden können. Zudem wurde bei ihrer Gestaltung darauf geachtet, dass sie sich nicht auf Anrieb respektive nach Augenschein als besonders effektiv respektive besonders ineffektiv zu erkennen geben (vgl. Weekley et al., 2006). Weitere Schritte der Gestaltung oder Korrektur sollten den unten beschriebenen Expertenworkshops vorbehalten bleiben.

3.2 Durchführung von Expertenworkshops

Wie bei den Interviews zur Anforderungsanalyse geschah die Auswahl der Teilnehmer an den Workshops durch die Personalchefin des Hotel Münster. Anders als bei den Interviews sollten in diese Phase der Testentwicklung aber explizit nur Abteilungsleiter beziehungsweise Mitarbeiter mit Führungsaufgaben beteiligt werden. Diese Festlegung auf eine bestimmte Hierarchieebene sollte eine stärkere Verankerung der Unternehmenswerte und -philosophie sowohl in dem Bewertungsschlüssel für den SJT als auch in der finalen Ausformulierung der Verhaltensalternativen ermöglichen. Die Teilnehmer an den Workshops waren darüber hinaus andere SME als diejenigen, welche in den Interviews befragt wurden. Auf diesem Wege sollte eine Unabhängigkeit der verschiedenen Entwicklungsschritte gewährleistet werden.

Die von der Personalchefin des Hotel Münster ausgesuchten SME erhielten vom Autor vorab, eine Woche vor Beginn des Workshops, die in dem jeweiligen Workshop zu besprechenden Situationen samt vier Verhaltensalternativen in einer Paper-Pencil-Version. Den Situationen und Verhaltensalternativen vorangestellt waren Informationen über den Sinn und Zweck der Workshops. Zur inhaltlichen Vorbereitung und Sicherstellung eines ökonomischen Ablaufs der Workshops bat der Autor die SME um die Erfüllung dreier Aufgaben: Zunächst sollten sie die ihnen dargebotenen Situationen auf einer fünfstufigen Likertskala (1 = *überhaupt nicht realistisch* bis 5 = *vollkommen realistisch*) hinsichtlich ihres Realitätsgrades einschätzen. Um den SME diese Einstufung zu erleichtern, wurden ihnen nur solche Situationen zugeteilt, die thematisch ihn ihr spezifisches

Berufsfeld, zum Beispiel Bankett, fielen. Zweitens sollten die beteiligten SME jene vier pro Situation aufgelisteten Verhaltensalternativen gemäß ihrer situativen Effektivität bei der Zufriedenstellung des Hotelgastes in eine Rangordnung bringen: An die ersten Stelle sollten sie diejenige Verhaltensalternative setzen, die sie als effektivste benennen würden, das heißt die den Kunden maximal zufrieden stellt, während die letzte Stelle der Rangordnung von der Verhaltensalternative eingenommen werden sollte, welche nach Meinung des SME am ineffektivsten wäre, das heißt einen maximal unzufriedenen Hotelgast hinterlässt. Dieser Schritt sollte der Festlegung eines Bewertungsschlüssels für den SJT dienlich sein. Eine dritte Anfrage des Autors bezog sich auf mögliche Verbesserungsvorschläge der SME sowohl hinsichtlich der Situationsbeschreibung als auch der Formulierung der Verhaltensalternativen. Abschließend bat der Autor die SME darum, ihre Urteile und Bewertungen nicht vor Beginn des Workshops untereinander auszutauschen. Zudem sollten die Situationen und Verhaltensweisen nicht an andere Mitarbeiter des Hotel Münster weitergeleitet werden.

Vom Autor wurden im Januar 2008 insgesamt zwei Expertenworkshops im Hotel Münster abgehalten. An dem ersten Workshop nahmen drei SME des Hotel Münster teil. Die Durchführung beanspruchte eine Zeit von dreieinhalb Stunden. An diesem Workshop nahmen SME mit verantwortlicher Tätigkeit in den Bereichen Rezeption und Bankett teil, die 16 Situationen besprechen sollten. Für den zweiten Workshop standen vier SME des Hotel Münster zur Verfügung. Zeitlich schlug dieser mit annähernd vier Stunden zu Buche. Die Teilnehmer an diesem Workshop waren in den Bereichen Housekeeping sowie Service (Restaurant/Bar) beschäftigt und hatten 25 Situationen als Diskussionsgrundlage.

Von den Teilnehmern an den Workshops waren zwei männlich und fünf weiblich. In Bezug auf Berufserfahrung erreichten die Teilnehmer einen durchschnittlichen Wert von rund 13 Jahren, wobei der Teilnehmer mit der meisten Berufserfahrung seit 30 Jahren und der Teilnehmer mit der geringsten Erfahrung seit sechseinhalb Jahren im Hotelfach angestellt waren.

Zu Beginn eines Workshops wurden die Teilnehmer nochmals über den Zweck des Zusammenkommens und über den bereits vollendeten ersten Schritt der Testentwicklung, die Interviews mit ihren Kollegen, infor-

miert. Daran schloss sich die Besprechung der Situationen und Verhaltensalternativen an. Bei der Besprechung der Situationen wurde vom Autor die folgende standardisierte Vorgehensweise beachtet: Anfangs teilten die SME dem Autor der Reihe nach ihre Einschätzungen des Realitätsgrades mit. Ergab sich für eine Situation eine über die Expertenurteile gemittelte Einschätzung von größer/gleich 3, wurde diese beibehalten. In den Workshops wurde acht Situationen im Schnitt ein geringerer Realitätsgrad bescheinigt, so dass diese ausschieden aus der weiteren Testentwicklung. Nach der Besprechung des Realitätsgrades der jeweiligen Situation erfolgte die Diskussion der a priori erstellten Rangordnungen der vier Verhaltensalternativen je Situation, um einen Bewertungsschlüssel für den SJT festzulegen. McDaniel und Nguyen (2001) bezeichneten die Herbeiführung eines Expertenkonsenses als eine geeignete Möglichkeit, um solch einen Bewertungsschlüssel anzufertigen.

Jeder SME¹⁹ trug in diesem Fall zunächst seine persönliche Rangordnung für die zu besprechende Situation vor. Über die beiden Workshops hinweg konnte für vier Situationen ein Expertenkonsens im Sinne übereinstimmender Rangordnungen hergestellt werden, so dass diese als erste Situationen für den Videodreh feststanden. Da sich bei den anderen 29 Situationen keine Übereinstimmung zwischen den Rangordnungen ergeben hatte, mussten die relevanten Verhaltensalternativen inhaltlich revidiert und umgestaltet werden (vgl. McDaniel & Nguyen, 2001). Für acht Situationen konnte selbst im Zuge einer inhaltlichen Revision kein Konsens zwischen den SME erreicht werden, so dass diese für den SJT nicht weiter in Frage kamen (vgl. Abschnitt 3.4). Diejenigen 25 Situationen, welche die Workshops überdauert hatten, verteilten sich wie folgt auf die verschiedenen Ausbildungstationen im Hotel Münster: Neun Situationen ereigneten sich an der Hotelrezeption, zwölf spielten sich im Bereich des Restaurants ab und vier im Bankett- und Veranstaltungsumfeld.

Am Ende eines Workshops wurde den SME die Möglichkeit eingeräumt, weitere Bedenken oder Änderungswünsche zu äußern. In diesem Verlauf wurden vor allem inhaltliche Anpassungen der Verhaltensalternativen vorgenommen, da manche Experten zum Beispiel bestimmte Formulie-

¹⁹ Nicht jeder SME konnte aus zeitlichen Gründen für alle Situationen vorab eine Rangordnung anfertigen.

rungen für unpassend hielten oder aber bestimmte Verhaltensalternativen als zu leicht einstuften.

Die 25 Items²⁰, welche den SJT nach Ablauf der beiden Expertenworkshops konstituieren, können den oben genannten fünf Anforderungsdimensionen Serviceorientierung, Beschwerdemanagement, Durchsetzungsfähigkeit, Handlungsfähigkeit und Gewissenhaftigkeit zugeteilt werden. In Orientierung an den Definitionen der fünf Dimensionen (s. Anhang A) und der inhaltlichen Gestaltung der betroffenen vier Verhaltensalternativen soll ein Item eindeutig derjenigen Anforderung zugeordnet werden, bei der eine hohe Ausprägung eine erfolgreiche Bewältigung der jeweiligen Situation verheißt (s. Abschnitt 3.4). Welches Item welcher Anforderungsdimension zugewiesen wird, ist der Tabelle 3-2 zu entnehmen. Die Nummer des Items entspricht der Position im SJT (s. Anhang A).

Tabelle 3-2: Zuordnung der 25 SJT-Items zu den 5 Anforderungsdimensionen

Anforderungsdimension	Items	Anzahl an Items
Serviceorientierung	2,6,8,9,17,18	6
Beschwerdemanagement	3,7,13,20,23	5
Durchsetzungsfähigkeit	4,14,19,25	4
Handlungsfähigkeit	1,5,11,16,22	5
Gewissenhaftigkeit	10,12,15,21,24	5

In späteren Studien soll eine empirische Überprüfung dieser auf qualitativem Wege vorgenommen Zuordnung erfolgen. Nach Kanning (2005) thematisieren die meisten SJTs das Bewältigen von Problemen und Konflikten im sozialen Kontext. Auch den 25 Situationen kann eine deutliche soziale Komponente zugebilligt werden, da in allen eine Begegnung zwischen einem Hotelmitarbeiter und einem Hotelgast respektive zwei Hotelgästen thematisiert wird. Für eine bessere Einsicht in die Inhalte der einzelnen Situationen sei auf den Anhang A verwiesen.

Bevor der Dreh der 25 Videos beginnen konnte, mussten einerseits kurze Skripte respektive Drehbücher durch den Autor verfasst und andererseits

²⁰ Der Begriff Item bezieht sich auf eine Situation und vier dazugehörige Verhaltensalternativen (s. Kapitel 4).

Rollenspieler für den Videodreh gefunden werden. Bei der Erstellung kurzer Drehbücher und Skripte dienten die in den Gesprächen mit SME gesammelten Ausdrücke und Sätze von Hotelgästen sowie gängige Begrüßungsformen und Fragen an Hotelgäste durch Hotelmitarbeiter zur Einleitung der Situationen als Vorlage. Neben der Aufzeichnung der unterschiedlichen Gesprächsinhalte wurde den Skripten hinzugefügt, in welcher Stimmung oder in welcher Gefühlslage sich der Hotelgast jeweils befinden sollte und wie dies durch entsprechende Gestik zu untermalen sei. Für den Hotelmitarbeiter wurde bei den einzelnen Skripten angemerkt, welche Aktionen parallel zu der Interaktion mit dem Hotelgast durchzuführen sind, zum Beispiel das Drucken einer Rechnung oder das Servieren von Speisen. Ein exemplarisches Drehskript ist in der Abbildung 3-6 nachzulesen.

Hotelmitarbeiter (HM): „Herzlich Willkommen! Wie kann ich ihnen weiterhelfen?“ (lächelt Hotelgast (HG) an)

HG: „Guten Tag. Ich würde gerne einen Tisch für Freitag, so ab 20 Uhr, im Wintergarten reservieren. Wir sind 10 Personen.“ (freundlich)

HM: „Oh, das tut mir leid. Für diesen Tag haben wir bereits zwei größere Gruppen im Wintergarten.“ (blickt HG etwas ratlos an)

HG: „Das ist aber ärgerlich. Was machen wir denn da?“ (wirft Hände in die Luft und schaut HM fragend an)

- HM blättert im Reservierungsbuch -

Abbildung 3-6: Beispiel für ein Drehskript

Alle erstellten Drehbücher und Skripte wurden vor dem Videodreh von der Personalchefin des Hotel Münster gegengelesen und hinsichtlich ihres Realitätsgrades begutachtet. Solche Skripte, an denen Mängel oder Versäumnisse festgestellt wurden, unterzog der Autor einer Korrektur.

3.3 Videoproduktion

Zur Verkörperung der Hotelmitarbeiter wurden alleinig tatsächliche Angestellte des Hotel Münster kontaktiert und für den Videodreh gewonnen. Auch diese Unternehmensangehörigen waren bei keiner der bisherigen Entwicklungsstufen des SJT in Erscheinung getreten. Hingegen wurden zur Darstellung der Hotelgäste unternehmensfremde Akteure mit mehrjähriger Bühnenerfahrung im Theater oder anderem Erfahrungshintergrund, zum Beispiel als Rollenspieler in Personalauswahlverfahren, einbezogen. Alle Rollenspieler erhielten vor Drehbeginn ein vom Autor angefertigtes Exposé über den Sinn und Zweck des Videodrehs sowie eine Danksagung für ihre Beteiligung an dem Projekt. Zudem erhielten sie die für ihren Drehtag notwendigen Skripte und Rollenspieleranweisungen, um sich mit ihrer Aufgabe vorab vertraut zu machen und den Videodreh auf ein ökonomisches Maß reduzieren zu können²¹.

Der Videodreh im Hotel Münster und die technische Nachbearbeitung der Videos wurde unterstützt und begleitet vom *Servicepunkt Film*, welcher dem Zentrum für Informationsverarbeitung (ZIV) der Westfälischen Wilhelms Universität Münster angehört. Die 25 Videos in ihrer Rohfassung wurden an insgesamt vier Drehtagen im März und April 2008 erstellt. Als Schauplätze dienten drei Bereiche des Hotel Münster: Ein Restaurant, die Rezeption und ein Veranstaltungssaal.

Zu Beginn eines jeden Drehtages mussten die Rollenspieler – sowohl die Beschäftigten des Hotel Münster als auch die Rollenspieler, welche Hotelgäste darstellen sollten – einer Einverständniserklärung zustimmen, dass die von ihnen erstellten Aufnahmen honorarfrei zum Zwecke der Testentwicklung vervielfältigt, veröffentlicht und verbreitet werden dürfen. Zudem informierte der Autor die beteiligten Rollenspieler vor Drehbeginn nochmals über den Zweck des Videodrehs und klärte etwaige Fragen seitens der Rollenspieler.

Nach Abschluss des letzten Drehtages wurden die Aufnahmen vom Autor unter Anleitung eines Mitarbeiters des Servicepunkts Film audiovisuell nachbearbeitet. Eine derartige Nachbearbeitung war vonnöten, da

²¹ Ein ökonomischer Zeitrahmen für den Videodreh war vor allem deshalb zu wählen, weil die Videos im Hotel Münster während des normalen Hotelbetriebes gedreht werden mussten.

zum Beispiel bei Szenen, welche an der Rezeption gedreht wurden, die Stimmen realer Hotelgäste oder das Klingeln eines Telefons stärker in den Hintergrund gemischt werden mussten, um nicht den Dialog der Rollen-
spieler zu überdecken. Der darauf folgende Schnitt der Videos wurde mit Hilfe des Programms AVID® vollzogen. Die durchschnittliche Länge eines Videoclips betrug rund 19 Sekunden, wobei die längste Szene ungefähr 35 Sekunden und die kürzeste Szene rund acht Sekunden andauerte. Zum Abschluss der technischen Nachbearbeitung wurde mittels des Programms ADOBE PHOTOSHOP® ein DVD-Menü erstellt, und die 25 Videos wurden für die spätere Präsentation im Rahmen der Testdurchführung auf DVD gebrannt. Eine DVD mit den 25 Videos ist dieser Arbeit beigelegt. Abschließend werden drei Abbildungen angeführt, welche die drei vorherrschenden Interaktionsmuster 22F zwischen Hotelmitarbeiter und Hotelgast beziehungsweise Hotelgästen darstellen.



*Abbildung 3-7: Interaktionsmuster I
(Hotelmitarbeiter steht, Gast steht)*

²² Bei einer verfilmten Situation stehen drei Hotelgäste in einer Schlange am Kuchenbuffet.



*Abbildung 3-8: Interaktionsmuster II
(Hotelmitarbeiter steht, Gast sitzt)*



*Abbildung 3-9: Interaktionsmuster III
(Hotelmitarbeiter steht, zwei Gäste sitzen)*

Nach der Vorstellung der methodischen Vorgehensweise bei der Entwicklung des SJT sollen zum Abschluss dieses Kapitels kritische Punkte und Anmerkungen genannt werden, welche bei den einzelnen Entwicklungsschritten auftraten.

3.4 Diskussion

Die Diskussion orientiert sich an dem in der Abbildung 3.1 einsehbaren Ablauf der Testentwicklung.

Zur Feststellung der Anforderungen im Hotelfach entschied sich der Autor für die CIT. Bereits Flanagan (1954) wies auf die Freiheitsgrade hin, welche der Anwender der CIT besäße: „It should be emphasized that the critical incident technique does not consist of a single rigid set of rules governing such data collection. Rather it should be thought of as a flexible set of principles which must be modified and adapted to meet the specific situation at hand” (S. 336). Insbesondere der Ableitung der Anforderungsdimensionen wurde ein erhöhtes Maß an Subjektivität zugeschrieben: „In the absence of an adequate theory of human behavior, this step is usually an inductive one and is relatively subjective” (S. 336). Um den Grad an Subjektivität bei den Interviews mit SME einzudämmen, wurden jeweils einheitliche Fragen im Sinne einer Standardisierung an die SME verwandt und eine einheitliche Reihenfolge der Fragen und Abschnitte im Interview (s. o.) eingehalten. Bei der Ableitung der Anforderungsdimensionen wurden zwei weitere Diplom-Psychologen neben dem Autor einbezogen. Auch dieser Schritt sollte das Ausmaß an Subjektivität minimieren, indem Maße der Inter-Rater-Übereinstimmung (vgl. Fisseni, 2004) stärker in den Blickpunkt rückten. Jedoch bleibt insbesondere die Benennung der Verhaltenscluster nicht frei von subjektiven Anteilen und Vorstellungen. Nach Di Salvo, Nikkel und Monroe (1989) stellen nämlich gerade die Klassifikation und weitere Analyse der gesammelten Ereignisse und Verhaltensweisen die größte Abkehr von einer Objektivierbarkeit der Ergebnisse dar.

Das zentrale Element der Interviews mit SME im Kontext der Anforderungsanalyse war die Nennung von Situationen, in denen der SME Einsicht erhielt in die Unzufriedenheit beziehungsweise Zufriedenheit von Hotelgästen. An dieser Stelle ist zu erwähnen, dass viele unzufriedene Kunden ihre Unzufriedenheit nicht direkt an Mitarbeiter eines Unternehmens weitergeben, sondern oftmals durch negative Mund-zu-Mund Propaganda an Freunde oder Bekannte weitergeben. Matilla (2001) konnte zum Beispiel in Studien aufzeigen, dass Personen schlechte Erfahrungen mit Unternehmen an zehn bis zwanzig Bekannte weiterleiten, aber nicht

an Mitarbeiter des betroffenen Unternehmens. Dementsprechend werden viele kritische Ereignisse von Mitarbeitern nicht als solche erfasst. Ferner ist davon auszugehen, dass die Mitarbeiter bestimmte Situationen selbst trotz Zusicherung der Anonymität und des Datenschutzes nicht genannt haben, da diese ihnen zum Beispiel peinlich oder unangenehm waren oder womöglich eine negative Außendarstellung des Hotel Münster evoziert hätten. Neben Situationen und Verhaltensweisen fragte der Autor SME nach möglichst genauen Sätzen oder Ausdrücken in Situationen. Dies sollte zum einen den Realitätsgrad der im SJT verwandten Situationen erhöhen und zum anderen das Schreiben der Skripte und Rollenanweisungen für den Videodreh erleichtern. Aufgrund der hier oftmals gegebenen zeitlichen Distanz zur geschilderten Situation sind Gedächtniseffekte oder Verzerrungen wie bei der CIT insgesamt nicht auszuschließen. Andererseits waren viele erfolgskritische Situationen, vor allem solche mit unzufriedenen Hotelgästen, nach Einschätzung des Autors den Gesprächspartnern oftmals wortwörtlich in Erinnerung. Dies könnte womöglich damit begründet werden, dass diese Situationen auch für den Hotelmitarbeiter sehr belastend oder unangenehm und damit leichter direkt abrufbar waren. Eine zusätzliche und aus Sicht des Autors nahe liegende Befragung von Hotelgästen konnte aufgrund von Bedenken seitens des Hotel Münster nicht erfolgen. Eine Einbeziehung von Kunden in die CIT findet sich zum Beispiel bei Bitner et al. (1994).

Dass für die beiden zur Entwicklung des SJT vollzogenen Expertenworkshops ausschließlich Führungskräfte des Hotel Münster akquiriert wurden, sollte dazu führen, dass die zumeist von oberen Hierarchieebenen vorgegebenen Unternehmenswerte einen starken Niederschlag in der Formulierung der Verhaltensalternativen und in der Festlegung eines Bewertungsschlüssels finden. Es ist nicht ausgeschlossen, dass Mitarbeiter ohne Führungsaufgaben einen anderen Bewertungsschlüssel präferiert hätten. Zudem soll nicht der Gedanke außer Acht gelassen werden, dass viele Führungskräfte eher selten in direkten Kundenkontakt treten und sich deshalb möglicherweise gedanklich weniger leicht in Kunden hineinversetzen können als frontline employees, welche tagtäglich mit Kunden umgehen müssen.

Nur für vier Situationen stellten die Experten unabhängig voneinander dieselbe Rangordnung hinsichtlich der situativen Effektivität von vier

vorgegebenen Verhaltensalternativen her. Als Erklärung für diesen niedrigen Prozentsatz (12.12 % der in den Expertenworkshops besprochenen Situationen) können mehrere mögliche Gründe angeführt werden. Erstens besteht die Möglichkeit, dass innerhalb des Hotel Münster keine einheitliche Festlegung von Unternehmenswerten besteht beziehungsweise dass Führungskräfte unterschiedlicher Arbeitsbereiche oder Berufserfahrung diese Werte unterschiedlich und variabel auslegen. Ein zweiter Grund ist in der inhaltlichen Gestaltung der Verhaltensalternativen zu sehen. So herrschte in nahezu allen Situationen Konsens zwischen den Experten bezüglich der besten und schlechtesten Verhaltensalternativen. Differenzen traten jedoch bei der Einordnung der mittleren Verhaltensalternativen zutage, die weder als richtig gut noch als richtig schlecht in der jeweiligen Situation zu bezeichnen waren. Dies deutet auf ein generelles Problem der Konstruktion von SJTs hin. Je mehr Verhaltensalternativen einer Situation zugeordnet werden, desto schwieriger wird sowohl die Wahrung der inhaltlichen Trennschärfe der Situationen untereinander als auch eine eindeutige Bestimmung der situativen Effektivität einzelner Verhaltensalternativen. Bei Situationen, für die a priori kein einheitlicher Bewertungsschlüssel seitens der Experten genannt wurde, kam es zu Diskussionen zwischen den Experten. In Gruppendiskussionen mit sehr vielen Mitgliedern besteht die Wahrscheinlichkeit, dass sich einzelne Mitglieder zurückhalten oder dass einzelne Mitglieder eine dominante Rolle einnehmen (vgl. Endres & Putz-Osterloh, 1994; Farmer & Roth, 1998). Um diesen möglichen Prozessverlusten entgegenzuwirken und den Beitrag eines jeden Mitglieds identifizieren zu können, beschränkte der Autor die Expertenrunden auf maximal vier Teilnehmer. Außerdem bat der Autor ruhigere Teilnehmer – falls notwendig – darum, ihre eigene Meinung zu äußern und sich zu beteiligen.

Für den Videodreh der 25 in Expertenworkshops festgelegten Situationen wurden Hotelmitarbeiter und semiprofessionelle Rollenspieler als Akteure eingesetzt. Bei vergleichbaren Testentwicklungen (z. B. DeCotiis, 1984) wurden sowohl die Hotelmitarbeiter als auch die Hotelgäste von professionellen Schauspielern verkörpert. Diesem Vorhaben stand bei dieser Arbeit der zu große finanzielle Aufwand im Wege.

Innerhalb der gedrehten Videoszenen herrschte ausschließlich die Interaktion und Kommunikation zwischen jeweils einem Hotelmitarbeiter und

einem einzigen Hotelgast vor. Zwar waren bei sieben der 25 Videoszenen andere Hotelgäste zu sehen, zum Beispiel indem sie an demselben Restaurantisch saßen oder in einer Warteschlange standen, jedoch traten sie dabei nicht aktiv in Erscheinung. Diese Beschränkung auf ein 1:1-Verhältnis wurde vom Autor aus Gründen leichter Nachvollziehbarkeit und Verständlichkeit gewählt. Im wahren Hotelalltag sind auf beiden Seiten, gerade in kritischen Situationen, mehrere handelnde Personen zugegen. Eine weitere künstliche Reduktion des Hotelalltags besteht darin, dass beide Parteien sich in den 25 Szenen stets haben ausreden lassen und sich nicht gegenseitig ins Wort gefallen sind. Auch dies ist trotz einer leichteren Verständlichkeit für die Rezipienten vor dem Hintergrund eines realistischen Hotelalltags zu hinterfragen.

3.5 Ausblick

Im nachfolgenden Kapitel wird eine Studie vorgestellt, in der Auszubildende im Hotelfach den aus 25 Situationen bestehenden SJT absolvieren. Zur Berechnung der Testwerte im SJT kommen zwei verschiedene scoring keys zur Anwendung. Neben dem SJT wird ein Verfahren zur Messung der sozialen Kompetenz eingesetzt, um Aussagen über die Konstruktvalidität des SJT treffen zu können.

4. Die zweite Studie

In diesem Kapitel wird eine Studie besprochen, in welcher Auszubildende im Hotelfach den in Kapitel 3 vorgestellten SJT sowie einen Persönlichkeitsfragebogen zur Messung der sozialen Kompetenz durchlaufen. Zunächst werden die eingesetzten Verfahren beschrieben, bevor der Ablauf der Datenerhebung und die Charakteristika der Stichprobe dargestellt werden. Daran schließt eine Auswertung der an der Stichprobe gewonnenen Daten an. Einen Abschluss erhält das Kapitel einerseits durch eine Diskussion der methodischen Vorgehensweise sowie der dadurch hervorgerufenen Ergebnisse und andererseits durch einen Ausblick auf die im nachfolgenden Kapitel vorgestellte Studie.

4.1 Beschreibung der eingesetzten Instrumente und Fragebögen

Das Ziel dieses Abschnitts ist die Erläuterung der Zusammensetzung der in dieser Studie verwendeten Testbatterie. Alle eingesetzten Instrumente wurden für die Teilnehmer der Studie in Form eines einzigen Fragebogens zusammengefasst²³. Die einzelnen Teilabschnitte dieses Fragebogens sind samt ihrer Seitenzahl in der Tabelle 4-1 enthalten.

Tabelle 4-1: Aufbau des Fragebogens im Rahmen der zweiten Studie

Teilabschnitte	Seite(n) im Fragebogen
Persönlicher Code	1
Inventar sozialer Kompetenzen – Kurzversion (33 Items)	2 – 4
Situational Judgment Test (25 Items)	5 – 30
Demographische Angaben (6 Items)	31

Nachfolgend werden die verschiedenen Testverfahren genauer vorgestellt. In Übereinstimmung mit dem Aufbau des Fragebogens (s. Tabelle 4-1) wird mit einer Darstellung des Inventars sozialer Kompetenzen in der Kurzversion begonnen. Daran schließt sich ein Überblick über den einge-

²³ Dieser Vorgang sollte eine leichtere Handhabbarkeit für die Teilnehmer gewährleisten. Innerhalb des Fragebogens wurden die verschiedenen Tests jedoch als eigenständige Verfahren behandelt.

setzten SJT ein. Abgeschlossen wird dieser Abschnitt durch die Nennung der erhobenen demographischen Variablen.

4.1.1 Das Inventar sozialer Kompetenzen in der Kurzversion

Von Kanning (2009) wurde das Inventar sozialer Kompetenzen (ISK) zur Messung allgemeiner sozialer Kompetenzen sowohl in einer Langversion, welche 108 Items der Selbstbeschreibung umfasst, als auch in einer Kurzversion mit 33 Items (ISK-K) vorgelegt. Ein exemplarisches Item des ISK-K befindet sich in der Abbildung 4-1.

	trifft gar nicht zu	trifft eher nicht zu	trifft eher zu	trifft sehr zu
Wenn ich mich mit anderen Menschen unterhalte, dann schweifen meine Gedanken oft ab.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 4-1: Beispielitem aus dem ISK-K

Da die vorgegebenen Rahmenbedingungen der Datenerhebung einen Einsatz der Langversion verhindern²⁴, wird im Folgenden ausschließlich der für diese Studie relevante Stand der Forschung zum ISK-K dargelegt. Dem ISK-K liegt eine definitorische Unterscheidung zwischen sozial kompetentem Verhalten auf der einen Seite und sozialer Kompetenz auf der anderen Seite zugrunde. Die Abbildung 4-2 greift die beiden von Kanning (2002) vorgegebenen Definitionen auf.

²⁴ Die genauen Bedingungen, welche diesen Umstand herbeiführen, werden im nächsten Abschnitt 4.2.1 *Ablauf der Datenerhebung und Beschreibung der Stichprobe* ausführlicher beschrieben.

Sozial kompetentes Verhalten
Verhalten einer Person, das in einer spezifischen Situation dazu beiträgt, die eigenen Ziele zu verwirklichen, wobei gleichzeitig die soziale Akzeptanz des Verhaltens gewahrt wird.
Soziale Kompetenz
Gesamtheit des Wissens, der Fähigkeiten und Fertigkeiten einer Person, welche die Qualität eigenen Sozialverhaltens – im Sinne der Definition sozial kompetenten Verhaltens – fördert.

Abbildung 4-2: Definitionen sozial kompetenten Verhaltens und sozialer Kompetenz

Gemäß Kanning (2009) stellt das Vorhandensein sozialer Kompetenzen eine notwendige, aber nicht hinreichende Bedingung für sozial kompetentes Verhalten dar. Während soziale Kompetenzen globaler, situationsübergreifender Natur sind, wird konkretes Verhalten erst durch den jeweiligen Kontext oder die konkrete Situation als sozial kompetent oder aber als sozial inkompetent festgelegt.

Mit dem ISK-K werden vier Sekundärfaktoren der allgemeinen sozialen Kompetenz gemessen, welchen wiederum 17 Primärfaktoren²⁵ der sozialen Kompetenz zugrunde liegen. Kanning (2009) betitelt die vier Sekundärfaktoren mit Soziale Orientierung, Offensivität, Selbststeuerung und Reflexibilität. Die Faktoren Offensivität und Reflexibilität werden jeweils über sieben Items gemessen, während zur Messung des Faktors Selbststeuerung acht Items herangezogen werden. Zehn Items werden eingesetzt, um den Faktor Soziale Orientierung der Messbarkeit zuzuführen. Von Kanning (2009, S. 18 ff.) werden kurze Definitionen der vier Sekundärfaktoren vorlegt. Demnach bedeutet eine hohe Ausprägung auf dem Faktor Soziale Orientierung, dass eine Person ihrem Gegenüber in sozialen Interaktionen Aufmerksamkeit und Wertschätzung entgegenbringt. Ferner steht eine hohe Ausprägung dafür, dass sich eine Person in andere Personen hineinversetzt und abweichende Meinungen akzeptiert. Hingegen impliziert eine hohe Ausprägung des Faktors Offensivität, dass eine Person aktiv den Kontakt zu anderen Personen aufsucht und Netzwerke

²⁵ Jene 17 Primärfaktoren werden ausschließlich mit dem ISK gemessen, jedoch nicht mit dem ISK-K.

knüpft. Zudem ist eine starke Beschaffenheit dieses Faktors damit verbunden, dass eine Person notwendigen Konflikten nicht ausweicht. Aus einer starken Ausprägung des Faktors Selbststeuerung ergibt sich nach Kanning (2009), dass solche Personen sich selbst als verantwortlich für ihr Handeln und die daraus erwachsenden Konsequenzen ansehen. Sie bewahren auch bei Konflikten die Ruhe und kontrollieren ihre Emotionen. Außerdem können sich Personen mit einem hohen Grad an Selbststeuerung schnell auf sich verändernde Situationen einstellen. Zuletzt sei genannt, was sich gemäß Kanning (2009) hinter dem Faktor Reflexibilität verbirgt. Personen, die hohe Werte auf diesem Faktor erreichen, reflektieren ihr eigenes Verhalten in sozialen Interaktionen unter Beachtung der Reaktionen ihres Gegenübers. In solchen Situationen streben sie sowohl nach einer vorteilhaften Selbstdarstellung als auch nach einer Zufriedenstellung des Interaktionspartners. Die Abbildung 4-3 enthält die vier Sekundärfaktoren mit jeweils einem Itembeispiel aus dem ISK-K.

Sekundärfaktor im ISK-K	Itembeispiel
Soziale Orientierung	Auch wenn meine Zeit äußerst knapp bemessen ist, habe ich immer ein offenes Ohr für andere.
Offensivität	Ich gehe selbstsicher auf Menschen zu, wenn ich sie kennen lernen möchte.
Selbststeuerung	In meinem Leben kommt es sehr selten vor, dass ich mich hilflos
Reflexibilität	Ich habe ein ausgezeichnetes Gespür dafür, wenn Reden und Denken eines Menschen nicht übereinstimmen.

Abbildung 4-3: Sekundärfaktoren des ISK-K und Itembeispiele

Die Selbsteinschätzung der Teilnehmer wird auf allen 33 Items über eine vierstufige Ratingskala mit den Skalenendpunkten 1 = *trifft gar nicht zu* respektive 4 = *trifft sehr zu* eingeholt. Eine neutrale Antwortkategorie existiert nicht. Für die Durchführung des ISK-K veranschlagt Kanning (2009) eine Zeit von ungefähr zehn Minuten.

Für den ISK-K wurden an der untersuchten Gesamtstichprobe von 4208 Personen Reliabilitätswerte in Form von internen Konsistenzen erhoben

(Kanning, 2009; S. 42). Die Werte – jeweils als Cronbach's Alpha vorliegend – decken einen Bereich von .69 (Offensivität) bis .75 (Soziale Orientierung) ab. Neben Angaben zur internen Konsistenz liegt mit Retest-Reliabilitäten eine weitere Bestimmungsgröße der Verlässlichkeit des ISK-K vor. An einer studentischen Stichprobe wurden Retest-Reliabilitäten zwischen .80 (Offensivität) und .86 (Soziale Orientierung/Selbststeuerung) festgestellt.

Zur Bestimmung der Validität des ISK-K wurden bislang 19 Validierungsstudien durchgeführt, welche Zusammenhänge zu ausgewählten Persönlichkeitsskalen, den Big Five, Selbstwirksamkeit, Selbstsicherheit, Arbeits- und Lebenszufriedenheit, Beanspruchung in Studium und Beruf, Arbeitsleistung, biographischen Daten, political skills, der sozialen Integration, Netzwerkbildung sowie zur sozialen Anerkennung nachweisen.

4.1.2 Der eingesetzte SJT

Für die vorliegende Studie wurde ein SJT mit 25 Items eingesetzt, dessen Konstruktion im dritten Kapitel dieser Arbeit nachvollzogen werden kann. Der Begriff Item bezieht sich im Falle dieses SJT auf die Kombination aus einer Situation, welche als Video illustriert wird, und vier dazugehörigen Verhaltensalternativen im Paper-Pencil-Format. Demnach stellt die Durchführung des SJT ein Wechselspiel zwischen der Darbietung videogestützter Stimuli und der Bearbeitung von Paper-Pencil-Elementen dar. Um die Darstellung zu vereinfachen, werden die eingesetzten Items nachfolgend mit Situation 1,..., Situation 25 benannt. Mit dem SJT soll das Ziel verfolgt werden, die fünf in Kapitel 3 aufgestellten Anforderungsdimensionen des Hotelfachs zu erfassen. Es wird die Hypothese aufgestellt:

Hypothese I-1

Dem SJT liegt eine fünffaktorielle Struktur zugrunde. Die fünf Faktoren entsprechen den Anforderungsdimensionen Serviceorientierung, Beschwerdemanagement, Durchsetzungsfähigkeit, Handlungsfähigkeit und Gewissenhaftigkeit. Den fünf Anforderungsdimensionen wiederum liegt als Faktor 2. Ordnung Kundenorientierung zugrunde.

Zur Vorbereitung auf den eigentlichen Test wird den Teilnehmern vor der Durchführung des SJT eine *schriftliche Instruktion* gegeben, welche zum einen den Ablauf des SJT thematisiert und zum anderen die an die Teilnehmer gestellten Aufgaben erklärt. Bei jeder der 25 Situationen sollen die Teilnehmer aus einer Liste von vier im forced-choice Format dargebotenen Verhaltensalternativen einerseits die Verhaltensalternative bestimmen, die sie am wahrscheinlichsten zeigen würden, und andererseits diejenige Option benennen, welche sie am allerwenigsten ausführen würden. Die Art dieser Verhaltensinstruktion entspricht einer Instruktion zur Verhaltenstendenz (vgl. Abschnitt 2.3). Durch den Ausdruck von Verhaltenstendenzen und damit Präferenzen für bestimmte Verhaltensalternativen entwickeln sich Ziele und Absichten (Latham & Skarlicki, 1995), welche sich einerseits in Plänen, wie in bestimmten Situationen zu handeln ist (Gollwitzer, 1999), und andererseits in Form tatsächlichen Verhaltens umsetzen lassen (Latham & Saari, 1984). In der Abbildung 4-4 findet sich ein beispielhaftes Item aus dem SJT.

Videogestützte Situationspräsentation:



Verhaltensalternativen (Paper-Pencil)

- A. Sie erklären den Gästen, dass die Größe ihrer Speisen der normalen Größe in Ihrem Hause entspricht und bieten ihnen weitere Beilagen an, falls sie später noch Hunger haben sollten.
- B. Sie versichern den Gästen höflich, dass sie sich ausnahmsweise irren müssen, da in ihrem Hause stets gleichgroße Portionen ausgegeben werden.
- C. Sie nehmen den Fisch wieder mit und bestellen in der Küche einen kräftigen Nachschlag.
- D. Sie nehmen die Teller der Gäste wieder mit in die Küche und versuchen abzuklären, ob beim Auffüllen der Speisen etwas schiefgelaufen ist.

Verhaltensinstruktion (Paper-Pencil)

Am wahrscheinlichsten würde ich mich verhalten wie bei Buchstabe: ____

Am allerwenigsten würde ich mich verhalten wie bei Buchstabe: ____

Abbildung 4-4: Beispielitem aus dem SJT

Als Maßstab für die Einordnung der Antworten fungiert der von SME in Expertenworkshops entwickelte *Bewertungsschlüssel* (s. Kapitel 3; vgl. Anhang A). In diesen Workshops wurde für jede der 25 Situationen eine Rangordnung der vier zugeteilten Verhaltensalternativen gemäß ihrer situativen Effektivität erstellt. Die von den SME an Rangplatz 1 gesetzte Verhaltensalternative ist gleichzusetzen mit derjenigen Verhaltensalterna-

tive, welche in der jeweiligen Situation am wahrscheinlichsten durch die Experten gezeigt werden würde. Im Gegensatz dazu repräsentiert die auf dem vierten und letzten Rangplatz positionierte Verhaltensalternative diejenige Option, welche nach Ansicht der SME am allerwenigsten ausgeführt werden sollte. Bei jeder Situation wurde die Anordnung der vier Verhaltensalternativen im Übrigen per Permutation variiert, um zu vermeiden, dass sich zum Beispiel an der ersten Position einer Auflistung immer die von den Experten als am effektivsten oder ineffektivsten charakterisierte Verhaltensalternative befände. Um die Angaben der Teilnehmer zu einem Testwert zusammenfassen zu können, werden zwei verschiedene Formen des *Scorings* kontrastiert (vgl. Abschnitt 2.3). Bei der einen Variante erhalten die Teilnehmer einen Punkt (+1), wenn sowohl die Verhaltensalternative, welche sie am wahrscheinlichsten zeigen würden, als auch die Verhaltensalternative, welche sie am allerwenigsten zeigen würden, mit den von Experten bestimmten Verhaltensalternativen übereinstimmen. Liegt dagegen keine vollständige Kongruenz vor, werden dem Teilnehmer keine Punkte gutgeschrieben. Der Gesamtwert für den SJT ergibt sich als Summe der 25 Einzelscores. Im Falle einer maximalen Übereinstimmung – das heißt bei allen 25 Items – mit den SME erhielte ein Teilnehmer demnach 25 Punkte. Diese Art des Scorings orientiert sich an den Arbeiten von Stevens und Champion (1999) sowie Hunter (2003). Im weiteren Verlauf wird diese Form des Scorings mit *Pick best – pick worst* betitelt. Die zweite Form des Scorings basiert auf eine Studie von Motowidlo et al. (1990). In der Abbildung 4-5 wird dargestellt, wie sich die Punktevergabe nach Motowidlo et al. (1990) gestaltet.

- + 2 Punkte: Der Teilnehmer wählt als Verhaltensalternativen, welche er am wahrscheinlichsten und am allerwenigsten durchführen würde, dieselben Verhaltensalternativen, welche Experten gewählt haben.
- + 1 Punkt: Es stimmt entweder diejenige Verhaltensalternative, welche der Teilnehmer am wahrscheinlichsten zeigen würde, oder die Verhaltensalternative, welche er am allerwenigsten zeigen würde, mit der von Experten bestimmten Verhaltensalternative überein.
- 2 Punkte: Der Teilnehmer wählt als Verhaltensalternative, welche er am wahrscheinlichsten durchführen würde, diejenige Verhaltensalternative, welche Experten am allerwenigsten zeigen würden. Zudem sucht er als Verhaltensalternative, die er am allerwenigsten an den Tag legen würde, jene aus, welche Experten am wahrscheinlichsten performen würden.
- 1 Punkt: Entweder würde der Teilnehmer diejenige Verhaltensalternative, welche Experten am wahrscheinlichsten ausüben würden, am allerwenigsten offenbaren oder er würde die Verhaltensalternative, welche Experten am allerwenigsten zeigen würden, am wahrscheinlichsten an den Tag legen.
- 0 Punkte: Weder die Verhaltensalternative, welche der Teilnehmer am wahrscheinlichsten ausüben würde, noch diejenige, die er am allerwenigsten performen würde, stimmen mit den von Experten auserkorenen Optionen überein.

Abbildung 4-5: Scoring key in Anlehnung an Motowidlo et al. (1990)

Zur Berechnung eines Gesamtscores für den SJT wird bei dieser Variante des Scorings zunächst die Summe der Einzelscores für die 25 Items gebildet. In einem zweiten Schritt wird der Summenwert durch die Anzahl der Items dividiert. Aus dieser Berechnungsregel ergibt sich ein maximal erreichbarer Gesamtscore von +2. Im Folgenden firmiert diese Variante des Scorings unter der Bezeichnung *Motowidlo-Schlüssel*. Mit diesem Schlüssel wird ein Wertebereich von -2 bis +2 abgedeckt.

Ein Scoring der Teilnehmerantworten mit dem Motowidlo-Schlüssel respektive in Befolgung der Pick best – pick worst Variante wird aus ökonomischen Gesichtspunkten vorgenommen. Die Auswertung des SJT soll bei einer möglichen späteren Anwendung in der Praxis auch schnell und fehlerfrei von Laien durchführbar sein.

4.1.3 Demographische Variablen

Die Befragung der Teilnehmer endet mit der Erfassung verschiedener demographischer Angaben. Es werden das Alter, das Geschlecht, der höchste erreichte Schulabschluss, der Notendurchschnitt auf dem Schulabgangszeugnis, das Ausbildungsjahr sowie zuletzt die Muttersprache der Teilnehmer erfragt. Für eine Erfragung des Notendurchschnittes spricht eine von Schuler (2000) durchgeführte Meta-Analyse, in der sich für Haupt- und Realschulabschlussnoten eine mittlere korrigierte Validität von .41 bei der Vorhersage des Ausbildungserfolges ergab. Mit Hilfe der Abiturnote konnte sogar eine mittlere korrigierte Validität von .46 bei der Prognose des Studienerfolges erreicht werden.

4.2 Ablauf der Datenerhebung und Beschreibung der Stichprobe

4.2.1 Ablauf der Datenerhebung

Die Datenerhebung fand in Kooperation mit dem Adolph-Kolping-Berufskolleg in Münster statt. An dieser Berufsschule werden Fachkräfte im Gastgewerbe, Hotelfachleute, Hotelkaufleute, Fachleute für Systemgastronomie, Restaurantfachleute und Köche ausgebildet. Dabei erstreckt sich die Lehrzeit im Hotelfach in der Regel über drei Jahre. Als Stichprobe sollten mindestens 100 Berufsschüler aus dem ersten beziehungsweise zweiten Ausbildungsjahr getestet werden (s. Abschnitt 4.5).

Um die Datenerhebung mit Berufsschülern durchzuführen zu können, suchte der Autor zunächst den Kontakt zu der zuständigen schulinternen Abteilungsleiterin für das Hotel- und Gastgewerbe. In einem darauf folgenden Informationsgespräch stellte der Autor der zuständigen Abteilungsleiterin den SJT sowie den ISK vor. Zudem klärte der Autor die Abteilungsleiterin über den Sinn der Erhebung und die Zusicherung des Datenschutzes auf. Als Nutzen für die beteiligten Berufsschüler sicherte der Autor der Abteilungsleiterin die Erstellung eines schriftlichen Leistungsprofils für jeden Teilnehmer sowohl in Hinblick auf die Performanz im SJT als auch in dem ISK-K (s. Anhang B) zu. Nach der Zustimmung der Abteilungsleiterin zur Datenerhebung wurde ein Termin vereinbart, um die Erhebung zu realisieren. In diesem Zusammenhang wies die Abteilungsleiterin den Autor auf die Tatsache hin, dass die Datenerhebung nur

während der normalen Unterrichtszeit stattfinden könne. Zudem sollte die Stichprobe auf zwei Gruppen von jeweils rund 55-60 Schülern aufgeteilt werden, um eine störungsfreie und angenehme Atmosphäre während der Datenerhebung zu gewährleisten. Als zeitlichen Rahmen setzte sie jeweils 75 Minuten für die beiden Schülergruppen fest. Aufgrund dieser zeitlichen Beschränkung entschied sich der Autor für den Einsatz des ISK in der Kurz- statt in der Langversion (s. o.), da er allein für die (erstmalige) Durchführung des SJT rund 60 Minuten einplante.

Die *Datenerhebung* wurde im Juni 2008 vom Autor in einem Großgruppenraum des Berufskollegs durchgeführt. Die Präsentation der Videos erfolgte über eine sich an der Rückwand des Raumes befindende Leinwand. Die Stühle für die Schüler wurden in Form einer parlamentarischen Ordnung vor der Leinwand aufgestellt. Eine erste Schülergruppe wurde von 7:50 bis 9:20 Uhr getestet und die zweite Gruppe zwischen 9:40 und 11:10 Uhr. An beiden Teilerhebungen nahmen sowohl Schüler des ersten als auch des zweiten Ausbildungsjahres teil. Neben den Schülern waren stets auch die jeweiligen Kurslehrer in dem Großgruppenraum anwesend.

Bei den beiden Teilerhebungen ging der Autor jedes Mal nach demselben Muster vor. Zunächst erfolgte eine Selbstvorstellung des Autors und eine kurze Erläuterung über den Zweck der Datenerhebung, den Ablauf der Datenerhebung mit dem Wechselspiel zwischen Video und Fragebogen beim SJT (s. o.) sowie die Gewährleistung des Datenschutzes. Daran anschließend teilte der Autor die zu bearbeitenden Tests an die Berufsschüler aus. An dieser Stelle bat er die Schüler, zunächst ihren persönlichen Code – eine Kombination aus Buchstaben und Zahlen – einzutragen (s. Anhang B). Der Code sollte die Erstellung eines individuellen Leistungsprofils ermöglichen. Um die Erstellung des Codes zu demonstrieren, vollzog der Autor dies am Beispiel seiner eigenen Person.

Nach der Erstellung eines persönlichen Codes konnten die Schüler fortfahren mit der Bearbeitung des ISK-K. Für das Lesen der Instruktion sowie die Bearbeitung der 33 Items wurde den Schülern ein Zeitraum von 12 Minuten eingeräumt. Hatten alle Schüler den ISK-K bearbeitet, bat der Autor sie, in einem nächsten Schritt die Instruktion für den SJT zu lesen. Im Anschluss an das Lesen der Instruktion erinnerte der Autor die Schüler nochmals daran, dass kein Video zweimal gezeigt werden könne und dass sie ihre Antworten möglichst schnell und ohne lange nachzudenken in den

Fragebogen eintragen sollten. Zudem wies der Autor die Schüler an, nicht zwischen den Seiten des Fragebogen hin und her zu blättern, sondern sich voll und ganz auf die jeweils zu bearbeitende Situation zu konzentrieren. Nach diesen Erläuterungen startete die videogestützte Präsentation der ersten Situation über die Leinwand im Großgruppenraum. An die Videopräsentation schloss sich jeweils die Bearbeitung der zur relevanten Situation gehörigen Aufgabenstellung im Fragebogen an (s. o.).

Nach der Bearbeitung der letzten Situation ersuchte der Autor die Schüler, nochmals alle ihre Angaben sowohl beim ISK-K als auch beim SJT in Bezug auf ihre Vollständigkeit und Eindeutigkeit zu überprüfen. Im Anschluss an die Erledigung dieser Aufgabe wurden die Schüler dazu angeleitet, die abschließenden sechs demographischen Fragen zu beantworten. War auch dies erledigt, bedankte sich der Autor für die Teilnahme und verabschiedete die Schüler.

4.2.2 Beschreibung der Stichprobe

Von den 97 Berufsschülern, die teilgenommen haben, waren 23 männlich (23.7 %) und 74 weiblich (76.3 %). 44 Berufsschüler (45.4 %) befanden sich im ersten Ausbildungsjahr, während 53 (54.6 %) das zweite Ausbildungsjahr absolvierten. Die jüngsten Teilnehmer an der Studie waren 17 Jahre alt und der älteste Teilnehmer 34 Jahre. Das durchschnittliche Alter der Stichprobe belief sich auf 20.88 Jahre. Mit 88 Nennungen (85.36 %) gab die Mehrheit der Berufsschüler Deutsch als Muttersprache an. Die verbliebenen neun Nennungen verteilten sich auf Türkisch (4), Polnisch (3) und Russisch (2). In der Tabelle 4-2 sind die von den Berufsschülern vor Antritt der Berufsausbildung erreichten Schulabschlüsse aufgelistet.

Tabelle 4-2: Erreichte Schulabschlüsse der Berufsschüler (N = 97)

Erreichter Schulabschluss	Anzahl	In Prozent (%)
Abitur	31	32.00
Fachhochschulreife	27	27.80
Mittlere Reife	29	29.98
Hauptschulabschluss	7	7.23
Qualifizierter Hauptabschluss	1	1.07
Sonstiges	2	2.12

In Hinblick auf den Notendurchschnitt des Schulabgangszeugnisses ergab sich für die Stichprobe ein Mittelwert von 2.65. Die Werte streuten in einem Bereich von 1.20 bis 3.90.

4.3 Deskriptive Statistik

Für die 33 Items umfassende Kurzversion des ISK wurden Mittelwerte und Standardabweichungen berechnet. In der Tabelle 4-3 werden diese Werte separat für die vier Hauptskalen des ISK-K angeführt. Zusätzlich finden sich in derselben Tabelle Angaben zu Minimal- beziehungsweise Maximalausprägungen auf den jeweiligen Skalen.

Tabelle 4-3 Deskriptive Statistiken der vier Hauptskalen des ISK-K (N = 97)

Skala	MIN	MAX	MW	SD
Soz. Orientierung	2.20	3.80	2.96	32
Offensivität	1.63	4.00	2.77	41
Selbststeuerung	1.38	3.88	2.72	.49
Reflexibilität	1.71	3.87	2.85	.41

Anmerkung: N = Anzahl der Teilnehmer; MW = arithmetisches Mittel; SD = Standardabweichung; MIN = minimal erreichter Wert; MAX = maximal erreichter Wert

Wie der Tabelle 4-3 zu entnehmen ist, liegen alle vier berechneten Mittelwerte (MW) knapp unter einem Wert von 3. Für die Skala Soziale

Orientierung wird der höchste Mittelwert ($MW = 2.96$) errechnet und für die Skala Selbststeuerung der niedrigste ($MW = 2.72$). Die Streuungen (SD) der Werteverteilungen bewegen sich zwischen $SD = .32$ (Soziale Orientierung) und $SD = .49$ (Selbststeuerung). Für die vier Hauptskalen des ISK-K wurden folgende interne Konsistenzen (Cronbach's Alpha) festgestellt: Soziale Orientierung ($\alpha = .58$), Reflexibilität ($\alpha = .59$), Offensivität ($\alpha = .64$) und Selbststeuerung ($\alpha = .73$). Eine genauere Erörterung, wie die interne Konsistenz einer Skala zu deuten ist, erfolgt später in diesem Kapitel.

4.4 Inferenzstatistische Analysen und Hypothesentestung

Aufgrund der Bedeutung der einzelnen eingesetzten Methoden für nachfolgende Studien wird im Folgenden einmalig der Nennung von Annahmen, Voraussetzungen und Anwendungsmöglichkeiten dieser Verfahren der notwendige Raum eingeräumt. Statistische Analysen werden im Rahmen dieser Arbeit mit dem Programmpaket *SPSS 18*© durchgeführt. Eine Ausnahme bildet die in Abschnitt 4.4.1 beschriebene konfirmatorische Faktorenanalyse. Für diesen Vorgang kommt das Programm *AMOS 16*© zur Anwendung.

4.4.1 Konfirmatorische Faktorenanalyse

Die konfirmatorische Faktorenanalyse (CFA) wird als hypothesenprüfendes Verfahren angesehen, mit dessen Hilfe festgestellt werden kann, ob eine hinreichende Übereinstimmung (Modellfit) zwischen empirischen Daten und einem a priori aufgestellten theoretischen Modell besteht (Moosbrugger & Schermelleh-Engel, 2007). Dies geschieht über die statistische Abschätzung der Abweichung einer beobachteten Kovarianzmatrix von einer implizierten Kovarianzmatrix (Bühner, 2006). Es sei an dieser Stelle vorweggenommen, dass sich ein mit Hilfe des *Pick best – pick worst* Bewertungsschlüssels erstellter Datensatz aufgrund zu geringer (Ko-)Varianzen nicht für eine CFA eignet (s. Abschnitt 4.5). Die weiteren Ausführungen im Kontext der CFA beziehen sich demzufolge alleinig auf den *Motowidlo-Schlüssel*.

Zur Überprüfung der Hypothese I-1 sollen zwei aufeinander aufbauende CFAs durchgeführt werden. Die erste widmet sich der Frage, ob die in Kapitel 3 vorgenommene Zuordnung der SJT-Items zu den fünf Faktoren 1. Ordnung empirisch legitimiert werden kann (s. Tabelle 3-2). Verläuft diese erste CFA erfolgreich im Sinn einer Bestätigung der Zuordnung, soll daran anschließend aufgedeckt werden, ob sich die angenommenen korrelativen Beziehungen zwischen den fünf Faktoren 1. Ordnung auf einen gemeinsamen Faktor 2. Ordnung – Kundenorientierung – zurückführen lassen.

Zunächst soll das sogenannte *Mess- und Strukturmodell* für beide geplanten CFAs aufgestellt werden. Das Messmodell legt fest, welche beobachteten Variablen (Items des SJT) Indikatoren einer latenten Variable

(Anforderungsdimensionen im Hotelfach) darstellen. Hingegen zielt das Strukturmodell auf die Definition der Beziehungen zwischen latenten Variablen ab. Im Falle der vorliegenden Arbeit wird in diesem Zusammenhang erstens angenommen, dass sich zwischen den fünf Faktoren 1. Ordnung, welche die Anforderungsdimensionen repräsentieren, korrelative Beziehungen ergeben. Zweitens wird vermutet, dass sich die Zusammenhänge zwischen den fünf Faktoren 1. Ordnung und dem Faktor 2. Ordnung – Kundenorientierung – als partielle Regressionsgewichte deuten lassen²⁶. Vor der Durchführung einer CFA ist weiterhin eine Identifikation des Modells vorzunehmen, welche gewährleistet, dass die Anzahl empirisch gegebener (Ko-)Varianzen die Anzahl frei zu schätzender Parameter übersteigt. Im Falle dieser Arbeit wird beispielsweise bei jedem latenten Faktor 1. Ordnung die Ladung einer Indikatorvariablen gleich 1 festgesetzt. Das gesamte Mess- und Strukturmodell samt der vorab vorgenommenen Identifikation von Parametern wird in Form der Abbildung 4-6 dargestellt. Aus Gründen der Übersichtlichkeit sind korrelative Beziehungen zwischen den fünf Faktoren 1. Ordnung (s. o.) nicht in die Abbildung 4-6 aufgenommen worden.

²⁶ In diesem Szenario würden die Faktoren 1. Ordnung selbst zu manifesten Indikatorvariablen mutieren.

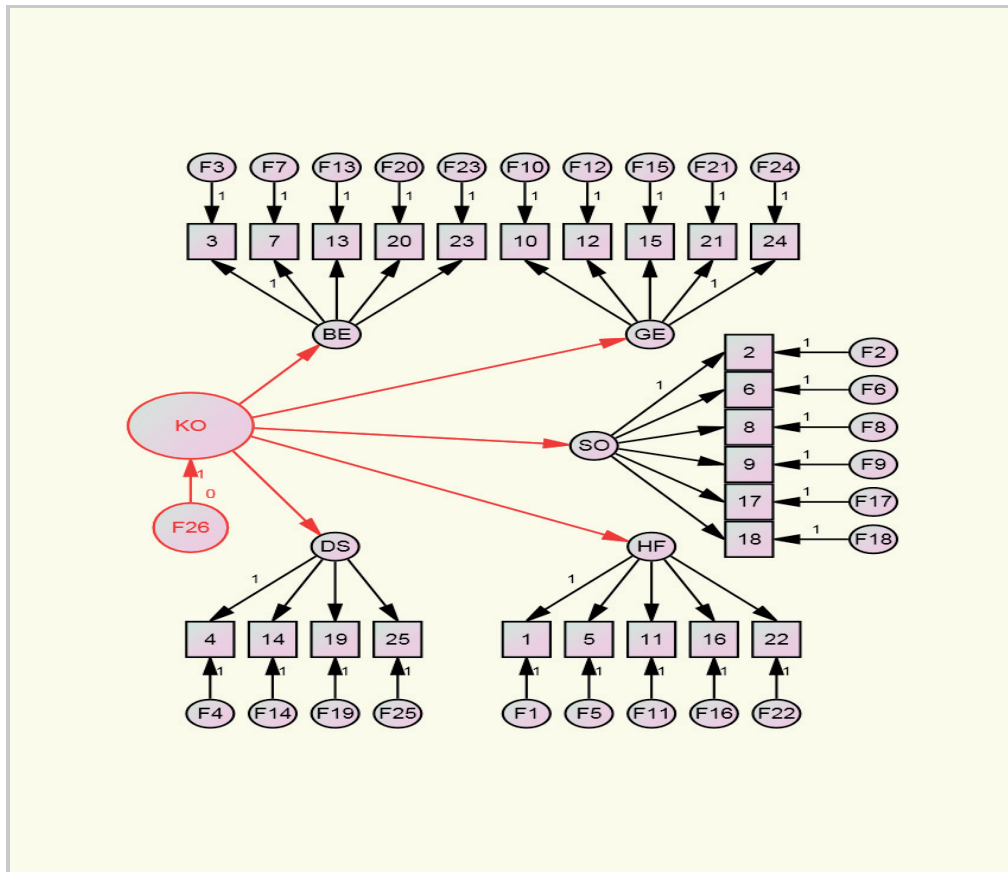


Abbildung 4-6: Mess- und Strukturmodell der konfirmatorischen Faktorenanalysen

Anmerkung: Quadratische Variablen repräsentieren die 25 Items des SJT (1-25), ovale Variablen deuten die Faktoren 1. bzw. 2. Ordnung und relevante Fehleranteile (F1-F26) an, Einfachpfeile (\rightarrow) symbolisieren Pfade. KO = Kundenorientierung (Faktor 2. Ordnung)/ BE = Beschwerdemanagement/ GE = Gewissenhaftigkeit/ SO = Serviceorientierung/ HF = Handlungsfähigkeit/ DS = Durchsetzungsstärke (Faktoren 1. Ordnung). Rot hervorgehoben ist das Strukturmodell.

Zur *Abschätzung des Modellfits* kommt die Maximum-Likelihood-Methode (ML) zum Einsatz. Die ML erfordert, dass die Indikatorvariablen intervallskaliert und multivariat normalverteilt sind (Moosbrugger & Schermelleh-Engel, 2007). Da der Einsatz des Kolmogorov-Smirnov-Tests nahelegt, dass für keines der 25 Items des SJT eine Normalverteilung der Werte vorliegt (s. Anhang B), wird die Bollen-Stine-Bootstrap-Methode zur Korrektur eingesetzt. Das Intervallskalen-Niveau kann hingegen als gegeben betrachtet werden (vgl. Abbildung 4-5).

Bei Durchführung der ML zur Abschätzung, inwiefern die vorgenommene Zuweisung der SJT-Items zu den Faktoren 1. Ordnung empirisch

gerechtfertigt ist, ergibt sich ein Chi-Quadrat-Wert von 308.59 mit einer Irrtumswahrscheinlichkeit von $p = .03$. Diese Prüfgröße für die Abschätzung des Modellfits ist derart zu interpretieren, dass nicht das in der Hypothese I-1 formulierte Modell, sondern ein alternatives, bisher nicht expliziertes Modell eine adäquate Passung zu den Daten aufweist²⁷. Da kein ausreichender Modellfit für die Zuweisung der SJT-Items zu den Faktoren 1. Ordnung erbracht werden konnte, erübrigt sich eine zweite CFA zum Nachweis eines Faktors 2. Ordnung. Das vollständige Pfaddiagramm mit den erzielten Schätzergebnissen der ersten CFA ist im Anhang B einsehbar.

Eine endgültige Ablehnung der Hypothese I-1 soll zu diesem Zeitpunkt noch nicht vorgenommen werden, sondern es werden zunächst die nachfolgend beschriebenen Resultate einer explorativen Faktorenanalyse (EFA) abgewartet.

4.4.2 Exploration der faktoriellen Struktur des SJT

Bevor die Exploration der faktoriellen Struktur des SJTs vonstatten gehen kann, soll die Eignung der codierten Daten, welche die Ausgangsvariablen (Items) repräsentieren, für anschließende faktorenanalytische Berechnungen überprüft werden. Als Maßstab zur Überprüfung dieser Eignung werden nach Backhaus, Erichson, Plinke und Weiber (2008) mehrere Prüfgrößen angesehen, zum Beispiel der Bartlett-Test oder die Untersuchung der Anti-Image-Korrelations-Matrix. Für die vorliegende Studie wurde die Prüfgröße *measure of sampling adequacy* (MSA) berechnet, was nach Backhaus et al. (2008) „vor der Durchführung einer Faktorenanalyse auf jeden Fall zu empfehlen ist“ (S. 276). Der MSA-Koeffizient²⁸ stützt sich auf die Beschaffenheit der Anti-Image-Korrelationsmatrix und zeigt in diesem Fall an, in welchem Ausmaß Zusammenhänge zwischen den Ausgangsvariablen bestehen. Damit dient der MSA-Koeffizient als Indikator dafür, inwieweit eine Faktorenanalyse sinnvoll durchgeführt werden kann (Bühner, 2006). Der MSA-Koeffizient bewegt sich in den Grenzen von Null und Eins. Dabei gilt ein Wert von $< .50$ als ungenü-

²⁷ Eine Überprüfung der Fit-Indizes ist aufgrund der geringen Stichprobengröße nicht notwendig (Bühner, 2006; S. 259).

²⁸ In der relevanten Literatur wird das Kaiser-Meyer-Olkin-Kriterium zuweilen als Synonym für den MSA-Koeffizienten verwandt.

gend, während ein hoher Wert ($> .80$) eine gute Eignung der Testwerte für die Faktorenanalyse impliziert (Bühner, 2006). Der MSA-Koeffizient kann einerseits für die gesamte Anti-Image-Korrelationsmatrix und andererseits für einzelne Variablen oder Items bestimmt werden.

Für den mit dem Motowidlo-Schlüssel codierten Datensatz ergibt sich ein MSA-Koeffizient für die gesamte Anti-Image-Korrelationsmatrix in Höhe von $.55$. Dieser Wert ist als „*schlecht*“ (Bühner, 2006) respektive „*kläglich*“ (Backhaus et al., 2008) einzustufen. Aus der Anti-Image-Korrelationsmatrix sind in einem zweiten Schritt die MSA-Koeffizienten für einzelne Items zu entnehmen, welche Partialkorrelationen zwischen dem jeweiligen Item und den anderen einbezogenen Items entsprechen (Bühner, 2006). Damit erteilen diese MSA-Koeffizienten Auskunft über die Verwendbarkeit einzelner Items in der Faktorenanalyse. Die MSA-Koeffizienten für die 25 Items bewegen sich in einem Bereich von $.40$ (Situationen 1, 2 und 4) bis $.73$ (Situation 17). Die MSA-Koeffizienten von neun Items sind nach den Entscheidungsregeln von Bühner (2006) und Backhaus et al. (2008) als untragbar, von neun als kläglich, von sechs als mittelmäßig und von einem Item als ziemlich gut einzustufen (s. Anhang B).

Für den nach dem *Pick best – pick worst Ansatz* codierten Datensatz konnten aufgrund unzureichender Varianzen keine MSA-Koeffizienten und keine Anti-Image-Korrelations-Matrix berechnet werden (s. Abschnitt 4.5). Die folgenden Rechenschritte beziehen sich demgemäß abermals ausschließlich auf die über den Motowidlo-Schlüssel bestimmten Testwerte der Teilnehmer.

Die EFA stellt ein Verfahren zur Überprüfung der Dimensionalität komplexer Merkmale dar (Bortz, 2005). In der vorliegenden Arbeit wird zur *Extraktion der Faktoren* die Hauptkomponentenanalyse²⁹ eingesetzt. Die Hauptprämisse dieses Analyseverfahrens betont, dass die Varianz eines Items oder einer Variablen gänzlich durch die zu extrahierenden Komponenten (Faktoren) erklärt werden kann. Vorgefundene Zusammenhänge zwischen Variablen oder Items sollen auf eine möglichst geringe Anzahl von Faktoren zurückgeführt werden (Backhaus et al., 2008). Ein

²⁹ Die Hauptkomponentenanalyse wird in vielen Lehrbüchern als eigenständiges Analyseverfahren *neben* der Faktorenanalyse angeführt.

generelles Problem der Faktorenanalyse liegt darin begründet, dass es kein allgemeingültiges *Abbruchkriterium* während der Faktorenextraktion gibt: Theoretisch könnten genauso viele Faktoren ermittelt werden, wie der SJT Items umfasst, also 25. Da dies dem Postulat der Datenreduktion diametral gegenübersteht, können verschiedene Kriterien eingesetzt werden, um die Extraktion auf ein sinnvolles Maß bedeutsamer Faktoren zu beschränken. Als einer inhaltlich sinnvoll interpretierbaren Faktorenlösung besonders zuträglich erscheint eine Kombination mehrerer Kriterien. Die in dieser Arbeit eingesetzten Kriterien werden nachfolgend kurz beschrieben.

Nach dem *Kaiser-Guttman-Kriterium* (z. B. Kaiser & Dickmann, 1959) werden nur solche Faktoren als bedeutsam erachtet, welche einen Eigenwert größer Eins aufweisen. Falls ein Faktor dieses Kriterium erfüllt, klärt er selbständig mehr Varianz auf als ein einzelnes Item. Allerdings kann das Befolgen dieses Kriteriums eine Überschätzung der Anzahl bedeutsamer Faktoren nach sich ziehen (Gorsuch, 1983; Zwick & Velicer, 1986). Der *Scree-Test* (vgl. Cattell, 1966) liefert ein Eigenwertdiagramm (Scree-Plot), in welchem die Eigenwerte der Faktoren entsprechend ihrer Größe angeordnet werden. In den meisten Fällen weisen solche Eigenwertverläufe einen „Knick“ auf. Ein „Knick“ zeigt sich an der Stelle, an der die Differenz zwischen den Eigenwerten zweier Faktoren maximal wird (Backhaus et al., 2008). Nach Tabachnik und Fidell (2004) und Bortz (2005) sollte nur solchen Faktoren weitere Aufmerksamkeit zuteil werden, deren Eigenwerte sich *vor* dem Knick wieder finden. Allerdings zeigen sich oftmals mehrere Knicke im Eigenwerteverlauf (Kline, 1997), was die Interpretation des Scree-Plots erschwert. Ein weiteres Kriterium zur Begrenzung der Anzahl extrahierter Faktoren wurde von Horn (1965) vorgeschlagen: In dem als *Parallelanalyse* bezeichneten Verfahren werden die Verläufe von empirisch ermittelten und zufällig erzeugten Eigenwerten kontrastiert. Zu diesem Zweck werden auf Zufall basierende Korrelationsmatrizen erstellt, welche dem empirischen Datensatz in Bezug auf die determinierenden Kenngrößen, zum Beispiel der Variablenzahl oder dem Stichprobenumfang, entsprechen (Bortz, 2005). Den Selektionsprozess überstehen in diesem Fall alleinig solche empirisch realisierten Eigenwerte, die *über* zufällig generierten Eigenwerten liegen, und somit als *nicht zufällig* charakterisiert werden können. Gemäß Moosbrugger und

Schermelleh-Engel (2007) trägt dieses Prinzip ausdrücklich dem Umstand Rechnung, dass einige Eigenwerte aus Stichprobendaten eine inhaltliche Relevanz nur vorspiegeln, indem sie auch dann > 1 ausfallen können, wenn die Variablen in der Population in Wahrheit unkorreliert sind, dagegen aber in der Stichprobe Zufallskorrelationen ungleich Null aufweisen.

Im Rahmen des *Minimum Average Partial-Test* (Velicer, 1976; MAP-Test) wird genau die Anzahl an Faktoren extrahiert, mit der sich Varianzanteile zwischen Items vollständig aufklären lassen (Bühner, 2006). Nach O'Connor (2000) wird der MAP-Test in zwei Stufen vollzogen: Zunächst erfolgt eine Hauptkomponentenanalyse und im zweiten Schritt eine Analyse von Partialkorrelationen. Bei der Analyse der Partialkorrelationen werden die im ersten Schritt ermittelten Hauptkomponenten der Reihe nach aus den Korrelationen zwischen den Items – im Fall der vorliegenden Arbeit 25 – auspartialisiert. Nun wird die Anzahl an Faktoren extrahiert, bei der die mittlere quadrierte Partialkorrelation zwischen den Items nach Auspartialisierung der relevanten Hauptkomponenten am geringsten ausfällt (Bühner, 2006). O'Connor (2000) fasst das Vorgehen wie folgt zusammen: „*Statistically, components are retained as long as the variance in the correlation matrix represents systematic variance. Components are no longer retained when there is proportionately more unsystematic variance than systematic variance*” (S. 2). Im Nachfolgenden werden die Anwendung der vier beschriebenen Extraktionskriterien und deren Implikationen dargelegt. Dazu werden in der Abbildung 4-7 zunächst der Scree-Plot und die Anzahl der Faktoren, welche das Kaiser-Guttman Kriterium erfüllen, präsentiert.

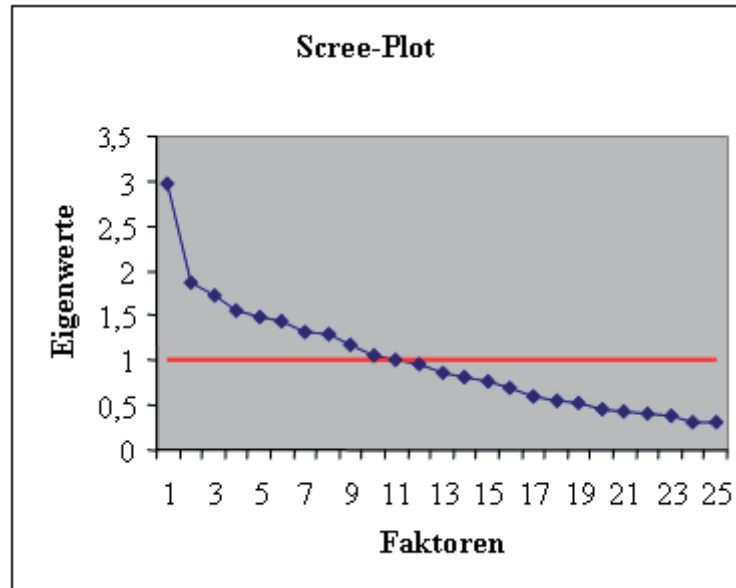


Abbildung 4-7: Scree-Plot

Anmerkung: Die rote Linie in der Abbildung repräsentiert das Kaiser-Guttman-Kriterium

Würde das Kaiser-Guttman-Kriterium als alleiniges Extraktionskriterium herangezogen, so müssten gemäß dem Scree-Plot (s. Abbildung 4-7) elf Faktoren extrahiert werden, da all jene einen Eigenwert (λ) > 1 aufweisen: $\lambda_1 = 2.98$; $\lambda_2 = 1.88$; $\lambda_3 = 1.72$; $\lambda_4 = 1.56$; $\lambda_5 = 1.49$; $\lambda_6 = 1.44$; $\lambda_7 = 1.33$; $\lambda_8 = 1.30$; $\lambda_9 = 1.18$; $\lambda_{10} = 1.06$; $\lambda_{11} = 1.01$. Hingegen legt der Scree-Test die Befolgung einer einfaktoriiellen Lösung nahe, da sich der erste Knick in dem im Scree-Plot abgebildeten Eigenwerteverlauf *nach* dem ersten Faktor zeigt.

In der Abbildung 4-8 werden der empirische (blaue Linie) und ein im Rahmen der Parallelanalyse zufällig generierter Eigenwerteverlauf (rote Linie) kontrastiert.

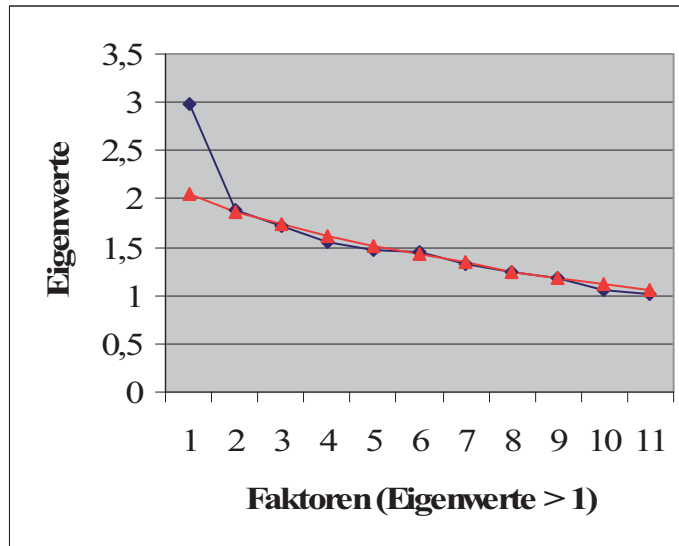


Abbildung 4-8: Grafische Gegenüberstellung eines empirisch mit einem zufällig generierten Eigenwerteverlauf

Die Tabelle 4-4 enthält eine numerische Kontrastierung der in der Abbildung 4-8 dargestellten empirisch und zufällig ermittelten Eigenwerte (λ).

Tabelle 4-4: Ergebnisse der Parallelanalyse

Zu extrahierende Faktoren	Empirische Eigenwerte		Zufällige Eigenwerte
1	$\lambda_1 = 2.98$	>	$\lambda_1 = 1.78$
	$\lambda_2 = 1.88$	>*	$\lambda_2 = 1.86$
	$\lambda_3 = 1.72$	<	$\lambda_3 = 1.74$
	$\lambda_4 = 1.56$	<	$\lambda_4 = 1.62$
	$\lambda_5 = 1.49$	<	$\lambda_5 = 1.51$
	$\lambda_6 = 1.44$	=	$\lambda_6 = 1.44$
	$\lambda_7 = 1.33$	<	$\lambda_7 = 1.34$
	$\lambda_8 = 1.24$	<	$\lambda_8 = 1.25$
	$\lambda_9 = 1.18$	=	$\lambda_9 = 1.18$
	$\lambda_{10} = 1.06$	<	$\lambda_{10} = 1.11$
	$\lambda_{11} = 1.01$	<	$\lambda_{11} = 1.05$

Anmerkung: * = nicht signifikant ($p > .05$)

Der Tabelle 4-4 kann entnommen werden, dass nur der erste empirisch beobachtete Eigenwert signifikant größer ist als sein zufällig erzeugtes

Äquivalent. Demnach ist als Ergebnis der Parallelanalyse festzuhalten, dass nur ein Faktor ausreichende statistische Bedeutsamkeit genießt, um extrahiert zu werden.

Als viertes Kriterium zur Bestimmung der Anzahl zu extrahierender Faktoren wird der MAP-Test angewandt. Die Ergebnisse sind in der Tabelle 4-5 enthalten.

Tabelle 4-5: Ergebnisse des MAP-Tests

Faktor	Mittlere Partialkorrelation (r^2)
1	.013
2	.014
3	.016
4	.018
5	.020
6	.022
7	.025
8	.027
9	.031
10	.035
11	.041

Aus der Tabelle 4-5 wird ersichtlich, dass die für den ersten Faktor ermittelte mittlere Partialkorrelation den niedrigsten Wert ($r^2 = .013$) aufweist. Demzufolge legt in Übereinstimmung mit dem Ausgang des Scree-Tests und der Parallelanalyse auch das Ergebnis des MAP-Tests die Extraktion eines einzigen Faktors nahe.

Werden die Ergebnisse, welche die vier Extraktionsmethoden hervor gebracht haben, zusammengefasst, sollten im Rückgriff auf das Kaiser-Guttman-Kriterium elf Faktoren und bei Berücksichtigung des Scree-Tests, der Parallelanalyse und des MAP-Tests ein Faktor extrahiert werden. Aufgrund dieser Befunde und vor dem Hintergrund, dass der Parallelanalyse und dem MAP-Test ein höherer Grad an Objektivität zugesprochen wird als dem Kaiser-Guttman-Kriterium (z. B. Bühner, 2006), soll in dieser Studie *ein* Faktor extrahiert werden. Die Hypothese I-1, dass der SJT auf einer fünffaktoriellen Struktur basiert, deren Faktoren mit Servi-

ceorientierung, Beschwerdemanagement, Durchsetzungsfähigkeit, Handlungsfähigkeit und Gewissenhaftigkeit benannt werden können, kann daher auch nicht auf Basis einer EFA aufrechterhalten werden (s. Abschnitt 4.5). Daher muss die Hypothese I-1 an dieser Stelle eine endgültige Ablehnung erfahren. Die im Weiteren dargestellten inferenzstatistischen Analysen konzentrieren sich auf eine einfaktorielle Struktur.

Mit dem über eine Hauptkomponentenanalyse extrahierten Faktor kann eine Varianz von 11.92 % erklärt werden. Die elf Items, welche auf diesem Faktor substantielle Ladungen $> .30$ (vgl. Kline, 1997; s. Abschnitt 4.5) erreichen, sind in der Tabelle 4-6 einzusehen³⁰. Zur Erinnerung, welche konkreten Situationen sich hinter den Itemnummern verbergen, sei abermals auf den Anhang A verwiesen.

Tabelle 4-6: Faktorladungen auf dem extrahierten Faktor

Item	Faktorladung
Situation 5	.51
Situation 8	.39
Situation 10	.60
Situation 13	.32
Situation 16	.49
Situation 17	.57
Situation 18	.65
Situation 19	.65
Situation 20	.35
Situation 22	.49
Situation 24	.32

Nach der Tabelle 4-6 werden für die Situationen 18 und 19 die höchsten Faktorladungen verzeichnet. Im Gegensatz dazu müssen den Situationen 13 und 24 die niedrigsten Faktorladungen zugeschrieben werden. Die mittlere Faktorladung beläuft sich auf .49.

Im weiteren Verlauf dieses Kapitels soll die Skala, welche aus den elf in der Tabelle 4-6 dargestellten Items besteht, mit *SJT_kurz* benannt und

³⁰ Items mit negativen Ladungen finden keine Berücksichtigung.

weiteren Skalenanalysen unterzogen werden. Dabei ergibt sich der Skalenwert des SJT_kurz, welcher für weitere Rechenschritte Bedeutung besitzt, als Mittelwert der Summe der elf einbezogenen Itemscores und liegt in einem Wertebereich von -2 bis +2.

4.4.3 Zentrale Statistiken des SJT_kurz

Die Tabelle 4-7 bietet eine Zusammenfassung der Mittelwerte, Standardabweichungen und Schiefen der Skala SJT_kurz sowie der elf konstituierenden Items.

Tabelle 4-7: Zentrale Statistiken der Skala SJT_kurz und deren Items

	MW	SD	Schiefe
SJT_kurz (Gesamtwert)	1.09	.31	-1.01
Situation 5	-.07	.98	-.06
Situation 8	1.3	.94	-1.41
Situation 10	1.2	.78	-1.48
Situation 13	.76	1.06	-.59
Situation 16	1.58	.56	-.87
Situation 17	1.12	.70	-.93
Situation 18	1.30	.72	-.69
Situation 19	1.26	.67	-.1.21
Situation 20	.82	.76	-1.12
Situation 22	1.51	.63	-1.16
Situation 24	.98	.14	-6.85

Anmerkung: MW = arithmetisches Mittel; SD = Standardabweichung

Aus der Tabelle 4-7 geht hervor, dass sich die Mittelwerte in einem Wertebereich von -.07 (Situation 5) bis 1.58 (Situation 16) bewegen. Für die Gesamtskala resultiert ein durchschnittlicher Testwert von 1.09 Punkten. Die in der vierten Spalte der Tabelle 4-7 aufgelisteten Schiefen der elf Items fallen allesamt negativ aus. Nach Bortz (2005) liegen rechtssteile Werteverteilungen vor, falls die Schiefe < 0 ausfällt.

4.4.4 Itemanalyse: Trennschärfe und Schwierigkeit

Auf statistischer Ebene verbirgt sich hinter der Trennschärfe eine Korrelation zwischen einem einzelnen Itemwert und dem Gesamtwert der da-

zugehörigen Skala (Lienert & Raatz, 1998). Ein inhaltlicher Schluss, der aus der Trennschärfe eines Items gezogen werden kann, ist, wie stark ein einzelnes Item eine Skala repräsentiert (Bühner, 2006). Zudem gestattet die Höhe der Trennschärfe Aussagen darüber, inwieweit mit Hilfe eines Items zwischen leistungsstarken Probanden auf der einen Seite und leistungsschwachen Probanden auf der anderen Seite differenziert werden kann (Lienert & Raatz, 1998). Als Mindestwert für die Trennschärfe eines Items wird in dieser Arbeit ein Wert von $r_{it} = .30$ (vgl. Backhaus et al., 2008) festgesetzt und für die elf Items der Skala SJT_kurz berechnet. In der Tabelle 4-8 können die Trennschärfekoeffizienten der elf Items eingesehen werden. Dabei liegen die Trennschärfen part-whole korrigiert vor, um zu vermeiden, dass ein Item zweimal – sowohl als einzelner Itemwert als auch als Teil des Gesamtwertes – in die Berechnungen einfließt (Fis-seni, 2004).

Tabelle 4-8: Trennschärfen der Items aus der Skala SJT_kurz

Skala SJT_kurz (11 Items)	Trennschärfe (r_{it})
Situation 5	.40
Situation 8	.27
Situation 10	.47
Situation 13	.28
Situation 16	.25
Situation 17	.39
Situation 18	.13
Situation 19	.42
Situation 20	.17
Situation 22	.42
Situation 24	.20

Anmerkung: Kursiv gehalten sind Trennschärfen $< .30$.

Die Berechnung der Trennschärfen erbringt für sechs Items (Situation 8, Situation 13, Situation 16, Situation 18, Situation 20, Situation 24) Werte $< .30$. Für die Situation 10 resultiert hingegen die höchste Trennschärfe ($r_{it} = .47$). Es ergibt sich eine mittlere Trennschärfe von $r_{it} = .32$. Nach Lienert und Raatz (1998; S. 105 ff.) ist die Trennschärfe eines Items nicht losgelöst von seiner Schwierigkeit zu betrachten. Demnach besitzen Items mitt-

lerer Aufgabenschwierigkeit, das heißt solche Items, die von rund 50 % der Probanden richtig und von 50 % der Probanden falsch gelöst werden, bessere Voraussetzungen für hohe Trennschärfen als solche Items, welche zu leicht oder zu schwierig sind. Aufgrund dieser gegebenen Interdependenz sollen in einem nächsten Schritt die Aufgabenschwierigkeiten berechnet werden.

Für die Berechnung der Schwierigkeit des Items wird die nachstehende Formel verwendet (vgl. Fisseni, 2004; S.34):

$$(1) P_m = \frac{\sum X}{\sum X_{\max}}, \text{wobei}$$

- $X = \text{Itemscore}$
- $P_m = \text{Schwierigkeitsindex für mehrstufige Items}$
- $X_{\max} = \text{Maximaler Itemscore; } X \text{ läuft von } 0 \text{ bis } X_{\max}$
- $\sum X_{\max} = \text{Summe von } X_{\max} \text{ über alle Probanden}$

Extrem leichte ($P > .80$) und extrem schwere Items ($P < .20$) besitzen geringe Aussagekraft, da sie nicht ausreichend zwischen den Probanden differenzieren (Fisseni, 2004) beziehungsweise nur für bestimmte Probandengruppen geeignet sind. Um dies vermeiden zu können, sollten alle Items eine Schwierigkeit in dem Bereich von $.20 < P < .80$ aufweisen. Die Tabelle 4-9 fasst die Schwierigkeitsindizes von denjenigen elf Items zusammen, welche die Skala SJT_kurz bilden.

Tabelle 4-9: Schwierigkeitsindizes der Items aus der Skala SJT_kurz

Item	Schwierigkeits- Indizes (P)
Situation 5	.05
Situation 8	.63
Situation 10	.61
Situation 13	.38
Situation 16	.79
Situation 17	.56
Situation 18	.65
Situation 19	.63
Situation 20	.41
Situation 22	.75
Situation 24	.49

Anmerkung: Kursiv gehalten sind Aufgabenschwierigkeiten $< .20$.

Bis auf eine Ausnahme (Situation 5) liegen alle Aufgabenschwierigkeiten P in dem empfohlenen Wertebereich von $.20 < P < .80$. Mit einer Aufgabenschwierigkeit von $P = .05$ ist dieses Item – Situation 5 – als extrem schwierig einzustufen. Als mittlere Aufgabenschwierigkeit lässt sich ein P von .54 angeben.

Zusätzlich zur Berechnung der Aufgabenschwierigkeit soll als weitere Urteilsgrundlage in die Skalenanalyse des SJT_kurz einfließen, von wie vielen Teilnehmern jedes der elf Items im Sinne der oben skizzierten Scoringmethode – Motowidloschlüssel – komplett richtig (+2 Punkte) beziehungsweise komplett falsch (-2 Punkte) beantwortet wurde (vgl. Abbildung 4-5). Diese Häufigkeiten liegen in Form der Tabelle 4-10 vor.

Tabelle 4-10: Komplett richtig- und Komplett falsch-Antworten je Item

Item	Komplett richtig beantwortet (Anzahl/Prozent)	Komplett falsch beantwortet (Anzahl/Prozent)
<i>Situation 5</i>	0/0	3/3.09
Situation 8	49/50.51	1/1.03
Situation 10	35/36.08	1/1.03
Situation 13	26/26.80	1/1.03
Situation 16	59/60.82	0/0
Situation 17	27/27.83	1/1.03
Situation 18	43/44.32	0/0
Situation 19	33/34.02	0/0
Situation 20	11/11.34	0/0
Situation 22	55/56.70	0/0
<i>Situation 24</i>	0/0	0/0

Anmerkung: Kursiv gehalten sind Items, welche nicht komplett richtig gelöst wurden.

Sechs Items des SJT wurden gemäß der Tabelle 4-10 von den Teilnehmern der Studie nie komplett falsch beantwortet. Zwei Items – Situationen 5 und 24 – hat währenddessen kein Teilnehmer komplett richtig gelöst (s. Abschnitt 4.5). Am häufigsten komplett richtig gelöst wurde die Situation 16 (60.82 % der Stichprobe). Die meisten komplett falschen Antworten (3.09 %) sind hingegen bei der Situation 5 zu verzeichnen.

In Anbetracht der soeben dargestellten Analysen erfüllen vier (Situationen 10, 17, 19, 22) der elf Items, welche die Skala SJT_kurz darstellen, die Mindestanforderungen in Bezug auf die Trennschärfe und die Aufgabenschwierigkeit. Für diese vier Situationen muss jedoch angefügt werden, dass sie nie (Situationen 19, 22) beziehungsweise nur von jeweils einem Teilnehmer komplett falsch (Situationen 10, 17) gelöst wurden.

Da die Reliabilität und Validität einer Skala als besonders wichtig für einen späteren Einsatz der Skala in der Praxis erachtet werden, erfährt zunächst kein Item einen Ausschluss aus der Skala. Dementsprechend besteht die Skala SJT_kurz weiterhin aus elf Items.

4.4.5 Reliabilitätsanalyse

Als Maß der Reliabilität wird der *Alpha-Koeffizient* (Cronbach, 1951) zur Schätzung der internen Konsistenz berechnet. Analysen der internen Konsistenz werden vor allem bei solchen Skalen oder Tests empfohlen, welche homogene Merkmale zu messen beabsichtigen (Bühner, 2006; s. Abschnitt 4.5). Nach Lienert und Raatz (1998) kennzeichnet die interne Konsistenz die Leistungsfähigkeit eines Tests und ermöglicht damit sowohl Aussagen über seine Qualität als auch über seine Anwendbarkeit. Krauth (2005) betrachtet Cronbach's Alpha als untere Grenze der Reliabilität. Die Reliabilitätsanalyse eignet sich, um die relevanten Skalen hinsichtlich ihrer psychometrischen Zuverlässigkeit zu erforschen. Gemäß Lienert und Raatz. (1998) eignen sich Reliabilitäten zwischen $\alpha = .50$ und $\alpha = .70$ nur für Gruppenvergleiche, nicht jedoch für eine individuelle Differenzierung. Eine Reliabilität von unter $\alpha = .60$ wird bei ihnen als sehr niedrig bezeichnet. Auch nach Hair, Black, Babin, Anderson und Tatham (2008) sollte die interne Konsistenz mindestens $\alpha = .70$ betragen. Für explorative Zwecke wurde von diesem Forscherteam mindestens ein Wert von $\alpha = .60$ eingefordert. Kanning (2004) nannte ebenfalls Werte $>.70$ als Mindestanforderung, um eine Reliabilität als akzeptabel zu bezeichnen. Bühl (2008) bewertete Reliabilitäten $>.78$ als „*recht hoch*“ (S. 461). Als Mindestanspruch an die Reliabilität(-en) soll zusammengefasst an dieser Stelle $.70$ eingeführt werden.

Mit einem Wert von $.64$ (Cronbach's Alpha) erreicht die interne Konsistenz der Skala SJT_kurz nicht den geforderten Wert von $.70$ (s. Abschnitt 4.5). Um den Beitrag eines einzelnen Items zur internen Konsistenz der Skala abschätzen zu können, soll untersucht werden, wie Cronbach's Alpha ausfiele, falls das entsprechende Item aus der Skala SJT_kurz entfernt würde. Die entsprechenden Werte sind Bestandteil der Tabelle 4-11.

Tabelle 4-11: Veränderungen der internen Konsistenz bei Entfernung eines Items

Item	Cronbach`s Alpha, falls Item entfernt
Situation 5	.60
Situation 8	.63
Situation 10	.58
Situation 13	.63
Situation 16	.63
Situation 17	.60
Situation 18	.65
Situation 19	.60
Situation 20	.60
Situation 22	.64
Situation 24	.60

In Anlehnung an die Tabelle 4-11 hätte die Entfernung keines der elf Items einen bedeutsamen Anstieg der internen Konsistenz zur Folge. Allerdings würde eine Wegnahme der Situation 10 einen Abfall der internen Konsistenz von .64 auf .58 nach sich ziehen. Dieses Item erzielte obendrein die höchste Trennschärfe unter den elf Items der Skala (vgl. Tabelle 4-8). Ähnlich wie Cronbach`s Alpha fällt die *mittlere Interitemkorrelation* als Homogenitätsindex mit einem Wert von .15 niedrig aus, da Bühner (2006) erst Interitemkorrelationen in einem Bereich von .20 und .40 als Ausdruck für die gegebene Homogenität einer Skala ansieht. Aus der Varianz der mittleren Interitemkorrelation ist nach Cortina (1993) die *Präzision von Alpha* zu errechnen. Erreicht die Präzision von Alpha einen hohen Wert, liegen heterogene Interitemkorrelationen vor und die Annahme einer eindimensionalen Skala ist nicht zu rechtfertigen. Bei einer gegebenen Varianz der Interitemkorrelationen in Höhe von .014 erzielt die Präzision von Alpha – berechnet als Quadratwurzel der Varianz – einen Wert vom .12. Dieser Wert spricht zusammen mit der niedrigen mittleren Iteminterkorrelation nach den Vorgaben von Cortina (1993) gegen eine ausreichende Homogenität der Skala.

4.4.6 Validitätsanalyse

Zunächst sollen die Ergebnisse von *bivariaten Korrelationen* zwischen der Skala SJT_kurz und den vier Hauptskalen des ISK-K – Soziale Orientierung, Selbststeuerung, Offensivität und Reflexibilität – wiedergegeben werden. Da alle relevanten Daten intervallskaliert sind, kommt jeweils die Produkt-Moment-Korrelation zur Anwendung. Aufgrund der weiter oben beschriebenen Unterscheidung von sozialer Kompetenz auf der einen Seite und sozial kompetentem Verhalten auf der anderen Seite wird keine Hypothese bezüglich des Zusammenhangs zwischen SJT_kurz und ISK-K verfolgt (s. Abschnitt 4.5).

Die Produkt-Moment-Korrelationen zwischen der Skala SJT_kurz und den vier Hauptskalen des ISK-K sind in dieser Studie als Maße der Konstruktvalidität zu verstehen. Korrelative Zusammenhänge können durch einzelne *Extremwerte* (Ausreißer) verfälscht werden (Bortz, 2005). Anhand von Stängel-Blatt-Diagrammen und Boxplots (s. Anhang B) lassen sich auf Seiten der Skala SJT_kurz zwei Teilnehmer mit Extremwerten identifizieren. Diese entsprechen Skalenwerten im SJT_kurz, welche $< .3$ ausfallen. Auf den Skalen Soziale Orientierung und Reflexibilität werden jeweils sieben Extremwerte und auf der Skala Offensivität vier Extremwerte festgestellt. In der Tabelle 4-12 wird eine Zusammenfassung der korrelativen Berechnungen gegeben.

Tabelle 4-12: Korrelationen zwischen der Skala SJT_kurz und den Hauptskalen des ISK-K

	Soz. Orientierung (mit/ohne Ausreißer*)	Offensivität (mit/ohne Ausreißer**)	Selbststeuerung	Reflexibilität (mit/ohne Ausreißer***)
SJT_kurz	.02/.00	.06/.21****	.09	.01/.08

Anmerkung: Es wurde jeweils 2-seitig getestet. * n = 88 ** n = 94 *** n = 90 **** p < .05

Bei einer Bereinigung der Datensätze um die oben erwähnten Extremwerte ergibt sich für die Skala SJT_kurz eine signifikante Korrelation in Höhe von $r = .21$ ($p < .05$) mit der Skala Offensivität aus dem ISK-K. Al-

le anderen Korrelationen verfehlen jedoch ungeachtet der Tatsache, ob Extremwerte eliminiert wurden oder nicht, deutlich das notwendige Maß an statistischer Bedeutsamkeit.

Neben Extremwerten beeinflussen die Reliabilitäten der einbezogenen Tests die Höhe einer Korrelation (Lienert & Raatz, 1998). Um abzuschätzen können, wie hoch die Korrelationen und damit die Validitäten ausfallen würden, falls das Kriterium – in dieser Studie die vier Hauptskalen des ISK-K – vollkommen reliabel – also ohne Messfehler – wäre, kann eine *einfache Minderungskorrektur* durchgeführt werden. Durch die Negation eines Messfehlers spiegelt die einfache Minderungskorrektur idealtypische Verhältnisse wider, die in der Realität nahezu ausgeschlossen sind (Lienert & Raatz, 1998). Daher sollten ihre Resultate mit Bedacht analysiert werden (Bühner, 2006).

Die einfache Minderungskorrektur wird nach folgender Formel (vgl. Lienert & Raatz, 1998; S. 257) berechnet:

$$(2) \frac{r^*}{\sqrt{r\#}}, \text{ wobei}$$

- r^* = Korrelation zwischen Test und Kriterium (Validitätskoeffizient)
- $r\#$ = Reliabilität des Kriteriums (in dieser Studie: Cronbach's Alpha)

Bei Inanspruchnahme der einfachen Minderungskorrektur ergibt sich für keine der berechneten Validitäten³¹ ein praktisch bedeutsamer Anstieg (s. Anhang B). Es bleibt festzuhalten, dass die Skala SJT_kurz eine signifikant positive Korrelation mit der Hauptskala Offensivität, welche dem ISK-K entstammt, erzielen kann. Auf Basis der vorliegenden Daten besteht jedoch kein korrelativer Zusammenhang zwischen der Skala SJT_kurz und den drei anderen Hauptskalen des ISK-K.

Abgesehen von Korrelationen zwischen der Skala SJT_kurz und den Hauptskalen des ISK-K können auch Korrelationen zwischen einzelnen

³¹ Es wurden jeweils die um Extremwerte bereinigten Korrelationen beachtet.

Items der Skala SJT_kurz und den vier Hauptskalen des ISK-K von Bedeutung sein. Bühner (2006) bezeichnet diese Art von Korrelationen als *Fremdtrennschärfe*. Der Begriff der Fremdtrennschärfe ist abzugrenzen von der in Tabelle 4-8 einsehbaren (Eigen-)Trennschärfe von Items, welche Bedeutung besitzt bei Betrachtung der Homogenität einer Skala. Da im Zuge der Berechnung von Fremdtrennschärfe eine Vielzahl von Einzelkorrelationen durchgeführt wird, muss die Irrtumswahrscheinlichkeit (α -Fehler) mit Hilfe der Bonferoni-Korrektur (vgl. Bortz, 2005; S. 260 ff.) a priori entsprechend adjustiert werden. Es konnte auf diesem Wege keine signifikante Korrelation zwischen einem Item der Skala SJT_kurz und einer der vier Hauptskalen des ISK-K festgestellt werden.

In dem folgenden Analyseschritt soll der Frage nachgegangen werden, ob die vier Hauptskalen des ISK-K einen statistisch bedeutsamen Beitrag bei der Vorhersage der Skala SJT_kurz leisten können. Um diese Frage beantworten zu können, wird eine *multiple Regression* berechnet, in welcher die Hauptskalen des ISK-K als Prädiktoren fungieren, während die Skala SJT_kurz das Kriterium darstellt.

Die Methode der Regressionsanalyse ist neben einem notwendigen Intervallskalenniveau der einbezogenen Variablen an die Erfüllung bestimmter Voraussetzungen gebunden: Linearität in den Parametern, Vollständigkeit des Modells, Homoskedastizität der Störgrößen, Unabhängigkeit der Störgrößen, fehlende lineare Abhängigkeit zwischen den unabhängigen Variablen und Normalverteilung der einbezogenen Variablen (Backhaus et al., 2008). Allerdings wird diese Methode als „*recht unempfindlich gegenüber kleineren Verletzungen der Annahmen*“ eingestuft und stellt damit ein „*äußerst flexibles und vielseitig anwendbares Analyseverfahren*“ dar (Backhaus et al., 2008; S. 93). Aufgrund dessen soll eine Überprüfung der Voraussetzungen nicht explizit vorgenommen werden.

Die vier Hauptskalen des ISK-K werden schrittweise (engl.: Stepwise) in das Regressionsmodell aufgenommen, wobei in diesem Fall die Rückwärts-Methode (vgl. Bühl, 2008) herangezogen wird. Hierbei werden anfangs alle Prädiktoren in die Regressionsgleichung eingebunden, um in den darauf folgenden Schritten Prädiktoren mit den kleinsten partiellen Korrelationskoeffizienten auszuschließen. Die Ergebnisse der schrittweisen Regression werden in der Tabelle 4-13 präsentiert.

Tabelle 4-13: Ergebnisse der schrittweisen Regression

Modell	β	R	R ²	ΔR^2
1 Offensivität Selbststeuerung Reflexibilität Soziale Orientierung	.14 .14 .11 -.07	.24	.06	-
2 Offensivität Selbststeuerung Reflexibilität	.14 .13 .11	.23	.05	-.00
3 Offensivität Selbststeuerung	.17 .09	.21	.04	-.01
4 Offensivität	.19	.19	.04	-.01

Anmerkungen: β = standardisiertes Regressionsgewicht; R = Multiple Korrelation; R² = Determinationskoeffizient; ΔR^2 = Aufklärungsverlust bei Hinwegnahme des Prädiktors

Werden alle vier Hauptskalen des ISK-K gemeinsam zur Vorhersage der Skala SJT_kurz herangezogen, erklären sie gemeinsam approximativ 6 % an Varianz. Wie es aufgrund der korrelativen Analysen zu erwarten war, erweist sich die Hauptskala Offensivität als nützlichster Prädiktor und erklärt alleine 4 % an Varianz auf Seiten der Skala SJT_kurz.

Weiterhin soll in diesem Abschnitt überprüft werden, ob ein systematischer Zusammenhang besteht zwischen der Skala SJT_kurz und dem Notendurchschnitt auf dem Schulabgangszeugnis. In diesem Fall wird abermals eine bivariate Korrelation (Produkt-Moment) herangezogen. Um einen möglichen Moderatoreffekt auszuschließen, wird der höchste erreichte Schulabschluss der Teilnehmer auspartialisiert. Zwischen der Skala SJT_kurz und dem Notendurchschnitt ergibt sich bei einer 2-seitigen Testung kein statistisch bedeutsamer Zusammenhang, sondern eine Korrelation nahe Null ($n = 94$)³². Daher kann nicht geschlussfolgert werden, dass Teilnehmer mit einem besseren Notendurchschnitt in der Skala SJT_kurz

³² Bei den Korrelationen zwischen der Skala SJT_kurz und dem Notendurchschnitt wurden wie gehabt Extremwerte vorab identifiziert und nicht in die Berechnungen einbezogen.

erfolgreicher oder erfolgloser abschneiden als Teilnehmer mit einem schlechteren Notenschnitt. Die Leistung in der Skala SJT_kurz erweist sich viel mehr als unabhängig vom Notendurchschnitt. Eine lineare Regression des Notendurchschnitts auf die Leistung im SJT_kurz erbringt im Übrigen eine Varianzaufklärung von nahezu 0 %, was die oben getätigte Aussage unterstützt.

Für eine Ausbildung im Hotelfach, zum Beispiel im Hotel Münster, ist heutzutage vielfach mindestens ein Realschulabschluss (Mittlere Reife) erforderlich. Werden bei der Korrelation zwischen der Skala SJT_kurz und dem Notendurchschnitt nur Teilnehmer, welche diese Anforderung erfüllen ($n = 84$), berücksichtigt, ergibt sich eine Korrelation mit negativer Tendenz ($r = -.11$), welche nicht den Status statistischer Bedeutsamkeit erreicht. Bei Durchführung einer linearen Regression des Notendurchschnitts auf die Skala SJT_kurz wird 1 % an Varianz durch den Notendurchschnitt vorhergesagt. Die Unabhängigkeit von SJT_kurz und Notendurchschnitt ist demnach auch für diese Substichprobe gegeben.

4.4.7 Zusammenhänge mit demographischen Variablen

Bevor die Ergebnisse dieser Studie diskutiert werden, sollen Vergleiche bezüglich der Leistung im SJT_kurz zwischen verschiedenen Substichproben stattfinden.

Aufgrund der stark unterschiedlichen Stichprobengrößen bei Teilnehmern mit Deutsch als *Muttersprache* ($n = 88$) auf der einen Seite und Teilnehmern, welche eine andere Muttersprache als Deutsch angegeben haben ($n = 9$), auf der anderen Seite, können diese beiden Substichproben nicht verglichen werden.

Männliche vs. weibliche Teilnehmer

Unter Zuhilfenahme eines *t-Tests für unabhängige Stichproben* wird überprüft, ob ein signifikanter Unterschied zwischen männlichen ($n = 24$) und weiblichen Teilnehmern ($n = 73$) hinsichtlich der Leistung im SJT_kurz besteht. Es sei vor der Darstellung der Ergebnisse vorangestellt, dass die männliche Stichprobe nicht die geforderte Mindestgröße von $n = 30$ erreicht.

Der Levene-Test zur Überprüfung der Varianzhomogenität wird beim Vergleich zwischen männlichen und weiblichen Teilnehmern in Bezug

auf die Skala SJT_kurz nicht erfüllt. Hier liegen heterogene Varianzen vor. Auf Basis des durchgeführten t-Tests für unabhängige Stichproben lässt sich für die Leistung im SJT_kurz kein bedeutsamer Unterschied zwischen den beiden relevanten Substichproben feststellen. Damit hat das Geschlecht des Teilnehmers keinen Einfluss auf das Abschneiden im SJT_kurz. Ein Punkt, in welchem die männliche und weibliche Stichprobe jedoch differieren, bezieht sich auf die ungleiche interne Konsistenz der Skala SJT_kurz. Für die männliche Stichprobe ergibt sich ein α von .79, während für die weibliche Stichprobe ein bedeutend niedrigeres α von .52 resultiert.

Teilnehmer mit unterschiedlichem schulischem Hintergrund

Mittels einer *einfaktoriellen Varianzanalyse* wird untersucht, ob sich signifikante Unterschiede zwischen Teilnehmern unterschiedlichen schulischen Hintergrunds aufzeigen lassen. Als abhängige Variable fungiert der SJT_kurz. Die unabhängige Variable, der höchste erreichte schulische Abschluss der Teilnehmer, ist dreifach abgestuft: Teilnehmer mit Abitur (n = 31), Teilnehmer mit Fachhochschulreife (n = 27) und Teilnehmer mit Mittlerer Reife (n = 29). Da sich insgesamt nur zehn Teilnehmer fanden, welche keinen der drei genannten Abschlüsse erlangt hatten, wurden diese aus der Varianzanalyse ausgeschlossen. Auf Basis des Ergebnisses der einfaktoriellen Varianzanalyse ist zu konstatieren, dass sich zwischen Teilnehmern mit unterschiedlichen Schulabschlüssen keine signifikanten Unterschiede in Bezug auf das Abschneiden im SJT_kurz aufzeigen lassen ($F = 2.96$; $p = .06$). Die Leistung im SJT_kurz ist damit als statistisch unabhängig von der vorhergehenden Schulbildung der Teilnehmer anzusehen.

Teilnehmer im ersten Ausbildungsjahr vs. Teilnehmer im zweiten Ausbildungsjahr

Wie bei dem Vergleich zwischen männlichen und weiblichen Teilnehmern wird an dieser Stelle ebenfalls ein *t-Test für unabhängige Stichproben* angewandt, um festzustellen, ob sich Teilnehmer im ersten Ausbildungsjahr (n = 44) in Bezug auf die Leistung im SJT_kurz signifikant von Teilnehmern im zweiten Ausbildungsjahr (n = 53) unterscheiden. Seitens der Überprüfung der Voraussetzungen für t-Tests bestätigt ein Levene-

Test das Vorliegen homogener Varianzen bei den zu vergleichenden Substichproben. Es ist festzuhalten, dass Teilnehmer, welche sich im zweiten Ausbildungsjahr befinden, im SJT_kurz nicht signifikant leistungsstärker sind als Teilnehmer mit einem Jahr weniger Berufserfahrung (vgl. Kapitel 7). Zwischen Teilnehmern im ersten und zweiten Ausbildungsjahr lässt sich jedoch in Bezug auf die interne Konsistenz der Skala SJT_kurz ein Unterschied notieren. Während die Reliabilität dieser Skala einen Wert von .54 erreicht, falls nur Teilnehmer mit einem Jahr Berufserfahrung berücksichtigt werden, wird für Teilnehmer aus dem zweiten Ausbildungsjahr gar ein zufrieden stellender Reliabilitätswert in Höhe von .70 erreicht.

Als Resümee dieses Abschnitts kann ausgesagt werden, dass ein Abschneiden im SJT_kurz nicht in statistisch bedeutsamer Weise beeinflusst wird vom Geschlecht, dem schulischen Hintergrund und dem Ausbildungsjahr der Teilnehmer. Zudem konnte anhand einer bivariaten Korrelation (Produkt-Moment), welche kein signifikantes Ergebnis ($r = -.07$) erbrachte, ebenfalls eine Unabhängigkeit des SJT_kurz vom Alter der Teilnehmer erkannt werden³³.

4.5 Diskussion

Für die Datenerhebung an einer Stichprobe von 97 Berufsschülern wurden vom Autor nur Berufsschüler berücksichtigt, die sich zum Testzeitpunkt im ersten beziehungsweise zweiten Ausbildungsjahr befanden. Berufsschüler, welche sich im dritten Lehrjahr befanden, wurden dagegen als zu erfahren und damit als einer Bewerberstichprobe zu unähnlich angesehen, um mit dieser Datenerhebung Erkenntnisse zu gewinnen über die Nützlichkeit des SJT in der Auswahl neuer Auszubildender für das Hotelfach. Es bleibt zu hinterfragen, ob nicht bereits Schüler, deren zweites Ausbildungsjahr sich dem Ende zuneigte³⁴, als zu erfahren angesehen werden müssen (s. Abschnitt 4.6).

Bei Gruppentestungen, welche sich im Rahmen dieser Studie ereignet haben, ist eine störungsfreie Atmosphäre nicht ohne weiteres herzustellen

³³ Bei dieser Korrelation wurden neben Extremwerten auf der Skala SJT_kurz auch Extremwerte bezüglich des Alters ausgeschlossen. Dies betraf drei Teilnehmer, welche älter als 27 Jahre waren.

³⁴ Die Datenerhebung fand im Juni statt. Ein Ausbildungsjahr beginnt turnusmäßig im September.

beziehungsweise für die gesamte Testdauer (75 Minuten) aufrechtzuerhalten. Des Weiteren war bedingt durch die beschränkte Größe der Räumlichkeiten, in welchen die Datenerhebung stattfand, nicht auszuschließen, dass Teilnehmer Lösungen – vor allem im SJT – von ihren Nachbarn übernehmen oder sich untereinander über den SJT austauschen. Um den Anteil dieser Verzerrungen zu reduzieren, mag das Inaussichtstellen eines individuellen Leistungsprofils eine Hilfe gewesen sein. Zudem sollte durch die inhaltliche Gestaltung der Verhaltensalternativen verhindert werden, dass auf Antrieb zu erkennen war, welche Verhaltensalternative die beste beziehungsweise schlechteste ist (vgl. Kapitel 3).

Im Rahmen der Datenerhebung ereignete sich das für empirische Studien ungewöhnliche Phänomen, dass hinsichtlich der Bearbeitung des SJT keine fehlenden Werte auftraten. Dies kann nach Meinung des Autors auf zwei mögliche Ursachen zurückgeführt werden: Zum Einen kann das Nicht-Auftreten fehlender Werte auf die mehrmalige Instruktion – sowohl mündlicher als auch schriftlicher Natur –, die eigenen Angaben auf Vollständigkeit zu überprüfen, zurückzuführen sein. Andererseits mag auch hier die Aussicht auf ein individuelles Leistungsprofil einen suffizienten Anreiz dargestellt haben, die dargebotenen Tests komplett zu bearbeiten.

Die Angaben der Teilnehmer im SJT wurden unter Berücksichtigung eines mit SME generierten Bewertungsschlüssels auf zweierlei Weise codiert. Zum einen wurde auf ein Codierverfahren von Motowidlo et al. (1990; vgl. Abbildung 4-5) zurückgegriffen und zum anderen wurde eine Codierung nach der Pick best – pick worst Variante vollzogen. Bei der letztgenannten Alternative erhielten die Teilnehmer als Itemscores entweder eine 0 (falsch gelöst) oder eine 1 (richtig gelöst). Auf diesem Wege resultierten unzureichende Varianzen, so dass weiterführende faktorenanalytische Untersuchungen – sowohl konfirmatorischer als auch explorativer Natur – nicht durchgeführt werden konnten. Eine weitere Anwendung der Pick best – pick worst Variante beim Codieren der Angaben im SJT erscheint aus diesem Grund trotz ihrer leichten Handhabbarkeit und Ökonomie nicht empfehlenswert. Hingegen ermöglichte ein nach den Vorgaben von Motowidlo et al. (1990) codierter Datensatz weiterführende Analysen des SJT.

Eine CFA wurde durchgeführt, obwohl die Stichprobengröße von $N = 97$ unter dem in der Literatur (z. B. Backhaus et al., 2008) genannten mi-

nimal erforderlich Stichprobenumfang von $N > 100$ lag. Das mit ihr hervorgebrachte Ergebnis, dass die in der Hypothese I-1 formulierte theoretische Rahmenstruktur nicht mit den empirischen Gegebenheiten vereinbar sei, sollte daher mit Bedacht als aussagekräftig bezeichnet werden. Allerdings wiesen auch die Ergebnisse der explorativen Faktorenanalyse, welche geringere Stichprobengrößen voraussetzt, in die entsprechende Richtung (s. u.).

Vor der faktorenanalytischen Exploration des SJT wurden die MSA-Koeffizienten für die 25 Items sowie für die Anti-Image-Korrelationsmatrix ermittelt. Die Ergebnisse implizierten eine fehlende Eignung der Daten für nachfolgende Faktorenanalysen. Nach Bühner (2006) ist es jedoch nicht ratsam, allein aufgrund der MSA-Kennwerte Schlüsse über die Qualität eines Datensatzes zu ziehen, so dass spätere Faktorenanalysen trotz dieses Ergebnisses erfolgten. Die Frage, wie viele Faktoren bei einer Analyse des SJT zu extrahieren sind, wurde mittels eines Scree-Tests, einer Parallelanalyse und des MAP-Tests beantwortet. Neben den bereits oben genannten Einwänden zum Scree-Test existieren auch für die Anwendung der Parallelanalyse Kritikpunkte. So stellte Beauducel (2001) zum Beispiel fest, dass durch eine Parallelanalyse die Zahl bedeutsamer und damit zu extrahierender Faktoren überschätzt werden kann. Zudem merkte er an, dass bei Vorliegen einer sehr dominanten Hauptkomponente die Zahl bedeutsamer Faktoren womöglich zu gering gesetzt wird. Das Vorliegen einer dominanten Hauptkomponente kann für diese Studie bestätigt werden, wenn der Scree-Plot (s. Abbildung 4-7) ein weiteres Mal betrachtet wird. Für den ersten Faktor wurde ein Eigenwert von $\lambda_1 = 2.98$ ermittelt, während der zweite Faktor nur noch einen Eigenwert von $\lambda_2 = 1.88$ gutgeschrieben bekam. Allerdings legte neben der Parallelanalyse auch der Ausgang des MAP-Tests die Extraktion eines einzigen Faktors nahe, so dass nicht davon auszugehen ist, dass die Anzahl bedeutsamer Faktoren unterschätzt wurde. Neben der Parallelanalyse und dem MAP-Test votiert auch der Ausgang einer Maximum-Likelihood-Faktorenanalyse (MLF)³⁵ (vgl. Rost, 1996) gegen die weitergehende Analyse einer fünffaktoriellen Struktur, welche in der Hypothese I-1 themati-

³⁵ Dieser Ansatz ist nicht zu verwechseln mit dem Einsatz der ML als Schätzmethode im Rahmen der CFA.

siert wurde (s. o.). Mit einem Chi-Quadrat = 146.68 ($p = .98$) verfehlte ein fünffaktorielles Modell deutlich eine Passung zu dem relevanten empirischen Datensatz. Eine Annahme der Hypothese I-1 musste aus diesen Gründen und aufbauend auf dem Ausgang der CFA abgelehnt werden.

Bortz (2005) stellte in Abhängigkeit von ermittelten Faktorladungen weitere Kriterien zur Verfügung, mit deren Hilfe die Bedeutsamkeit eines Faktors beurteilt werden kann. Auf dem extrahierten Faktor wiesen, wie der Tabelle 4-8 zu entnehmen ist, weniger als zehn Items – nämlich sechs – eine Faktorladung $> .40$ auf, so dass dieser nach Bortz (2005) nicht sinnvoll inhaltlich interpretiert werden könne. Die Frage nach der inhaltlichen Interpretierbarkeit eines Faktors, welcher sich bei der Analyse eines SJTs ergibt, ist bei dieser Testgattung ohnehin schwierig zu beantworten (vgl. Abschnitt 2.5), da mit einem SJT die Messung eines multidimensionalen Konstrukts verfolgt wird. Aufgrund der Tatsache, dass die Stichprobe aus weniger als 300 Personen bestand, ist nach Bortz (2005) weiterhin mit „zufälligen Ladungsstrukturen“ (S. 535) zu rechnen. Der Autor erachtete alleinig solche Items als bedeutsam, welche Faktorladungen $>.30$ aufweisen konnten. In der relevanten Literatur besteht Uneinigkeit darüber, ab welcher Höhe eine Faktorladung als gehaltvoll zu interpretieren ist. Kline (1997) erachtet beispielsweise Faktorladungen $>.30$ als praktisch bedeutsam, während Backhaus et al. (2008) $.50$ als Mindesthöhe festsetzen. Im Fall dieser Studie hätte eine Skala, bestehend aus Items mit Faktorladungen $> .50$ (7 Items), eine interne Konsistenz von $.64$ erbracht und wäre damit nur wenig reliabler als die ausgewählte Skala SJT_kurz mit elf Items. Ein weiterer Punkt des Bedenkens hätte gegen die Verwendung einer Skala bestehend aus nur sieben Items gesprochen. Aus Sicht des Autors würden sieben Situationen zukünftigen Bewerbern nur ein unvollständiges Bild über den späteren Alltag im Hotel liefern, da nicht alle Ausbildungsstationen gleichwertig einflößen. Die Frage, ob elf Situationen einen erschöpfenden oder zumindest repräsentativen Einblick in das Berufsfeld eines Hotelfachmannes bieten, ist ebenfalls nicht eindeutig zu bejahen (s. Abschnitt 4.6).

Bei der Überprüfung der Werteverteilungen für die Skala SJT_kurz anhand eines Stängel-Blatt-Diagramms (vgl. Bühl, 2008) und eines Boxplots (s. Anhang B) identifizierte der Autor zwei Teilnehmer mit extremen Werten (Ausreißern). Bei diesen Teilnehmern wurde jeweils ein Gesamt-

wert für die Skala SJT_kurz $< .30$ errechnet. Solche Ausreißerwerte können niedrige Trennschärfen und Reliabilitäten bedingen (Bühner, 2006). Eine Entfernung der beiden Datensätze aus weiteren Analysen zur Itemselektion hätte im Fall dieser Studie jedoch sowohl in einer geringeren internen Konsistenz ($\alpha = .56$) als auch in geringeren Trennschärfen (.06 bis .39) der elf Items gemündet. Aufgrund dessen wurden die Datenreihen dieser beiden Teilnehmer weiterhin in die Analysen zur Itemselektion einbezogen.

Der Tabelle 4-7 ist zu entnehmen, dass die Werteverteilungen für alle elf Items der Skala SJT_kurz sowie für den Gesamtwert der Skala linksschief sind. Nach Lienert und Raatz (1998) kann unter anderem das Fehlen von Teilnehmern mit geringen Testwerten eine asymmetrische Verteilungsform bedingen. Die Verteilungen der Werte je Item – und damit die Streuungen je Item – stellen Funktionen der Aufgabenschwierigkeiten dar. Im Rahmen der Berechnung der Aufgabenschwierigkeiten fiel auf, dass selbst Personen mit Berufserfahrung die Situationen 5 und 24 nicht richtig gelöst haben (s. Tabelle 4-10). Dies führt dazu, dass diejenigen Verhaltensalternativen, welche für die entsprechenden Situationen zur Verfügung standen, einer inhaltlichen Korrektur unterzogen werden sollen. Hingegen wird für die Situationen, welche nie falsch gelöst wurden, zunächst keine inhaltliche Revision angestrebt, da das Ausbleiben falscher Lösungen in diesem Fall womöglich auf die gegebene Berufserfahrung der Teilnehmer zurückgeführt werden kann. Weiterhin fiel bei der Analyse der Itemschwierigkeiten auf, dass die meisten Situationen keine mittleren Schwierigkeitsniveaus erreichten, sondern dass die Aufgabenschwierigkeiten viel mehr in einem breiten Bereich zwischen .05 und .79 streuten. Im Rückgriff auf Lienert und Raatz (1998) können stark variierende Aufgabenschwierigkeiten, welche $\neq .5$ ausfallen, sowohl geringere Streuungen der Itemwerte (vgl. Tabelle 4-7), geringere Trennschärfen der Items als auch damit verbunden niedrigere Reliabilitäten hervorrufen. Zur Bestimmung der Verlässlichkeit der Skala SJT_kurz wurde die interne Konsistenz berechnet. Mit einem Wert von .64 liegt die interne Konsistenz unter dem in der relevanten Literatur aufgestellten Mindestwert von .70 (z. B. Kanning, 2004). An dieser Stelle ist allerdings neben der oben angesprochenen gegebenen Streuung der Aufgabenschwierigkeiten die Multidimensionalität, welche SJTs innewohnt, als Einflussgröße zu erwähnen (vgl. Abschnitt

2.5.2), da jene ebenfalls das Auffinden einer zufrieden stellend homogenen Skala erschweren kann. Wird in Form von Reliabilitäten, welche für andere videobasierte SJTs berichtet wurden, ein differenter Vergleichsmaßstab herangezogen, ergibt sich folgendes Bild: Der in dieser Studie erzielte Reliabilitätswert liegt über den von Smiderle et al. (.47; 1994) oder auch Kanning (.57; 2005) verzeichneten Werten. Zudem liegt der für die Skala SJT_kurz vorgefundene Wert über der internen Konsistenz, welche durchschnittlich für situative Interviews (vgl. Klehe & Latham, 2006) verbucht werden kann.

Für drei der vier Hauptskalen des ISK-K ergaben sich bivariate Korrelationen nahe Null mit der Skala SJT_kurz. In einer Studie von Kanning (2005) ergab sich eine ähnliche Konstellation: Für drei von vier allgemeinen Dimensionen sozialer Kompetenz Soziale Orientierung, Selbstbehauptung, Selbstkontrolle und Soziale Wahrnehmung wurden bivariate Korrelationen zwischen $-.04$ und $.02$ mit einem videobasierten SJT angegeben. Nur die Dimension Selbstbehauptung erzielte mit dem videobasierten SJT in der Studie von Kanning (2005) eine signifikante Korrelation in Höhe von $.13$. Bei der vorliegenden Studie resultierte zwischen der Skala Offensivität und der Skala SJT_kurz eine statistisch bedeutsame Korrelation ($p < .05$). Die Hauptskala Offensivität des ISK-K ähnelt auf inhaltlicher Ebene der Dimension Selbstbehauptung, welche Kanning (2005) derart charakterisierte: „[...]*eigene Interessen gegenüber anderen Menschen durchsetzen*“ (S. 141). Dass für die Skala SJT_kurz keine signifikanten Korrelationen zu drei Hauptskalen des ISK-K nachgewiesen werden konnten, mag in dem strukturellen Unterschied zwischen *sozialen Kompetenzen*, welche der ISK-K abfragt, und *sozial kompetentem Verhalten* in konkreten Situationen, welches der SJT simuliert, zu suchen sein. Nach Rescke (1995) wird das Handeln in konkreten sozialen Situationen, zum Beispiel dem Umgang mit einem schwierigen Hotelgast, zugleich von allgemeinen sozialen Kompetenzen, welche das Ergebnis eines Sozialisationsprozesses sind, und spezifischen sozialen Kompetenzen determiniert. Letztere unterliegen nach Kanning (2009) einer Situationsspezifität und entspringen stärker individuellen Lernerfahrungen. Die Umsetzung von sozialen Kompetenzen in Verhaltenabsichten respektive darauf aufbauendem Verhalten erfolgt nicht in jeder Situation auf die gleiche Art und Weise, sondern gestaltet sich viel mehr als situationsspezifisches Zu-

sammenspiel von verschiedenen Kompetenzen. Ferner unterliegt sie zahlreichen situativen Einflussfaktoren (Kauffeld & Grote, 2000). Eine hohe Ausprägung sozialer Kompetenzen ermöglicht sozial kompetentes Verhalten und erhöht dessen Auftretenswahrscheinlichkeit, garantiert es jedoch nicht, da situative Einflüsse die Überführung in konkretes Verhalten verhindern können. Ob das gezeigte Verhalten als sozial kompetent angesehen wird, kann weiterhin nur in Abhängigkeit von der spezifischen Situation bestimmt werden (Erpenbeck, 1997): Der jeweilige Kontext bestimmt die Angemessenheit der jeweiligen Verhaltensweise und damit ihre Bewertung als sozial kompetent oder inkompetent (Huck & Schade, 2003). Nach Kanning (2005) ist die Verhaltensbewertung sozial kompetenten Verhaltens als „richtig“ oder „falsch“ von verschiedenen Bezugspunkten abhängig. Der soziale Bezugspunkt berücksichtigt die Gegebenheit, dass Sozialverhalten immer in einen sozialen Kontext eingebettet ist und nur vor dessen Hintergrund verstanden und beurteilt werden kann. Im Falle dieses SJT haben SME festgelegt, welches Verhalten als kompetent und welches als weniger kompetent zur Bewältigung sozialer Interaktionen in einem realen Kontext zu interpretieren ist (vgl. Kapitel 3). Der durch die SME festgelegte Bewertungsschlüssel für den SJT stellt damit in gewisser Weise den normativen Bezugspunkt dar, an dem die Angaben der Teilnehmer gemessen werden.

Zwischen dem Notendurchschnitt der Teilnehmer auf ihrem Schulabgangszeugnis und ihrer Leistung in der Skala SJT_kurz ergab sich eine nicht-signifikante Korrelation nahe Null. Dieses Ergebnis vermag nicht zu überraschen, da der Notendurchschnitt den Mittelwert mehrerer Einzelnoten darstellt, welche wiederum nach Ingenkamp (1997) sehr wohl den Erreichungsgrad kognitiver Lernziele widerspiegeln, nicht jedoch den Erreichungsgrad sozialer Lernziele. Zudem zeigen Untersuchungen zu einzelnen Schulnoten weitgehend übereinstimmend, dass diese psychometrischen Gütekriterien nicht gerecht werden (Ingenkamp, 1997). Weiß (2006) wies zur diagnostischen Güte von Schulnoten darauf hin, dass ihr Zustandekommen von so unterschiedlichen Faktoren wie der Leistungsmotivation des Schülers, dem Bezugssystem und der Qualifikation des Lehrers sowie der Schulform und deren curricularen Vorgaben abhängt. Bei einem Selbstbericht des Notendurchschnitts ist ferner nicht die Möglichkeit von der Hand zu weisen, dass sich ein Teilnehmer der Studie

womöglich entweder nicht an seinen exakten Notendurchschnitt erinnert oder sich aber einen leicht positiveren Notendurchschnitt zugeschrieben hat.

4.6 Ausblick

Mit Schülern, welche Interesse an einer Ausbildung im Hotelfach hegen und kurz vor dem Ende ihrer Schullaufbahn stehen, sollen in der nachfolgenden Studie Personen getestet werden, welche realen Bewerbern stärker ähneln als Personen, die sich bereits in einer Ausbildung im Hotelfach befinden und damit auf einen gewissen Erfahrungsschatz im Umgang mit Hotelgästen zurückblicken können. Zudem setzt der Autor im Rahmen der nächsten Studie verschiedene Verfahren zur Messung der Intelligenz und berufsbezogenen Persönlichkeit ein, um auf diesem Wege weiterführende Aussagen über die Konstruktvalidität des SJT treffen zu können. Da die Stimuluspräsentation im SJT in der nächsten Studie ebenfalls variiert werden soll (Video vs. Paper-pencil), kann zudem festgestellt werden, inwiefern dies die Gütekriterien des SJT beeinflusst. Durch den Einsatz zweier verschiedener Ratingverfahren und damit verbunden fünf differierender scoring keys soll der Versuch unternommen werden, eine SJT-Skala zu erhalten, welche im Vergleich zum SJT_kurz einerseits mehr Items umfasst und andererseits eine höhere Reliabilität zu erzielen vermag.

5. Die dritte Studie

Im Rahmen der dritten Studie durchlaufen Probanden *ohne* vorherige Erfahrung im Hotelfach den SJT. Neben dem SJT werden Verfahren zur Messung der sozialen Validität, verschiedener Facetten der Intelligenz und unterschiedlicher Aspekte der berufsbezogenen Persönlichkeit herangezogen.

5.1 Beschreibung der eingesetzten Instrumente und Fragebögen

5.1.1 Der SJT: Video - vs. Paper-Pencil

In der vorliegenden Studie soll der SJT sowohl in einem bereits in der vorhergehenden Studie erprobten Videoformat als auch in einem für diese Studie entworfenen Paper-Pencil-Format eingesetzt werden. Zur Erinnerung sei nochmals erwähnt, dass beim SJT, welcher im Videoformat dargeboten wird, die Situationen als kurze Videos präsentiert werden, während die Verhaltensalternativen und Instruktionen für die Teilnehmer in Paper-Pencil Form vorgegeben werden. Hingegen liegen beim SJT im Paper-Pencil-Format sämtliche beteiligte Komponenten – Situationen, Verhaltensalternativen und Instruktion – in Textform (Paper-Pencil) vor. In Anbetracht der im vorangegangenen Kapitel vorgestellten Ergebnisse wurden die Verhaltensalternativen bei Items, welche selbst von Berufsschülern mit Erfahrung nicht richtig gelöst wurden, vor Beginn der in diesem Kapitel thematisierten Datenerhebung für beide in dieser Studie eingesetzten Formate einer *inhaltlichen Revision* unterzogen (vgl. für ein analoges Vorgehen Weekley & Jones, 1997). Hier wurde vermutet, dass die Verhaltensalternativen, welche von SME an Platz 1 der Rangordnungen gesetzt wurden (vgl. Kapitel 3), als zu unrealistisch von den Probanden eingestuft wurden. Zur Revision der Verhaltensalternativen wurde im Dezember 2008 ein weiteres Gespräch mit Mitarbeitern des Hotel Münster geführt, an dem neben dem Autor die Personalchefin und der für die Rezeption des Hotels zuständige Abteilungsleiter teilnahmen. Betroffen von den inhaltlichen Modifikationen waren die Situationen 5 und 24 (vgl. Anhang A). Bei der Situation 5 war bisher die Verhaltensalternative A, welche die SME an die oberste Position ihrer Rangordnung gesetzt haben, inhaltlich wie folgt beschrieben: *Sie stellen dem Gast auf Kosten des Hauses eine Monatskarte für Ihre Garage aus, damit ihm beim nächsten Be-*

such der Ärger über die Garagengebühr erspart bleibt (s. Anhang A). Nach der Revision ergibt sich folgende veränderte inhaltliche Ausgestaltung: *Sie berechnen dem Gast in diesem Fall nicht die Parkgebühr. Allerdings weisen Sie ihn zusätzlich darauf hin, dass die Benutzung der Garage der Inanspruchnahme einer kostenpflichtigen Dienstleistung entspricht.* Bei der Situation 24 war ebenfalls die Verhaltensalternative A diejenige Option, welche die SME präferiert haben: *Sie bitten den Gast um Geduld, da Sie ihren Vorgesetzten einschalten möchten, um diese Sache zu klären.* Die durchgeführte Modifikation führte zu folgenden inhaltlichen Ergänzungen: *Sie bieten dem Gast an, Ihren Vorgesetzten hinzuziehen, so dass der Gast mit diesem den Zeitplan und etwaige Verbesserungswünsche durchsprechen kann.*

Für die textgebundene Umschreibung der Situationen für die Paper-Pencil-Version des SJT griff der Autor einerseits auf die im Rahmen der Testkonstruktion (vgl. Kapitel 3) in Gesprächen mit SME gesammelten Notizen und andererseits auf die für den Videodreh geschriebenen Drehbücher zurück. Auf diesen beiden Quellen fußend verfasste der Autor erste Situationsbeschreibungen. Mit diesen Texten sollten die in den 25 Videos dargestellten Inhalte und Zusammenhänge möglichst passgenau wiedergegeben werden. Die verfassten Situationsbeschreibungen wurden zum einen von der Personalchefin des Hotel Münster hinsichtlich ihres Grades an Realismus gegengelesen und zweitens in Bezug auf ihre Passung zum Inhalt der Videos von einem weiteren Diplom-Psychologen überprüft. Beanstandete Texte verbesserte der Autor. Der SJT im Paper-Pencil-Format mit allen 25 Situationsbeschreibungen ist im Anhang C einzusehen.

In dieser Studie sehen sich die Teilnehmer im SJT mit zwei zu bearbeitenden *Aufgabenstellungen* konfrontiert. Dabei wird den Teilnehmern in beiden Versuchsbedingungen (Video vs. Paper-Pencil) im Anschluss an die Präsentation respektive Beschreibung der relevanten Situation zunächst abverlangt, die Wahrscheinlichkeit, dass sie die vier genannten Verhaltensweisen jeweils in der realen Situation zeigen würden, einzuschätzen. Zur Abgabe der vier Einschätzungen – jede der vier Verhaltensalternativen wird separat eingestuft (vgl. Chan & Schmitt, 2002) – steht jeweils eine siebenstufige Ratingskala (1 = *sehr unwahrscheinlich*/7 = *sehr wahrscheinlich*) zur Verfügung.

Die Einschätzungen der Teilnehmer müssen in Beziehung gesetzt werden zu dem von SME generierten Bewertungsschlüssel. In Kapitel 3 wurde dargestellt, dass SME die vier Verhaltensalternativen je Situation gemäß ihrer situativen Effektivität in eine Rangordnung gebracht haben. Da von den SME jedoch keine Einschätzungen auf Ratingskalen eingeholt wurden, müssen die Rangordnungen der Experten in das Ratingformat übersetzt werden (s. *Diskussion*). Zur Gewährleistung dieses Transformationsschrittes wird angenommen, dass SME die Verhaltensalternative, welche den ersten Platz in ihrer Rangfolge einnimmt, auf einer Ratingskala mit 7 (sehr wahrscheinlich) eingestuft hätten. Dagegen hätten sie die Verhaltensalternative, welche dem vierten und letzten Rangplatz zugeordnet wurde, mit 1 (sehr unwahrscheinlich) bewertet. Für die mittleren Rangplätze 2 und 3 hätten sie die Skalenpunkte 5 respektive 3 gewählt. Die Abbildung 5-1 bietet ein Beispiel für die Transformation anhand der Situation 17. Bei dieser Situation haben die beteiligten SME die Verhaltensalternative B an den 1. Rang, die Alternative C an den 2. Rang, die Verhaltensalternative D an den 3. Rang und die Verhaltensoption A an den 4. und letzten Rang gesetzt (vgl. Anhang A).

A. Sie erklären dem Gast, dass Sie ihm leider kein Zimmer anbieten können, wenn er den Kreditkartenabzug ablehnt.	⊗-②-③-④-⑤-⑥-⑦ 4. Rang
B. Sie erklären dem Gast, dass die Geschäftsbedingungen Ihres Hotels einen Kreditkartenabzug bei Neukunden vorschreiben und dass dies ein einmaliger Vorgang ist.	①-②-③-④-⑤-⑥-⊗ 1. Rang
C. Sie verweisen den Gast auf die Tatsache, dass der Kreditkartenabzug bei Neukunden eine Anweisung Ihrer Vorgesetzten ist, an die Sie sich zu halten haben.	①-②-③-④-⊗-⑥-⑦ 2. Rang
D. Sie verzichten auf den Kreditkartenabzug, da Sie dem Gast nicht das Gefühl geben wollen, ihm zu misstrauen.	①-②-⊗-④-⑤-⑥-⑦ 3. Rang

Abbildung 5-1: Umrechnung der Rohwerte im SJT (I)

Da nun die Rangordnungen der Experten übersetzt in das Ratingformat vorliegen, müssen in einem zweiten Schritt die Ratings der Teilnehmer mit diesem Maßstab verglichen werden, um ihnen einen Testwert für den SJT zuteilen zu können. Dies geschieht über die Bildung absoluter Diffe-

renzen zwischen dem Rating des Teilnehmers und dem transformierten Rating der Experten. Dieses Vorgehen soll in Form der Abbildung 5-2 illustriert werden.

A. Sie erklären dem Gast, dass Sie ihm leider kein Zimmer anbieten können, wenn er den Kreditkartenabzug ablehnt.	6-5-4-3-2-1-0
B. Sie erklären dem Gast, dass die Geschäftsbedingungen Ihres Hotels einen Kreditkartenabzug bei Neukunden vorschreiben und dass dies ein einmaliger Vorgang ist.	0-1-2-3-4-5-6
C. Sie verweisen den Gast auf die Tatsache, dass der Kreditkartenabzug bei Neukunden eine Anweisung Ihrer Vorgesetzten ist, an die Sie sich zu halten haben.	0-1-2-3-4-3-2
D. Sie verzichten auf den Kreditkartenabzug, da Sie dem Gast nicht das Gefühl geben wollen, ihm zu misstrauen.	2-3-4-3-2-1-0

Abbildung 5-2: Umrechnung der Rohwerte im SJT (II)

Stimmt ein Teilnehmer hinsichtlich seiner Einschätzung bei jeder der vier Verhaltensalternativen mit den Expertenratings überein, beträgt sein Itemscore gemäß der Abbildung 5-2 20 Punkte (6 + 6 + 4 + 4). Derart können für alle 25 Situationen Itemscores berechnet werden. Der Gesamtwert für den SJT ergibt sich weiterhin als $\sum \text{Situation}_{1,\dots,25} / 25$. Im weiteren Verlauf erhält dieser scoring key zur Vereinfachung den Titel *Rating-Schlüssel*.

Die Anwendung von Ratingskalen zum Einholen der Einschätzungen von Teilnehmern wird von mehreren Forschergruppen nahe gelegt (z. B. McDaniel & Nguyen, 2001; Ployhart & Holtz, 2006; vgl. Kapitel 2). Ratingskalen sollen demnach beispielsweise eine dem Erzielen angemessener Reliabilitäten und Validitäten zuträgliche Varianz erzeugen. Zudem wird der Einsatz von Ratingskalen mit der Aussicht, einzelne, inhaltlich homogene Konstrukte messen zu können, verbunden (z. B. Motowidlo, Crook, Kell & Naemi, 2009).

Im Anschluss an die Bewertung der Verhaltensalternativen auf Ratingsskalen werden die Teilnehmer dieser Studie dazu aufgefordert, die vier Verhaltensalternativen gemäß der Wahrscheinlichkeit, mit welcher sie die

vier genannten Verhaltensweisen in der jeweils vorab beschriebenen (Paper-Pencil) oder gezeigten (Video) Situation zeigen würden, anzuordnen. Zu diesem Zweck sollen sie jede der vier Verhaltensalternativen (A – D) einem von vier Rangplätzen (1. – 4.) *eindeutig*³⁶ zuordnen. Den Probanden die Festlegung einer Rangordnung abzuverlangen, entspricht nach Weekley et al. (2006) einer Variante des forced-choice Formats zum Einholen der Probandenantworten im SJT (vgl. Kapitel 2). Ein analoges Vorgehen findet sich zum Beispiel in Studien von Weekley, Ployhart und Harold (2004) sowie Kanning et al. (2006). Dem 1. Rang der Reihenfolge ist diejenige Verhaltensweise zuzuweisen, welche die Teilnehmer am wahrscheinlichsten in der realen Situation zeigen würden, und dem 4. Rangplatz entsprechend diejenige Verhaltensweise, welche am unwahrscheinlichsten gezeigt werden würde. Die präferierte Rangordnung wird dann in eine vorgegebene Tabelle eingetragen. Die Bewertung der von den Probanden erstellten Rangordnung zur Berechnung eines Scores für den SJT setzt voraus, dass diese mit der von SME erstellten Rangordnung (vgl. Kapitel 3) verglichen wird. Für die aus diesem Vergleich resultierende Punktevergabe sollen in dieser Studie vier verschiedene Schlüssel eingesetzt werden, welche nachfolgend samt ihrer Entscheidungsregeln eingeführt werden sollen.

1. Falls die Probanden eine Rangordnung erstellen, welche komplett – das heißt auf allen vier Positionen – mit der von SME gebildeten Rangordnung übereinstimmt, erhalten Sie 1 Punkt. Überschneidet sich die Rangordnung der Probanden hingegen nicht vollständig, sondern zum Beispiel nur an drei oder weniger Positionen, mit der von SME festgelegten Rangordnung, werden dem Probanden keine Punkte gutgeschrieben. Diese Art der Punkteverteilung soll fortan mit *Komplett-Schlüssel* titulierte werden. Bei Anwendung dieses scoring keys können die Teilnehmer der Studie als Testwert im SJT entweder 0 Punkte oder 1 Punkt erreichen.
2. Haben die Probanden an den ersten Platz und zugleich an den letzten Platz ihrer Rangordnung dieselben Verhaltensalternativen gesetzt wie die SME, erhalten sie zwei Punkte (+2). Sollten die Pro-

³⁶ Zuweisungen von zwei Verhaltensalternativen zu einem einzigen Rangplatz sind nicht zugelassen.

banden jedoch an den ersten Platz die von SME am schlechtesten eingestufte Verhaltensalternative und an den letzten Platz die von den Arbeitsplatzexperten am besten eingestufte Verhaltensalternative setzen, werden ihnen im Gegenzug zwei Punkte abgezogen (-2). Eine Übereinstimmung in der Rangordnung hinsichtlich entweder des ersten Platzes oder aber des letzten Platzes mit der Rangordnung der SME, bringt den Probanden einen Punkt ein (+1). Das Setzen der von Experten als beste Verhaltensalternative auserkorene an den letzten Platz oder der von den SME als am schlechtesten eingestufte Alternative an den ersten Platz führt zur Subtraktion eines Punktes (-1). Null Punkte erhalten solche Rangordnungen, deren erster und letzter Platz mit solchen Verhaltensalternativen besetzt sind, welche von SME weder als beste noch als schlechteste Verhaltensalternative eingestuft wurden. In der von SME präferierten Rangordnung nehmen diese Verhaltensoptionen die Rangplätze 2 respektive 3 ein. Aufgrund der konzeptionellen Nähe zu dem von Motowidlo et al. (1990) entwickelten Bewertungsschlüssel soll diese Form der Punkteverteilung im weiteren Verlauf als *Motowidlo-Schlüssel* bezeichnet werden. Falls dieser scoring key zur Anwendung kommt, bewegen sich die Testwerte der Teilnehmer im SJT zwischen -2 und +2 Punkten.

3. Die Rangordnung der Probanden wird auf allen vier Positionen mit der von SME angefertigten Rangordnung verglichen. Haben die Probanden eine Verhaltensalternative an dieselbe Position gesetzt wie die SME, verbuchen sie dafür einen Punkt (+1). Im Vergleich zur oben skizzierten Alternative 1 werden dementsprechend bereits Übereinstimmungen hinsichtlich einzelner Positionen belohnt. Unterscheiden sich die Verhaltensalternativen, welche Probanden und SME für eine bestimmte Position bestimmt haben, wird den Probanden kein Punkt angerechnet. Im Folgenden firmiert dieser scoring key unter der Bezeichnung *Einzelvergleich-Schlüssel*. Die Testwerte der Teilnehmer bewegen sich bei Zuhilfenahme dieses scoring keys zwischen 0 und +4 Punkten.
4. Auch bei diesem Verfahren wird die Rangordnung der Probanden bezüglich aller vier Positionen mit der von SME angefertigten

Rangordnung kontrastiert. Hier erhalten die Probanden vier Punkte, wenn ihre Rangordnung auf allen vier Positionen der von SME erstellten Rangordnung entspricht. Bei einer vorliegenden Nicht-Kongruenz wird bewertet, wie weit die Position, auf welche eine Verhaltensalternative von den Probanden gesetzt wurde, abweicht von der Position, welche die SME für die entsprechende Verhaltensalternative auserkoren haben. Das Vorgehen sei am Beispiel der Verhaltensalternative erklärt, welche Experten an Platz der 1 Rangordnung gesetzt haben. Probanden, welche diese Alternative nur auf Rang 2 setzen, erhalten -1 Punkt, während Probanden, welche die Verhaltensalternative auf die Plätze 3 respektive 4 gesetzt haben, -2 beziehungsweise -3 Punkte erhalten, da sie sich derart immer weiter von der Rangordnung der SME entfernen. Diese Differenz wird analog für die verbleibenden Verhaltensalternativen errechnet. Als Benennung dieses scoring keys wird nachfolgend der Name *Distanz-Schlüssel* verwandt werden. Testwerte, welche anhand dieses scoring keys berechnet werden, liegen zwischen -8 und +4 Punkten.

Im Anhang C sind Rechenbeispiele für Punkteverteilungen bei allen vier Schlüsseln zu finden.

Anhand der ermittelten Testwerte soll folgende Hypothese II-1 überprüft werden (vgl. Kapitel 4):

Hypothese II-1

Dem SJT liegt eine fünffaktorielle Struktur zugrunde. Die fünf Faktoren entsprechen den Anforderungsdimensionen Serviceorientierung, Beschwerdemanagement, Durchsetzungsfähigkeit, Handlungsfähigkeit und Gewissenhaftigkeit. Den fünf Anforderungsdimensionen wiederum liegt als Faktor 2. Ordnung Kundenorientierung zugrunde.

5.1.2 Skala zur Messung der sozialen Validität und Akzeptanz des SJT

Um die wahrgenommene soziale Validität (z. B. Schuler & Stehle, 1983) und Akzeptanz des SJT erfassen zu können, kommt eine von Kanning (2008) entwickelte und erprobte, aus zehn Items bestehende Skala zur Anwendung. Der Entwicklungsprozess der Skala stützt sich alleinig auf

an Bewerberstichproben (N = 273) generierten Daten. Die Bewerber aus verschiedenen Berufsfeldern, zum Beispiel aus dem Polizeidienst oder dem Bankgewerbe, durchliefen ein Assessment Center und mussten die von Kanning (2008) konstruierte Skala nach jeder einzelnen Übung des AC, beispielsweise einem Interview oder einem Leistungstest, bearbeiten. Die Skala soll von den Teilnehmern dieser Studie direkt im Anschluss an das Durchlaufen des SJTs bearbeitet werden (s. u.).

Mit seiner Skala ermöglicht Kanning (2008) die Messung dreier Dimensionen der sozialen Validität: Informationsgrad, Beeinflussbarkeit und Anforderungsbezug. Auf eine Erfassung der Dimension Feedback wird verzichtet. Zur inhaltlichen Vergegenwärtigung der Konstrukte sei auf den Abschnitt 2.5.3 verwiesen. Aus der Skala dienen jeweils zwei Items zur Messung des Informationsgrades und der Beeinflussbarkeit, fünf Items sind hingegen zur Messung des Anforderungsbezugs vorgesehen. Für die Messung der Akzeptanz – als ergänzendes Merkmal zur sozialen Validität – wird zudem ein einziges Item herangezogen, bei welchem die Teilnehmer angeben sollen, ob sie das eingeschätzte Verfahren selbst in der Personalauswahl einsetzen würden. Die Einschätzungen für die zehn Items geben die Teilnehmer mit Hilfe siebenstufiger Ratingskalen (1 = *stimme gar nicht zu* bis 7 = *stimme voll zu*) ab.

Für die drei einbezogenen Faktoren der sozialen Validität ermittelte Kanning (2008) folgende Reliabilitätskoeffizienzen (Cronbach's Alpha): Informationsgrad (.85), Beeinflussbarkeit (.63) sowie Augenscheinvalidität (.73).

Als weitere Variablen erhob Kanning (2008) Selbsteinschätzungen der Teilnehmer bezüglich der Leistung in verschiedenen Assessment Center-Übungsformen ein. Diese wurden jeweils über folgendes zusätzliche Item eingeholt: „*Was glauben Sie, wie gut haben Sie in Übung x abgeschnitten?*“ (Kanning, 2008; S. 23). Die höchste mittlere Korrelation (.52) zwischen den Skalen der sozialen Validität und Akzeptanz auf der einen Seite sowie der selbst eingeschätzten Leistung auf der anderen Seite trat auf bei der Übungsform Planungsaufgabe. Dagegen zeigte sich die niedrigste mittlere Korrelation (.17) in Bezug auf die Übungsform Leistungstest. Über alle Übungen hinweg ergeben sich für den Faktor Beeinflussbarkeit die stärksten korrelativen Zusammenhänge mit Selbsteinschätzungen der Übungsleistung.

Aufgrund der Ergebnisse aktueller Studien, in denen sich bei SJTs mit variierter Situationspräsentation (Video vs. Paper-Pencil) keine signifikanten Unterschiede in Bezug auf die soziale Validität ergaben (z. B. Richman et al., 2000; Kanning et al., 2006; vgl. Abschnitt 2.5.3), wird folgende Hypothese aufgestellt:

Hypothese II-2

Zwischen dem SJT im Videoformat und dem SJT im Paper-Pencil Format ergeben sich in Bezug auf die Einschätzung der sozialen Validität keine bedeutsamen Unterschiede.

5.1.3 Demographische Variablen

Auf die Einschätzung der sozialen Validität und Akzeptanz des SJT folgt die Erhebung demographischer Variablen. Begonnen wird mit der Erfassung des Alters, des Geschlechts und der Muttersprache der Teilnehmer. Daran schließt sich das Einholen der letzten Zeugnisnoten in den drei schulischen Hauptfächern Deutsch, Englisch und Mathematik an. Des Weiteren bat der Autor die Teilnehmer anzugeben, ob sie zum Zeitpunkt der Datenerhebung bereits über Berufserfahrung im Hotel- oder Gaststättengewerbe verfügten. Zum Abschluss geben die Teilnehmer ihre Pläne nach Erlangung des Schulabschlusses an. Zur Auswahl stehen der Erwerb höherer Schulabschlüsse oder der Beginn einer Berufsausbildung. Von den Teilnehmern, welche eine berufliche Ausbildung absolvieren möchten, ist ein weiterer Punkt zu beantworten. Hier müssen jene angeben, welche Art von beruflicher Ausbildung ergriffen werden soll: Eine Ausbildung im kaufmännischen Bereich, im technischen Bereich, im sozialen Bereich oder im künstlerisch-kreativen Sektor. Für jede Art der Ausbildung werden illustrative Beispiele genannt, zum Beispiel Kindergärtner für den sozialen Bereich.

5.1.4 Die Grundintelligenztest Skala 2-Revision

Die Grundintelligenztest Skala 2-Revision (Culture Fair Test; CFT-20-R; Weiß, 2006) intendiert die Erfassung des allgemeinen intellektuellen Niveaus (Grundintelligenz) im Sinne des Konzeptes der fluiden Intelligenz (general fluid ability). Grundsätzlich kann diese Facette der Intelligenz als Fähigkeit gedeutet werden, figurale Beziehungen und formal-

logische Denkprobleme unterschiedlichen Komplexitätsgrades auch in neuartigen Situationen zu erkennen und innerhalb einer vorgegebenen Zeitspanne zu verarbeiten (vgl. Cattell, 1968). Aufgrund des letzten Punktes ist der CFT-20-R als speed-Test (vgl. Lienert & Raatz, 1998) anzusehen. Der Testautor Weiß (2006) betont die Möglichkeit, das allgemeine Intelligenzniveau sprachfrei und milieunabhängig mit dem CFT-20-R messen zu können (s. u. zum Vergleich mit dem Wortschatztest; Weiß, 2007). Eingeordnet in das Intelligenzstrukturmodell von Jäger (1982) rückt Weiß (2006) den CFT-20-R hinsichtlich der zu seiner Bearbeitung erforderlichen Prozesse (operative Ebene) in die Nähe des Faktors Verarbeitungskapazität. In Bezug auf die inhaltliche Gestaltung der Items (s. u.) sieht Weiß (2006) einen hohen Überschneidungsgrad mit der figuralen Dimension des Jägerschen Modells. Im praktischen Gebrauch verankert Weiß (2006) den CFT-20-R in der Schullaufbahnberatung, wobei er sowohl bei der Diagnose von Hochbegabungen als auch von Entwicklungsdefiziten hilfreich sein könne.

Der CFT-20-R setzt sich aus zwei Testteilen zusammen, von denen der erste 56 Items umfasst und der zweite 45. Jeder Testteil wiederum besteht aus vier unterschiedlichen Subtests: Reihenfortsetzen, Klassifikationen, Matrizen und Topologische Schlussfolgerungen. Die vier Subtests werden in den beiden Testteilen nach derselben Reihenfolge durchlaufen. Innerhalb eines Subtests wächst die Itemschwierigkeit stetig an. Unabhängig von der Art des Subtests stellt jedes der 101 Items im CFT-20-R eine sprachfreie, in zeichnerischer Form dargestellte Aufgabe dar. Zur Lösung der Aufgabe müssen die Probanden stets aus fünf im Multiple-Choice Format vorgegebenen Alternativen die richtige Antwort wählen. Die Anzahl richtig gelöster Aufgaben entspricht dem Testscore (Rohwert), wobei Scores entweder getrennt für die beiden einzelnen Testteile oder aber für den gesamten Test (101 Items) berechnet werden können.

Für den CFT-20-R werden von Weiß (2006) einerseits Retest-Reliabilitäten und andererseits split-half Reliabilitäten angegeben. Je nach Zeitabstand zwischen den beiden Testdurchführungen und Schulform der Probanden nennt Weiß (2006) Retest-Reliabilitäten zwischen .80 und .82. Die split-half Reliabilität – als Ausdruck der internen Konsistenz – für den gesamten CFT-20-R (1. + 2. Testteil) beläuft sich gemäß Weiß (2006) auf einen Wert von .95 ($n = 144$). Bei der Exploration der Validität des

CFT-20-R rückte Weiß (2006) die faktorielle Validität in den Vordergrund. Hier fand der Testautor in mehreren Untersuchungen substantielle Ladungen aller vier Subtests des CFT-20-R (s. o.) auf dem Generalfaktor der fluiden Intelligenz vor. Daneben widmete sich Weiß (2006) auch einer Analyse der inneren kriterienbezogenen Validität des CFT-20-R. In diesem Zusammenhang ermittelte er Korrelationen zwischen dem CFT-20-R und anderen kognitiven Leistungstests. Es resultierte eine mittlere Korrelation von .64.

5.1.5 Der Wortschatztest

Der Wortschatztest (WS; Weiß, 2007) wird von Weiß (2007) als Ergänzung zum oben beschriebenen CFT-20-R betrachtet. Im Gegensatz zum CFT-20-R, welcher die Messung der fluiden Intelligenz anstrebt, zielt der WS auf die Messung der kristallinen Intelligenz (general crystallized ability). Einer Definition von Cattell (1973; S. 268) folgend repräsentiert die kristalline Intelligenz „[...]die Sammlung gelernter Kenntnisse, die sich ein Mensch angeeignet hat, indem er sein flüssige Intelligenz beim Lernen in der Schule anwandte“. Nach Weiß (2007) erfordert ein erfolgreiches Bearbeiten des WS von den Probanden die „Kenntnis von Wörtern und passive Verfügbarkeit über den Wortschatz der Umgangssprache, der über den Grundwortschatz der deutschen Sprache hinausgeht“ (S. 20) Im Verbund mit dem Zahlenfolgetest ermöglichen CFT-20-R und WS nach Weiß (2007) die Messung des zentralen Intelligenzfaktors Verarbeitungskapazität aus dem Intelligenzstrukturmodell von Jäger (1982; s. o.). Jäger (1982; S. 213) charakterisiert diesen Intelligenzfaktor wie folgt: „Verarbeitung komplexer Informationen bei Aufgaben, die nicht auf Antrieb zu lösen sind, sondern Heranziehen, Verfügbarhalten, vielfältiges Beziehungsstiften, formallogisch exaktes Denken und sachgerechtes Beurteilen von Informationen erfordern“. Im Zuge der Schullaufbahnberatung sieht Weiß (2007) in dem WS ein probates Mittel zur Feststellung des Entwicklungsstandes im verbalen Bereich. Jedoch sei die Leistung im Wortschatztest nach Weiß (2006) in starkem Maße abhängig von den bisherigen Lernerfahrungen und dem kulturellen Hintergrund der Probanden.

Der WS liegt in zwei Parallelförmigen, welche jeweils aus 30 Items zusammengesetzt sind, vor. Jedes Item im WS entspricht einer Mehrfachwahlaufgabe, bei der die Probanden einem vorgegebenen Schlüsselwort

aus einer Auswahl von fünf Wörtern dasjenige Wort mit gleicher oder ähnlicher Bedeutung zuweisen sollen. Der Testwert (Rohwert) ergibt sich im WS als Summe der richtig gelösten Items. Die Iteminhalte sind nach Weiß (2007; S. 11) elf verschiedenen Lebensbereichen, zum Beispiel Essen/Trinken oder Wirtschaft/Zahlungsverkehr, entliehen.

Aussagen über die Zuverlässigkeit des WS sind mit Hilfe von Weiß (2007) berichteter split-half und Retest-Reliabilitäten möglich. In Bezug auf die split-half Reliabilität gibt Weiß (2007) von der Schulform der Probanden abhängende Werte zwischen .79 bis .90 an. Die über alle Schulformen (N = 608) gemittelte split-half Reliabilität, als Maß der internen Konsistenz, beziffert er mit .85. An Schülern aus 4. Grundschulklassen und 9. Hauptschulklassen gewann Weiß (2007) Auskünfte über die Stabilität der Messungen. Die mittlere Retest-Reliabilität liegt im Falle des WS bei .83. Hinsichtlich der Validität des WS konnte Weiß (2007) aufzeigen, dass sich mit der Deutschnote eine mittlere Korrelation von .48 (N = 689 mit Schülern unterschiedlicher Jahrgänge und Schulformen) sowie mit der Englischnote eine mittlere Korrelation von .41 ergab (N = 390 mit Schülern unterschiedlicher Jahrgänge und Schulformen). Dabei wies Weiß (2007) auf die „*erhebliche Schwankungsbreite*“ (S. 45) der Ergebnisse je nach untersuchter Klasse und Schulform hin.

Neben Schulnoten wählte Weiß (2007) verschiedene andere Kriterien zur konvergenten und divergenten Validierung des WS. Bei der konvergenten Validierung ergaben sich für den WS durchgängig, das heißt klassen- und schulformübergreifend, sehr signifikante Korrelationen (.62 bis .69; $p < .01$) mit Tests zur Messung verwandter Konstrukte, zum Beispiel einem anderen Wortschatztest und einem kognitiven Fähigkeitstest. Hingegen erzielte der WS im Rahmen der divergenten Validierung mit von Weiß (2007) als „*konstruktfernen*“ (S. 48 ff.) bezeichneten Tests, zum Beispiel einem Matrizen- oder einem Test zur Messung der Aufmerksamkeit, keinerlei signifikante Korrelation.

5.1.6 Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung

Das von Hossiep und Paschen (2003) entworfene Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP) erlaubt „[...]die standardisierte Erfassung des Selbstbildes eines Testkandidaten in Hin-

lick auf relevante persönlichkeitsorientierte Beschreibungsdimensionen aus dem Berufsleben“ (S. 14). Für die Entwicklung des BIP stand keine umfassende theoretische Ausgangsposition Pate. Hossiep und Paschen (2003) orientierten sich viel mehr an den Ergebnissen von Validitätsstudien zu veröffentlichten Persönlichkeitstests, um besonders erfolgsrelevante Dimensionen identifizieren zu können. Daneben suchten sie das Gespräch mit erfahrenen Personalern, welche ihnen erfolgskritische Dimensionen aus dem Berufsalltag übermitteln sollten. Auf diesem Wege – Sichtung von Validitätsstudien und Befragung von SME – legten sich die Testautoren auf vier übergeordnete persönliche Eignungsvoraussetzungen fest: Berufliche Orientierung, Soziale Kompetenzen, Arbeitsverhalten und Psychische Konstitution. Den vier übergeordneten Eignungsvoraussetzungen wiederum ordneten sie qualitativ 14 Persönlichkeitsdimensionen zu, deren Messung beim BIP fokussiert wird. Zu der Beruflichen Orientierung zählen die Dimensionen Leistungsmotivation, Gestaltungsmotivation und Führungsmotivation, zu den Sozialen Kompetenzen werden Sensitivität, Kontaktfähigkeit, Soziabilität, Teamorientierung und Durchsetzungsstärke gezählt, dem Arbeitsverhalten werden Gewissenhaftigkeit, Flexibilität und Handlungsorientierung zugerechnet und die Psychische Konstitution umfasst Emotionale Stabilität, Belastbarkeit und Selbstbewusstsein. Einen inhaltlichen Einblick in die 13 für diese Studie³⁷ relevanten Dimensionen gestattet die Tabelle 5-1. Weiterführende Erläuterungen zur theoretischen Inspiration der Dimensionen sind bei Hossiep und Paschen (2003; S. 27 ff.) nachzulesen.

³⁷ Aufgrund der Zusammensetzung der in dieser Studie untersuchten Stichprobe (s.u.) wurde von einer Erhebung der Dimension Führungsmotivation Abstand genommen.

Tabelle 5-1: Kurzdefinitionen der mit dem BIP erfassten Konstrukte nach Hossiep & Paschen (2003; S. 22)

Dimension	Anzahl der Items	Kurzdefinition
Leistungsmotivation	14	Bereitschaft zur Auseinandersetzung mit einem hohen Gütemaßstab; Motiv, hohe Anforderungen an die eigene Leistung zu stellen; große Anstrengungsbereitschaft; Motiv zur fortwährenden Steigerung der eigenen Leistungen
Gestaltungsmotivation	12	Ausgeprägtes Motiv, subjektiv erlebte Missstände zu verändern sowie Prozesse und Strukturen nach eigenen Vorstellungen gestalten zu wollen; ausgeprägte Bereitschaft zur Einflussnahme und zur Verfolgung eigener Auffassungen
Gewissenhaftigkeit	14	Sorgfältiger Arbeitsstil; hohe Zuverlässigkeit; detailorientierte Arbeitsweise; hohe Wertschätzung konzeptionellen Arbeitens; Hang zum Perfektionismus
Flexibilität	14	Hohe Bereitschaft und Fähigkeit, sich auf neue oder unvorhergesehene Situationen einzustellen und Ungewissheit zu tolerieren; Offenheit für neue Perspektiven und Methoden; hohe Veränderungsbereitschaft
Handlungsorientierung	14	Fähigkeit und Wille zur raschen Umsetzung einer Entscheidung in zielgerichtete Aktivität sowie zur Abschirmung einer gewählten Handlungsalternative gegenüber
Sensitivität	12	Gutes Gespür auch für schwache Signale in sozialen Situationen; großes Einfühlungsvermögen; sichere Zuordnung und Interpretation der Verhaltensweisen anderer

5. Die dritte Studie

Kontaktfähigkeit	16	Ausgeprägte Fähigkeit und Präferenz des Zugehens auf bekannte und unbekannte Menschen und des Aufbaus sowie der Pflege von Beziehungen; aktiver Aufbau und Aufrechterhaltung von beruflichen wie privaten Netzwerken
Soziabilität	15	Ausgeprägte Präferenz für Sozialverhalten, welches von Freundlichkeit und Rücksichtnahme geprägt ist; Großzügigkeit in Bezug auf Schwächen der Interaktionspartner; ausgeprägter Wunsch nach einem harmonischen Miteinander
Teamorientierung	13	Hohe Wertschätzung von Teamarbeit und Kooperation; Bereitschaft zur aktiven Unterstützung von Teamprozessen; bereitwillige Zurücknahme eigener Profilierungsmöglichkeiten zugunsten
Durchsetzungsstärke	12	Tendenz zur Dominanz in sozialen Situationen; Bestreben, die eigenen Ziele auch gegen Widerstände nachhaltig zu verfolgen; hohe
Emotionale Stabilität	16	Ausgeglichene und wenig sprunghafte emotionale Reaktionen; rasche Überwindung von Rückschlägen und Misserfolgen; ausgeprägte Fähigkeit zur Kontrolle eigener emotionaler Reaktionen
Belastbarkeit	13	Selbsteinschätzung als (psychophysisch) hoch widerstandsfähig; starke Bereitschaft, sich auch außergewöhnlichen Belastungen auszusetzen und diesen
Selbstbewusstsein	16	(Emotionale) Unabhängigkeit von den Urteilen anderer; hohe Selbstwirksamkeitsüberzeugung; großes Selbstvertrauen bezüglich der eigenen Fähigkeiten und

Für die Messung der berufsbezogenen Persönlichkeit werden im BIP 210 Items aufgewendet. Um die relevanten Selbstbeschreibungen der Probanden einzuholen, setzen Hossiep und Paschen (2003) sechsfach abgestufte (1 = *trifft voll zu*/6 = *trifft überhaupt nicht zu*) Ratingskalen ein. Für die Berechnung der vierzehn Skalenwerte (s. o.) werden die entsprechenden einzelnen Itemscores aufsummiert.

Für alle Dimensionen des BIP liegen verschiedene Reliabilitätskoeffizienten vor (Hossiep & Paschen, 2003; S. 30 ff.). Die internen Konsistenzen liegen – gemessen über Cronbach's Alpha – in einem Bereich von .74 (Gestaltungsmotivation) bis .91 (Belastbarkeit). Wird die interne Konsistenz über die split-half Methode berechnet, bewegen sich die Reliabilitätswerte zwischen .72 (Gestaltungsmotivation) und .91 (Belastbarkeit). Neben den gerade genannten Kennwerten zur internen Konsistenz des BIP sind in Form von Retest-Reliabilitäten Erkenntnisse zur Merkmalskonstanz verfügbar. Bei einem zeitlichen Abstand von fünf bis 36 Monaten zwischen den beiden Durchführungen des BIP ergeben sich Reliabilitätskoeffizienten von .69 (Handlungsorientierung) bis .80 (Kontaktfähigkeit und Soziabilität). Studien zur kriterienbezogenen Validität des BIP beschäftigten sich zum einen mit korrelativen Zusammenhängen zwischen den BIP-Dimensionen und verschiedenen Indikatoren beruflichen Erfolgs. So ergaben sich um Alterseffekte bereinigte Korrelationen im Rahmen von -.09 (Soziabilität) und .32 (Führungsmotivation) zwischen Dimensionen des BIP und dem Entgelt der Probanden. Nach Hossiep und Paschen (2003) sind rund 15 % der Varianz bezüglich des Entgelts durch unterschiedliche Ausprägungen der im BIP abgefragten Dimensionen erklärbar. Ein ähnlicher Befund ist für das Kriterium Hierarchische Position feststellbar. In diesem Fall erklären die 14 Dimensionen gemeinsam 16 % an Kriteriumsvarianz. Die Spannbreite der altersbereinigten Korrelationen reicht von -.10 (Soziabilität) bis .35 (Führungsmotivation). Zudem zeigten sich positive Zusammenhänge zwischen allen Dimensionen des BIP und zwei subjektiven Kriterien, der eigenen Berufserfolgseinschätzung und der selbst eingestuften Tätigkeitszufriedenheit. Als statistisch bedeutsamste Prädiktoren traten hierbei jeweils die Dimensionen Emotionale Stabilität und Durchsetzungsstärke hervor.

Zur Erklärung von Unterschieden in der Abiturnote oder Leistungen im ersten Studienabschnitt (Vordiplom) leisten die Dimensionen des BIP nur

einen geringen Beitrag. Hossiep und Paschen (2003) deuten dies im Sinne einer gegebenen divergenten Validität des BIP, da Schulnoten und universitäre Noten zumeist auf Basis kognitiver Leistungstests erteilt würden. Bestätigungen für die konvergente Validität des BIP liegen in Form von Korrelationen mit verschiedenen Persönlichkeitstests vor, zum Beispiel dem NEO-FFI (Borkenau & Ostendorf, 2008). Hier ergeben sich jeweils die höchsten Korrelationen zwischen thematisch ähnlich konzipierten Dimensionen.

5.2 Ablauf der Datenerhebung und Beschreibung der Stichprobe

5.2.1 Ablauf der Datenerhebung

Für die Datenerhebung konnten Schüler der Paul-Gerhardt Realschule (PGR) in Münster gewonnen werden. Zwischen dem Hotel Münster und der PGR existiert seit mehreren Jahren eine Kooperation, so dass der Kontakt zu der Schule und die Informationsweitergabe über das geplante Vorhaben durch die Personalchefin des Hotel Münster vorgenommen wurde. Es sollten für diese Datenerhebung nur Schüler als Teilnehmer akquiriert werden, welche sowohl Interesse an einer Ausbildung im Hotelfach hatten als auch über das entsprechende Alter – Jahrgangsstufen Neun und Zehn – verfügten, um in Kürze die Realschule abzuschließen. Zudem sollten mindestens 100 Schüler zu dem Termin eingeladen werden, um eine genügend große Stichprobe testen zu können. Eine finale Zusammenstellung der Stichprobe erfolgte über die zuständige Verbindungslehrerin an der PGR.

Vorab wurde mit der Verbindungslehrerin vereinbart, dass den partizipierenden Schülern einerseits die Gelegenheit gegeben werden sollte, einen wirklichkeitsnahen Einblick in eine Personalauswahlsituation zu erhalten und andererseits ein Stück weit am Berufsalltag eines Hotelfachmannes teilhaben zu können. Um beide Vorhaben umsetzen zu können, wurde sowohl der Einsatz der oben vorgestellten Tests als auch eine Führung durch das Hotel Münster mit seinen Arbeitsbereichen beschlossen. Als Gratifikation für die Teilnahme sollte den Schülern eine Übersicht über ihr Abschneiden in den verschiedenen Tests zugestellt werden. Wie bei der in Kapitel 3 dargestellten Studie geschah dies in Form eines Leistungsprofils (s. Anhang C).

Die Datenerhebung mit den Schülern der PGR wurde am 20.01.2009 in zwei Veranstaltungssälen des Hotel Münster durchgeführt. Eine Kontrolle der technischen Voraussetzungen, insbesondere der Möglichkeit, die videobasierten Items im SJT audiovisuell angemessen präsentieren zu können, hatte bereits wenige Tage zuvor stattgefunden. Bei der eigentlichen Datenerhebung wurde der Autor ganztätig von einem weiteren Diplompsychologen unterstützt. Ferner waren über den gesamten Zeitraum drei

Lehrer der PGR vor Ort. Der in der Tabelle 5-2 aufgeführte Zeitplan lag der Datenerhebung zugrunde.

Tabelle 5-2: Zeitplan am Tag der Datenerhebung

Testform	Zeitdauer
SJT ₃₈ (2 Versuchsbedingungen)	8:15 – 9:30 Uhr (75 Minuten)
CFT-20-R	9:45 – 10:45 Uhr (60 Minuten)
Wortschatztest	10:45 – 11:00 Uhr (15 Minuten)
Führung durch das Hotel Münster	11:15 – 12:15 Uhr (60 Minuten)
BIP	12:15 – 13:00 Uhr (45 Minuten)

Zwischen den Übungen und Tests³⁹ wurden den Schülern, wie der Tabelle 5-2 entnommen werden kann, Pausen eingeräumt. Die Durchführungszeiten für den CFT-20-R⁴⁰, den Wortschatztest und den BIP wurden den Vorgaben der jeweiligen Testautoren angepasst.

Vor Beginn des SJTs fand eine kurze Begrüßung und Vorstellung des Vorhabens durch den Autor statt. Die anwesenden Schüler wurden in diesem Zusammenhang darum gebeten, den Tag als Chance anzusehen, Bestandteile eines realistischen Auswahlverfahrens kennen zu lernen und sich dementsprechend anzustrengen. Zudem wurden die strenge Wahrung des Datenschutzes und der Erhalt eines individuellen Leistungsprofils⁴¹ (s. Anhang C) nach Auswertung der Ergebnisse hervorgehoben. Nach dieser allgemeinen Einführung wurden die Schüler in zufälliger Weise zwei Versuchsbedingungen zugeteilt: Die eine Gruppe sollte den SJT im Videoformat bearbeiten, während die andere Gruppe den SJT als Papier-Pencil Version zugeteilt bekam. Nach der randomisierten Zuteilung fand die Durchführung des SJT statt – aufgrund der unterschiedlichen Versuchsbedingungen in zwei räumlich getrennten Sälen.

³⁸ Der SJT wurde mit „Fragebogen zum Verhalten gegenüber Hotelgästen“ benannt.

³⁹ Gegenüber den Teilnehmern wurde der Begriff „Test“ vermieden und durch Übung oder Aufgabe ersetzt.

⁴⁰ Der CFT-20-R wurde ohne Testzeitverlängerung durchgeführt (vgl. Weiß, 2006; S. 24 ff.).

⁴¹ Nach Rücksprache mit den Lehrern des PGR wurde für das BIP kein individuelles Profil erstellt. Dieses Verfahren wurde den Schülern im Rahmen der Rückmeldung in allgemeiner Form erklärt.

Als weiterer Einschub vor Start des SJT forderte der Autor die Teilnehmer dazu auf, einen persönlichen Code aus Buchstaben und Zahlen zu bilden (s. Kapitel 4). Bei der Erklärung des Codes und der daran anschließenden Durchführung des videobasierten SJT orientierte sich der Autor an der bereits in Kapitel 4 skizzierten Vorgehensweise. Zusätzlich erklärte der Autor den Teilnehmern vorab an einer bereitgestellten Tafel die Verwendung der Ratingskalen sowie die Erstellung der Rangordnungen (s. o.) an einem fiktiven Beispiel. Daneben gewährte der Autor den Teilnehmern wegen ihres Alters und ihrer angenommenen Testunerfahrenheit bei dem ersten Item eine Wiederholung des Videos sowie Rückfragen zur Richtigkeit der Aufgabendurchführung. Zudem fragte der Autor die Teilnehmer nach jedem Item, ob sie auch alle vier Verhaltensalternativen eingeschätzt und eine Rangordnung eingetragen haben. Die parallel stattfindende Testung unter der Paper-Pencil-Bedingung wurde von dem zweiten Diplom-Psychologen begleitet und beruhte auf einer vollständig standardisierten schriftlichen Instruktion. Für die Instruktion der Teilnehmer bei der Durchführung des CFT-20-R und des WS befolgte der Autor die von Weiß (2006; S. 20 ff.) respektive Weiß (2007; S. 14) vorgegebenen formalisierten Erläuterungen und Anweisungen. Da ein Abgucken beim CFT-20-R oder auch WS stärkere Verzerrungen in den Testwerten hervorrufen kann als beispielsweise beim SJT oder BIP, wurden die Teilnehmer bei diesen Tests über den gesamten Saal verteilt. Im Übrigen wurden die Teilnehmer vor jedem dem SJT nachfolgenden Test gebeten, ihren individuellen Code auf die Front ihrer jeweiligen Testbögen einzutragen. Die Führung der Teilnehmer durch das Hotel Münster wurde in Form von fünf Kleingruppen realisiert, welche der Leitung von Auszubildenden des Hotels unterstanden. Einen Abschluss fand die Datenerhebung durch die Durchführung des BIP. Die Instruktion der Teilnehmer erfolgte über eine von Hossiep und Paschen (2003) entwickelte schriftliche Instruktion. Daneben stellte der Autor, nachdem alle Teilnehmer diese Instruktion gelesen hatten, angesichts des Alters der Teilnehmer besonders die Wichtigkeit heraus, dass sich diese ihr Verhalten in solchen Situationen, die sie bislang noch nicht erlebt hatten, möglichst genau vorstellen sollten. Zudem stand der Autor den Teilnehmern die Möglichkeit zu, jederzeit nachzufragen, falls ihnen Aussagen oder Inhalte unklar respektive unbekannt wären. Nach Beendigung des BIP bedankte sich der Autor bei

den Teilnehmern und erinnerte sie nochmals an die Aushändigung eines individuellen Leistungsprofils sowie die Sicherung der Anonymität.

5.2.2 Beschreibung der Stichprobe

Von den 62 Realschülern, die teilgenommen haben, waren 34 männlich (54.84 %) und 28 (45.16 %) weiblich. Zehn Berufsschüler (16.13 %) wiesen zum Zeitpunkt der Datenerhebung erste Erfahrungen im Hotel- und Gaststättengewerbe auf, während 52 (83.87 %) noch auf keinerlei Erfahrung in diesem Arbeitsbereich zurückblicken konnten. Die jüngsten Teilnehmer an der Studie waren 14 Jahre alt und die ältesten Teilnehmer 17 Jahre. Das durchschnittliche Alter der Stichprobe beträgt 15.02 Jahre. Bei der Frage nach der Muttersprache der Teilnehmer rangiert Deutsch mit 56 Nennungen (90.32 % der Stichprobe) an erster Stelle. Weitere angegebene Muttersprachen waren Türkisch (4 Nennungen), Polnisch (1) und Englisch (1). Nach Beendigung der Realschule streben 33 Teilnehmer (53.23 %) den Beginn einer Ausbildung an und 27 Teilnehmer (43.55 %) planen die Erreichung höherer Schulabschlüsse. Zwei Schüler (3.22 %) konnten sich nicht festlegen, wie ihre weitere berufliche oder schulische Zukunft aussehen soll. Aus der Gruppe an Teilnehmern, welche eine Ausbildung im Anschluss an die Realschule ins Auge fassen, präferieren elf (33.33 %) eine kaufmännische Ausbildung⁴², fünf (15.15 %) eine technische Ausbildung, sechs (18.18 %) eine Ausbildung im sozialen Bereich und acht (24.24 %) eine Ausbildung mit künstlerisch-kreativem Anstrich. Unentschieden hinsichtlich der Art der Ausbildung waren drei Teilnehmer (9.09 %). In den drei abgefragten Schulfächern⁴³ Mathematik, Deutsch und Englisch wurden durchschnittliche Noten von 3.41 (Mathematik), 3.23 (Deutsch) und 3.36 (Englisch) ermittelt. Eine nach Versuchsbedingungen (Video vs. Paper-Pencil) getrennte Beschreibung der Demographie bleibt aufgrund der geringen Stichprobengrößen (jeweils 31 Teilnehmer) aus.

⁴² Eine Ausbildung im Hotelfach wird den kaufmännischen Berufen zugerechnet.

⁴³ Ein Teilnehmer enthielt sich bei der Angabe der Schulnoten.

5.3 Deskriptive Statistik

Das Ziel dieses Abschnitts ist die Darstellung deskriptiver Statistiken in Bezug auf die eingesetzten Fragebögen und Leistungstests (s. Abschnitt 5.1). An dieser Stelle noch keine Erwähnung findet eine Darstellung deskriptiver Statistiken zum SJT. Diese erfolgt erst im Anschluss an noch ausstehende faktorenanalytische Untersuchungen (s. u.).

5.3.1 Fluide Intelligenz

Im Folgenden stehen deskriptive Statistiken im Zusammenhang mit dem CFT-20-R im Vordergrund. Einen Überblick über zentrale Kennwerte verschafft die Tabelle 5-3.

Tabelle 5-3: Deskriptive Statistiken des CFT-20-R (N = 62)

Teil des CFT-20-R	MW	SD	Minimalwert	Maximalwert
1. Testteil (56 Items)	40.60	6.03	25	54
2. Testteil (45 Items)	29.71	5.31	18	42
Gesamttest (101 Items)	70.26	10.12	44	94

Zur Interpretation der Mittelwerte (MW) sei daran erinnert, dass die Testwerte der Teilnehmer im CFT-20-R der Anzahl richtig gelöster Aufgaben entsprechen (s. o.). Exemplarisch hätte sich für den Gesamttest (101 Items) ein – illusorischer – Mittelwert von 101 ergeben, falls alle 62 Teilnehmer alle 101 Aufgaben richtig gelöst hätten. Im Schnitt wurden gemäß der Tabelle 5-3 rund 70 Aufgaben von den Teilnehmern richtig gelöst. Der CFT-20-R bietet die Möglichkeit, die Testrohre in IQ-Werte umzuwandeln. Zur Umrechnung der Rohwerte in IQ-Werte wurde die dem jeweiligen Alter des Teilnehmers entsprechende Altersnorm (Weiß, 2006; S. 137 ff.) herangezogen. Weiß (2006) rät davon ab, IQ-Werte für den Gesamttest zu berechnen, falls sich zwischen dem IQ-Wert für den 1. Testteil und dem IQ-Wert für den 2. Testteil eine Differenz von

12 IQ-Punkten ergibt. Da dies bei keinem der Teilnehmer eintritt, können für alle 62 Teilnehmer IQ-Werte für den Gesamtttest ermittelt werden. Für die untersuchte Stichprobe ($N = 62$) resultierte ein mittlerer IQ-Wert von 105, welcher im Vergleich zum Mittelwert der standardnormierten IQ-Skala um 5 IQ-Punkte nach oben abweicht (vgl. Lienert & Raatz, 1998). Weiß (2006, S. 97) gibt aufbauend auf Studien zur Normierung des CFT-20-R als durchschnittlichen IQ-Wert für Realschüler der 9. Klasse 102 an. Als Richtwert „für erfolgreiche Realschüler“ (Weiß, 2006, S. 96) bezeichnet er IQ-Werte zwischen 100 und 110. Der niedrigste IQ-Wert, welcher in dieser Studie aufgetreten ist, beläuft sich auf 73, während der höchste IQ-Wert 150 beträgt. 16 Teilnehmern (25.81 % der Stichprobe) kann ein IQ-Wert ≥ 115 zugeschrieben werden und 5 Teilnehmer (8.06 %) erhalten aufgrund ihrer Leistung im CFT-20-R sogar einen IQ-Wert ≥ 130 . Auf der anderen Seite fällt der IQ-Wert von drei Teilnehmern (4.84 % der Stichprobe) ≤ 85 aus. Die Majorität der Stichprobe (69.35 %) befindet sich in Bezug auf den IQ-Wert in einem Bereich zwischen 86 und 114. Aufgrund dieses Verteilungsmusters kann die Distribution der IQ-Werte als annähernd normal verteilt bezeichnet werden⁴⁴.

5.3.2 Kristalline Intelligenz

In dem WS erreicht die Stichprobe von 62 Teilnehmern einen durchschnittlichen Rohwert von 25.68 Punkten, so dass im Schnitt mehr als 83 % der 30 Items im WS richtig gelöst wurden. Dieser Mittelwert entspricht dem von Weiß (2007; S. 34) angegebenen durchschnittlichen Rohwert für Schüler des neunten Jahrgangs an Realschulen. Ein Teilnehmer hat im Rahmen der vorliegenden Studie sogar alle 30 Aufgaben (maximal erreichter Rohwert) richtig gelöst, während ein anderer Teilnehmer nur 20 Aufgaben (minimal erreichter Rohwert) erfolgreich bearbeiten konnte. Die Streuung der Rohwerte beläuft sich auf einen Wert von 2.35 und fällt geringer aus als eine von Weiß (2007) an einer vergleichbaren Stichprobe errechnete Standardabweichung. Wie bei dem CFT-R-20 gestattet der WS die Transformation der Rohwerte in IQ-Werte. Für diesen Zweck wird auf eine Altersnormtabelle von Weiß (2007; S. 65) zurückgegriffen. Es ergibt

⁴⁴ Die Annahme einer Normalverteilung wird untermauert durch das Ergebnis eines K-S-Tests ($Z = .96$; $p = .32$).

sich ein durchschnittlicher IQ-Wert von 100.03. Der höchste erreichte IQ-Wert ist 124, während der niedrigste ermittelte IQ-Wert 82 beträgt. 51 Teilnehmer (82.25 % der Stichprobe) erreichen einen IQ-Wert zwischen 85 und 115. Dagegen liegen sieben Teilnehmer (11.29 %) mit ihrem IQ-Wert über der Marke von 115. Ein IQ-Wert von 85 wird von vier Probanden (6.45 %) verfehlt.

5.3.3 Soziale Validität und Akzeptanz

Im Folgenden werden deskriptive Ergebnisse zur Einschätzung der sozialen Validität und Akzeptanz des SJT durch die Teilnehmer dargestellt. Dabei werden die Resultate getrennt für beide eingesetzten Versuchsbedingungen in der Tabelle 5-4 aufgeführt.

Tabelle 5-4: Mittelwerte der Dimensionen sozialer Validität

Dimension der sozialen Validität	Video-Bedingung (n = 31)	Paper-Pencil-Bedingung (n = 31)
<i>Informationsgrad (2 Items)</i>	5.58	5.34
<i>Beeinflussbarkeit (2 Items)</i>	5.98	6.14
<i>Anforderungsbezug (5 Items)</i>	4.73	4.41
Allgemeine Akzeptanz (1 Item)	5.03	4.89

Anmerkung: Kursiv gehalten sind die drei gemessenen Dimensionen der sozialen Validität.

Der Tabelle 5-4 kann entnommen werden, dass alle acht Mittelwerte über dem Skalenmittelwert von 4 (siebenstufige Ratingskalen) liegen. In beiden Bedingungen werden auf der Dimension Beeinflussbarkeit die jeweilig höchsten Mittelwerte erzielt. Dagegen wird in der Video-Bedingung der niedrigste Mittelwert in Bezug auf die Dimension Augenscheinvalidität erreicht, während der niedrigste Mittelwert in der Paper-Pencil-Bedingung der Dimension Allgemeine Akzeptanz zugeordnet werden kann. Ein Vergleich der Mittelwerte unter Zuhilfenahme von t-Tests für unabhängige Stichproben (2-seitige Testung/ $p = .20$) erbringt für eine der drei Dimensionen sozialer Validität, die Dimension *Anforderungsbe-*

zug, einen signifikanten Unterschied zwischen der Video-Bedingung und der Paper-Pencil-Bedingung ($p = .14$; s. Anhang C). Dem SJT wird in der Video-Bedingung ein höherer Anforderungsbezug beigemessen als in der Paper-Pencil-Bedingung. Aufgrund dessen kann die Hypothese II-2 nur teilweise als bestätigt angesehen werden. Der SJT im Videoformat wird von den Teilnehmern in Bezug auf zwei von drei Dimensionen der sozialen Validität nicht besser eingestuft als der SJT im Paper-Pencil-Format.

In Analogie zu zwei der drei Dimensionen sozialer Validität ergibt sich auch hinsichtlich der Abwägung der allgemeinen Akzeptanz kein signifikanter Unterschied zwischen den beiden Versuchsbedingungen. Für die Überprüfung dieses Mittelwertunterschieds wurde ebenfalls ein t-Test für unabhängige Stichproben herangezogen. Teilnehmer aus beiden Bedingungen wurden im Anschluss an die Bearbeitung des SJT dazu aufgefordert, auf einer siebenstufigen Ratingskala (1 = *sehr schlecht* bis 7 = *sehr gut*) einzuschätzen, wie gut sie selbst in dem SJT abgeschnitten haben. Hinsichtlich der Einschätzung der eigenen Leistung im SJT wurde in Rückgriff auf einen weiteren t-Test für unabhängige Stichproben zwischen den beiden Bedingungen, Video (MW = 4.87) und Paper-Pencil (MW = 4.86), kein statistisch bedeutsamer Unterschied ($p > .20$) festgestellt. Ferner ist festzuhalten, dass der Mittelwert in beiden Bedingungen über dem Skalenmittelwert von 4 liegt, so dass die Teilnehmer in beiden Bedingungen ihre Leistungen als eher gut bezeichnet haben.

Zum Abschluss dieses Abschnitts soll analysiert werden, ob sich zwischen der Selbsteinschätzung der eigenen Leistung und den Dimensionen der sozialen Validität sowie Akzeptanz bivariate Korrelationen statistischer Bedeutsamkeit ergeben. Für die Teilnehmer der Videobedingung zeigt sich in diesem Fall eine signifikant positive Korrelation ($r = .37$; $p < .05$) zwischen der Dimension Informationsgrad und der Einschätzung der eigenen Leistung im SJT. Allerdings können für den SJT in der Paper-Pencil-Bedingung keine signifikanten Korrelationen ausfindig gemacht werden.

5.3.4 Zentrale Dimensionen der berufsbezogenen Persönlichkeit

Die Darstellung der deskriptiven Statistiken wird beschlossen durch die Auflistung zentraler Kennwerte für die relevanten Dimensionen des BIP. Diese werden in der Tabelle 5-5 aufgeführt.

Tabelle 5-5: Zentrale Skalenstatistiken des BIP (N = 62)

Dimension	MW	SD	Minimalwert	Maximalwert
Leistungsmotivation	2.97	.45	1.71	4.00
Gestaltungsmotivation	2.66	.48	1.50	3.58
Gewissenhaftigkeit	2.86	.66	1.07	4.50
Flexibilität	2.68	.45	1.86	3.79
Handlungsorientierung	2.59	.76	1.57	4.29
Sensitivität	3.20	.54	1.58	4.58
Kontaktfähigkeit	3.11	.70	1.81	4.38
Soziabilität	2.68	.54	1.27	3.80
Teamorientierung	2.75	.70	1.00	4.31
Durchsetzungsstärke	3.17	.65	1.75	4.75
Emotionale Stabilität	2.72	.73	1.00	4.56
Belastbarkeit	2.72	.73	1.15	4.69
Selbstbewusstsein	2.66	.72	1.50	4.13

In Anbetracht der in der Tabelle 5-5 dargestellten Ergebnisse kann als Fazit erfolgen, dass die von den Teilnehmern dieser Studie erreichten Mittelwerte (MW) in keiner der dreizehn Dimensionen über dem Skalenmittelwert von 3.5 (s. o.: Sechsfach abgestufte Ratingskala) liegen (s. Abschnitt 5.5). Die in dieser Studie empirisch ermittelten Mittelwerte unterschreiten ebenfalls die von Hossiep und Paschen (2003; S. 33) angeführten Mittelwerte. Dafür ergeben sich in dieser Studie im Schnitt geringere Standardabweichungen als bei Hossiep und Paschen (2003). Der höchste Mittelwert (3.17) tritt in dieser Studie bei der Dimension Durchsetzungsstärke auf. Mit einem Wert von 2.59 wird für die Dimension Handlungsorientierung der niedrigste Mittelwert erzielt.

5.4 Inferenzstatistische Analysen und Hypothesentestung

5.4.1 Exploration der faktoriellen Struktur des SJT

Da an dieser Studie mitwirkende Teilnehmer zwei Versuchsbedingungen zugewiesen wurden, müssten zwei separate CFA vorgenommen werden. Aufgrund der zu geringen Stichprobengröße je Versuchsbedingung ($n = 31$) muss davon jedoch Abstand genommen und die faktorenanalytische Untersuchung auf ein exploratives Vorgehen beschränkt werden (vgl. Abschnitt 5.5).

Als Ziel der vorliegenden Arbeit wurde die Entwicklung eines videobasierten SJT ausgegeben. Daher fokussiert die nachfolgend beschriebene faktorenanalytische Untersuchung zunächst auf den unter der Video-Bedingung gewonnenen Datensatz ($n = 31$). Nach der Auffindung einer Struktur, welche psychometrischen Gütekriterien gerecht wird, soll ein Vergleich zwischen den beiden Versuchsbedingungen (Video vs. Paper-Pencil) bezüglich zentraler Eigenschaften dieser Struktur vollzogen werden.

Der faktoriellen Analyse des SJT in der Video-Bedingung vorausgehend sollen zunächst die Erfüllung bestimmter Vorbedingungen überprüft werden. Dazu zählen die Inspektion der Normalverteilung der Daten, die Berechnung der MSA-Koeffizienten sowie die Durchführung von Bartlett's Test auf Spharizität. Alle Angaben beziehen sich anfangs auf die anhand des Ratingsschlüssels erstellten SJT-Scores der Teilnehmer. Die Ergebnisse statistischer Analysen zu den vier auf Rangordnungen basierenden scoring keys werden im Abschnitt 5.5 ausführlich erörtert. Eine Überprüfung der Normalverteilung der 25 Itemscores mittels des Kolmogorov-Smirnov-Tests bringt hervor, dass die Werteverteilung bei keinem der 25 Items signifikant ($p < .05$) von einer Normalverteilung abweicht (s. Anhang C). Weder der MSA-Koeffizient für die vollständige Anti-Korrelationsmatrix noch die MSA-Koeffizienten für die einzelnen Items belegen eine ausreichende Eignung der ermittelten Daten für faktorenanalytische Prozesse (s. Anhang C). Sowohl der Wert für die gesamte Matrix (.23) als auch die Werte für die einzelnen Items (.09 - .50) sind allesamt als unzureichend zu bezeichnen (Bühner, 2006; Backhaus et al., 2008). Die Frage, ob die Interkorrelationen zwischen den 25 Items signifikant

von Null abweichen, kann mit Hilfe des Bartlett's Test auf Spharizität beantwortet werden. Die Prüfgröße Chi-Quadrat erreicht einen sehr signifikanten ($p < .01$) Wert von 410.03 bei 300 Freiheitsgraden. Demzufolge kann geschlussfolgert werden, dass die Höhe der Interkorrelation in statistisch bedeutsamer Weise von Null differiert. Nach Bühner (2006) ist dieser Umstand als förderlich für durchzuführende Faktorenanalysen anzusehen.

Die Anzahl zu extrahierender Faktoren soll vorab im Zuge eines Rückgriffs auf den *Scree-Test* und den *MAP-Test* (s. Kapitel 4 zur Beschreibung der Verfahren) abgeschätzt werden. Eine weitere Absicherung dieser Schätzung soll in Form einer MLF erfolgen. Mit diesen Schritten wird einer Überprüfung der Hypothese II-1 nachgegangen, welche das Postulat einer fünffaktoriellen Struktur aufstellte (s. o.). Eine Einbeziehung der Parallelanalyse in den Methodenkanon kann aufgrund der geringen Stichprobengröße in der Videobedingung ($n = 31$) nicht erfolgen (s. Abschnitt 5.5). Der bei Anwendung des *Scree-Test* resultierende Scree-Plot befindet sich im Anhang C. Um die Übersichtlichkeit zu erhöhen, wird in dem Scree-Plot nur Faktoren mit einem Eigenwert > 1 Platz eingeräumt.

Eine Auseinandersetzung mit dem im Scree-Plot dargestellten Eigenwertverlauf gestattet kein objektivierbares Urteil, wie viele Faktoren extrahiert werden sollten, da ein Knick nicht eindeutig hervortritt. Bei Berücksichtigung des in Kapitel 4 bereits kritisch gewürdigten Kaiser-Guttman-Kriteriums sollten im Übrigen zehn Faktoren extrahiert werden. Im Anschluss an die Durchführung des MAP-Tests, dessen Resultate im Anhang C einsehbar sind, offenbart sich folgendes Bild: Nach Auspartialisierung des ersten Faktors steigt die Höhe der Partialkorrelation wieder an, so dass gemäß der Entscheidungsregeln dieses Verfahrens (vgl. Kapitel 4) ein einziger Faktor extrahiert werden sollte.

Unter Vorgabe der Extraktion eines einzigen Faktors bringt eine Hauptkomponentenanalyse einen Faktor hervor, welcher eine Varianzaufklärung von 16.97 % zu leisten vermag. Auf diesem Faktor substantiell ($> .30$) ladende Items werden in der Tabelle 5-6 vorgestellt.

Tabelle 5-6: Faktorladungen auf dem extrahierten Faktor

Item	Faktorladung
Situation 1	.61
Situation 5	.61
Situation 10	.66
Situation 11	.60
Situation 13	.31
Situation 17	.53
Situation 18	.41
Situation 19	.33
Situation 24	.68
Situation 25	.48

Der Tabelle 5-6 ist zu entnehmen, dass zehn Items aus dem SJT auf dem extrahierten Faktor substantiell laden und die Bildung einer Skala erlauben. Die Höhe der Ladungen erstreckt sich über Werte von .31 (Situation 13) bis .68 (Situation 24). Als mittlere Ladungshöhe kann ein Wert von .52 berichtet werden.

Neben den oben beschriebenen Verfahren zur Extraktion von Faktoren kann auch die Betrachtung einer unabhängig erstellten varimaxrotierten Ladungsmatrix bei Vorgabe fünf zu extrahierender Faktoren (s. Anhang C) kein legitimes Votum für die Annahme einer fünffaktoriellen Struktur liefern. Gegen die Annahme und Weiterverfolgung einer fünffaktoriellen Struktur spricht weiterhin das Ergebnis einer MLF (vgl. Rost, 1996), deren Prüfgröße Chi-Quadrat bei Vorgabe von fünf Faktoren keine statistische Signifikanz erlangt. Währenddessen wird einem einfaktoriellen Modell eine dem Datensatz gerecht werdende Passung zugebilligt (Chi-Quadrat = 348.94; sig. = .002). Damit muss an dieser Stelle die Hypothese II-1 bei Berücksichtigung des Scree-Plots, des MAP-Tests und der MLF eindeutig als falsifiziert betrachtet werden.

Als ein Ziel der aktuellen Studie wurde formuliert, eine Skala aus Items zu bilden, welche den Arbeitsplatz eines Hotelfachmannes in Hinblick auf Interaktionen mit Hotelgästen angemessen repräsentieren. Die in Tabelle 5-6 aufgezählten zehn Items werden unter diesem Gesichtspunkt als nicht

erschöpfend angesehen und sollen im weiteren Verlauf durch weitere Items bereichert werden (s. u.; s. Abschnitt 5.5). Welche Kriterien für die Aufnahme weiterer Items gelten, soll an einer späteren Stelle dargelegt werden. Zunächst wird die Skala, welche aus zehn Items besteht, weiteren Analysen unterzogen. Diese Skala wird im Folgenden mit *SJT_Hotelfach* benannt. Der Gesamtwert für den *SJT_Hotelfach* entspricht dem Mittelwert der Summe der zehn Itemscores ($\sum Situation_{1,...,25}/10$) und kann Werte zwischen 0 (Minimum) und 20 (Maximum) annehmen. Ein Gesamtwert von 20 würde bedeuten, dass ein Proband bei allen zehn Situationen jede der vier Verhaltensalternativen exakt wie die bei der Entwicklung des SJT beteiligten SME eingeschätzt hat (s. Abschnitt 5.5).

Neben der Ermittlung zentraler Kennwerte und Gütekriterien für die Skala *SJT_Hotelfach* und der oben angesprochenen Selektion von Items zur Ergänzung der Skala wird des Weiteren bei jedem nachfolgenden Analyseschritt einem Vergleich des *SJT_Hotelfach* in der Video-Bedingung auf der einen Seite und dem *SJT_Hotelfach* in der Paper-Pencil-Bedingung auf der anderen Seite Platz eingeräumt.

5.4.2 Zentrale Statistiken des *SJT_Hotelfach*

Einen Überblick der Mittelwerte, Standardabweichungen und Schiefen der Skala *SJT_Hotelfach* sowie der zehn in ihr enthaltenen Items erlaubt die Tabelle 5-7. Dabei werden die relevanten Kennwerte getrennt für die beiden Versuchsbedingungen – Video vs. Paper-Pencil – angeführt.

*Tabelle 5-7: Zentrale Statistiken der Skala *SJT_Hotelfach**

	MW (Video/Paper- Pencil)	SD (Video/Paper- Pencil)	Schiefe (Video/Paper- Pencil)
SJT_Hotelfach (Ge-	12.28/13.06	1.50/1.18	-.25/-.14
Situation 1	13.81/12.94	2.95/3.07	-1.41/-.26
Situation 5	10.90/11.71	2.72/3.52	-.25/.08
Situation 10	13.13/13.32	2.40/2.53	.39/-.24
Situation 11	11.81/12.77	2.59/2.55	.35/-.90
Situation 13	11.94/12.71	3.03/2.76	-.07/-.39
Situation 17	12.87/13.03	1.80/2.02	-.16/.19
Situation 18	13.61/14.94	2.80/2.97	-.64/-.80
Situation 19	13.74/14.26	2.54/2.63	.14/-.58
Situation 24	6.94/10.06	3.14/3.01	-.23/.20
Situation 25	14.10/14.90	3.14/2.04	-2.04/.34

Der Tabelle 5-7 ist zu entnehmen, dass sich der höchste Mittelwert (MW) in der Video-Bedingung bei der Situation 1 ergibt, während der niedrigste Mittelwert bei der Situation 24 verzeichnet wird. In der Paper-Pencil-Bedingung scoren die Probanden im Schnitt am höchsten bei der Situation 18, während der niedrigste Durchschnittswert ebenfalls bei der Situation 24 vorliegt. Ein Vergleich der Mittelwerte in Hinblick auf einerseits den Gesamtwert für den SJT_Hotelfach und andererseits die Mittelwerte für die zehn Items der Skala mit Hilfe von t-Tests für unabhängige Stichproben⁴⁵ mündet in dem Ergebnis, dass sich die Mittelwerte bei der Situation 24 in sehr signifikanter ($p < .01$) Weise unterscheiden. Die Teilnehmer in der Paper-Pencil-Bedingung erzielen hier einen Mittelwert, welcher den Mittelwert in der Video-Bedingung in statistisch bedeutsamer Weise übertrifft. Für den Gesamtwert im SJT_Hotelfach lässt sich diese Aussage nicht tätigen. Auch wenn der Mittelwert in der Paper-Pencil-Bedingung augenscheinlich höher ausfällt als in der Video-Bedingung, erlangt dieser Unterschied keine statistische Signifikanz. Dennoch kann der Trend festgestellt werden, dass bei neun der zehn Items unter der Paper-Pencil-Bedingung höhere Durchschnittswerte erreicht werden als unter der Video-Bedingung.

Für beide Bedingungen sollen Extremwerte (Ausreißer) identifiziert werden. Dies geschieht durch die Erstellung von Stängel-Blatt-Diagrammen und Boxplots (s. Anhang C). In der Video-Bedingung werden Gesamtwerte im SJT_Hotelfach < 9.6 ($n = 2$) respektive > 15.6 ($n = 1$) als Extremwerte charakterisiert. Hingegen gelten Gesamtwerte < 10.6 ($n = 2$) beziehungsweise > 16.2 ($n = 1$) in der Video-Bedingung als Ausreißer (s. Abschnitt 5.5). Den Extremwerten soll verstärkt Beachtung geschenkt werden, wenn im späteren Verlauf dieses Kapitels bivariate Korrelationen bei der Analyse der Validität berechnet werden.

Da jede der vier Verhaltensalternativen (A – D) pro Item über Ratings eingestuft wurde, können auch für jede Verhaltensoption mittlere Ratings getrennt nach Versuchsbedingungen berechnet werden. Eine relative Gewichtung der vier Mittelwerte je Situation gestattet in einem nächsten

⁴⁵ Bei zu vollziehenden elf Vergleichen wird das Signifikanzniveau mittels Bonferroni-Korrektur angepasst. Homogene Varianzen können auf Basis der Ergebnisse von Levene-Tests bei allen t-Tests angenommen werden.

Schritt das Aufstellen mittlerer empirischer Rangordnungen⁴⁶ für die zehn Items der Skala SJT_Hotelfach. An die erste Stelle der Rangordnung wird die Verhaltensalternative mit dem höchsten Mittelwert gesetzt, die anderen Rangplätze werden mit den nächst höheren Mittelwerten besetzt. In der Tabelle 5-8 sind sowohl die Mittelwerte für die vier Verhaltensalternativen je Situation als auch die daraus ableitbaren Rangordnungen einsehbar.

⁴⁶ Diese Rangordnungen sind keinesfalls zu verwechseln mit den von Teilnehmern individuell vorgenommenen Rangordnungen.

Tabelle 5-8: Vergleich der Rangordnungen (Video/Paper-Pencil)

	A	B	C	D	Teilnehmer-rang-ordnung	Experten-rang-ordnung
Situation 1	2.48 <i>/2.94</i>	4.45 <i>/3.74</i>	4.48 <i>/3.77</i>	2.39 <i>/2.48</i>	CBAD <i>/CBAD</i>	CADB
Situation 5	2.58 <i>/2.87</i>	2.61 <i>/2.94</i>	2.87 <i>/2.88</i>	2.84 <i>/3.03</i>	CDBA <i>/DBCA</i>	ACBD
Situation 10	2.29 <i>/2.84</i>	3.45 <i>/3.47</i>	2.52 <i>/2.45</i>	4.87 <i>/4.29</i>	DBCA <i>/DBAC</i>	BCAD
Situation 11	4.42 <i>/4.03</i>	4.71 <i>/4.13</i>	2.74 <i>/3.03</i>	2.65 <i>/2.81</i>	BACD <i>/BACD</i>	BCDA
Situation 13	3.74 <i>/3.81</i>	3.00 <i>/3.00</i>	2.61 <i>/2.84</i>	2.58 <i>/3.06</i>	<i>ABCD</i> <i>/ADBC</i>	ABCD
Situation 17	2.61 <i>/3.10</i>	4.90 <i>/4.45</i>	2.64 <i>/2.77</i>	2.62 <i>/2.71</i>	<i>BCDA</i> <i>/BACD</i>	BCDA
Situation 18	2.45 <i>/2.71</i>	3.32 <i>/3.00</i>	4.97 <i>/5.13</i>	2.87 <i>/2.94</i>	CBDA <i>/CBDA</i>	CDAB
Situation 19	3.23 <i>/3.74</i>	2.83 <i>/2.97</i>	2.58 <i>/2.71</i>	5.10 <i>/4.84</i>	DABC <i>/DABC</i>	ABCD
Situation 24	2.61 <i>/3.45</i>	2.35 <i>/3.29</i>	.38 <i>/1.23</i>	1.58 <i>/2.10</i>	ABDC <i>/ABDC</i>	ADCB
Situation 25	4.16 <i>/4.45</i>	2.45 <i>/2.84</i>	4.81 <i>/4.77</i>	2.67 <i>/2.85</i>	CADB <i>/CADB</i>	ABDC

Anmerkung: Kursiv gehalten sind solche Rangordnungen von Teilnehmern, welche mit Expertenrangordnungen übereinstimmen.

Ein Vergleich der Rangordnungen, welcher über die Tabelle 5-8 hergestellt werden kann, bringt hervor, dass sich in der Video-Bedingung bei zwei Items (Situationen 13 und 17) übereinstimmende Rangordnungen zwischen Teilnehmern und Experten ergeben würden. Indessen kann für die Rangordnungen der Teilnehmer aus der Paper-Pencil-Bedingung keine Kongruenz mit Rangordnungen der SME diagnostiziert werden. Weiter-

hin fällt auf, dass sich die Rangordnungen der Teilnehmer aus den beiden Versuchsbedingungen bei sechs der zehn Items (Situationen 1,11,18,19,24 und 25) vollständig entsprechen.

5.4.3 Itemanalyse: Trennschärfe und Schwierigkeit

Die nachfolgenden Analysen der Skala SJT_Hotelfach beginnen mit der Berechnung der part-whole korrigierten Trennschärfen. Die resultierenden Werte r_{it} sind Bestandteil der Tabelle 5-9.

Tabelle 5-9: Trennschärfen der Items aus der Skala SJT_Hotelfach

Skala SJT_Hotelfach (10 Items)	Trennschärfe (r_{it}) (Video/Paper-Pencil)
Situation 1	.50/.00
Situation 5	.44/.18
Situation 10	.48/.28
Situation 11	.47/.27
Situation 13	.35/.13
Situation 17	.42/.54
Situation 18	.21/.17
Situation 19	.22/.03
Situation 24	.64/.29
Situation 25	.36/.36

Anmerkung: Kursiv gehalten sind Trennschärfen < .30.

Nach der Tabelle 5-9 verfehlen in der Video-Bedingung die Trennschärfen von zwei Items (Situationen 18 und 19) die Mindesthöhe von .30 (z. B. Backhaus et al., 2008). Verglichen mit der Paper-Pencil-Bedingung liegen bei der Video-Bedingung jedoch für alle zehn Items höhere Trennschärfen vor. Die Situation 24 erzielt dabei mit einem Wert von .64 die höchste Trennschärfe. Für die zehn Items ergibt sich in der Video-Bedingung eine mittlere Trennschärfe von .41. In der Paper-Pencil-Bedingung überspringen hingegen nur zwei Items (Situationen 17 und 25) den erforderlichen Wert von .30. Die zehn Items aus der Skala SJT_Hotelfach erreichen unter dieser Bedingung im Mittel eine Trennschärfe von .23.

Als weiterer Itemkennwert sollen die Aufgabenschwierigkeiten der zehn Items aus der Skala SJT_Hotelfach ermittelt werden. Dafür wird zunächst

der bereits in Kapitel 4 eingeführte Schwierigkeitsindex nach Fisseni (2004) gebildet. Die resultierenden Indizes sind in der Tabelle 5-10 aufgelistet.

Tabelle 5-10: Schwierigkeitsindizes der Items aus der Skala SJT_Hotelfach

Item	Schwierigkeitsindizes (P) (Video/Paper-Pencil)
Situation 1	.44/.42
Situation 5	.30/.34
Situation 10	.43/.44
Situation 11	.35/.41
Situation 13	.36/.40
Situation 17	.41/.42
Situation 18	.46/.56
Situation 19	.47/.51
Situation 24	.12/.25
Situation 25	.50/.56

Anmerkung: Kursiv gehalten sind Aufgabenschwierigkeiten < .20.

Aus den in Tabelle 5-10 aufgeführten Daten kann abgeleitet werden, dass alle Items bis auf Situation 1 in der Video-Bedingung einen höheren Schwierigkeitsgrad aufweisen als die entsprechenden Items in der Paper-Pencil-Bedingung. Dabei entpuppt sich die Situation 24 in beiden Bedingungen als schwierigstes Item und liegt in der Video-Bedingung außerdem unter dem empfohlenen Richtwert von .20 (z. B. Lienert & Raatz, 1998). In der Video-Bedingung liegt die mittlere Aufgabenschwierigkeit bei .38, in der Paper-Pencil-Bedingung bei .43.

5.4.4 Reliabilitätsanalyse

Die von der Skala SJT_Hotelfach in der Video-Bedingung erreichte interne Konsistenz von .74 (Cronbach's Alpha) übertrifft den geforderten Grenzwert von .70, so dass diese Skala Einzelvergleiche zwischen den Probanden gestattet (vgl. Fisseni, 2004). Indessen fällt die interne Konsistenz für die Skala SJT_Hotelfach in der Paper-Pencil Bedingung mit einem Wert von .50 deutlich niedriger als in der Video-Bedingung aus und verfehlt den kritischen Wert von .70. Ein korrespondierendes Muster offenbart sich bei der Gegenüberstellung der *Homogenitätsindizes*, berechnet über die mittlere Interitemkorrelation: In der Video-Bedingung ergibt sich mit .23 eine höhere mittlere Interitemkorrelation als in der Paper-Pencil-Bedingung (.10). Bei der Berechnung der Präzision von Alpha (Cortina, 1993) ergeben sich für die beiden Versuchsbedingungen jedoch nahezu identische Ergebnisse: Die *Präzision von Alpha* beläuft sich in beiden Bedingungen auf .17 – jeweils berechnet als Wurzel aus der Varianz der Interitemkorrelationen. Nach von Bühner (2006) vorgetragenen Kriterien ist die Skala SJT_Hotelfach in der Videobedingung als homogen zu betrachten, da sie eine angemessen hohe mittlere Interitemkorrelation mit einem Wert $> .20$ bei gleichzeitig niedriger Präzision von Alpha aufweisen kann. Welchen Wert ein einzelnes Item für die Beschaffenheit der internen Konsistenz besitzt, kann der Tabelle 5-11 entnommen werden. Diese Tabelle listet auf, wie Cronbach's Alpha ausfallen würde, falls das jeweilige Item kein Bestandteil der Skala wäre.

Tabelle 5-11: Veränderungen der Reliabilität bei Eliminierung eines Items

Item	Cronbach's Alpha, falls Item entfernt (Video/Paper-Pencil)
Situation 1	.71/.55
Situation 5	.72/.49
Situation 10	.71/.46
Situation 11	.73/.46
Situation 13	.72/.51
Situation 17	.75/.41
Situation 18	.75/.45
Situation 19	.68/.43
Situation 24	.73/.49
Situation 25	.71/.52

Bei einer Wegnahme der Situation 19 sinkt die interne Konsistenz des SJT_Hotelfach in der Videobedingung gemäß der Tabelle 5-11 von .74 auf .68. Dieses Item scheint einen wichtigen Beitrag zur Homogenität der Skala zu liefern. Andererseits würde die Elimination keines Items aus der Skala SJT_Hotelfach in der Video-Bedingung die interne Konsistenz auf einen bedeutend höheren Wert steigen lassen. Die Rolle der Situation 19 in der Video-Bedingung übernimmt für die Paper-Pencil-Bedingung die Situation 17, da ein Fehlen des Items in der Skala die interne Konsistenz auf einen Wert von .41 abrutschen lassen würde. Bemerkenswerterweise würde dasselbe Item (Situation 17) bei Nicht-Aufnahme in die entsprechende Skala in der Video-Bedingung die interne Konsistenz sogar leicht ansteigen lassen. Dieses Phänomen ereignet sich ebenfalls in Zusammenhang mit der Situation 18 – in der Videobedingung führt ein Fehlen zu einer geringfügig höheren internen Konsistenz, in der Paper-Pencil-Bedingung dagegen zu einer niedrigeren internen Konsistenz. Der umgekehrte Fall zeigt sich bei den Situationen 1 und 25. Eine Entfernung der beiden Items lässt die interne Konsistenz in der Videobedingung fallen, aber in der Paper-Pencil-Bedingung anwachsen.

5.4.5 Validitätsanalyse

Dieser Abschnitt rückt die Konstruktvalidität des SJT_Hotelfach in den Vordergrund. Da mit dem CFT-20-R, dem WS sowie mit dem BIP Verfahren zur Messung unterschiedlicher Konstrukte eingesetzt wurden, kann untersucht werden, ob der SJT_Hotelfach substantielle Anteile dieser Merkmale zu messen vermag.

Die Beantwortung der Frage, ob die Ausprägung der allgemeinen kognitiven Leistungsfähigkeit einen bedeutsamen Einfluss auf die Leistung im SJT_Hotelfach besitzt, soll mittels bivariater Produkt-Moment-Korrelationen beantwortet werden. Die Ergebnisse der Korrelationen werden durch die Tabelle 5-12 mitgeteilt.

*Tabelle 5-12: Korrelationen zwischen dem SJT_Hotelfach und
Maßen der Intelligenz*

	CFT-20-R (1. Test- teil)	CFT-20-R (2. Test- teil)	CFT-20-R (Gesamt- wert)	WS (Gesamt- wert)
SJT_Hotelfach (Video- Bedingung)	.18	.04	.26	.05
SJT_Hotelfach (Paper- Pencil-Bedingung)	.02	-.08	.12	.02

Für den SJT-Hotelfach ergeben sich gemäß der Tabelle 5-12 weder in der Video-Bedingung noch in der Paper-Pencil-Bedingung signifikante Korrelationen zu den drei mit dem CFT-20-R erhobenen Testwerten (Rohwerte). Diese Gegebenheiten ändern sich auch dann nicht, wenn aus den Korrelationen zwischen dem SJT_Hotelfach und dem CFT-20-R Extremwerte entfernt werden (vgl. Anhang C). Aufgrund dieser Ergebnisse kann zunächst der Schluss gezogen werden, dass die Leistung im SJT-Hotelfach, ungeachtet der Versuchsbedingung, unabhängig ist von der Ausprägung der fluiden Intelligenz gemessen durch den CFT-20-R. Bivariaten Korrelationen⁴⁷ zwischen den einzelnen Items des SJT-Hotelfach einerseits und dem Score für den ersten Teil des CFT-20-R, dem Score für den zweiten Teil des CFT-20-R und dem Gesamtscore für den CFT-20-R andererseits kann ebenfalls weder in der Video- noch in der Paper-Pencil-Bedingung statistische Signifikanz bescheinigt werden (s. Anhang C).

Dass der SJT_Hotelfach auf statistischer Ebene Unabhängigkeit vom CFT-20-R besitzt, zeigt sich darüber hinaus in den Ergebnissen zweier multipler Regressionen, in denen die drei Testwerte des CFT-20-R als Prädiktoren fungieren und der Gesamtwert des SJT_Hotelfach, einmal in der Video-Bedingung und einmal in der Paper-Pencil-Bedingung, das Kriterium darstellt. Bei Anwendung der schrittweisen Regression nach der

⁴⁷ Bei den in diesem Abschnitt dargestellten Korrelationen wurde das Signifikanzniveau jeweils einer der Anzahl an Einzelvergleichen gerecht werdenden Bonferoni-Korrektur unterzogen.

Rückwärts-Methode (Bühl, 2008) erbringen die drei Prädiktoren in beiden Versuchsbedingungen eine zusammengefasste Varianzaufklärung von 9 % (Video-Bedingung) respektive 6 % (Paper-Pencil-Bedingung) (s. Anhang C). Werden die drei Prädiktoren hinsichtlich ihres individuellen Beitrags zur Varianzaufklärung gewichtet, so offenbart sich der Gesamtwert im CFT-20-R als derjenige Prädiktor, welcher in beiden Bedingungen den größten Anteil aufklären kann.

Die in der Tabelle 5-12 dargestellten Ergebnisse zeigen, dass sich für den SJT_Hotelfach in beiden Versuchsbedingungen keine statistisch bedeutsamen Korrelationen mit dem Rohwert im WS ergeben. Eine Berechnung der bivariaten Korrelationen mit vorhergehender Eliminierung von Extremwerten auf Seiten des SJT_Hotelfach und des WS bringt ebenfalls für beiden Bedingungen Korrelationskoeffizienten nahe Null hervor (vgl. Anhang C). Was für die Gesamtskala SJT_Hotelfach gilt, kann auch für die zehn ihr entnommenen Items konstatiert werden: Es bestehen weder in der Video- noch in der Paper-Pencil-Bedingung statistisch bedeutsame Korrelationen zu der Leistung im WS (vgl. Anhang C).

Soll der Gesamtwert im SJT_Hotelfach mittels linearer Regressionen durch den Rohwert im WS vorhergesagt werden, so erzielt der WS eine Varianzaufklärung von 0 % seitens des Gesamtwertes im SJT_Hotelfach erhoben in der Video-Bedingung. Eine Vorhersage des unter der Paper-Pencil-Bedingung gemessenen Gesamtscores im SJT_Hotelfach führt hingegen zu einer Varianzaufklärung von rund 3 % durch den WS (s. Anhang C). Nichtsdestotrotz untermauern die Ergebnisse der linearen Regressionen die oben postulierte Unabhängigkeit der Leistung im SJT_Hotelfach von der mit Hilfe des WS erhobenen kristallinen Intelligenz.

Wie oben bereits dargestellt wurde, ist neben dem CFT-20-R und dem WS mit dem BIP ein Verfahren zur Messung der persönlichen Eignungsvoraussetzung eingesetzt worden. An dieser Stelle werden nun die Ergebnisse statistischer Analysen zur Aufschlüsselung des Zusammenhangs zwischen dem SJT_Hotelfach sowie dem BIP mitsamt seinen verschiedenen Dimensionen präsentiert. Dabei werden zunächst bivariate Korrelationen zwischen dem SJT_Hotelfach und den Dimensionen des BIP berechnet. Die Ergebnisse werden separat für die vier von Hossiep und Paschen (2003) genannten übergeordneten Eignungsvoraussetzungen – Be-

berufliche Orientierung⁴⁸, Arbeitsverhalten, Psychische Konstitution und Soziale Kompetenzen – vorgestellt. Zu welchen Ergebnissen bivariate Korrelationen zwischen den Dimensionen des BIP auf der einen Seite und dem SJT_Hotelfach auf der anderen Seite geführt haben, ist der Tabelle 5-13 zu entnehmen.

Tabelle 5-13: Bivariate Korrelationen zwischen dem SJT_Hotelfach und den Eignungsvoraussetzungen des BIP

<i>Berufliche Orientierung</i>		
	Video	Paper-Pencil
Gestaltungsmotivation	-.07	-.37*
Leistungsmotivation	-.12	.03
<i>Arbeitsverhalten</i>		
	Video	Paper-Pencil
Handlungsorientierung	.49**	.11
Flexibilität	.39*	-.26
Gewissenhaftigkeit	.13	-.03
<i>Psychische Konstitution</i>		
	Video	Paper-Pencil
Emotionale Stabilität	.09	-.12
Belastbarkeit	.35	-.21
Selbstbewusstsein	.16	-.08
<i>Soziale Kompetenz</i>		
	Video	Paper-Pencil
Sensitivität	.33	.34
Kontaktfähigkeit	-.16	-.02
Soziabilität	.09	-.18
Teamorientierung	.03	.31
Durchsetzungsstärke	-.06	-.18

Anmerkung: * = signifikant ($p < .05$); ** = sehr signifikant ($p < .01$)

⁴⁸ Es sei nochmals erwähnt, dass die Ausprägung der Dimension Führungsmotivation nicht ermittelt wird.

Während sich gemäß der Tabelle 5-13 zwischen dem SJT_Hotelfach in der Video-Bedingung und den Dimensionen Leistungs- beziehungsweise Gestaltungsmotivation keine statistisch bedeutsamen Korrelationen auffinden lassen, ergibt sich für den SJT_Hotelfach in der Paper-Pencil-Bedingung eine signifikant negative Korrelation ($p < .05$) mit der Dimension *Gestaltungsmotivation*. Aus diesem Ergebnis kann der Schluss gezogen werden, dass hohe Testwerte im SJT_Hotelfach (Paper-Pencil-Bedingung) mit einer niedrigen Ausprägung der Gestaltungsmotivation einhergehen. Für die Leistung im SJT_Hotelfach in der Video-Bedingung ist die Gestaltungsmotivation hingegen aufgrund der vorliegenden Korrelation nahe Null als unbedeutend zu erachten. Ein anderes Bild ergibt sich bei den zu der Eignungsvoraussetzung Arbeitsverhalten zu zählenden Dimensionen. Hier weist der SJT_Hotelfach in der Video-Bedingung eine sehr signifikante ($p < .01$) – positive – Korrelation mit der Dimension *Handlungsorientierung* und eine signifikante ($p < .05$) – positive – Korrelation mit der Dimension *Flexibilität* auf. Es ist demnach festzustellen, dass Probanden mit einer stärkeren Ausprägung hinsichtlich Handlungsorientierung und Flexibilität unter der Video-Bedingung signifikant höhere Testscores im SJT_Hotelfach erzielen als Probanden mit einer geringeren Ausprägung in Bezug auf diese beiden Dimensionen. Weiterhin ist zu berichten, dass sich für die Skala SJT_Hotelfach in der Paper-Pencil-Bedingung keine statistisch bedeutsamen Korrelationen mit den drei Dimensionen des Arbeitsverhaltens ermitteln lassen. Zudem ist anzufügen, dass sich für die Dimension Flexibilität je nach Versuchsbedingung Korrelationen unterschiedlicher Valenz ergeben: In der Video-Bedingung tritt wie oben dargestellt eine positive Korrelation zutage, während in der Paper-Pencil-Bedingung eine negative Korrelation (nicht signifikant) feststellbar ist. Für die anderen einbezogenen und ausgewerteten Dimensionen des BIP finden sich weder mit dem SJT_Hotelfach in der Video-Bedingung noch mit dem SJT_Hotelfach in der Paper-Pencil-Bedingung berichtenswerte signifikante Korrelationen.

Hinsichtlich der Ausprägung auf den Dimensionen Sensitivität, Gestaltungsmotivation, Gewissenhaftigkeit, Handlungsorientierung und Selbstbewusstsein können mittels Stängel-Blatt-Diagrammen und Boxplots (vgl. Anhang C) für manche Teilnehmer in der Video-Bedingung Extremwerte festgestellt werden. Eine Entfernung der entsprechenden Datensätze aus

den korrelativen Analysen hat jedoch keine Veränderung der in Tabelle 5-13 aufgeführten Verhältnisse zur Folge. Analog können auch für den innerhalb der Paper-Pencil Bedingung gewonnenen Datensatz *Extremwerte* erkannt werden (vgl. Anhang C). Hier erreichen einzelne Teilnehmer extreme Scores in Bezug auf die Dimensionen Sensitivität, Leistungsmotivation, Gestaltungsmotivation, Handlungsorientierung, Belastbarkeit, Selbstbewusstsein und Emotionale Stabilität. Eine Neuberechnung der Korrelationen mit dem SJT_Hotelfach unter Auslassung der jeweils relevanten Extremwerte führt zu zwei von den in der Tabelle 5-13 dargestellten Werten differierenden Ergebnissen. Nun erlangt die Korrelation zwischen dem SJT_Hotelfach und der Dimension Sensitivität den Status statistischer Signifikanz ($r = .36 / p < .05$), während die Korrelation zwischen Gestaltungsmotivation und dem SJT_Hotelfach nicht mehr als statistisch bedeutsam einzustufen ist ($r = -.32 / p > .05$).

Welche korrelativen Beziehungen resultieren, falls *einfache Minderungskorrekturen* (vgl. Lienert & Raatz, 1998) auf Seiten der Kriterien – das heißt der Dimensionen des BIP – durchgeführt werden, ist aus Gründen der Übersichtlichkeit vollständig im Anhang C einsehbar. Neben den bereits ohne Minderungskorrektur bedeutungsvollen Korrelationen des SJT_Hotelfach im Videoformat zu den Dimensionen Handlungsorientierung und BIP gewinnen nun auch die Korrelationen zu *Belastbarkeit* ($r_{\text{cor.}} = .41$) und *Sensitivität* ($r_{\text{cor.}} = .44$) an Bedeutung. Seitens des SJT_Hotelfach im Paper-Pencil-Format profitieren vor allem die statistischen Beziehungen zu *Sensitivität* ($r_{\text{cor.}} = .43$) und *Teamorientierung* ($r_{\text{cor.}} = .37$) vom Prozess der einfachen Minderungskorrektur. Es lassen sich nicht nur signifikante Korrelationen zwischen der Skala SJT_Hotelfach und den unterschiedlichen Dimensionen des BIP auffinden, sondern es stechen auch statistisch bedeutsame Korrelationen zwischen einzelnen Items aus dem SJT_Hotelfach und verschiedenen Dimensionen des BIP hervor⁴⁹. So können der Situation 1 in der Video-Bedingung zwei signifikante positive Korrelationen mit den Dimensionen Handlungsorientierung respektive Soziale Sensitivität sowie eine signifikante negative Korrelation mit der Dimension Teamorientierung zugestanden werden. Für die Situation

⁴⁹ Bei den anfallenden korrelativen Berechnungen wurde das angelegte Signifikanzniveau abermals einer Bonferoni-Korrektur unterzogen.

13 können ebenfalls zwei signifikante Korrelationen mit Dimensionen des BIP verbucht werden, in diesem Fall mit Flexibilität und Gewissenhaftigkeit. Auch die Situation 18 erzielt eine signifikant positive Korrelation mit der Dimension Gewissenhaftigkeit. Zudem korreliert sie signifikant positiv mit der Dimension Handlungsorientierung. Den Situationen 19 beziehungsweise 25 kann jeweils nur eine signifikant positive Korrelation mit einer einzigen Dimension des BIP attestiert werden. Während die Situation 19 signifikant mit der Dimension Soziale Sensitivität korreliert, gilt dies bei der Situation 25 für die Dimension Handlungsorientierung. Zusammengefasst unterhalten drei Items (Situationen 1, 18 und 25) aus dem SJT_Hotelfach in der Video-Bedingung statistisch bedeutsame korrelative Beziehungen zu der Dimension Handlungsorientierung und jeweils zwei Items zu den Situationen Gewissenhaftigkeit und Soziale Sensitivität. Für einzelne Items aus dem SJT_Hotelfach ergeben sich auch in der Paper-Pencil-Bedingung signifikante Korrelationen mit Dimensionen des BIP. Dabei sind für die Situation 13 drei signifikant negative Korrelationen mit Gestaltungsmotivation, Emotionaler Stabilität und Durchsetzungsstärke zu verzeichnen. Daneben korreliert die Situation 5 signifikant negativ mit Flexibilität, die Situation 10 signifikant negativ mit Durchsetzungsstärke, die Situation 11 signifikant positiv mit Teamorientierung, die Situation 17 signifikant negativ mit Gestaltungsmotivation und zuletzt die Situation 19 signifikant negativ mit Kontaktfähigkeit.

Neben der Aufdeckung korrelativer Beziehungen zwischen der Skala SJT_Hotelfach und den Dimensionen des BIP soll nachvollzogen werden, ob sich die Testscores im SJT_Hotelfach durch die jeweils zu einer Eignungsvoraussetzung des BIP gehörenden Dimensionen vorhersagen lassen. Dazu werden drei *schrittweise Regressionen* (Rückwärts-Methode) angestrengt, in denen die Dimensionen des BIP in ihrer Funktion als Prädiktoren betrachtet werden. Es werden jedoch nur solche Prädiktoren einbezogen, welche bivariate Korrelationen ungleich Null mit dem SJT_Hotelfach aufweisen können. Demnach kann für die Dimensionen der Beruflichen Orientierung keine multiple Regression berechnet werden, da jeweils eine der zwei Dimensionen in beiden Versuchsbedingungen eine Nullkorrelation mit dem SJT_Hotelfach unterhält. Bei der Vorhersage des Scores im SJT_Hotelfach (Video-Bedingung) leisten die drei Dimensionen, welche im BIP dem Arbeitsverhalten zugerechnet werden

(s. Tabelle 5-13), eine Varianzaufklärung von rund 35 % ($R = .59$; $R^2 = .35$). Im Vergleich dazu bieten die Dimensionen Handlungsorientierung und Flexibilität nur eine gemeinsame Varianzaufklärung von 9 % ($R = .30$; $R^2 = .09$) bei der Prognose des Gesamtwertes im SJT_Hotelfach (Paper-Pencil-Bedingung). Dabei werden allein circa 7 % durch die Dimension Flexibilität aufgeklärt. In Bezug auf die Eignungsvoraussetzung Psychische Konstitution erklären zwei Dimensionen, Belastbarkeit und Selbstbewusstsein, gemeinsam einen Varianzanteil von 12 % ($R = .35$; $R^2 = .12$) auf Seiten des Gesamtscores im SJT_Hotelfach unter der Video-Bedingung. Da dieser Wert dem Anteil entspricht, welcher separat durch die Dimension Belastbarkeit erklärbar ist (vgl. Tabelle 5-13), kann die Bedeutung der Dimension Selbstbewusstsein für die multiple Regression als gering eingestuft werden. Für die multiple Regression auf den Gesamtwert im SJT_Hotelfach in der Paper-Pencil Bedingung besitzen die Dimensionen Emotionale Stabilität und Belastbarkeit Relevanz. Gemeinsam klären sie einen Varianzanteil von 14 % auf ($R = .37$; $R^2 = .14$). Zuletzt soll der Blick auf die Eignungsvoraussetzung Soziale Kompetenz gerichtet werden. Die beiden diesem Eignungsbereich entlehnten Dimensionen Sensitivität und Kontaktfähigkeit vollbringen zusammengefasst eine Varianzaufklärung von circa 19 % ($R = .43$; $R^2 = .19$) seitens des Scores für den SJT_Hotelfach in der Video-Bedingung. Der Hauptanteil der Varianz (rund 11 %) wird in diesem Fall durch die Dimension Sensitivität prognostiziert. Für die Vorhersage des Gesamtwertes im SJT_Hotelfach unter der Paper-Pencil-Bedingung können sogar vier Dimensionen – Sensitivität, Teamorientierung, Soziabilität und Durchsetzungsstärke – als Prädiktoren eingesetzt und in die multiple Regression einbezogen werden. Sie leisten insgesamt eine Varianzaufklärung von approximativ 32 % ($R = .57$; $R^2 = .32$).

5.4.6 Zusammenhänge mit demographischen Variablen

Im Anschluss an die Bearbeitung des SJT wurden die Teilnehmer in beiden Versuchsbedingungen gebeten, demographische Angaben abzugeben (s. o.). Ob die Ausprägung oder Beschaffenheit dieser Variablen einen Einfluss auf die Leistung im SJT_Hotelfach auszuüben imstande ist, wird nun exploriert. Aufgrund der geringen Stichprobengrößen in den konträren Versuchsbedingungen kann der Einfluss des Geschlechts, der Mut-

tersprache, der beruflichen Vorerfahrung im Hotelfach und der zukünftigen Berufs- oder Ausbildungsziele jedoch nicht verlässlich abgeschätzt werden (s. Abschnitt 5.5).

Auf Basis zweier bivariater Korrelationen kann festgehalten werden, dass die Leistung im SJT_Hotelfach in beiden Versuchsbedingungen nicht beeinflusst wird durch das *Alter der Probanden*. Auch die *Leistung in drei Schulfächern* – Mathematik, Deutsch, Englisch – besitzt bei Rückgriff auf die Ergebnisse bivariater Korrelationen weder in der Video- noch in der Paper-Pencil-Bedingung einen nicht-zufälligen Einfluss auf die Performanz im SJT_Hotelfach (s. Abschnitt 5.5).

5.4.7 Erweiterung des SJT_Hotelfach

Da die zehn Items, welche die Skala SJT_Hotelfach konstituieren, als nicht ausreichend angesehen werden, um den Alltag im Hotelfach adäquat inhaltlich widerzuspiegeln, sollen der Skala weitere Items aus dem verbleibenden Pool (15 Items) beigelegt werden⁵⁰. Um in die Skala SJT_Hotelfach Aufnahme zu erhalten, sollten die Items zwei Kriterien erfüllen. Zum einen sollten sie wie die Skala SJT_Hotelfach auf statistischer Ebene vollständige Unabhängigkeit von sowohl der kristallinen (WS) als auch der fluiden (CFT-20-R) Intelligenz besitzen. Zum anderen sollten sie statistisch bedeutsame positive Korrelationen mit mindestens drei verschiedenen Dimensionen aus dem BIP erzielen (vgl. Abschnitt 5.5). Mit der zweiten Forderung soll gewährleistet werden, dass die Bedeutung von Items, welche zufällig mit einer Dimension aus dem BIP korrelieren, nicht überschätzt wird.

Ob die Items des SJT das Kriterium der statistischen Unabhängigkeit von Facetten der allgemeinen kognitiven Leistungsfähigkeit erfüllen, wird mittels bivariater Korrelationen überprüft. Es zeigt sich hierbei, dass die Situation 14 negativ mit dem Testwert im 2. Teil des CFT-20-R korreliert und dass die Situation 15 signifikant negativ mit dem Testwert im WS korreliert. Für diese beiden Items ist eine vollständige statistische Unabhängigkeit von der Intelligenz nicht gegeben (s. o.), so dass bereits an dieser Stelle zu vermerken ist, dass sie keine Aufnahme in die Skala

⁵⁰ Für die nachfolgenden statistischen Betrachtungen sind nur Daten von Belang, welche unter der Video-Bedingung generiert wurden (n = 31).

SJT_Hotelfach erhalten können. In der Tabelle 5-14 sind die Einzelkorrelationen zwischen den verbleibenden Items aus dem SJT und den Dimensionen des BIP enthalten.

Tabelle 5-14: Einzelkorrelationen zwischen Items des SJT und Dimensionen des BIP

Dimension des BIP	Positiv korrelierende Items
Sensitivität	21*/22*/23**
Soziabilität	-
Kontaktfähigkeit	22*/23*
Teamorientierung	6**
Durchsetzungsstärke	23*
Leistungsmotivation	3*
Gestaltungsmotivation	-
Gewissenhaftigkeit	2*/8**
Flexibilität	12*/23*
Handlungsorientierung	-
Selbstbewusstsein	21**/22**/23**
Emotionale Stabilität	-
Belastbarkeit	21*

Anmerkung: * = signifikant ($p < .05$); ** = sehr signifikant ($p < .01$); 1-seitige Testung

Auf Grundlage der in der Tabelle 5-14 wiedergegebenen Ergebnisse sollen die Situationen 21, 22 und 23 der Skala SJT_Hotelfach hinzugefügt werden. Diese genannten Items korrelieren jeweils entweder in signifikanter ($p < .05$) oder sehr signifikanter ($p < .01$) Weise positiv mit mindestens drei unterschiedlichen Dimensionen des BIP. Im Falle der Situation 21 sind dies die Dimensionen Sensitivität, Selbstbewusstsein und Belastbarkeit, die Situation 22 korreliert mit den Dimensionen Kontaktfähigkeit, Selbstbewusstsein und der Situation 23 können sogar statistisch bedeutsame Korrelationen mit fünf Dimensionen – Sensitivität, Kontaktfähigkeit, Durchsetzungsstärke, Flexibilität sowie Selbstbewusstsein – nachgewiesen werden.

Im Zuge der Durchführung multipler Korrelationen kann erstens festgestellt werden, dass die Situationen 21, 22 und 23 mit der Dimension Sensitivität zu $R = .63$ korrelieren und gemeinsam rund 39 % Prozent der Varianz seitens dieser Dimension des BIP vorherzusagen vermögen. Eine multiple Korrelation zwischen den Situationen 22 und 23 auf der einen Seite und Kontaktfähigkeit auf der anderen Seiten bringt zweitens einen Korrelationswert R in Höhe von $.53$ hervor ($R^2 = .28$). Schlussendlich können mit den Situationen 21, 22 und 23 rund 53 % der Varianz auf Seiten der Dimension Selbstbewusstsein aufgeklärt werden ($R = .73$; $R^2 = .53$).

Für die um die Situationen 21, 22 und 23 erweiterte Skala SJT_Hotelfach beträgt die interne Konsistenz (Cronbach's Alpha) $.72$. Gegenüber der internen Konsistenz, welche die Skala SJT_Hotelfach (10 Items) erzielt hat (s. o.), bedeutet dies einen nur geringen Abfall von $.02$ Punkten. Die Aufnahme der drei genannten Items führt dazu, dass die Skala SJT_Hotelfach, welche sich nun zusammensetzt aus 13 Items, nicht nur wie bislang sehr signifikant positiv ($p < .01$) mit Handlungsorientierung ($r = .55$) und signifikant positiv ($p < .05$) mit Flexibilität ($r = .43$) korreliert (vgl. Tabelle 5-14), sondern außerdem auch noch sehr signifikant positiv mit der Dimension Sensitivität ($r = .47$) und signifikant positiv mit der Dimension Belastbarkeit ($r = .44$) korreliert. Gewahrt bleibt hingegen die statistische Unabhängigkeit von den eingesetzten Maßen der Intelligenz (s. Anhang C).

5.5 Diskussion

Im Rahmen dieser Diskussion werden zu Beginn methodische Einschränkungen und Makel der dargestellten Studie genannt und erörtert. Daran schließt sich eine zusammenfassende Gegenüberstellung der für beide Versuchsbedingungen (Video vs. Paper-Pencil) hervorgebrachten Ergebnisse an. Weiterhin wird die in dieser Studie zu Tage getretene Evidenz in Bezug auf die Konstruktvalidität des SJT_Hotelfach zusammengefasst.

Für die in diesem Kapitel dargestellte Studie war vormals eine Stichprobengröße von mindestens 100 Personen anvisiert worden (s. o.). Aufgrund krankheitsbedingter Ausfälle und nach Angaben der anwesenden Lehrerinnen unentschuldigtem Fernbleiben erschienen jedoch lediglich 62 Schüler zu dem vereinbarten Termin. Dementsprechend konnten nur jeweils 31 Teilnehmer einer der beiden Versuchsbedingungen (SJT im Videoformat vs. SJT im Paper-Pencil Format) zugewiesen werden. Mit einem n dieser Größenordnung sind laut Hair et al. (2008) vor allem zwei mögliche Nachteile und Einschränkungen verbunden: „[...] (1) *too little statistical power to for the test to realistically identify significant results* or (2) *too easily “overfitting” the data such that the results are statistically good because they fit the sample yet not provide generalizability*” (S. 21). Der in dem letzten Zitat aufgeworfenen Frage nach der Generalisierbarkeit beziehungsweise Repräsentativität von Ergebnissen, welche an kleinen Stichproben gewonnen wurden, widmete sich auch Streiner (2006; S. 259):

The reason that larger values of sample size result in smaller estimates of the standard error is that it is assumed that errors are random and normally distributed, with a mean of zero. Consequently, when there are more people in the groups, there is greater opportunity for the errors to cancel each other out, resulting in less bias in the estimates of any parameter in which we’re interested. When the sample sizes are under, say, 30 in any one group, though, we need to be concerned about this issue.

Den Äußerungen von Streiner (2006) folgend bedingen kleine Stichprobengrößen ein Anwachsen des Standardfehlers und damit einhergehend

breitere, weniger genaue Konfidenzintervalle für beobachtete Werte. Ferner verbietet eine Stichprobengröße von $n = 31$ je Versuchsbedingung wie oben beschrieben die Durchführung bestimmter multivariater Verfahren, zum Beispiel einer konfirmatorischen Faktorenanalyse (vgl. Bühner, 2006) oder einer Parallelanalyse. Für explorative Faktorenanalysen wird ein $N \geq 50$ als absolutes Minimum erachtet (de Winter, Dodou & Deringa, 2009). Bei Beachtung dieser Gegebenheiten muss die Generalisierbarkeit der in diesem Kapitel erzielten Resultate in Frage gestellt werden (vgl. Kapitel 7).

Im Umgang mit den vier einer Situation zugeordneten Verhaltensalternativen waren die Teilnehmer auf zweierlei Art und Weise gefordert. Von den Teilnehmern wurden einerseits Einschätzungen eingeholt, wie wahrscheinlich sie jede der vier Verhaltensalternativen in einer realen Situation ausführen würden. Diese Einschätzungen wurden über siebenstufige Ratingskalen erhoben. Andererseits sollten die Teilnehmer die vier Verhaltensalternativen beginnend mit derjenigen Verhaltensalternative, welche sie in einer realen Situation am wahrscheinlichsten zeigen würden, in Gestalt einer Rangordnung gruppieren.

Das Scoring der als Ratings vorliegenden Teilnehmerantworten erforderte einen vorausgehenden Schritt. Zunächst mussten die von SME festgelegten Rangordnungen (vgl. Kapitel 3) zur Einstufung der situativen Effektivität von vier Verhaltensalternativen pro Situation auf das Ratingformat übersetzt werden. Dem Rangplatz 1 wurde hierbei auf einer 7-stufigen Ratingskala stets die 7 (*sehr wahrscheinlich*) zugeordnet, dem Rangplatz 2 die 5, dem Rangplatz 3 die 3 und zuletzt dem vierten Rangplatz bei allen 25 Situationen eine 1 (*sehr unwahrscheinlich*). Für die Verankerung des ersten und letzten Rangplatzes auf den beiden Endpunkten der Ratingskala spricht die Entstehungsgeschichte des SJT. In den Expertenworkshops mit SME wurden jene dazu angehalten, an die erste Stelle einer Rangordnung ausschließlich diejenige Verhaltensalternative zu setzen, die sie eindeutig als effektivste benennen würden, während die letzte Stelle der Rangordnung von der Verhaltensalternative eingenommen werden sollte, welche nach ihrer Meinung ohne Zweifel am ineffektivsten wäre (vgl. Kapitel 3). Dass die mittleren Rangplätze bei jeder Situation den Skalenpunkten 3 respektive 5 gleichgesetzt wurden, ist dagegen nicht ohne Einwände zu legitimieren. Bereits bei der Festlegung der

Rangordnungen ergaben sich zwischen den SME Dispute bezüglich der mittleren (2 und 3) Rangplätze, da diese Verhaltensalternativen widerspiegeln, welche weder besonders gut noch besonders schlecht sind. Nach Meinung des Autors hätte an dieser Stelle je Situation erörtert werden müssen, welcher Rangplatz welchem Skalenpunkt auf der Ratingskala zugeordnet wird. Beispielsweise hätte der Rangplatz 2 auf diesem Wege vielleicht in der einen Situation nur ein Rating von 4 erhalten, in der nächsten Situation aber sogar eine 6. Zu der beschriebenen differenzierten Gestaltung des Ratingschlüssels hätten vorab in den Expertenworkshops Ratings von den SME eingeholt werden müssen. Dabei hätte dies entweder durch die Herbeiführung eines Gruppenkonsenses oder durch eine separate Einschätzung der SME mit anschließender Mittelwertbildung vollzogen werden können. Im Rückbezug auf diese methodischen Einwände gegenüber dem Ratingschlüssel wurde der Weg, die Teilnehmer wie die SME Rangordnungen der vier Verhaltensalternativen je Situation erstellen zu lassen, von dem Autor favorisiert. Für den Vergleich der von Teilnehmern erstellten Rangordnungen mit den von SME konsensuell festgelegten Rangordnungen und die daraus resultierende Bewertung ist im Gegensatz zum Ratingschlüssel kein intermittierender Transformationsschritt vonnöten. Gegen eine Bevorzugung der über Rangordnungen ermittelten SJT-Scores gegenüber dem Ratingschlüssel sprachen verschiedene Gründe, welche nun im Einzelnen in Bezug auf die vier erprobten scoring keys für Rangordnungen (s. o.) aufgezeigt werden sollen⁵¹.

Für ein Heranziehen des Komplet-Schlüssels (s. o.) bei der Punktevergabe sprach zunächst eine ökonomische Handhabbarkeit. Allerdings gestattete ein über diesen scoring key generierter Datensatz⁵² aufgrund unzureichender Varianzen keine faktorenanalytischen Untersuchungen, so dass von einer weiteren Beschäftigung mit diesem scoring key abzuweichen war. Dieser Befund entspricht im Übrigen den Aussagen, welche im Zusammenhang mit der Pick best – pick worst Methode (vgl. Kapitel 4) getätigt wurden.

⁵¹ Die Ergebnisse zu den vier Schlüsseln basieren auf Daten aus der Video-Bedingung.

⁵² Der Begriff Datensatz bezieht sich im Zusammenhang mit den verschiedenen scoring keys auf die zusammengefassten Testwerte der Teilnehmer im SJT (n = 31).

Eine explorative Faktorenanalyse⁵³ des mit Hilfe des Motowidlo-Schlüssels (s. o.) erstellten Datensatzes führte zum Hervortreten zweier dominanter Hauptskalen. Die erste Skala (7 Items) erreicht eine interne Konsistenz von $\alpha = .67$. Sie korreliert außerdem signifikant positiv mit dem Testwert im 1. Teil des CFT-20-R ($r = .37$; $p < .05$) und signifikant positiv mit der Dimension Handlungsorientierung ($r = .40$; $p < .05$) aus dem BIP. Eine zweite hervorstechende Skala setzt sich zusammen aus 5 Items und verfügt über eine interne Konsistenz von $\alpha = .60$. Für diesen Hauptfaktor lassen sich weder zu eingesetzten Maßen der Intelligenz noch zu den erhobenen Dimensionen des BIP Korrelationen mit dem Status statistischer Bedeutsamkeit auffinden. Verglichen mit der Skala SJT_Hotelfach weisen die beiden Faktoren niedrigere Reliabilitäten und nur eine einzige signifikante Korrelation mit einer Dimension aus dem BIP auf. Auffällig ist zudem, dass der erste Faktor zum Teil durch die Ausprägung der fluiden Intelligenz beeinflusst wird.

Neben dem Komplett-Schlüssel und dem Motowidlo-Schlüssel wurde auch der Einzelvergleich-Schlüssel (s. o.) bei der Punktevergabe für die Rangordnungen erprobt. Wird ein über diese Methode erstellter Datensatz Faktorenanalysen unterzogen, zeichnet sich eine vierfaktorielle Struktur ab. Der an erster Stelle extrahierte Faktor umfasst 6 Items. Die interne Konsistenz dieser Skala beträgt $.62$. Signifikante Korrelationen mit dem CFT-20-R oder dem WS ergeben sich für diesen Faktor nicht – als Vorgriff sei erlaubt zu sagen, dass auch die drei anderen extrahierten Faktoren nicht in statistisch bedeutsamer Weise mit diesen Maßen der allgemeinen kognitiven Leistungsfähigkeit korrelieren. Stattdessen ergeben sich für den ersten Faktor vielfältige Korrelationen mit unterschiedlichen Dimensionen des BIP: Sehr signifikant positive Korrelationen ($p < .01$) ergeben sich mit den Dimensionen Handlungsorientierung ($r = .56$) und Belastbarkeit ($r = .50$). Weitere signifikant positive Korrelationen ($p < .05$) können festgestellt werden mit den Dimensionen Selbstbewusstsein ($r = .41$), Gewissenhaftigkeit ($r = .38$), Emotionale Stabilität ($r = .38$) und Kontaktfähigkeit ($r = .36$). Auf dem an zweiter Stelle extrahierten Faktor laden drei Items substantiell. Als Reliabilitätswert ergibt sich ein Cronbach's

⁵³ Die Schritte der EFA entsprachen jeweils den in Abschnitt 5.4.1 beschriebenen Prozeduren.

Alpha von .58. Berichtenswerte Korrelationen mit Dimensionen des BIP ergeben sich für diesen Faktor nicht. Dasselbe gilt für den dritten Faktor, welcher sechs Items auf sich vereint. Diesem Faktor kann eine Reliabilität von $\alpha = .49$ zugeschrieben werden. Der vierte und letzte Faktor (3 Items) erzielt mit einem α in Höhe von .28 die mit Abstand niedrigste Reliabilität. Auch für diesen Faktor sind keine signifikanten Korrelationen mit Dimensionen des BIP zu berichten. Die vier soeben charakterisierten Faktoren verfügen durchgehend über zum Teil erheblich niedrigere interne Konsistenzen als die Skala SJT_Hotelfach. Dem ersten Faktor, welcher sich durch seine umfangreichen statistischen Beziehungen zu Dimensionen des BIP auszeichnet, müssen im Vergleich zur Skala SJT_Hotelfach nicht nur eine niedrigere Reliabilität, sondern auch eine geringere Anzahl substantiell ladender Items zugewiesen werden.

Um die Auseinandersetzung mit den vier scoring keys für Rangordnungen abzuschließen, wird nun dargestellt, welche faktorielle Struktur sich bei Berücksichtigung des Distanz-Schlüssels manifestiert hätte. Dabei ist zunächst anzuführen, dass eine Anwendung dieses scoring keys in der Praxis mit dem höchsten Aufwand und der höchsten Fehlerquote verbunden sein dürfte. Faktorenanalytische Prozeduren legen die Annahme einer zweifaktoriellen Struktur nahe. Dem ersten Faktor lassen sich fünf Items zuweisen. Die durch diese Items gebildete Skala weist eine Reliabilität in Höhe von $\alpha = .53$ auf. Für diesen Faktor ergeben sich zwar keine signifikanten Korrelationen mit dem CFT-20-R oder dem WS, dafür sind aber zwei sehr signifikant positive ($p < .01$) Korrelationen mit den Dimensionen Handlungsorientierung und Gewissenhaftigkeit aus dem BIP zu nennen. An der Zusammensetzung des zweiten Faktors sind fünf Items beteiligt. Es ergibt sich eine Reliabilität in Höhe von .58 (Cronbach's Alpha). Dieser Faktor erreicht weder mit den beiden eingesetzten Maßen der Intelligenz noch mit den verschiedenen Persönlichkeitsdimensionen des BIP statistisch bedeutsame Korrelationen. Auch diesen beiden Faktoren ist jeweils eine im Vergleich zur Skala SJT_Hotelfach niedrigere interne Konsistenz zu bescheinigen. Zudem erfassen die beiden Faktoren mit insgesamt zehn Items keine höhere Itemzahl als die Skala SJT_Hotelfach.

Zusammengefasst lässt sich eine Fokussierung auf die Skala SJT_Hotelfach vor allem durch die für SJTs ungewöhnlich hohe interne Konsistenz legitimieren. Ferner sprechen die vorgefundenen Validitäten –

insbesondere im Zusammenhang mit dem BIP – und die höhere Anzahl konstituierender Items für diesen Schritt. Dass der Einsatz unterschiedlicher Bewertungsschlüssel wie aufgezeigt gänzlich unterschiedliche Faktorenstrukturen, Reliabilitäten und auch Validitäten herbeiführen kann, zeigte sich in vergleichbaren Studien, welche ebenfalls die Nützlichkeit verschiedener scoring keys erforscht haben (z. B. Weekley & Jones, 1997). In Bezug auf eine spätere Anwendung des Ratingschlüssels in der Praxis ist diese Art des scorings als aufwändiger zu betrachten als zum Beispiel der in Kapitel 4 verwandte scoring key nach Motowidlo et al. (1990).

Die vorliegende Arbeit setzt sich die Entwicklung eines videobasierten SJT zum Ziel. Innerhalb der in diesem Kapitel nachvollzogenen Studie wurde der SJT jedoch in zwei verschiedenen Formaten durchgeführt und kontrastiert: Einem Videoformat und einem Paper-Pencil-Format (s. o.). Ob sich je nach Versuchsbedingung unterschiedliche Reliabilitäten und Validitäten ergeben haben, wird nun zusammengefasst dargestellt⁵⁴. Zunächst ist im Zuge des Vergleichs der beiden unterschiedlichen Formate respektive Versuchsbedingungen festzuhalten, dass die Leistung im SJT_Hotelfach (10 Items) – berechnet anhand des Ratingschlüssels – statistisch unabhängig ist von der Art der Situationspräsentation (Video vs. Paper-Pencil). Für die einzelnen Items lässt sich dies nicht durchgängig feststellen, da die Teilnehmer, welche den SJT im Paper-Pencil-Format bearbeitet haben, in der Situation 24 signifikant höhere Testwerte zustande bringen als die Teilnehmer in der Video-Bedingung. Bei allen anderen Items aus der Skala SJT_Hotelfach ergeben sich jedoch wie beim Gesamtwert keine statistisch gewichtigen Differenzen bezüglich der Itemscores. Dies spricht dafür, dass die oben beschriebene Umsetzung der bis dato ausschließlich im Videoformat vorliegenden Situationen in Textform gelungen ist und dass den Teilnehmern in beiden Bedingungen der erforderliche Informationsgehalt zur Bearbeitung der Situationen geliefert werden konnte. Des Weiteren ergeben sich für die Items aus der Skala SJT_Hotelfach in der Video-Bedingung stets höhere Trennschärfen als in der Paper-Pencil-Bedingung. In der Paper-Pencil-Bedingung werden von

⁵⁴ Eine Diskussion der sozialen Validität - verglichen für die beiden Versuchsbedingungen - erfolgt weiter unten.

acht Items sogar nur solche Trennschärfen erreicht, welche als weit unzureichend ($< .30$) zu erachten sind. Hinsichtlich der Aufgabenschwierigkeit lässt sich konstatieren, dass die Items unter der Paper-Pencil-Bedingung über eine leicht niedrigere mittlere Aufgabenschwierigkeit verfügen als in der Video-Bedingung. Als schwierigstes Item entpuppt sich in beiden Bedingungen übereinstimmend die Situation 24, wobei die für dieses Item resultierende Aufgabenschwierigkeit nur in der Video-Bedingung als zu hoch ($P < .20$) einzustufen ist. Hinsichtlich der Reliabilität – angegeben in Form der internen Konsistenz – resultieren für den SJT_Hotelfach in den beiden Versuchsbedingungen Werte deutlich unterschiedlicher Höhe. Während die Reliabilität in der Paper-Pencil-Bedingung mit $.50$ nicht zufrieden stellend ausfällt, ist die Reliabilität in der Video-Bedingung mit einem Wert von $.74$ als zufrieden stellend zu betrachten. Weder für das Alter der Teilnehmer, die Schulnoten in den Fächern Deutsch, Mathematik und Englisch noch für zwei Facetten der allgemeinen kognitiven Leistungsfähigkeit – fluide und kristalline Intelligenz – ist in beiden SJT-Formaten ein signifikanter Zusammenhang mit der Performanz im SJT_Hotelfach zu berichten.

Da sich die faktorenanalytische Untersuchung, welche die Grundlage darstellte für alle nachfolgenden statistischen Berechnungen (s. o.), vorrangig an dem SJT im Video-Format orientierte, soll nun kurz dargestellt werden, welche faktorenanalytische Struktur sich beim SJT im Paper-Pencil-Format ergäbe⁵⁵. Im Gegensatz zum SJT im Video-Format hätte eine MLF in diesem Fall die theoretische Annahme einer fünffaktoriellen Struktur empirisch legitimiert ($\text{Chi}^2 = 217.85$; $p = .04$). Die inhaltlich-qualitative Zuordnung der 25 SJT-Items zu den fünf Anforderungsdimensionen, welche in Kapitel 3 einsehbar ist, kann jedoch empirisch nicht vorgefunden werden. Die internen Konsistenzen (Cronbach's Alpha) der fünf Faktoren wären wie folgt ausgefallen: Faktor 1 mit $\alpha = .79$, Faktor 2 mit $\alpha = .69$, Faktor 3 mit $\alpha = .64$, Faktor 4 mit $\alpha = .29$ und Faktor 5 mit $\alpha = .63$. Damit fielen die Reliabilitäten für vier der fünf Faktoren zum Teil sowohl deutlich höher (Faktor 1) als auch sehr viel niedriger (Faktor 4) aus als bei der Skala SJT_Hotelfach im Videoformat ($\alpha = .74$). Zusam-

⁵⁵ Als Basis dienen abermals mit Hilfe des Ratingschlüssels berechnete Testwerte im SJT.

mengefasst ist zu sagen, dass sich abhängig vom Format des SJTs unterschiedliche faktorielle Strukturen ergeben.

Wie bei der im vorhergehenden Kapitel 4 beschriebenen Studie konnte eine dem SJT hypothetisch zugrunde gelegte fünffaktorielle Struktur nicht in den Daten aufgefunden werden⁵⁶. Die Ergebnisse eines MAP-Tests und einer Überprüfung in Form einer MLF argumentierten viel mehr übereinstimmend für die weitere Fokussierung auf eine eindimensionale Struktur (s. o.). Ist dem SJT respektive der Skala SJT_Hotelfach aufgrund dessen aber die Fähigkeit abzusprechen, Kundenorientierung zu messen? Die Frage ist auf zweierlei Wegen zu beantworten. Es ist zunächst zu verneinen, dass sich mit dem SJT die in Kapitel 3 eingeführte theoretische Struktur messen lässt (vgl. auch Kapitel 4). In dieser Struktur fungieren Serviceorientierung, Beschwerdemanagement, Durchsetzungsfähigkeit, Handlungsfähigkeit und Gewissenhaftigkeit als Faktoren 1. Ordnung. Diesen Faktoren übergeordnet wird als Faktor 2. Ordnung Kundenorientierung gesetzt (vgl. Abbildung 3-4). Zu bejahen ist aber nach Ansicht des Autors, dass mit dem SJT festzustellen ist, inwieweit Probanden einen von SME festgelegten Maßstab, wie kundenorientiertes Verhalten für sie beschaffen sein sollte (vgl. Kapitel 3), erfüllen. Insofern ist dem SJT respektive der Skala SJT_Hotelfach die Fähigkeit, eine stark organisationsintern definierte, dafür weniger theoretischen Gesichtspunkten entspringende, Kundenorientierung zu messen, zuzusprechen. (vgl. Kanning et al., 2006).

Für den Zusammenhang zwischen dem SJT_Hotelfach und der allgemeinen kognitiven Leistungsfähigkeit konnte vor dem Hintergrund der relevanten Forschung keine eindeutige Hypothese formuliert werden. Es zeigten sich in einigen Studien signifikante Korrelationen geringer bis moderater Höhe zwischen SJTs im Video-Format und Tests zur Messung der allgemeinen kognitiven Leistungsfähigkeit (z. B. Smiderle et al., 1994; Weekley & Jones, 1997), während in anderen Studien (z. B. Kanning, 2005) Korrelationen nahe Null vorgefunden wurden. Die Ergebnisse der in diesem Kapitel besprochenen Studie (s. o.) sprechen für eine statis-

⁵⁶ Fünffaktorielle Strukturen konnten im Übrigen auch nicht solchen empirischen Datensätzen zugrunde gelegt werden, welche der Bewertung von Rangordnungen entsprungen sind.

tische Unabhängigkeit des SJT_Hotelfach von sowohl der fluiden (CFT-20-R) als auch der kristallinen Intelligenz (WS). In einer Meta-Analyse stellten Frei und McDaniel (1998; vgl. Abschnitt 2.6) fest, dass sich für verschiedene Verfahren zur Messung der Kundenorientierung keine statistisch bedeutsamen Zusammenhänge mit Maßen der allgemeinen kognitiven Leistungsfähigkeit nachweisen lassen. Wird der Gesamtwert im SJT_Hotelfach als Ausdruck einer unternehmensintern definierten Kundenorientierung (s. o.) aufgefasst, so deckt sich die in dieser Studie gewonnene Erkenntnis mit dem von Frei und McDaniel (1998) gefundenen Resultat.

Ähnlich wie bei der Relation zur allgemeinen kognitiven Leistungsfähigkeit wurde auch für die Beziehung zwischen SJT und berufsbezogener Persönlichkeit a priori keine spezifische Hypothese formuliert. In der entsprechenden Forschungsliteratur dominieren Studien (z. B. McDaniel & Nguyen, 2001), welche Zusammenhänge zwischen SJTs und globaleren Persönlichkeitsfaktoren erforschen. Dem SJT_Hotelfach im Videoformat (13 Situationen) konnten signifikant positive Korrelationen zu vier Dimensionen des BIP – Handlungsorientierung, Flexibilität, Belastbarkeit und Sensitivität – nachgewiesen werden. Bei einer Vergegenwärtigung der Definitionen dieser Dimensionen (s. Tabelle 5-1) kann diese statistische Beziehung derart interpretiert werden, dass höhere Scores im SJT_Hotelfach einhergehen mit der Fähigkeit, sich rasch unbekanntem Situationen anzupassen (Flexibilität). Weiterhin fördert die Befähigung, Entscheidungen schnell zu treffen und in konkretes Handeln umzusetzen, eine höhere Leistung im SJT_Hotelfach (Handlungsorientierung). Teilnehmer, welche ein feines Gespür für die Interessen und Erwartungen ihres Gegenübers besitzen, schneiden im SJT_Hotelfach besser ab als Teilnehmer, welche nicht über diese Fähigkeit verfügen (Sensitivität). Die statistische bedeutsame Korrelation mit Belastbarkeit ist so zu interpretieren, dass Teilnehmer, denen es leichter fällt, mit starkem Stress und anspruchsvollen Situationen umzugehen, im SJT_Hotelfach höhere Testwerte erreichen als Teilnehmer, denen dies schwerer fällt. Mit der oben beschriebenen signifikanten Korrelation mit der Dimension Sensitivität ergibt sich für die Skala SJT_Hotelfach nur eine einzige systematische Beziehung zu Selbsteinschätzungen der sozialen Kompetenz, welche im BIP insgesamt entlang fünf verschiedener Dimensionen (s. o.) vorgenommen

werden. Dieses Ergebnis deckt sich mit den in Kapitel 4 getätigten Aussagen zur Beziehung zwischen dem ISK-K (Kanning, 2009) und dem SJT_kurz.

Der Einsatz des BIP an einer schulischen Stichprobe ist aufgrund des zum Teil stark berufsbezogenen Inhalts als diskussionswürdig zu erachten. Um dem Alter der Teilnehmer gerecht zu werden, wurde auf eine Erhebung der Dimension Führungsmotivation verzichtet. Zudem wurden den Teilnehmern bei Unklarheiten verstärkt Hilfestellungen seitens des Autors angeboten. Die von den Teilnehmern dieser Studie erreichten Mittelwerte auf den verschiedenen Dimensionen des BIP (s. Tabelle 5-5) lagen durchgängig unter sowohl dem Skalenmittelwert als auch unter von Hossiep und Paschen (2003) empirisch ermittelten Mittelwerten. In einer Studie, in denen ebenfalls Schüler das BIP bearbeitet haben (Krämer, 2009), wurden hingegen durchweg Mittelwerte erreicht, welche über dem Skalenwert von 3.5 lagen.

In Abhängigkeit von der Versuchsbedingung resultieren Korrelationen zwischen dem SJT_Hotelfach (13 Items) und *unterschiedlichen* Dimensionen des BIP (s. o.). Während in der Video-Bedingung signifikante Korrelationen mit vier Dimensionen des BIP hervortreten, liegt in der Paper-Pencil-Bedingung nur eine einzige nicht-zufällige Korrelation vor, nämlich mit der Dimension Sensitivität. Bei der Dimension Flexibilität ergibt sich weiterhin, dass je nach Format des SJT Korrelationen unterschiedlicher Valenz (positiv vs. negativ) vorliegen. Das in dieser Studie vorgefundene Ergebnis, dass sich die Validität eines SJT in Abhängigkeit seines Format verändert, stimmt überein mit einem Befund aus einer Studie von Levens und Sackett (2006), welche ebenfalls einen SJT in einem Video- und einem Paper-Pencil-Format verwandten und auf unterschiedliche Beziehungen zu verschiedenen Merkmalen stießen.

Hinsichtlich der Einschätzung der sozialen Validität des SJT (alle 25 Items) ergaben sich auf den Dimensionen Informationsgrad und Beeinflussbarkeit zwischen den beiden Versuchsbedingungen (Video vs. Paper-Pencil) *keine* signifikanten Unterschiede. Hingegen zeigte sich hinsichtlich der Dimension Anforderungsbezug ein nicht-zufälliger Unterschied zwischen den beiden Bedingungen. Der Anforderungsbezug des SJT wurde in der Video-Bedingung im Schnitt höher eingeschätzt als in der Paper-Pencil-Bedingung (s. Anhang C). Damit konnte die Hypothese II-2 nur

teilweise bestätigt werden. In vergleichbaren Studien (z. B. Richman et al., 2000; Kanning et al., 2006) zeigte sich, dass positivere Einschätzungen der sozialen Validität erst dann systematisch hervortreten, wenn *alle* Bestandteile des SJTs – sowohl die Situationen als auch die Verhaltensalternativen – multimedial und interaktiv präsentiert werden (vgl. Abschnitt 2.3).

Die 31 Teilnehmer der Video-Bedingung erzielten im Schnitt einen Gesamtscore von 12.54 Punkten im SJT_Hotelfach (13 Items). Dieser Wert könnte in späteren konkreten Auswahl-situationen als cut-off Wert dienen, das heißt als derjenige kritische Wert, den Bewerber im SJT_Hotelfach mindestens erreichen müssen, um diesen zu bestehen, falls der Rating-schlüssel als scoring key dient. Dass der mittlere Gesamtscore im SJT_Hotelfach < 20 Punkte (maximal erreichbarer Wert) ausfällt, ist w-möglich vor dem Hintergrund zu sehen, dass mehr als drei Viertel der Stichprobe in der Video-Bedingung zum Testzeitpunkt über keinerlei Erfahrung im Hotelfach verfügten (s. o.). Ob die Länge und auch Art der Berufserfahrung tatsächlich einen statistisch bedeutsamen Einfluss auf die Leistung im SJT_Hotelfach besitzt, wird jedoch erst im Anschluss an die nächste Studie geklärt werden können (s. Abschnitt 5.6). Davon abgesehen erreichten verschiedene Teilnehmer bei einzelnen Items des SJT_Hotelfach durchaus Werte nahe 20. So erlangten beispielhaft sieben Teilnehmer bei der Situation 25 einen Score ≥ 17 .

Die interne Konsistenz der Skala SJT_Hotelfach (Video-Bedingung) beläuft sich auf einen Wert von .72 (Cronbach`s Alpha). Mit einem Reliabilitätswert dieser Größenordnung übertrifft sie sowohl die Reliabilität der Skala SJT_Kurz (.64; s. Kapitel 4) als auch solche Reliabilitäten, welche in ähnlichen Studien (z. B. Dalessio, 1994; Kanning, 2005) berichtet wurden. Bühner (2006) rät dazu, Extremwerte vor der Berechnung von Trennschärfen zu identifizieren, da diese die Höhe der Korrelation zwischen Itemscore und Gesamtscore verzerren können. Bei der vorliegenden Studie hätte ein Entfernen der Extremwerte aus dem Datensatz der Video-Bedingung erstens niedrigere Trennschärfen (.04 – .49) und zweitens eine deutlich geringere interne Konsistenz der Skala SJT_Hotelfach (Cronbach`s Alpha = .60) bedingt. Ein ähnliches Muster zeigte sich bei einem analogen Vorgehen in der Paper-Pencil-Bedingung.

Als Argument für die Einbeziehung dreier Schulnoten in die Demographie wird die realistische Simulation eines Auswahlverfahrens im Hotelfach angeführt. Nach Angaben der Personalchefin des Hotel Münster dienen die Noten in Deutsch, Englisch und Mathematik als ein zentrales Vorauswahlkriterium in ihrem Hause. Gegen die Erhebung eines Notendurchschnitts sprach in den Augen der Personalchefin auch, dass Schüler im 9. oder 10. Schuljahr gemeinhin keinen Notendurchschnitt erhalten. Die in den Augen des Autors vorhandene psychometrische Überlegenheit des Notendurchschnitts wurde bereits an anderer Stelle dieser Arbeit dokumentiert (s. Kapitel 4).

Werden die Items, welche die Skala SJT_Hotelfach bilden, verglichen mit denjenigen Items, welche Bestandteile des SJT_kurz (vgl. Kapitel 4) waren, so kann festgestellt werden, dass sich acht Items (Situationen 5, 10, 13, 17, 18, 19, 22 und 24) in beiden Skalen wieder finden lassen. Allerdings sei daran erinnert, dass sich die sprachliche Gestaltung der Verhaltensalternativen bei der Situation 5 in den beiden Studien unterscheidet (s. o.).

Eine Anzahl von zehn Situationen als Itembasis für den SJT_Hotelfach wurde vom Autor als nicht ausreichend erachtet, um den Hotelalltag beziehungsweise das Beziehungsgeflecht mit Hotelgästen realistisch wiederzugeben. Die Frage, wie viele Situationen mindestens vonnöten sind, um eine angemessene Repräsentativität zu gewährleisten, kann nicht eindeutig beantwortet werden. In vergleichbaren Studien findet sich zumeist eine größere Anzahl an Situationen, z. B. 40 Situationen bei Jones und DeCotiis (1986) oder 65 Situationen bei Smiderle et al. (1994). In den zitierten Studien (vgl. Abschnitt 2.4) wurden aber im Gegensatz zu der vorliegenden Arbeit weder faktorenanalytische Untersuchungen noch Analysen der Trennschärfe oder Aufgabenschwierigkeit, welche der Itemselektion dienen, durchgeführt. Zudem ist die Produktion weiterer Videos mit einem gesteigerten finanziellen, personellen und zeitlichen Aufwand verbunden, welcher für die vorliegende Arbeit nicht zu realisieren gewesen wäre. Damit Items als sinnvolle Ergänzung zum SJT_Hotelfach in Frage kamen, mussten sie zwei Kriterien erfüllen. Erstens sollten sie sich wie die Skala SJT_Hotelfach als statistisch unabhängig von sowohl dem CFT-20-R als auch dem WS erweisen. Zweitens sollten sie mit mindestens drei Dimensionen des BIP positiv korrelieren. Mittels der Forderung, mit min-

destens drei Dimensionen signifikant positiv korrelieren zu müssen, sollte die Bedeutung von Items, die mit nur einer Dimension statistisch bedeutsam korrelieren, nicht überschätzt werden. Den genannten Anforderungen gerecht wurden drei Items (Situationen 21,22 und 23). Ein Hinzufügen dieser drei Items ist angesichts des Zuwachses an signifikant positiven Korrelationen der Skala SJT_Hotelfach mit Dimensionen aus dem BIP legitimiert, da dies einen weiteren Beitrag zu der Aufklärung der Konstruktvalidität des SJT_Hotelfach liefert. In der klassischen Testtheorie wird die Verlängerung von Tests durch das Hinzufügen von Items als „*via regia der Reliabilitätsverbesserung*“ (Lienert & Raatz, 1998; S. 209) angesehen. Nach Yousfi (2005) ist dieser reliabilitätssteigernde Effekt jedoch nicht ohne weiteres, sondern nur dann gegeben, wenn bestimmte strenge Ansprüche an die Items, zum Beispiel Rasch-Homogenität, erfüllt sind. Auch in der vorliegenden Studie führte eine Verlängerung des Tests nicht zu einem Anstieg der Reliabilität, sondern zu einem – wenn auch geringen – Abfall.

5.6 Ausblick

Als Stichprobe für die nachfolgende Studie sollen Teilnehmer mit (langjähriger) Berufserfahrung herangezogen werden, um abzuschätzen zu können, ob die Leistung im SJT_Hotelfach beeinflusst wird durch den Grad an (spezifischer) Berufserfahrung. Weiterhin wird in der nächsten Studie neben dem SJT_Hotelfach (13 Items) mit dem NEO-FFI (Borkenau & Ostendorf, 1993) ein Verfahren eingesetzt, welches Persönlichkeitsdimensionen auf einem globaleren, abstrakteren Niveau misst, als es das in dieser Studie eingesetzte BIP vermag. Aufgrund der oben geäußerten Bedenken zur Stichprobengröße und der damit verbundenen Frage nach der Generalisierbarkeit der Ergebnisse wird die psychometrische Güte der Skala SJT_Hotelfach in der nachfolgenden Studie einer weiteren Prüfung unterzogen.

6. Die vierte Studie

Im Vergleich zur vorhergehenden Studie nehmen an dieser Datenerhebung ausschließlich Teilnehmer mit Berufserfahrung im Hotelfach teil. Neben dem SJT setzt der Autor Verfahren zur Messung sowohl verschiedener Aspekte der Berufserfahrung als auch allgemeiner Persönlichkeitsfaktoren ein.

6.1 Beschreibung der eingesetzten Instrumente und Fragebögen

6.1.1 Der eingesetzte SJT

In dieser Studie kommt der SJT wieder alleinig im Videoformat – die Situationen liegen als Videos vor, die dazugehörigen Verhaltensalternativen im Paper-Pencil-Format – zum Einsatz. Dabei bleiben die den Teilnehmern zur Auswahl gestellten Verhaltensalternativen hinsichtlich ihrer sprachlichen Ausgestaltung im Vergleich zur letzten Studie unverändert. Bei der Auseinandersetzung mit den vier Verhaltensalternativen je Situation werden von den Teilnehmern in Übereinstimmung mit der vorhergehenden Studie Einschätzungen auf Ratingskalen eingeholt⁵⁷. Von der zusätzlichen Festlegung einer Rangordnung wird im Unterschied zur Vorgängerstudie abgesehen (s. Abschnitt 5.5).

Da an der Studie Mitarbeiter desjenigen Unternehmens teilnehmen, mit dessen Hilfe der SJT konstruiert wurde (s. u.), wird aus Gründen der Transparenz und um den Mitarbeitern zu zeigen, was in Zusammenarbeit mit ihrem Hause entstanden ist, der komplette SJT - alle 25 Items – durchgeführt. Die weiter unten beschriebenen statistischen Analysen widmen sich jedoch ausschließlich der Skala SJT_Hotelfach, welche 13 der ursprünglich 25 Items enthält. Zu dieser Skala zählen die Situationen 1, 5, 10, 11, 13, 17, 18, 19, 21, 22, 23, 24 und 25. Es sei daran erinnert, dass die Situationen 21, 22 und 23 dem SJT_Hotelfach nachträglich hinzugefügt wurden, da sie eine berichtenswerte Steigerung der Validität herbeiführten (s. Kapitel 5).

An die Bearbeitung des SJT schloss sich die Einschätzung der eigenen Leistung im SJT an, welche auch in der im vorhergehenden Kapitel be-

⁵⁷ Als scoring key wird dementsprechend der im letzten Kapitel eingeführte Rating-schlüssel eingesetzt.

handelten Studie von den Teilnehmern eingeholt wurde. Zur Einstufung der eigenen Leistung erhalten die Teilnehmer eine siebenstufige Rating-skala (1 = *sehr schlecht*/ 7 = *sehr gut*). Ergänzt wird die Einschätzung der eigenen Leistung durch zwei Items, welche der Autor aus der in Kapitel 5 vorgestellten Skala (Kanning, 2008) zur Messung verschiedener Dimensionen der sozialen Validität entnommen hat. Es handelt sich um folgende Aussagen: „*Ich glaube, dass es mittels des Fragebogens möglich ist, die Leistung zukünftiger Mitarbeiter vorherzusagen.*“ und „*Wenn ich alleine zu entscheiden hätte, würde ich den situativen Test zur Auswahl neuer Mitarbeiter einsetzen*“. Diese Aussagen wurden mit aufgenommen, da einige der Teilnehmer zum Zeitpunkt der Datenerhebung in Führungspositionen tätig waren (s. u.). Ihre Urteile geben die Teilnehmer bei diesen beiden Items anhand von siebenstufigen Ratingskalen (1 = *stimme gar nicht zu*/ 7 = *stimme voll zu*) ab.

6.1.2 Demographische Variablen

In Übereinstimmung mit den beiden in den vorhergehenden Kapiteln beschriebenen Studien werden die Teilnehmer auch an dieser Stelle zunächst nach ihrem Alter, ihrem Geschlecht und ihrer Muttersprache befragt. Der Zusammenstellung der Stichprobe Rechnung tragend (s. u.) knüpfen daran verschiedene Fragen zur Messung der auf Seiten der Teilnehmer vorhandenen Berufserfahrung an. Zur Erfassung der Berufserfahrung kommen fünf verschiedene Fragen zum Einsatz (s. Tabelle 6-1).

Tabelle 6-1: Fragen zur Messung der Berufserfahrung

Frage	Inhalt
1	<i>Seit wie vielen Jahren sind Sie im Hotelfach beschäftigt?</i>
2	Seit wie vielen Jahren sind Sie in ihrer jetzigen Position beschäftigt?
3	<i>Was ist der längste Zeitraum, den Sie für ein und dasselbe Unternehmen gearbeitet haben?</i>
4	<i>Für wie viele verschiedene Unternehmen waren Sie bereits im Bereich des Hotelfachs tätig?</i>

Anmerkung: Kursiv gehalten sind Fragen, welche von Weekley und Jones (1997; 1999; 2005) übernommen wurden. Frage 2 wurde vom Autor entwickelt.

Im englischsprachigen Original entstammen die in der Tabelle 6-1 enthaltenen Fragen 1, 3 und 4 Studien von Weekley und Ployhart (1997;

1999; 2005). Eine Übertragung ins Deutsche wurde vom Autor vorgenommen. In den oben genannten Forschungsarbeiten wurden die drei Fragen auf fünfstufigen Likertskalen beantwortet und nicht wie im Falle der vorliegenden Studie geschehen im offenen Antwortformat. Die Verwendung eines offenen Antwortformats erfordert, dass die Teilnehmer die relevanten Zeitangaben bei den betreffenden Fragen eintragen. Nach der Erhebung der Berufserfahrung wird die Auskunft eingeholt, ob die Teilnehmer zum Zeitpunkt der Datenerhebung Führungsverantwortung tragen oder nicht. Beschlossen wird der demographische Part durch ein Einholen des Ausbildungshintergrundes der Teilnehmer. Hier stehen den Teilnehmern verschiedene Möglichkeiten zur Auswahl: Eine Ausbildung im Hotelfach, eine Ausbildung in einem anderen beruflichen Feld, der Erwerb eines Hochschul- oder Fachhochschulabschlusses.

6.1.3 Das Neo-Fünf-Faktoren-Inventar nach Costa und McCrae

Das NEO-Fünf-Faktoren-Inventar nach Costa und McCrae (NEO-FFI; Borkenau & Ostendorf, 2008) ist als multidimensionales Persönlichkeitsinventar konzipiert, welches die wichtigsten Bereiche individueller Merkmalsunterschiede abdeckt. Es stellt eine deutschsprachige Kurzversion des NEO-PI-R (Costa & McCrae, 1992) dar. Als theoretische Grundlage des NEO-FFI ist ein Fünf-Faktoren-Modell der Persönlichkeit anzusehen, welches in verschiedenen Datensätzen unabhängig voneinander faktorenanalytisch repliziert werden konnte (z. B. Ostendorf, 1990). Seinen Ursprung hat dieses Fünf-Faktoren-Modell wiederum in dem psycholexikalischen Ansatz von Klages (1926) und später Cattell (1943). Hinsichtlich des Einsatzes des NEO-FFI weisen Borkenau und Ostendorf (2008; S. 10) darauf hin, dass das NEO-FFI vor allem dann anzuwenden ist, „[...]wenn weniger an spezifischen Persönlichkeitskonstrukten Interesse besteht als vielmehr an einer groben aber vollständigen Erfassung der Persönlichkeitsunterschiede“. Die durch das NEO-FFI erfassten Merkmalsbereiche der Persönlichkeit werden von Borkenau und Ostendorf (2008) benannt mit Neurotizismus, Extraversion, Offenheit für Erfahrung, Verträglichkeit und Gewissenhaftigkeit. Zur inhaltlichen Anreicherung der fünf Faktoren gibt die Tabelle 6-2 einen Überblick über von Borkenau und Ostendorf (2008; S. 7) mit Hilfe von Adjektiven vorgenommener Kurzbeschreibungen.

Tabelle 6-2: Kurzbeschreibungen der fünf Persönlichkeitsfaktoren

Dimension	Anzahl der Items	Kurzbeschreibung
Neurotizismus	12	Nervös, ängstlich, traurig, unsicher, verlegen, sich sorgend, überfordert durch Stress
Extraversion	12	Gesellig, aktiv, gesprächig, personenorientiert, herzlich, optimistisch, interessiert an Anregungen und Aufregungen
Offenheit für Erfahrung	12	Wissbegierig, kreativ, phantasievoll, unabhängig im Urteil, vielfältig interessiert, interessiert an Abwechslungen
Verträglichkeit	12	Altruistisch, mitfühlend, verständnisvoll, wohlwollend, kooperativ, nachgiebig, vertrauensvoll, harmoniebedürftig
Gewissenhaftigkeit	12	Ordentlich, zuverlässig, hart arbeitend, diszipliniert, pünktlich, penibel, ehrgeizig, systematisch

Wie der Tabelle 6-2 zu entnehmen ist, erfasst das NEO-FFI jeden der fünf Persönlichkeitsfaktoren über zwölf Items. Demnach setzt sich das NEO-FFI insgesamt aus 60 Items zusammen. Bei jedem Item werden die Teilnehmer gebeten zu bewerten, inwieweit die dargebotene Aussage auf sie selbst zutrifft oder nicht. Der Grad der Zustimmung respektive Ablehnung gegenüber einer Aussage wird auf einer fünfstufigen Likertskala eingetragen (1 = *starke Ablehnung*/ 5 = *starke Zustimmung*). Der Skalenpunkt 3 ist einer neutralen Antwortkategorie vorbehalten.

Um Aussagen über die Zuverlässigkeit des NEO-FFI treffen zu können, wurden von Borkenau und Ostendorf (2008) einerseits interne Konsistenzen und andererseits Retest reliabilitäten für die fünf Faktoren berechnet. Die internen Konsistenzen lagen nach Borkenau und Ostendorf (2008; S. 18) in einem Bereich von $\alpha = .72$ (Verträglichkeit) bis $\alpha = .87$ (Neurotizismus). Als Stichprobe diente ein N von 11.724 Personen. Zur Berechnung der Retest reliabilitäten füllten 1730 Personen das NEO-FFI im Abstand von fünf Jahren zweimal aus. Es kamen Retest reliabilitäten zustan-

de, welche einen Wertebereich von .71 (Verträglichkeit) bis .82 (Extraversion) abdeckten. Zur Validierung des NEO-FFI wurden von Borkenau und Ostendorf (2008; S. 25 ff.) sowohl Übereinstimmungen mit Selbsteinschätzungen der Probanden auf 179 Adjektivskalen als auch Übereinstimmungen mit Fremdeinschätzungen durch Bekannte eingeholt. In Bezug auf die Relation zu Selbsteinschätzungen resultierten Korrelationen zwischen .64 (Verträglichkeit) und .80 (Gewissenhaftigkeit). Beim Vergleich mit Fremdeinschätzungen (Zwillingsstudien) traten Korrelationen in einem Wertebereich von .49 (Verträglichkeit) bis .61 (Extraversion) zutage.

Da von Borkenau und Ostendorf (2008) keine Studien mit externen Validierungskriterien durchgeführt wurden, sei an dieser Stelle darauf verwiesen, dass sich eine Zusammenfassung der Erkenntnisse zu Zusammenhängen zwischen den fünf Persönlichkeitsfaktoren und externen Kriterien wie der Arbeitsleistung in Abschnitt 2.5.2 finden lässt.

Kanning (2005) berichtete in einer Studie, in welcher unter anderem ein videobasierter SJT und das NEO-FFI zum Einsatz kamen, zwei signifikant positive Korrelationen zwischen den Faktoren Offenheit für Erfahrungen beziehungsweise Verträglichkeit auf der einen Seite und dem SJT auf der anderen Seite. Aufgrund dessen soll in dieser Studie folgende Hypothese III-1 überprüft werden:

Hypothese III-1

Höhere Ausprägungen auf den Faktoren Offenheit für Erfahrungen und Verträglichkeit aus dem NEO-FFI gehen einher mit höheren Scores im SJT_Hotelfach.

In einer Meta-Analyse von Frei und McDaniel (1998)⁵⁸ wurde unter anderem festgestellt, dass Maße der Kundenorientierung in moderater Höhe mit drei der oben genannten Persönlichkeitsfaktoren – Vertrauenswürdigkeit, Gewissenhaftigkeit und Neurotizismus – korrelieren. Aufgrund dieses vorgefundenen Zusammenhangs wird erwartet, dass sich die Leistung im SJT_Hotelfach – als Ausdruck der Kundenorientierung des Teilneh-

⁵⁸ Es ist anzumerken, dass in die Meta-Analyse von Frei und McDaniel (1998) keine Studien mit dem NEO-FFI eingeflossen sind.

mers – vorhersagen lässt durch die drei Faktoren Verträglichkeit, Gewissenhaftigkeit und Neurotizismus. Dies führt zu folgender Hypothese III-2:

Hypothese III-2

Die Leistung im SJT_Hotelfach kann vorhergesagt werden durch die individuelle Ausprägung der drei Faktoren Verträglichkeit, Gewissenhaftigkeit und Neurotizismus aus dem NEO-FFI.

6.2 Ablauf der Datenerhebung und Beschreibung der Stichprobe

6.2.1 Ablauf der Datenerhebung

Diese Studie bündelt die Informationen aus drei verschiedenen Datenerhebungen, in denen die oben beschriebenen Messinstrumente ihre Anwendung erfuhren. Für die Studie wurden einerseits Beschäftigte des Hotel Münster als Teilnehmer rekrutiert, andererseits wurden auch Angestellte des Hotel Essen, welches zu derselben Hotelkette wie das Hotel Münster zählt, für die Studie gewonnen. Bei der Auswahl der Teilnehmer durch die jeweilige Personalchefin des Hotels sollten ausschließlich solche Mitarbeiter angesprochen werden, die in direktem, am besten täglichem Kontakt zu Hotelgästen stehen. Bei der Erhebung im Hotel Münster wurde weiterhin darauf geachtet, dass nicht solche Personen teilnehmen, welche bereits an der Entwicklung des SJT (vgl. Kapitel 3) partizipiert haben. Zudem sollten, um Überschneidungen mit der Stichprobe aus der zweiten Studie (vgl. Kapitel 4) zu vermeiden, keine Auszubildenden an den Erhebungen teilnehmen. Die Durchführung der Datenerhebungen wurde vom Autor vor Ort an drei Tagen im Februar (Essen) respektive März (Münster) vorgenommen. Als Räume standen jeweils Tagungssäle zur Verfügung, welche die Möglichkeit boten, die Videos im SJT auf eine große Leinwand zu projizieren. Im Hotel Essen durchliefen 30 Personen den SJT und das NEO-FFI, während in Münster 20 Personen teilnahmen. Für die Durchführung des SJT und des NEO-FFI wurden vom Autor insgesamt 75 Minuten angesetzt. Als Durchführungszeit des NEO-FFI übernahm der Autor die von Borkenau und Ostendorf (2008) vorgeschlagene Zeit von zehn Minuten.

Die Durchführung des SJT entsprach der Vorgehensweise in den beiden vorhergehenden Studien und lässt sich an den entsprechenden Stellen in Erinnerung rufen⁵⁹. Als Anweisung für das NEO-FFI griff der Autor auf eine von Borkenau und Ostendorf (2008) vorgegebene schriftlich-standardisierte Instruktion zurück. Nach der Durchführung des NEO-FFI wurde den Teilnehmern Zeit eingeräumt, um Fragen zu stellen bezüglich der Entstehung des SJT und den möglichen zukünftigen Einsatzmöglichkeiten des Tests in der Personalauswahl.

6.2.2 Beschreibung der Stichprobe

An den drei einzelnen Datenerhebungen (s. o.) haben insgesamt 50 im Hotelfach beschäftigte Personen teilgenommen. Innerhalb dieser zusammengefassten Stichprobe belief sich der Altersdurchschnitt auf 29.47 Jahre. Der jüngste Teilnehmer war zum Zeitpunkt der Datenerhebungen 20 Jahre alt und der älteste Teilnehmer 54 Jahre. Weiterhin setzte sich die Stichprobe aus 33 weiblichen (66 %) und 17 männlichen Teilnehmern (34 %) zusammen. Als Muttersprache wurde von 49 Teilnehmern (98 %) Deutsch genannt. Eine einzige Teilnehmerin (2 %) gab Russisch als ihre Muttersprache an. Auf die *Berufserfahrung* der Teilnehmer Bezug nehmend wurde für die Stichprobe eine mittlere allgemeine Berufserfahrung von 9.39 Jahren ermittelt. Die kürzeste allgemeine Berufserfahrung wurde von zwei Teilnehmern mit anderthalb Jahren angegeben, während der erfahrenste Teilnehmer bereits seit 29 Jahren im Hotelfach beschäftigt war. Hinsichtlich der speziellen Berufserfahrung in der aktuellen Position des Teilnehmers ergab sich ein durchschnittlicher Wert von 3.94 Jahren. Zwei Teilnehmer hatten erst seit einem viertel Jahr (Minimum) ihre momentane Position inne, während ein Teilnehmer seine Position bereits seit 26 Jahren (Maximum) besetzte. Bei dem Zeitraum, welche die Teilnehmer für ein und dasselbe Unternehmen im Hotelgewerbe tätig waren, ergab sich eine mittlere Verweildauer von 8.72 Jahren. Fünf Teilnehmer gaben dabei einen Zeitraum von fünf Jahren (Minimum) als längste berufliche Bindung an das Unternehmen an und ein Teilnehmer hielt demselben Unternehmen bereits seit 26 Jahren (Maximum) die Treue. Mit einer Anzahl

⁵⁹ Bedingt durch das Ausbleiben einer formalen Rückmeldung der Testergebnisse blieb das Anlegen eines persönlichen Codes in dieser Studie außen vor.

von 41 Teilnehmern (82 %) war die Majorität der Teilnehmer zum Zeitpunkt der Datenerhebung erst bei einem einzigen Unternehmen angestellt. Acht Teilnehmer (16 %) waren bereits bei zwei Unternehmen und ein Teilnehmer sogar bei drei verschiedenen Unternehmen beschäftigt. Führungsverantwortung hatten 30 Teilnehmer (60 %) inne, während die restlichen 20 Teilnehmer (40 %) keine Führungsaufgaben wahrzunehmen hatten. Auf eine Ausbildung im Hotelfach konnten 33 Teilnehmer (66 %) zurückblicken, sieben Teilnehmer (14 %) gaben eine andere Berufsausbildung als Grundlage an. Zwei Personen (4 %) verfügten über einen Fachhoch- oder Hochschulabschluss. Von acht Personen lagen keine Angaben über den beruflichen Hintergrund vor.

6.3 Deskriptive Statistik

6.3.1 Die Skala SJT_Hotelfach

Zentrale statistische Kennwerte – Mittelwerte, Standardabweichungen und Schiefen –, welche für die Skala SJT_Hotelfach sowie die 13 konstituierenden Items berechnet wurden, werden in der Tabelle 6-3 aufgezeigt.

Tabelle 6-3: Deskriptive Statistiken der Skala SJT_Hotelfach

	MW	SD	Schiefe
SJT_Hotelfach (Gesamtwert)	14.32	.98	-.15
Situation 1	13.92	2.17	-.94
Situation 5	14.86	2.43	-1.09
Situation 10	13.50	2.92	-.32
Situation 11	14.52	2.24	-.62
Situation 13	12.52	3.20	-.43
Situation 17	14.60	2.43	.22
Situation 18	16.08	1.69	-.69
Situation 19	14.86	2.89	-.52
Situation 21	16.76	1.76	-1.12
Situation 22	12.94	2.68	-.38
Situation 23	14.54	2.62	-1.03
Situation 24	10.48	3.68	-.21
Situation 25	16.62	1.87	-1.03

Die mittleren Itemscores (MW) erstrecken sich getreu Tabelle 6-3 über einen Wertebereich von 10.48 (Situation 24) bis 16.76 (Situation 2). Hinsichtlich der Performanz in der Gesamtskala ergibt sich ein mittlerer Score von 14.32 (s. Abschnitt 6.5). Von den Streuungen (SD) der einzelnen Items werden Werte zwischen 1.76 (Situation 21) bis 3.68 (Situation 24) abgedeckt. Als mittlere Streuung der dreizehn Items aus der Skala SJT_Hotelfach kann ein Wert von 2.51 angegeben werden. Die negativen Schiefen, welche die Werteverteilungen bei zwölf Items und der Gesamt-

skala aufweisen, deuten auf linksschiefe Verteilungen hin (vgl. Bortz, 2005).

Ob die Verteilungen der Scores je Item und Skala signifikant von einer Normalverteilung abweichen, wird mittels Kolomogorov-Smirnov Tests überprüft. Diese bringen hervor, dass nur im Falle der Situation 5 eine signifikante Abweichung ($p < .05$) von der Normalverteilung besteht (s. Anhang D).

Auch in dieser Studie (vgl. Kapitel 5) können aufbauend auf den Ratings der Teilnehmer mittlere empirische Rangordnungen für die dreizehn Items der Skala SJT_Hotelfach aufgestellt werden. An die erste Stelle der Rangordnung wird die Verhaltensalternative mit dem höchsten Mittelwert gesetzt, die anderen Rangplätze werden mit den nächst höheren Mittelwerten besetzt. Die Tabelle 6-4 enthält sowohl die Mittelwerte für die vier Verhaltensalternativen je Situation als auch die daraus zu erschließenden Rangordnungen.

Tabelle 6-4: Vergleich der Rangordnungen

	A	B	C	D	Teilnehmer-rangordnung	Experten-rangordnung
Situation 1	2.24	5.24	4.16	2.28	BCAD	CADB
Situation 5	4.48	2.54	2.60	5.24	DACB	ACBD
Situation 10	2.64	3.40	2.50	4.96	DBAC	BCAD
Situation 11	5.32	4.28	2.10	2.12	ABDC	BCDA
Situation 13	3.20	2.34	2.36	4.62	DACB	ABCD
Situation 17	4.14	5.05	2.80	2.62	BACD	BCDA
Situation 18	2.60	5.44	5.60	2.44	CBAD	CDAB
Situation 19	3.84	2.64	2.48	5.90	DABC	ABCD
Situation 21	5.74	5.60	2.56	2.86	ABDC	BDCA
Situation 22	4.78	1.54	1.38	5.24	DABC	ABCD
Situation 23	4.18	2.42	5.26	2.68	CADB	CDBA
Situation 24	3.94	4.46	0.62	1.46	BADC	ADCB
Situation 25	5.38	2.54	5.80	2.90	CADB	ABDC

Ein über die Tabelle 6-4 möglicher Vergleich der mittleren empirischen Rangordnungen mit den vorgegebenen Rangordnungen der SME mündet in der Feststellung, dass sich bei keinem der dreizehn Items eine Übereinstimmung auffinden lässt (s. Abschnitt 6.5).

6.3.2 Fünf globale Persönlichkeitsfaktoren

Für die fünf mit dem NEO-FFI erfassten Persönlichkeitsfaktoren werden Mittelwerte und Standardabweichungen berechnet. Eine Übersicht über die resultierenden Kennwerte gibt die Tabelle 6-5.

Tabelle 6-5: Deskriptive Statistiken der fünf Faktoren aus dem NEO-FFI

Skala	MW	SD
Neurotizismus	26.70	6.50
Extraversion	28.70	5.04
Offenheit für Erfahrungen	20.70	6.26
Verträglichkeit	28.90	4.84
Gewissenhaftigkeit	30.73	6.39

In der Stichprobe von 50 berufserfahrenen Hotelmitarbeitern ergibt sich – wie aus der Tabelle 6-5 abzulesen ist – der höchste Mittelwert in Bezug auf die mit dem NEO-FFI erfassten Persönlichkeitsfaktoren bei dem Faktor Gewissenhaftigkeit. Für den Faktor Offenheit für Erfahrungen ergibt sich der niedrigste Mittelwert. Bei einem Vergleich mit den Mittelwerten, welche Borkenau und Ostendorf (2008; S. 31) an einer bevölkerungsrepräsentativen Quotenstichprobe (N = 871) ermittelt haben, ist festzustellen, dass die in dieser Studie (N = 50) berechneten Mittelwerte für die beiden Faktoren Neurotizismus und Extraversion über den von Borkenau und Ostendorf (2008) berichteten entsprechenden Skalenmittelwerten rangieren. Allerdings liegen die oben genannten Mittelwerte für die drei anderen Faktoren unter den korrespondierenden Werten der beiden deutschen Forscher. In Übereinstimmung mit der vorliegenden Studie weist der Faktor Gewissenhaftigkeit auch bei Borkenau und Ostendorf (2008) den größten Mittelwert auf. Die Streuungen für die fünf Faktoren umgrenzen gemäß Tabelle 6-5 einen Bereich von 4.84 (Verträglichkeit) und 6.26 (Offenheit für Erfahrungen). Bei einer Gegenüberstellung mit den von Borkenau und

Ostendorf (2008; S. 31) genannten Werten zeigt sich, dass die in dieser Studie zu Tage getretenen Streuungen – mit Ausnahme derjenigen des Faktors Gewissenhaftigkeit – geringer ausfallen als die vergleichbaren Streuungen bei dem deutschen Forscherteam.

6.4 Inferenzstatistische Analysen und Hypothesentestung

Im Falle der vorliegenden Studie entfallen vorangesetzte faktorenanalytische Untersuchungen, da sich die Skala SJT_Hotelfach, welche bereits im Rahmen der vorhergehenden Studie ermittelt wurde, abermals im Fokus der Analysen befindet. In der dieses Kapitel beschließenden Diskussion (Abschnitt 6.5) erfolgt im Übrigen ein Vergleich der Skala SJT_Hotelfach bezogen auf die Stichprobe der Realschüler ohne Berufserfahrung ($n = 31$ /Video-Bedingung) und Hotelmitarbeiter mit Berufserfahrung ($N = 50$).

6.4.1 Reliabilitätsanalyse

Die für die Skala SJT_Hotelfach an einer Stichprobe von 50 berufserfahrenen Teilnehmern ermittelte interne Konsistenz verfehlt die von zum Beispiel Kanning (2004) geforderte Mindesthöhe von $\alpha = .70$ mit einem Wert von $.52$ deutlich. Wie sich die Reliabilität der Skala verändern würde, falls einzelne Items ausschieden, gibt die Tabelle 6-6 wieder.

Tabelle 6-6: Veränderungen der Reliabilität bei Entfernung eines Items

Item	Cronbach's Alpha, falls Item entfernt
Situation 1	.49
Situation 5	.43
Situation 10	.49
Situation 11	.52
Situation 13	.50
Situation 17	.51
Situation 18	.55
Situation 19	.50
Situation 21	.49
Situation 22	.55
Situation 23	.51
Situation 24	.40
Situation 25	.47

Die in der Tabelle 6-6 aufgelisteten Ergebnisse sind derartig zu interpretieren, dass zum einen die Situation 24 eine wichtige Bedeutung für die Skala SJT_Hotelfach innehat, da eine Hinwegnahme dieses Items die interne Konsistenz stark abfallen lassen würde. Hingegen bewirke ein Entfernen der Situation 18 ein Anwachsen der Reliabilität von .51 auf .55. Jedoch würde auch eine Reliabilität dieser Höhe verglichen mit dem oben genannten Mindestwert (.70) als unzureichend einzustufen sein. Neben der geringen internen Konsistenz spricht auch eine vorgefundene *mittlere Interitemkorrelation* von .07 gegen eine ausreichende Homogenität der Skala SJT_Hotelfach (vgl. Bühner, 2006).

6.4.2 Itemanalyse: Trennschärfe und Itemschwierigkeit

Bevor den Aufgabenschwierigkeiten Aufmerksamkeit gewidmet wird, erhalten die part-whole korrigierten *Trennschärfen* der dreizehn Items Beachtung. Es sei daran erinnert, dass die Trennschärfe eines Items einen Indikator dafür darstellt, wie gut ein einzelnes Item die Gesamtskala widerspiegelt. Nach Fisseni (2004) stellt die Trennschärfe das wichtigste Kriterium zur Abschätzung der Nützlichkeit eines Items dar. Der Tabelle 6-7 können die errechneten Trennschärfen entnommen werden.

Tabelle 6-7: Trennschärfen der Items aus der Skala SJT_Hotelfach

Item	Trennschärfe (r_{it})
Situation 1	.27
Situation 5	.45
Situation 10	.27
Situation 11	.06
Situation 13	.20
Situation 17	.14
Situation 18	-.13
Situation 19	.17
Situation 21	.23
Situation 22	-.02
Situation 23	.15
Situation 24	.45
Situation 25	.36

Nach den in Tabelle 6-7 dargestellten Ergebnissen ist zu schlussfolgern, dass nur die Trennschärfen von drei Items (Situationen 5, 24 und 25) die empfohlene Mindesthöhe von .30 (z. B. Backhaus et al., 2008) überschreiten. Für zwei Items (Situationen 18 und 22) ergeben sich sogar Trennschärfen negativer Valenz. Dies bedeutet, dass diese beiden Items von Teilnehmern mit niedrigeren Gesamtwerten im SJT_Hotelfach häufiger richtig gelöst werden als von Teilnehmern mit höheren Gesamtwerten in der Skala (Kelava & Moosbrugger, 2007). Als mittlere Trennschärfe der dreizehn Items resultiert ein Wert von $r_{it} = .20$, was unbefriedigend ist.

Zur Berechnung der *Aufgabenschwierigkeiten* wird abermals auf die bereits in den vorhergehenden Studien (vgl. Kapitel 4 und 5) erwähnte Formel von Fisseni (2004) zurückgegriffen. Eine Auflistung der Aufgabenschwierigkeiten P findet in der Tabelle 6-8 statt.

Tabelle 6-8: Aufgabenschwierigkeiten der Items aus dem SJT_Hotelfach

Item	Aufgabenschwierigkeit (P)
Situation 1	.70
Situation 5	.74
Situation 10	.68
Situation 11	.73
Situation 13	.63
Situation 17	.73
Situation 18	.80
Situation 19	.74
Situation 21	.84
Situation 22	.65
Situation 23	.73
Situation 24	.52
Situation 25	.83

Bei Beachtung der Tabelle 6-8 liegen alle Aufgabenschwierigkeiten – ausgenommen diejenigen für die Situationen 21 und 25 – im geforderten Bereich zwischen $.20 < P < .80$. Demgegenüber sind die Situationen 21 und 25 mit Aufgabenschwierigkeiten $P > .80$ als zu leicht zu bezeichnen. Für die Situation 24 wird der verglichen mit den anderen Items höchste

Schwierigkeitsgrad verzeichnet, wobei für dieses Item absolut gesehen noch ein Wert im mittleren Schwierigkeitsbereich um .50 vorliegt. Über die 13 Items der Skala hinweg ergibt sich eine mittlere Aufgabenschwierigkeit von $P = .66$ (s. Abschnitt 6.5).

6.4.3 Validitätsanalyse

Neben der Skala SJT_Hotelfach wurden in dieser Studie verschiedene Items zur Messung der Berufserfahrung (s. o.) sowie mit dem BIP ein multidimensionales Persönlichkeitsinventar eingesetzt. Welche Validitäten die Skala SJT_Hotelfach im Zusammenspiel mit diesen beiden Kriterien zu erreichen vermag, wird mit Hilfe von Produkt-Moment-Korrelationen berechnet. Es wird begonnen mit der Darstellung des Zusammenhangs zwischen Berufserfahrung und SJT_Hotelfach.

Zwischen den vier in Tabelle 6-1 genannten Items zur Messung der *Berufserfahrung* und der Skala SJT_Hotelfach ergeben sich durchgängig nicht-signifikante Korrelationen (s. Anhang D). Darauf aufbauend ist festzuhalten, dass die Leistung in der Skala SJT_Hotelfach – ausgedrückt durch die Höhe des Gesamtscores –unabhängig ist von der allgemeinen und speziellen Berufserfahrung, der längsten Verweildauer bei demselben Unternehmen sowie der Anzahl der Unternehmen, bei denen die Teilnehmer einem Beschäftigungsverhältnis nachgegangen sind. Ob die Tatsache, dass ein Teilnehmer Führungsverantwortung besitzt, einen nicht-zufälligen Einfluss ausübt auf die Höhe seines Scores im SJT_Hotelfach, kann mit einem t-Test für unabhängige Stichproben überprüft werden. Das nicht-signifikante Ergebnis (s. Anhang D) führt zu dem Schluss, dass sich die Leistung von Teilnehmern mit Führungsverantwortung ($MW = 14.33$; $n = 30$) nur zufällig von Teilnehmern ohne Führungsverantwortung ($MW = 14.31$; $n = 20$) unterscheidet.

Eine Berechnung der bivariaten Korrelationen zwischen dem SJT_Hotelfach und den fünf Faktoren des NEO-FFI beinhaltet eine Überprüfung der Hypothese III-1, in welcher die Erwartung aufgestellt wurde, dass die Skala SJT_Hotelfach signifikant positiv mit den Faktoren *Offenheit für Erfahrungen* und *Verträglichkeit* korreliert. In der Tabelle 6-9 finden sich die Ergebnisse der bivariaten Korrelationen.

Tabelle 6-9: Korrelationen zwischen der Skala SJT_Hotelfach und den Hauptskalen des NEO-FFI

	NZ (mit/ohne Ausreißer*)	EV (mit/ohne Ausreißer**)	OFE	VT	GW (mit/ohne Ausreißer***)
SJT_Hotelfach	.04/.07	-.16/-.16	.14	-.02	.09/-.00

Anmerkung: Es wurde jeweils 1-seitig getestet. * n = 49 ** n = 49 *** n = 44 (N = 51); NZ = Neurotizismus; EV = Extraversion; OFE = Offenheit für Erfahrungen; VT = Verträglichkeit; GW = Gewissenhaftigkeit

Die in Tabelle 6-9 aufgeführten Resultate der Korrelationsanalysen gestatten die Feststellung, dass der SJT_Hotelfach keine statistisch bedeutsamen Beziehungen zu den fünf Faktoren des NEO-FFI aufweisen kann. Nachfolgend soll jedoch überprüft werden, ob sich Änderungen in den korrelativen Beziehungen ergeben, wenn *einfache Minderungskorrekturen* (vgl. Lienert & Raatz, 1998; s. Kapitel 4) auf Seiten der fünf Skalen des NEO-FFI durchgeführt werden. Die internen Konsistenzen der fünf Skalen aus dem NEO-FFI fallen wie folgt aus: Neurotizismus ($\alpha = .72$), Extraversion ($\alpha = .63$), Offenheit für Erfahrungen ($\alpha = .47$), Verträglichkeit ($\alpha = .45$) und Gewissenhaftigkeit ($\alpha = .79$). Aufgrund der Nullkorrelation zwischen dem SJT_Hotelfach und dem Faktor Gewissenhaftigkeit wird für diese Persönlichkeitsskala keine Minderungskorrektur unternommen. Nach Durchführung der Minderungskorrektur verbleibt die Höhe der Korrelationen zwischen dem SJT_Hotelfach und den Faktoren Neurotizismus sowie Verträglichkeit fortwährend auf einem niedrigen, statistisch nicht bedeutsamen Niveau. Für die Korrelation zwischen dem SJT_Hotelfach und dem Faktor Extraversion ergibt sich ein Anstieg des Korrelationskoeffizienten von -.16 auf -.20, die Stärke des Zusammenhanges zwischen dem SJT_Hotelfach und dem Faktor Offenheit für Erfahrungen wächst auf einen Wert von .20 an.

Nichtsdestotrotz ist auch bei dieser Sachlage die Hypothese III-1 abzulehnen. Für den Faktor Verträglichkeit ergibt sich erstens im Kontrast zu dem in der Hypothese III-1 vermuteten Zusammenhang eine negative Korrelation (nicht signifikant) mit dem SJT_Hotelfach. Zweitens bedingt

eine korrigierte Korrelation zwischen dem SJT_Hotelfach und dem Faktor Offenheit für Erfahrungen in Höhe von $r = .20$ (nicht signifikant) nur einen geringen Anteil gemeinsamer Varianz von 4 %. Insgesamt ist festzuhalten, dass die Leistung im SJT_Hotelfach als unabhängig zu betrachten ist von der individuellen Ausprägung in Bezug auf Neurotizismus, Extraversion, Offenheit für Erfahrungen, Verträglichkeit und Gewissenhaftigkeit (s. Kapitel 7).

Zur weiteren Überprüfung der Hypothese III-2 wird eine multiple Regression durchgeführt. In der aufgestellten Regressionsgleichung kommt den Faktoren *Verträglichkeit*, *Neurotizismus* und *Gewissenhaftigkeit* die Rolle von Prädiktoren zu. Durch die drei Prädiktoren soll als Kriterium die Leistung im SJT_Hotelfach vorhergesagt werden. Dabei werden die Prädiktoren schrittweise in die Regressionsgleichung aufgenommen. Welchen Aufklärungsbeitrag die einzelnen Prädiktoren sowohl einzeln als auch kombiniert leisten, wird in der Tabelle 6-10 zusammengefasst.

Tabelle 6-10: Schrittweise Regression zur Vorhersage der Leistung im SJT_Hotelfach

Modell	β	R	R ²	ΔR^2
1 Extraversion Gewissenhaftigkeit Neurotizismus	-.40 .18 .10	.21	.05	-
2 Extraversion Gewissenhaftigkeit	-.37 .21 .11	.20	.04	-.01
3 Extraversion	-.30	.16	.02	-.02

Anmerkungen: β = standardisiertes Regressionsgewicht; R = Multiple Korrelation; R² = Determinationskoeffizient; ΔR^2 = Aufklärungsverlust bei Hinzunahme des Prädiktors

Die multiple Korrelation zwischen den drei Faktoren aus dem NEO-FFI und dem SJT_Hotelfach erreicht einen Wert von $R = .21$ (s. Tabelle 6-10). Daraus ergibt sich ein Varianzanteil auf Seiten der Gesamtscores im SJT_Hotelfach, welcher durch die drei Faktoren des NEO-FFI erklärt

werden kann, von rund 5 %. Aufgrund dieses niedrigen Anteils erklärter Varianz muss die Hypothese III-2 als falsifiziert betrachtet werden. Der Testwert im SJT_Hotelfach lässt sich nicht über eine Kombination der drei Faktoren Extraversion, Gewissenhaftigkeit und Neurotizismus vorhersagen. Weiterhin geht aus der Tabelle 6-10 hervor, dass *Extraversion* als derjenige Prädiktor eingestuft werden kann, der für sich den größten Betrag an aufgeklärter Kriteriumsvarianz (2 %) beansprucht. Dass dieser Prädiktor über ein negatives β -Gewicht verfügt, überrascht vor dem Hintergrund der oben berichteten negativen Korrelation zwischen Extraversion und dem SJT_Hotelfach nicht. Indessen kann dem Prädiktor *Neurotizismus* der geringste Beitrag zur Vorhersage der Leistung im SJT_Hotelfach attestiert werden, da dies derjenige Prädiktor ist, welcher an erster Stelle aus der Regressionsgleichung ausscheidet.

6.4.4 Weitere Ergebnisse

An die Durchführung des SJT schlossen sich eine Einschätzung der eigenen Leistung sowie eine Beurteilung des Nutzens eines solchen Tests bei der Personalauswahl an. Da sich beide Aspekte auf den gesamten SJT (25 Situationen) beziehen, werden in diesem Zusammenhang festgestellte Resultate im Rahmen dieses Abschnitts zuerst angeführt. Hier zeigte sich, dass der Leistung im SJT (alle 25 Situationen) keine signifikante Korrelation mit der *Einschätzung der eigenen Leistung* in dem Verfahren nachzuweisen ist.

Hinsichtlich der *wahrgenommenen Nützlichkeit* des SJT bei der Vorhersage der Arbeitsleistungen möglicher neuer Mitarbeiter ergab sich eine mittlere Einschätzung von 4.42, welche den Skalenmittelwert von 4 übertrifft. Dem Einsatz des SJT bei der Personalauswahl stehen 56 % der Teilnehmer ($n = 28$) wohlwollend⁶⁰ gegenüber. Insgesamt resultiert ein über dem Skalenmittelpunkt von 4 liegender Mittelwert von 4.56. Von den Teilnehmern, welche zum Zeitpunkt der Datenerhebung als Führungskräfte tätig waren ($n = 30$), befürworteten sogar 18 Personen, das heißt 60 % der Führungskräfte, eine Verwendung dieses Instruments in der Berufseignungsdiagnostik.

⁶⁰ Als positive Einstellungen werden Ratings ≥ 5 gewertet.

Nachfolgend werden dem Leser ermittelte statistische Beziehungen zwischen bestimmten demographischen Variablen – *Alter* und *Geschlecht* – und der Leistung im SJT_Hotelfach näher gebracht. Bedingt durch das Ausbleiben ausreichend variierender Nennungen kann eine Analyse des Einflusses der Muttersprache und des Ausbildungshintergrundes nicht erfolgen (s. o.).

Die Leistung im SJT_Hotelfach unterliegt in keiner statistisch bedeutsamen Weise dem Einfluss des Alters ($r = -.01$). Ein t-Test für unabhängige Stichproben erbringt zudem den Nachweis, dass sich weibliche ($n = 33$) und männliche ($n = 17$) Teilnehmer nicht signifikant ($p. = .17$) hinsichtlich ihrer Leistung im SJT_Hotelfach unterscheiden.

6.5 Diskussion

Einen Anfang findet die Diskussion durch eine kritische Betrachtung der für die Skala SJT_Hotelfach ermittelten Itemkennwerte und Gütekriterien. Die Aufgabenschwierigkeiten von zehn der dreizehn Items bewegen sich innerhalb eines eng umgrenzten Wertebereiches von $.60 < P < .75$. Nach Fisseni (2004; S.36 ff.) kann dies eine hohe Homogenität des Skala herbeiführen. In dieser Studie bleibt dieser positive Effekt aus, da die interne Konsistenz als Maß der Homogenität der Skala mit $.51$ sehr niedrig ausfällt. Das niedrige Niveau der Trennschärfen (s. Tabelle 6-7) ist partiell zurückzuführen auf die niedrige mittlere Interitemkorrelation (s. o.). Nach Fisseni (2004) ist das Zustandekommen hoher Trennschärfen an das Vorliegen hoher Interitemkorrelationen gebunden.

Hinsichtlich des Zusammenhangs zwischen dem SJT_Hotelfach und Berufserfahrung konnten aufgrund des Mangels an Studien mit videobasierten SJTs a priori keine Erwartungen in Form von Hypothesen ausgedrückt werden (vgl. Abschnitt 2.5.2). Die Leistung im SJT_Hotelfach erwies sich in dieser Studie als statistisch unabhängig von unterschiedlichen Indikatoren der berufsrelevanten Erfahrung im Hotelfach. Bei der Auseinandersetzung mit der Berufserfahrung wurde auf die Berechnung von Extremwerten verzichtet. In Unternehmen ist stets eine Heterogenität bezüglich des Alters und der Berufserfahrung gegeben, so dass ein Ausschluss von statistisch als Ausreißer definierten Werten ein falsches Bild der Realität geliefert hätte. In Bezug auf solche Teilnehmer, welche bislang nur bei einem einzigen dem Hotelgewerbe zuzurechnenden Unternehmen beschäftigt waren, konnte zudem festgestellt werden, dass die längste berufliche Verweildauer bei demselben Unternehmen in der Regel der allgemeinen Berufserfahrung im Hotelfach entsprach.

Zwischen der Skala SJT_Hotelfach und denen mit dem NEO-FFI gemessenen Persönlichkeitsfaktoren wurde kein einziger korrelativer Zusammenhang statistischer Bedeutsamkeit ermittelt. Nichtsdestotrotz ist die vorgefundene negative Korrelation (nicht signifikant) zwischen dem SJT_Hotelfach und Extraversion gesondert hervorzuheben. Eine negative Korrelation ist derart auszulegen, dass Teilnehmer, welche sich selbst als eher introvertiert und verschlossen bezeichnen, tendenziell höhere Testwerte im SJT_Hotelfach vorweisen können als stärker extrovertierte Teil-

nehmer. Nerdinger (2003) sah aber gerade solche Personen prädestiniert für kundenorientiertes Verhalten, welche aktiv auf Kunden zugehen, das Gespräch mit ihnen suchen und Interesse an ihren Wünschen nach außen darstellen. Überschneidungen mit der Kurzdefinition von Extraversion durch Borkenau und Ostendorf (2008; s. Tabelle 6-2) sind offensichtlich.

In der in Kapitel 5 dargestellten Datenerhebung durchliefen 31 Realschüler⁶¹ ohne den Erfahrungshintergrund im Hotelfach, wie ihn die 50 Teilnehmer dieser Studie aufweisen können, den SJT_Hotelfach. Über einen t-Test für unabhängige Stichproben soll ermittelt werden, ob diese beiden Stichproben in Bezug auf ihr Abschneiden im SJT_Hotelfach signifikant differieren. Als Resultat ergibt sich ein sehr signifikanter ($p < .01$) Unterschied zwischen den Hotelmitarbeitern ($MW = 14.32$) und den Realschülern ($MW = 12.54$). Folglich erbringen Personen mit relevanter Berufserfahrung durchschnittlich bessere Leistungen im SJT_Hotelfach als Personen mit sehr geringer oder gänzlich fehlender Erfahrung im Hotelfach. Dass den Teilnehmern, welche in der Hotelbranche beschäftigt sind, der Umgang mit den Situationen des SJT_Hotelfach leichter fiel, zeigt sich ebenfalls in den oben aufgelisteten Aufgabenschwierigkeiten, welche für die dreizehn Items des SJT_Hotelfach festgestellt wurden. In diesem Zusammenhang resultierte in dieser Studie eine mittlere Aufgabenschwierigkeit von .66. Im Vergleich dazu trat in der Studie mit unerfahrenen Teilnehmern eine mittlere Aufgabenschwierigkeit von .38 zutage. Gemeinsam ist den beiden Studien hingegen, dass sich für die Situation 24 jeweils die relativ gesehen höchste Aufgabenschwierigkeit registrieren ließ.

Neben unterschiedlichen Aufgabenschwierigkeiten zeigen sich für die Items der Skala SJT_Hotelfach beziehungsweise die Gesamtskala je nach Stichprobe stark differierende Trennschärfen und Reliabilitäten. Während die Reliabilität des SJT_Hotelfach und die Trennschärfen der Items in der aus Schülern bestehenden Stichprobe zufrieden stellend ausfallen, sind dieselben Kennwerte ermittelt an einer berufserfahrenen Stichprobe als vollkommen unzureichend anzusehen. Die je nach Stichprobe differierende Ausprägung zentraler Itemkennwerte und Gütekriterien weist auf das der klassischen Testtheorie inhärente Problem der Stichprobenabhängig-

⁶¹ Weitere 31 Schüler bearbeiteten den SJT_Hotelfach im Paper-Pencil-Format.

keit hin (vgl. Kempf & Meder, 1993). Demnach können sich bereits bei geringen Variationen hinsichtlich der Stichprobenszusammensetzung unterschiedliche Reliabilitäten und andere Itemkennwerte ergeben. Insbesondere die Reliabilität eines Tests wird von verschiedenen Seiten (z. B. Klapproth, 1981) als besonders anfällig gegenüber spezifischen Charakteristika der Stichprobe erachtet (s. Kapitel 7).

Ein Vergleich der auf Basis von Ratings erstellten sogenannten empirischen Rangordnungen mit den von SME im Rahmen der Testentwicklung (vgl. Kapitel 3) aufgestellten Rangordnungen brachte hervor, dass sich bei keinem der dreizehn Items deckungsgleiche Rangordnungen ergeben (s. Tabelle 6-4). Die fehlende Übereinstimmung drückt sich vor allem darin aus, dass die Teilnehmer dieser Studie bei neun der 13 Items solche Verhaltensalternativen präferierten – ausgedrückt durch die relative Höhe des Mittelwerts –, welche SME (s. Kapitel 3) als besonders ineffektiv in der jeweiligen Situation betrachtet haben. Eine ähnliche Problematik aufgreifend wiesen Legree (1995) und auch Weekley et al. (2006) auf den Punkt hin, dass die Festlegung eines Bewertungsschlüssels stark von den ausgewählten SME abhängt und dass jederzeit die Eventualität zu beachten sei, dass verschiedene Expertengruppen bei demselben Test unterschiedliche scoring keys erzeugen könnten. Demgemäß besteht die Möglichkeit, dass auch bei der vorliegenden Arbeit eine Auswahl anderer SME zu einem abweichenden Bewertungsschlüssel geführt hätte. Dass sich zwischen unterschiedlichen Mitarbeitern desselben Unternehmens durchaus differierende Auslegungen, wie Kunden zu behandeln sind, ergeben, kann als Evidenz für die Existenz sogenannter differierender Konnotationen innerhalb einer Organisation ausgelegt werden (vgl. Bungard, 2004, S. 132 ff.). Umso überraschender erscheint das Fehlen kongruenter Rangordnungen zwischen berufserfahrenen Probanden und SME auch vor dem Hintergrund, dass sich bei Teilnehmern der vorherigen Studie, welche als unerfahren im Hotelfach einzustufen waren, durchaus deckungsgleiche Rangordnungen feststellen ließen (s. Kapitel 6).

Im Rahmen dieser Studie wurde davon Abstand genommen, die vollständige Skala zur Messung der sozialen Validität nach Kanning (2008) einzusetzen. Erhebungen der sozialen Validität sollten nach Ryan und Schmitt (2000) sowie Chan und Schmitt (2005) solchen Studien vorbehalten sein, an denen tatsächliche Bewerber teilnehmen, was in der vorlie-

genden Studie nicht der Fall war. Für wichtig wurde dennoch erachtet, ob die Teilnehmer dem SJT die Fähigkeit zusprechen, die spätere Leistungsfähigkeit von Bewerbern vorhersagen zu können und ob sie den SJT zur Auswahl zukünftiger Mitarbeiter einsetzen würden. In einer Studie von Schuler, Hell, Trapmann, Schaar und Boramir (2007) trat in diesem Zusammenhang die intuitiv nachvollziehbare Erkenntnis zutage, dass Personalverantwortliche dazu neigen, vermehrt solche Verfahren zur Personalauswahl einsetzen, welche sie selbst als hilfreich bei der Identifikation geeigneter Mitarbeiter erachten. Das Ergebnis der vorliegenden Studie, insbesondere die Einschätzungen der 30 Führungskräfte, kann eindeutig als Votum *für* eine spätere Anwendung des SJT in der Praxis ausgelegt werden.

7. Generelle Diskussion

Im siebten Kapitel dieser Arbeit wird in einem ersten Schritt ein Überblick geboten über die im Rahmen der vorliegenden Arbeit durchgeführten Studien und deren Ergebnisse. Ebenso sollen studienübergreifende Resultate und Auffälligkeiten hervorgehoben und besprochen werden. In einem zweiten Schritt zeigt der Autor Implikationen auf für sowohl die weitere Forschung mit SJTs als auch für den Einsatz des in dieser Arbeit entwickelten SJTs in der Praxis.

7.1 Zusammenfassung der Studien

In Interviews nach methodischen Vorgaben der CIT (Flanagan, 1954) sammelte der Autor sowohl erfolgsrelevante Situationen aus dem Berufsalltag eines Hotelfachmannes als auch dazugehörige situationsspezifische Verhaltensweisen, welche sich nach Ansicht der SME hinsichtlich ihrer Effektivität bei der Bewältigung der jeweiligen Situation unterscheiden ließen. Die auf diese Art und Weise zusammengestellten Verhaltensweisen wiederum stellten einerseits die Grundlage zur Ableitung von fünf Anforderungsdimensionen – Serviceorientierung, Beschwerdemanagement, Durchsetzungsfähigkeit, Handlungsfähigkeit und Gewissenhaftigkeit – und andererseits die Basis zur Durchführung zweier Expertenworkshops dar. In diesen Expertenworkshops trafen weitere, bisher nicht involvierte SME eine gemeinsame Entscheidung, welche Situationen in dem SJT abgebildet werden sollten. Ferner wurden jeder Situation vier verschiedene Verhaltensweisen zugeordnet, die sich in den Augen der SME eindeutig hinsichtlich ihres Grades an situativer Effektivität unterscheiden. In einem weiteren Teilschritt wurde in den Expertenworkshops ein für die spätere Einordnung der Probandenantworten relevanter Bewertungsmaßstab entwickelt. Zu diesem Zweck brachten die teilnehmenden SME die vier Verhaltensoptionen je Situation in eine Rangordnung. Der resultierende videobasierte SJT (25 Items) wurde an einer Stichprobe von 97 Auszubildenden im Hotelfach erprobt. Die Auszubildenden befanden sich zum Zeitpunkt der Datenerhebung im ersten beziehungsweise zweiten Lehrjahr. Neben dem SJT wurde mit dem ISK-K (Kanning, 2009) ein Fragebogen zur Selbstbeschreibung eingesetzt, welcher die Messung von vier Sekundärfaktoren der allgemeinen sozialen Kompetenz ermöglicht.

Die über den SJT eingeholten Angaben der Teilnehmer wurden über zwei verschiedene Bewertungsschlüssel – Pick best – pick worst respektive nach Vorgabe von Motowidlo et al. (1990) – mit dem Maßstab der SME in Beziehung gesetzt und zu Testwerten verrechnet. Als Anweisung für die Teilnehmer wurde wie in den beiden nachfolgenden Studien eine Instruktion zur Verhaltenstendenz gegeben. Da ein über den Pick best – pick worst-Schlüssel generierter Datensatz aufgrund unzureichender Varianzen keine statistischen Analysen zuließ, stützten sich weiterführende Untersuchungen des SJT alleinig auf Werte, welche über den Motowidlo-Schlüssel berechnet wurden. An dieser Stelle trat bei Durchführung einer explorativen Faktorenanalyse eine aus elf Items bestehende Skala hervor, welche von Autor mit SJT_kurz benannt wurde. Diese Skala erreichte an der Gesamtstichprobe (N = 97) eine interne Konsistenz von $\alpha = .64$. Die elf Items der Skala wiederum verfügten über eine mittlere Trennschärfe von $r_{it} = .32$ und mittlere Aufgabenschwierigkeit in Höhe von $P = .54$. In Bezug auf korrelative Zusammenhänge mit den Sekundärfaktoren des ISK-K ließ sich für den SJT_kurz eine statistisch bedeutsame Beziehung mit dem Faktor Offensivität ($r = .21$) verzeichnen. In Übereinstimmung mit dieser signifikanten Korrelation trat der Faktor Offensivität auch als wichtigster Prädiktor bei der gemeinsamen Vorhersage des Testwerts im SJT_kurz durch die Sekundärfaktoren des ISK-K in Erscheinung. Verschiedenen demographischen Variablen konnte kein nicht-zufälliger Einfluss auf die Leistung im SJT_kurz nachgewiesen werden.

In der darauf folgenden Studie wurden die Teilnehmer – Realschüler, welche sich eine Ausbildung im Hotelfach vorstellen können (N = 62) – zunächst einer von zwei Versuchsbedingungen zugeteilt: SJT im Videoformat vs. SJT im Paper-Pencil-Format. Unabhängig von der Versuchsbedingung waren alle Teilnehmer zum einen dazu aufgefordert, die Wahrscheinlichkeit, mit der sie die vier Verhaltensalternativen je Situation in einem realen Kontext an den Tag legen würden, auf einer siebenstufigen Ratingskala einzuschätzen. Zur Berechnung eines Testscores in diesem Szenario wurde der sogenannte Ratingschlüssel eingesetzt. Zum anderen sollten die Teilnehmer die vier Verhaltensalternativen in eine Rangordnung bringen. Für die Bewertung der Rangordnungen wurden vier unterschiedliche scoring keys angewandt: Komplett-Schlüssel, Motowidlo-Schlüssel, Einzelvergleich-Schlüssel und Distanz-Schlüssel. Als nützlich-

ste Grundlage zur Gewinnung einer reliablen und validen Skala in der Video-Bedingung erwies sich über Ratings ermittelter Datensatz (vgl. Abschnitt 5.5). Auf diesem Wege wurde die Skala SJT_Hotelfach aus zehn Items gebildet. Für diese wurde eine interne Konsistenz von $\alpha = .74$ berechnet. Die mittlere Aufgabenschwierigkeit betrug $P = .38$ und die mittlere Trennschärfe $r_{it} = .41$. Zudem ergaben sich für diese Skala Korrelationen statistischer Bedeutsamkeit mit den Dimensionen Handlungsfähigkeit und Flexibilität aus dem BIP (Hossiep & Paschen, 2003), welche von den beiden Testautoren der Eignungsvoraussetzung Arbeitsverhalten zugeordnet werden.

Für den SJT_Hotelfach in der Paper-Pencil-Version wurden durchgängig abweichende Ergebnisse im Bereich der Reliabilität, der Zusammenhänge mit dem BIP und in Bezug auf die Aufgabenschwierigkeit und die Trennschärfe der Items festgestellt. Dagegen konnte für den Testwert im SJT_Hotelfach kein statistisch bedeutsamer Unterschied zwischen den beiden Versuchsbedingungen ausfindig gemacht werden. Ebenso fanden sich in beiden Versuchsbedingungen nur zufällige Zusammenhänge mit Maßen der fluiden oder kristallinen Intelligenz sowie unterschiedlichen demographischen Variablen wie dem Alter oder den Noten in den Schulfächern Mathematik, Deutsch und Englisch. Eine Einschätzung der sozialen Validität, welche sich auf den gesamten SJT (25 Items) und nicht auf die Skala SJT_Hotelfach bezog, brachte hinsichtlich der Dimension Anforderungsbezug einen signifikanten Unterschied zugunsten der Videobedingung hervor. Der Skala SJT_Hotelfach wurden abschließend drei Items, welche unter der Video-Bedingung mindestens drei signifikant positive Korrelationen mit verschiedenen Dimensionen des BIP erzielten, hinzugefügt. Die nun aus 13 Items bestehende Skala SJT_Hotelfach⁶² wies eine interne Konsistenz von $\alpha = .72$ und im Gegensatz zur ursprünglichen Skala weitere statistisch bedeutsame Korrelationen mit den Dimensionen Sensitivität und Belastbarkeit aus dem BIP auf. Die statistische Unabhängigkeit von zwei Facetten der allgemeinen kognitiven Leistungsfähigkeit blieb erhalten.

⁶² Es sei nochmals darauf hingewiesen, dass sich die Skalen SJT_kurz (10 Items) und SJT_Hotelfach (13 Items) in Bezug auf acht Items entsprechen (vgl. Abschnitt 5.5).

Der videobasierte SJT_Hotelfach wurde in einer darauf folgenden Studie an einer Stichprobe von berufserfahrenen Teilnehmern ($N = 50$) eingesetzt. Zusätzlich zum SJT_Hotelfach sahen sich die Teilnehmer mit dem NEO-FFI (Borkenau & Ostendorf, 2008) konfrontiert, welches eine Erhebung fünf globaler Persönlichkeitsfaktoren – Neurotizismus, Extraversion, Offenheit für Erfahrungen, Verträglichkeit und Gewissenhaftigkeit – gestattet. Daneben wurde aufgrund der Zusammensetzung der Stichprobe die Berufserfahrung der Teilnehmer über vier verschiedene Indikatoren abgebildet und operationalisiert. Die Skala SJT_Hotelfach wies an dieser Stichprobe nur noch eine interne Konsistenz von $\alpha = .52$ auf (s. u.). Für die 13 Items des SJT_Hotelfach konnte eine sehr niedrige mittlere Trennschärfe von .20 und eine mittlere Aufgabenschwierigkeit von .66 vermerkt werden. Neben der niedrigeren mittleren Aufgabenschwierigkeit sprach auch die Feststellung, dass berufserfahrene Teilnehmer im SJT_Hotelfach signifikant besser abschneiden als Teilnehmer ohne diese Expertise dafür, dass die Auseinandersetzung mit dem SJT_Hotelfach für diese Art von Teilnehmern ein geringeres Anspruchsniveau als für die Teilnehmer der vorhergehenden Studie bot. Weder für die fünf Persönlichkeitsfaktoren des NEO-FFI noch für die vier unterschiedlichen Maße der Berufserfahrung konnten signifikante Korrelationen mit dem SJT_Hotelfach festgestellt werden. Zudem verblieben einbezogene demographische Variablen, zum Beispiel das Alter oder das Geschlecht, wie in den vorherigen Studien ohne bedeutsamen Einfluss auf die Leistung im SJT_Hotelfach.

Nach der Zusammenfassung zentraler Ergebnisse aus den einzelnen Studien werden im Folgenden *studienübergreifende Resultate* angesprochen und interpretiert. Dabei werden zunächst die Reliabilitäten der Skalen SJT_kurz und SJT_Hotelfach in den Blickpunkt gerückt. Danach werden generelle Ergebnisse zu Zusammenhängen mit sozialer Kompetenz, berufsbezogener Persönlichkeit, kognitiver Leistungsfähigkeit und Berufserfahrung diskutiert. Diese Aspekte sind allesamt dem Bereich der Konstruktvalidität zuzuordnen. An diese generelle Abhandlung der Konstruktvalidität schließen sich methodische Überlegungen zu Aspekten wie der Verzerrbarkeit der Testwerte im SJT an.

Über die verschiedenen Studien hinweg zeigte sich eine *Stichprobenabhängigkeit* zentraler Kennwerte – insbesondere der Reliabilität. So konnte für die Skala SJT_kurz an der Substichprobe von Berufsschülern im zweiten Ausbildungsjahr eine Reliabilität von $\alpha = .70$ ermittelt werden. Demgegenüber erreichte die interne Konsistenz derselben Skala nur einen Wert von $\alpha = .54$, falls als Substichprobe ausschließlich Berufsschüler herangezogen wurden, welche sich erst im Anfangsjahr ihrer Ausbildung befanden (s. Kapitel 3). Ein ähnliches Phänomen ereignete sich in Hinblick auf den SJT_Hotelfach, welcher sich aus dreizehn Items zusammensetzt. In der dritten Studie (s. Kapitel 4), an welcher Realschüler ohne Erfahrung im Hotelfach teilnahmen, wurde für diese Skala ein zufriedenstellender Reliabilitätswert von $.72$ errechnet. Jedoch ergab sich für den SJT_Hotelfach in der nachfolgenden Studie (s. Kapitel 6) mit erfahrenen Hotelmitarbeitern eine unzureichende Reliabilität in Höhe von $.52$. Dass sich speziell die interne Konsistenz – berechnet als Cronbach's Alpha – als anfällig für die Zusammensetzung der Stichprobe erweist, ist auf eine bestimmte Eigenart von SJTs zurückzuführen: Nach Kanning et al. (2006) zeichnet sich diese Art von Tests vor allem durch ihre kriterienbezogenen Validitäten aus und weniger durch ihre Möglichkeit, homogene Eigenschaften zu messen. Demnach ist die Berechnung von Cronbach's Alpha gemäß Whetzel und McDaniel (2009) als unangemessen bei SJTs zu betrachten. Die beiden Autoren sehen in der internen Konsistenz eine Unterschätzung der tatsächlichen Zuverlässigkeit eines SJT (S. 191). Sie bezeichnen die Bestimmung der Reliabilität über wiederholte Messungen (Retest-Reliabilität) oder den Vergleich von Parallelformen (Paralleltest-Reliabilität) als erstrebenswerter: „*In sum, test-retest and parallel forms reliability are the most appropriate reliability estimation methods for SJTs*“ (Whetzel & McDaniel, 2009; S. 191). Erste Evidenz für diese Annahme findet sich in den Studien von Chan und Schmitt (2002) beziehungsweise Kanning (2005), welche zeigten, dass eine Bestimmung der Reliabilität über einen Vergleich von Parallelformen (Chan & Schmitt, 2002) respektive wiederholte Messungen (Kanning, 2005) höhere Reliabilitäten – jeweils $> .70$ – ergab als die Berechnung von Cronbach's Alpha – jeweils $< .50$.

In der zweiten Studie (s. Kapitel 4) wurde untersucht, ob sich zwischen dem SJT_kurz und den vier Hauptskalen des ISK-K (Kanning, 2009) sta-

tistisch bedeutsame Beziehungen vorfinden lassen. Da sich nur zwischen der Skala Offensivität und dem SJT_kurz eine signifikante Korrelation ergab, wurde die Leistung im SJT_kurz als unabhängig erachtet von Selbsteinschätzungen entlang unterschiedlicher *Facetten der sozialen Kompetenz*. Im Rahmen der dritten Studie (s. Kapitel 5) wurde neben dem SJT_Hotelfach das BIP (Hossiep & Paschen, 2003) eingesetzt, welches ebenfalls die Messung der sozialen Kompetenz beinhaltet. Im Gegensatz zum ISK-K, welches eher verallgemeinernde Aussagen enthält, widmet sich das BIP einem stärkeren inhaltlichen Zuschnitt der Items auf den Berufskontext. Dennoch zeigte sich auch hier die Leistung im SJT als unbeeinflusst von Selbstbeschreibungen entlang vier der fünf Facetten sozialer Kompetenz. Speziell vor dem Hintergrund, dass SJTs vorrangig Situationen mit sozialem Charakter thematisieren (z. B. Kanning, 2005; Bledow & Frese, 2009), überrascht der nahezu gänzlich fehlende statistische Zusammenhang zwischen der Leistung im SJT und unterschiedlichen Facetten der sozialen Kompetenz (vgl. auch Abschnitt 4.5).

Bei Berücksichtigung der durchgeführten Studien konnte verschiedenen Indikatoren der *allgemeinen kognitiven Leistungsfähigkeit* ein nur zufälliger Einfluss auf die Leistung im SJT_kurz und auch dem SJT_Hotelfach nachgewiesen werden. So ergaben sich zwischen dem CFT-20-R (Weiß, 2006), welcher Aussagen gestattet über die individuelle Ausprägung der fluiden Intelligenz, und dem WS (Weiß, 2007), welcher eine Operationalisierung der kristallinen Intelligenz ermöglicht, keine statistisch bedeutsamen Zusammenhänge mit dem SJT_Hotelfach. Schulnoten können ebenfalls als Ausdruck kognitiver Leistungsfähigkeit begriffen. Weder für den SJT_kurz noch für den SJT_Hotelfach konnten jedoch statistisch bedeutsame Zusammenhänge ausfindig gemacht werden mit dem Notendurchschnitt respektive den Noten in den Fächern Mathematik, Deutsch und Englisch. Bei Deutung der beiden Skalen – SJT_Hotelfach und SJT_kurz – als Messung einer speziellen, organisationsintern definierten Kundenorientierung unterstreicht die studienübergreifend festgestellte Unabhängigkeit von allgemeiner kognitiver Leistungsfähigkeit das Ergebnis einer Meta-Analyse von Frei und McDaniel (1998). Diese beiden Forscher waren ebenfalls auf eine Unabhängigkeit von Kundenorientierung und Intelligenz gestoßen.

Hinsichtlich der Leistung im SJT_Hotelfach (13 Items) konnte ein nicht-zufälliger Unterschied zwischen berufserfahrenen Teilnehmern (N = 50) und Realschülern ohne vergleichbare Expertise in diesem Berufsfeld (n = 31) in Erfahrung gebracht werden. Es stellte sich heraus, dass berufserfahrene Teilnehmer im Schnitt signifikant höhere Scores im SJT_Hotelfach erzielten. Innerhalb der Stichprobe berufserfahrener Teilnehmer ließ sich jedoch keine statistisch bedeutsame Beziehung zwischen der *Berufserfahrung* im Hotelfach und dem Score im SJT_Hotelfach feststellen (s. Kapitel 6). Zum Beispiel schnitten Personen mit mehr als zehnjähriger Berufserfahrung im SJT_Hotelfach nicht signifikant besser ab als Personen, welche zum Zeitpunkt der Datenerhebung erst über drei oder vier Jahre Berufserfahrung verfügten. Dieses Ergebnis befindet sich in Einklang mit einem im Rahmen der zweiten Studie (vgl. Kapitel 4) hervorgebrachten Resultat: Berufsschüler im zweiten Ausbildungsjahr erreichten in der Skala SJT_kurz keine signifikant höheren Testwerte als Berufsschüler, welche sich erst im ersten Ausbildungsjahr befanden. Demzufolge kann geschlossen werden, dass es für die Leistung im SJT sehr wohl von Belang ist, ob der Teilnehmer Berufserfahrung im Hotelfach besitzt oder nicht. Indessen vermag die Menge an Berufserfahrung, für welche die Jahre an Berufserfahrung als Gradmesser dienen, keinen weiteren Einfluss auf die Performanz im SJT auszuüben. Diese Analyse deckt sich mit von McDaniel und Nguyen (2001) getätigten Äußerung, dass der erfolgskritische Anteil an berufsbezogenem Wissen in der Anfangszeit einer Berufstätigkeit aufgebaut werde, während jedes weitere Jahr an zusätzlicher Berufserfahrung das berufsbezogene Wissen nicht mehr entscheidend bereichere.

Sowohl in der zweiten als auch in der vierten Studie dieser Arbeit ergaben sich keine statistisch bedeutsamen Validitäten für den SJT_kurz beziehungsweise den SJT_Hotelfach, wenn breiter gefasste, globalere Maße wie die Skalen des ISK-K⁶³ oder die Persönlichkeitsfaktoren aus dem NEO-FFI als Korrelate hinzugezogen wurden. Dieser vorgefundene Sachverhalt erinnert an das sogenannte *bandwidth-fidelity-Dilemma* (Cronbach, 1960). Der Kern dieses Dilemmas besteht aus der bei dem Einsatz eines Tests zu treffenden Abwägung, ob mit dem Messinstrument ein um-

⁶³ Eine Ausnahme bildet die Hauptskala Offensivität aus dem ISK-K (s. Kapitel 3).

fassendes Bild über ein bestimmtes Merkmal (bandwidth) oder aber eher ein spezifischerer Ausschnitt des Merkmals (fidelity⁶⁴) erfasst werden soll (Schuler & Höft, 2006). Bowling und Gruys (2010; S. 59) fassen die Kontroverse wie folgt zusammen: „Typically, when there is greater fidelity, there is lower bandwidth, and when there is greater bandwidth, there is less fidelity“. Besonders hohe und bedeutsame Validitäten werden dann erwartet, wenn sowohl Prädiktor(-en) als auch Kriterium in Bezug auf ihren Grad an Allgemeinheit oder Spezifität übereinstimmen (Ones & Viswesvaran, 1996). Demnach sollten breiter gefasste Prädiktoren eine valide Vorhersage von breiter angelegten Kriterien ermöglichen, während enger gefasste Prädiktoren erst im Verbund mit ebenso eng umschriebenen Kriterien ihre Wirksamkeit entfalten können (vgl. die sogenannte Symmetriehypothese nach Cronbach & Gleser, 1965). Da sowohl der SJT_kurz und der SJT_Hotelfach als Messungen einer spezifischen Facette eines Konstrukts angelegt sind und damit einen hohen Grad an fidelity aufweisen, wird der unzureichende Zusammenhang mit globalen Maßen, welche folglich mit einem hohen Maß an bandwidth ausgestattet sind, erklärbar. Bei Beachtung der Tatsache, dass das Kriterium der allgemeinen beruflichen Leistung (vgl. Campbell et al., 1993) ebenfalls sehr global gefasst ist, um eine Anwendbarkeit auf zahlreiche Arbeitsbereiche zu ermöglichen, empfehlen Hogan und Roberts (1996) hier einen Rückgriff auf Prädiktoren mit einem entsprechend hohen Grad an bandwidth. Dieser Umstand bringt beachtenswerte Implikationen für die noch ausstehende Bestimmung der prognostischen Validität des SJT_Hotelfach mit sich, denn laut Motowidlo, Crook, Kell und Naemi (2009) gilt, dass „[...] SJT scores based on behavioral content that matches the behavioral content of a performance dimension will show the strongest correlation with that performance dimension“ (S. 287).

Die Bestimmung der Konstruktvalidität wurde in dieser Arbeit auf zweierlei Art und Weise verfolgt: Erstens wurden die bereits oben geschilderten Zusammenhänge mit Instrumenten wie dem ISK-K, dem CFT-20-R, dem WS, dem BIP und dem NEO-FFI hinsichtlich ihrer Bedeutung für die Konstruktvalidität des SJT erforscht. Zweitens wurde in Kapitel 3

⁶⁴ Dieses Konzept der fidelity ist nicht zu verwechseln mit demjenigen, welches in Kapitel 2 thematisiert wurde.

ein hierarchisches Modell zur Kundenorientierung aufgestellt, in welchem fünf empirisch ermittelte Anforderungsdimensionen ihrer Bestimmung als Faktoren 1. Ordnung zugeführt wurden. Zudem wurde Kundenorientierung in diesem Modell als Faktor 2. Ordnung festgesetzt (vgl. Abbildung 3-4). Um dieses Modell für eine empirische Überprüfung zu erschließen, wurden den relevanten Anforderungsdimensionen Items aus dem SJT qualitativ zugeordnet (vgl. Tabelle 3-2). Ein Abgleich dieser Zuordnung mit der Empirie mittels einer CFA und EFA schlug jedoch fehl (s. Kapitel 4). Bei der Entwicklung des SJT hätte ein anderer Weg beschritten werden können als derjenige, welcher in Kapitel 3 nachzuvollziehen ist. So hätte sich die Konstruktion des SJT auch einem vorliegenden theoretischen Modell zur Kundenorientierung orientieren können (vgl. Stevens und Campion, 1994, oder Mumford et al., 2008, für ein vergleichbares Vorgehen bei anderen Konstrukten). In Anbetracht des Fehlens eines solchen Modells speziell für die Kundenorientierung im Hotelfach und des möglicherweise geringen Realitätsgrades der auf diesem theoretischen Wege generierten Verhaltensalternativen (vgl. Abschnitt 2.3) wurde von diesem Vorhaben Abstand genommen.

Im zweiten Kapitel dieser Arbeit wurde unter anderem der aktuelle Forschungsstand zum Thema *Faking* in SJTs überblicksartig vorgestellt. In diesem Zusammenhang wagte der Autor die Überlegung, dass Probanden, welche zu einer verzerrten, positiveren Selbstdarstellung neigen, gleichzeitig eine höhere Anpassungsfähigkeit in sozial geprägten Situationen beweisen (vgl. Schuler & Höft, 2006). Übertragen auf den Kontext der Kundenorientierung im Hotelfach könnte sich eine stark ausgeprägte soziale Anpassungsfähigkeit beispielsweise in dem Umstand niederschlagen, mit Hotelgästen unterschiedlichster Couleur adäquat interagieren zu können. Dies impliziert, dass Personen mit stark ausgeprägter Anpassungsfähigkeit in einem SJT zur Messung von Kundenorientierung bessere Testwerte erreichen sollten als solche Personen, denen dieser Akt der Anpassung schwieriger gelingt. Dass in dem SJT_Hotelfach – vorliegend im Video-Format – tatsächlich solche Personen besser abschneiden, welche über einen höheren Grad an Anpassungsfähigkeit verfügen, wird in Form einer signifikant positiven Korrelation mit der Dimension Flexibilität aus dem BIP deutlich (s. Kapitel 4). Nach Hossiep und Paschen (2003) zeichnen sich flexible Personen dadurch aus, sich schnell auf neue Situa-

tionen und Personen einstellen zu können. Ein widersprechendes Bild ergibt sich jedoch für den SJT_kurz. Der Testwert in dieser Skala zeigt sich als nicht beeinflusst durch die Dimension Selbststeuerung aus dem ISK-K, welche Kanning (2009) unter anderem als Fähigkeit beschrieb, sich verändernden Situationen anzupassen. Zudem ergab sich für den SJT_Hotelfach in der Paper-Pencil Version eine negative, wenn auch nicht signifikante, Korrelation mit der Dimension Flexibilität aus dem BIP. Eine abschließende Aussage über den Einfluss des Fakings auf Testwerte im SJT_kurz oder SJT_Hotelfach ist bei Berücksichtigung dieser divergierenden Ergebnisse weiterhin zu erbringen.

Bei der Aufstellung der den einzelnen Studien zugrunde liegenden Hypothesen wurden einerseits solche formuliert, welche sich auf SJTs als Methode bezogen, und andererseits solche, welche auf das Konstrukt Kundenorientierung, das der SJT zu messen beansprucht, abzielen. Arthur und Villado (2008) wiesen, um die Validität eines Verfahrens richtig deuten zu können, auf eine notwendige Differenzierung von Methode und Konstrukt hin, denn „[...]it allows the isolation of variance due to predictor constructs from the variance due to predictor methods” (S. 436). Um dies zu gewährleisten schlugen sie vor, dass der Einfluss der Methode – in dem Fall dieser Arbeit des SJTs – abgeschätzt werden kann, indem das Konstrukt bei gleichzeitiger Variation der Methode konstant gehalten wird (vgl. Richman-Hirsch et al., 2000). Eine Umsetzung der Postulate von Arthur und Villado (2008) findet sich beispielhaft in der dritten Studie (s. Kapitel 4). In dieser Studie wurde der SJT sowohl im Video-Format als auch im Paper-Pencil-Format eingesetzt und mit verschiedenen Maßen, zum Beispiel der kristallinen Intelligenz, korreliert. Hypothesen, welche in dieser Arbeit das Konstrukt in den Vordergrund setzten, widmeten sich vorrangig dem Versuch der empirischen Legitimation eines in Kapitel 3 aufgestellten Modells und erforderten keine Überprüfung anhand korrelativer Analysen, so dass diese von den Anregungen Arthur und Villados (2008) ausgenommen sind. Insgesamt wurde in der vorliegenden Arbeit die von Arthur und Villado (2008) genannte Differenzierung von Methode und Konstrukt bei der Validitätsanalyse angemessen berücksichtigt.

7.2 Implikationen für die weitere Forschung und Praxis

7.2.1 Implikationen für die weitere Forschung

An dieser Stelle sollen sowohl Aspekte angesprochen werden, welche sich auf die Skala SJT_Hotelfach beziehen, als auch solche Punkte, welche die Forschung mit (videobasierten) SJTs im Allgemeinen adressieren.

Damit die Anwendung eines SJT in der Personalauswahl gerechtfertigt erscheint, muss ein Nachweis seiner Nützlichkeit bei der Vorhersage zukünftiger Arbeitsleistung vorliegen. Diese Bescheinigung prognostischer Validität wurde für den SJT_Hotelfach im Rahmen dieser Arbeit nicht erbracht⁶⁵. Erforderlich ist in diesem Zusammenhang zunächst die Anwendung des SJT_Hotelfachs an einer sehr viel größeren Stichprobe als im Rahmen der vorliegenden Arbeit geschehen (s. zum Problem der Stichprobenabhängigkeit auch Abschnitt 6.5). Dies würde gleichzeitig den Weg einschlagen zu einer ebenfalls noch ausstehenden Normierung des SJT. Neben einer Normierung böte das Vorliegen einer größeren Stichprobe auch die Möglichkeit, die Resultate einer CFA und der damit verbundenen Überprüfung eines in dieser Arbeit aufgestellten Rahmenmodells zur Kundenorientierung (s. Kapitel 3 und 4) auf ein belastbareres Fundament zu stellen. Als Stichprobe beziehungsweise Gros der Stichprobe sollten optimalerweise Bewerber gewonnen werden. Ein größeres Problem stellt sich jedoch vor dem Hintergrund der weiter oben geführten Diskussion zum bandwidth-fidelity-Dilemma in Form der Frage nach einem geeigneten (Leistungs-)Kriterium. Der Testwert im SJT_Hotelfach (Prädiktor) erfasst eine eng umgrenzte, unternehmensintern definierte Form der Kundenorientierung (vgl. Abschnitt 5.5). Ein geeignetes Kriterium für einen SJT als Prädiktor müsste ebenfalls sehr eng und spezifisch gefasst werden (z. B. Bartram, 2005; Christian, Edwards & Bradley, 2010). Lievens, Buyse und Sackett (2005) konnten in dieser Zusammenfassung nachweisen, dass ihr SJT zur Messung der Kompetenz von Studienbewerbern in der Arzt-Patienten-Kommunikation ein stärker sozial geprägtes Kriterium valider vorhersagen kann als ein Kriterium wie Stu-

⁶⁵ Der SJT_Hotelfach wurde nach Abschluss der vierten Studie einer Bewerberstichprobe vorgelegt. Mit einem N von 13 viel diese jedoch zu klein aus, um weitere Analysen zu rechtfertigen.

diennoten. Mögliche Ansatzpunkte bei einer Übertragung auf den SJT_Hotelfach wären zum Beispiel die Erfassung von Beschwerden oder die Messung der Zufriedenheit von Kunden respektive Gästen, welche ein Hotelmitarbeiter betreut hat. Es erscheint offensichtlich, dass diese Kriterien mit vielfältigen methodischen Makeln und Einschränkungen behaftet sind, zum Beispiel beschwerten sich nur wenige unzufriedene Kunden direkt beim betroffenen Unternehmen (s. Abschnitt 2.6).

In der dritten Studie (s. Kapitel 4), bei der die Teilnehmer zu Einschätzungen der Verhaltenswahrscheinlichkeiten auf Ratingskalen aufgerufen waren, zeigte sich bei manchen Teilnehmern, dass alle vier Verhaltensalternativen pro Situation dieselbe Einschätzung erfuhren. Dies manifestierte sich beispielsweise darin, dass sowohl diejenige Verhaltensalternative, welche von SME im Rahmen von Workshops als besonders ineffektiv beurteilt worden ist, als auch jene Verhaltensalternative, welche SME als außerordentlich effektiv benannt hatten, von den Teilnehmern mit einer 1 (sehr unwahrscheinliche Ausführung in der realen Situation) versehen wurden. Bei einem möglichen Einsatz des SJT in der Praxis könnten Bewerber mit einem solchen Antwortmuster bei einer bestimmten Anzahl von Items per se als ungeeignet im Sinne des Tests eingestuft werden.

Wie es bereits in der Einleitung (s. Kapitel 1) zu der vorliegenden Arbeit deutlich gemacht wurde, streben SJTs an, den Probanden in realistische Szenarien aus unterschiedlichen beruflichen Kontexten hineinzuzusetzen. Dabei wird der Grad an Realismus sowohl der geschilderten Situationen als auch des Einholens der Probandenreaktion auf die Erfordernisse der Situation unter dem Begriff *fidelity* diskutiert (s. Abschnitt 2.4). Weit weniger Aufmerksamkeit wird nach Ansicht des Autors jedoch dem Punkt zuteil, ob der Prozess der Urteilsbildung (engl.: *Judgment*), welcher auf Seiten des Probanden abläuft, in SJTs realistisch erfasst wird. Den Situationen im SJT wurden in dieser Arbeit zum Beispiel jeweils vier unterschiedliche Verhaltensoptionen zugewiesen, deren inhaltliche Ausgestaltung auf eine Anforderungsanalyse und Expertenworkshops (s. Kapitel 3) zurückverfolgt werden kann. Andere SJTs bieten ihren Probanden zum Teil bis zu 20 Verhaltensalternativen. Eine Verwendung mehrerer Verhaltensalternativen erscheint unter methodischen Gesichtspunkten, zum Beispiel der Prävention sozial erwünschter Antworten oder der Generierung ausreichender Varianz, gerechtfertigt. Dagegen bleibt neben der Frage

nach der inhaltlichen Trennschärfe zu erörtern, ob Personen, die sich der Situation in einem realen Berufskontext stellen, tatsächlich zwischen vier oder mehr Verhaltensalternativen abwägen oder ob sie sich nicht viel mehr bestimmter ökonomischer Entscheidungsregeln oder Heuristiken (z. B. Tversky & Kahneman, 1973) bedienen. Ein innovativer Ansatz zur Bereicherung dieser Diskussion stammt von Motowidlo et al. (2009), welche jedem Item ihres SJT nur eine einzige Verhaltensoption beifügten. Ihre auf diesem Weg generierten SJT-Items kennzeichneten sie wie folgt: „*The set of single-response items is a mixture of incidents that reflect a variety of performance categories in the performance domain that was referenced in instructions to SMEs who provided the incidents*” (S. 283).

Für einen SJT, welcher wie der SJT_Hotelfach die Messung von Kundenorientierung anstrebt, ist zu ergänzen, welche statistischen Beziehungen zwischen diesem SJT und anderen Instrumenten zur Erfassung von Kundenorientierung bestehen. Allerdings liegen im deutschsprachigen Raum bis dato keine veröffentlichten oder verfügbaren Verfahren zur Messung dieses Konstrukts vor. Als möglicher Ansatzpunkt für eine zukünftige Validierung des SJT_Hotelfach in diesem Bereich empfiehlt sich jedoch das noch in der Erprobung befindliche Bochumer Inventar zur Serviceorientierung (BISO) des Projektteams Testentwicklung unter der Leitung von Rüdiger Hossiep. Das als Verfahren zur Selbstbeschreibung angelegte BISO umfasst 130 Items, welche die Messung von zehn Dimensionen der Serviceorientierung, zum Beispiel Kundenfokussierung und Flexibilität, ermöglichen sollen. Gerade die beiden letztgenannten Skalen des BISO erscheinen viel versprechend bei einer weiterführenden Validierung des SJT_Hotelfach.

Während in der dritten Studie die Art der Situationspräsentation (Video vs. Paper-Pencil) variiert wurde (s. Kapitel 5), sind weitere Erkenntnisse, wie sich Veränderungen von anderen Komponenten auf die Gütekriterien des SJT_Hotelfach auswirken, noch zu erbringen. Zu denken wäre an eine Abwandlung der Instruktion im SJT, da bislang ausschließlich eine Instruktion zur Verhaltenstendenz zum Einsatz kam. Mit weiterem finanziellen und personellen Aufwand verbunden wäre eine Steigerung der fidelity des SJT_Hotelfach, falls die Verhaltensoptionen im SJT_Hotelfach entweder ebenfalls als Videos vorlägen (vgl. Funke & Schuler, 1998) oder wenn Verknüpfungen zwischen den verschiedenen Situationen des

SJT_Hotelfach hergestellt würden, was einer Zunahme der Interaktivität des SJT_Hotelfach gleichkäme (vgl. Richman et al., 2000). Mit einer wachsenden Interaktivität und Komplexität des SJT_Hotelfach würde jedoch die Frage nach einer angemessenen Homogenität und Konstruktvalidität des SJT_Hotelfach noch stärker als bislang in den Blickpunkt rücken und die Art des Testens würde sich von der klassischen Testtheorie abwenden und dem Bereich des adaptiven Testes annähern (vgl. für einen Überblick Frey, 2007).

In Kapitel 2 wurde angesprochen, dass aus dem deutschsprachigen Raum noch keine Studien hervorgegangen sind, welche überprüfen, ob Probanden unterschiedlicher Herkunft in SJTs Testwerte ähnlicher oder differierender Höhe erreichen. Die in dieser Arbeit vorgestellten Datenerhebungen adressierten diesen Punkt durch die Einbeziehung der Muttersprache der Teilnehmer als demographische Variable. Da jedoch stets rund 90 % der Teilnehmer Deutsch als ihre Muttersprache angaben, kann auch die vorliegende Arbeit die angesprochene Vakanz nicht adäquat ausfüllen.

Videobasierte SJTs erhielten wie bereits in Abschnitt 2.5 dargestellt keinen Einzug in die Meta-Analysen von McDaniel et al. (2001) und McDaniel et al. (2007). Eine umfassende und systematische Zusammenfassung der kriterienbezogenen Validität videobasierter SJTs steht demnach noch aus. Zudem fristeten SJTs im deutschsprachigen beziehungsweise nicht-englischsprachigen Raum bislang ein Schattendasein. Wenige Ausnahmen liegen zum Beispiel in Form der Arbeiten von Kanning (2005), Schubert et al. (2008), Bledow und Frese (2009) sowie Pöttker (2009) vor.

7.2.2 Implikationen für die Praxis

Mit dem SJT_Hotelfach steht Personalverantwortlichen im Hotelfach ein Instrument zu Verfügung, welches bei Beachtung der bestehenden Forschungslage (s. o.) derzeit nur als Screeninginstrument im Rahmen einer Vorauswahl der Bewerber eingesetzt werden sollte. Der videobasierte SJT_Hotelfach gestattet in diesem Fall eine ökonomische Gruppentestung und ist mit seinem Bestand von 13 Items innerhalb von 20 Minuten durchführbar. Bewerbern, welche im SJT_Hotelfach hohe Testwerte⁶⁶ errei-

⁶⁶ Als Cut-off könnte beispielsweise der Mittelwert (MW = 12.54), welchen die Stichprobe (Video-Bedingung) aus der dritten Studie (s. Kapitel 5) erreicht hat, hinzugezogen werden.

chen, kann bescheinigt werden, dass sie eine stark ausgeprägte Veränderungsbereitschaft und Anpassungsfähigkeit an den Tag legen. Zudem gelingt es diesen Bewerbern leicht, sich in soziale Interaktionspartner hineinzuversetzen und in Zeiten hoher Belastung ihre Widerstands- und Leistungsfähigkeit aufrechtzuerhalten (vgl. Korrelationen mit dem BIP in Kapitel 5). Den Bewerbern wiederum bietet sich in Form der im SJT_Hotelfach aufgezeigten Situationen, welche den Erfahrungen von verschiedenen Beschäftigten im Hotelfach entsprungen sind (s. Kapitel 3), ein realistischer und repräsentativer Einblick in den Hotelalltag beziehungsweise den zentralen Part des Umgangs mit Hotelgästen und deren Eigenarten.

8. Literaturverzeichnis

- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 267-299). New York: Academic Press.
- Allalouf, A. & Ben-Shakhar, G. (1998). The effects of coaching on the predictive validity of scholastic aptitude tests. *Journal of Educational Measurement*, 35, 31-47.
- Alliger, G. M. & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement*, 60, 59-72.
- Anderson, E. W. (1998). Customer satisfaction and word-of-mouth. *Journal of Service Research*, 1, 1-14.
- Anderson, E. W. & Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science*, 12, 125-43.
- Amabile, T. M. (1993). Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Human Resource Management Review*, 3, 185-201.
- Amabile, T. M., Hill, K. G., Hennessey, B. A. & Tighe, E. M. (1994). The Work Preference Inventory: Assessing Intrinsic and Extrinsic Motivational Orientations. *Journal of Personality and Social Psychology*, 66, 950-967.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington: American Educational Research Association.
- American Institutes for Research (AIR). (1998). The critical incident technique. *AIR Web Page* [Online]. Verfügbar unter: www.air.org/airweb/about/critical.html. [Oktober 2007]

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Andreasen, A. (1985). Consumer Responses to Dissatisfaction in Loose Monopolies. *Journal of Consumer Research*, 12, 135-141.
- Andreassen, T. W. (2000). Antecedents to satisfaction with service recovery. *European Journal of Marketing*, 34, 156-175.
- Arthur, W. & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, 93, 435-442.
- Arthur, W., Woehr, D. J. & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical reexamination of the assessment center construct-related validity paradox. *Journal of Management*, 26, 813-835.
- Arvey, R. D., Strickland, W., Drauden, G. & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43, 695-716.
- Asher, J. & Sciarrino, J. (1974). Realistic work sample tests: A review. *Personell Psychology*, 27, 519-533.
- Backhaus, K., Erichson, B. , Plinke, P. & Weiber, R. (2008). *Multivariate Analysemethoden - Eine anwendungsorientierte Einführung*. Berlin: Springer.
- Barrick, M. R. & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, 90, 1185-1203.

- Bauer, T. N., Maertz, C. P., Dolen, M. R. & Campion, M. A. (1998). A longitudinal assessment of applicant reactions to an employment test. *Journal of Applied Psychology*, 83, 892-903.
- Bauer, T. N., Truxillo, D. M., Paronto, M. E., Weekley, J. A. & Campion, 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.
- Beaty, J. C., Jr., Cleveland, J. N. & Murphy, K. R. (2001). The relation between personality and contextual performance in “strong” versus “weak” situations. *Human Performance*, 14, 125-148.
- Beauducel, A. (2001). Problems with parallel analysis in data sets with oblique simple structure. *Methods of Psychological Research*, 6, 141-157.
- Behrmann, M. (2007). Situational Judgment Tests. In H. Schuler & K. Sonntag (Hrsg.), *Handbuch der Personal- und Organisationspsychologie* (S. 483-489). Göttingen: Hogrefe.
- Bergman, M. E., Drasgow, F., Donovan, M. A. & Henning, J. B. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment*, 14, 223-235.
- Berman B. (2006). Developing an effective customer loyalty program. *California Management Review*, 2006, 49, 123-148.
- Bettencourt, L. A. & Brown, S. W. (2003). Role stressors and customer-oriented boundary-spanning behaviors in service organizations. *Journal of the Academy of Marketing Science*, 31, 394-408.
- Beutin, N. (2003). Verfahren zur Messung der Kundenzufriedenheit im Überblick. In C. Homburg (Hrsg.), *Kundenzufriedenheit. Konzepte – Methoden – Erfahrungen* (S. 116-151). Wiesbaden: Gabler.

- Bitner, M. J., Booms, B. H. & Mohr, L. A. (1994). Critical Service Encounters` Viewpoint. *Journal of Marketing*, 58, 95-106.
- Bitner, M. J., Booms, B. H. & Tetreault, M. S. (1990). The service encounter: diagnosing favorable and unfavorable incidents. *Journal of Marketing*, 54, 71-84.
- Bledow, R. & Frese, M. (2009). A Situational Judgment Test of personal initiative and its relationship to performance. *Personnel Psychology*, 62, 229-258.
- Bobko, P., Roth, P. L. & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors and job performance. *Personnel Psychology*, 52, 561-589.
- Bond, H. (1995). Frequent-guest programs build brand loyalty. *Hotel and Motel Management*, 210, 23-30.
- Bono, J. E. & Judge, T. A. (2003). Self-concordance at work: Toward understanding the motivational effects of transformational leaders. *Academy of Management Journal*, 46, 554-571.
- Borkenau, P. & Ostendorf, F. (2008). *NEO-FFI. Das NEO-Fünf-Faktoren-Inventar nach Costa und McCrae*. Göttingen: Hogrefe.
- Borman, W. C. & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.). *Personell selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.
- Bortz, J. (2005). *Statistik für Sozialwissenschaftler*. Berlin: Springer.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation*. Berlin: Springer.

- Boulding, W., Staelin, R., Kalra, A. & Zeithaml, V. (1993). A dynamic process model of service quality: From expectations to behavioral intentions. *Journal of Marketing Research*, 30, 7-27.
- Bowling, N. A. & Gruys, M. L. (2010). Overlooked issues in the conceptualization and measurement of counterproductive work behavior. *Human Resource Management Review*, 20, 54-61.
- Brown, G. H (1952). Brand loyalty - fact or fiction? *Advert Age*, 23, 53-55.
- Brown, T. J., Mowen, J. C., Donovan, T. & Licata, J. W. (2002). The Customer Orientation of Service Workers: Personality Trait Effects on Self- and Supervisor Performance Ratings. *Journal of Marketing Research*, 39, 110-119.
- Bruce, M. M. & Learner, D. B. (1958). A supervisory practices test. *Personell Psychology*, 11, 207-216.
- Bruhn, M. (2003). *Kundenorientierung. Bausteine für ein erfolgreiches Customer-Relationship-Management (CRM)*. München: Oldenbourg.
- Bruhn, M. (2009). Das Konzept der kundenorientierten Unternehmensführung. In H. H. Hinterhuber & K. Matzler (Hrsg.), *Kundenorientierte Unternehmensführung* (S. 33-68). Wiesbaden: Gabler.
- Bühl, A. (2008). *SPSS 16 - Einführung in die moderne Datenanalyse*. München: Pearson.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.
- Bungard, W. (2004). Organisationspsychologische Forschung im Anwendungsfeld. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 121-142). Bern: Huber.

- Cadotte, E. R. & Turgeon, N. (1988). Key factors in guest satisfaction. *Cornell Hotel and Restaurant Quarterly*, 28, 45-51.
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology: Vol. 1* (pp. 687-732). Palo Alto: Consulting Psychologists Press, Inc.
- Campbell, J. P., McCloy, R. A., Oppler, S. H. & Sager, C. E. (1993). *A theory of performance*. San Francisco: Jossey-Bass.
- Canter, J. R. R. (1951). A human relations trainings program. *Journal of Applied Psychology*, 35, 38-45.
- Cardall, A. J. (1942a). *Preliminary manual for the test of practical judgment*. Chicago: Science Research Associates.
- Cardall, A. J. (1942b). *Manual of Directions for the Practical Judgment Test*. Chicago: Science Research Associates.
- Carrington, D. H. (1949). Note on the Cardall practical judgment test. *Journal of Psychology*, 33, 29-30.
- Carter, G. C. (1952). Measurement of Supervisory Ability. *Journal of Applied Psychology*, 36, 393-395.
- Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology*, 1, 169-187.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Chan, D. & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143-159.

- Chan, D. & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15, 233-254.
- Chan, D. & Schmitt, N. (2005) Situational judgment tests. In A. Evers, N. Anderson & O. Voskuijil (Eds.), *Handbook of personnel selection* (pp. 219-246). Oxford: Blackwell.
- Cherry, B., Ordonez, L. D. & Gilliland, S. W. (2003). Grade expectations: The effects of expectations on fairness and satisfaction perception. *Journal of Behavioral Decision Making*, 16, 375-395.
- Christian, M. S., Edwards, B. D. & Bradley, J. C. (2010). Situational Judgment Tests: Constructs Assessed and a Meta-analysis of Their Criterion-related Validities. *Personell Psychology*, 63, 83-117.
- Clevenger, J. P., Pereira, G. M., Wiechmann, D., Schmitt, N. & Schmidt Harvey, V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410-417.
- Conrad, C. A., Brown, G. & Harmon, H. A. (1997). Customer satisfaction and and corporate culture: A profile deviation analysis of a relationship marketing outcome. *Psychology & Marketing*, 14, 663-674.
- Conway, J. M. & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology*, 13, 485-506.
- Cornelsen, J. (2000). *Kundenwertanalysen im Beziehungsmarketing*. Nürnberg: GIM-Verlag.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.
- Costa, P.T. & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI)*.

- Professional Manual*. Odessa: Psychological Assessment Resources.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.
- Cronbach, L. J. (1960). *Essentials of psychological testing*. New York: Harper & Row.
- Cronbach, L. J. & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. Oxford: U. Illinois Press.
- Cronin, J. J. & Taylor, S. A. (1992). Measuring service quality: a re-examination of extension. *Journal of Marketing*, *56*, 55-68.
- Cullen, M. J., Sackett, P. R. & Lievens, F. P. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment*, *14*, 142-155.
- Dabholkar, P. A. (1999). Expectancy value models. In P. E. Earl & S. Kemp (Eds.), *The Elgar Companion to Consumer Research and Economic Psychology* (pp. 201-208). Cheltenham: Edward Elgar.
- Dalessio, A. T. (1994). Predicting insurance agent turnover using a video-based situational judgment test. *Journal of Business and Psychology*, *1*, 23-32.
- Day, G. S. (1969). A two-dimensional concept of brand loyalty. *Journal of Advertising Research*, *9*, 29-35.
- Day, G. S. (2000). Managing market relationships. *Journal of Academy of Marketing Science*, *28*, 24-30.
- Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum.
- De Winter, J. C. F., Dodou, D. & Wieringa, P. A. (2009). Exploratory Factor Analysis With Small Sample Sizes. *Multivariate Behavioral Research*, *44*, 147-181.

- Dick, A. S. & Basu, K. (1994). Customer Loyalty: Toward an Integrated Conceptual Framework. *Journal of Academic Marketing Science*, 22, 99-113.
- DIN (2002). *DIN 33430. Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth Verlag.
- Di Salvo, V. S., Nikkel, E. & Monroe, C. (1989). Theory and practice: A field investigation and identification of group members' perceptions of problems facing natural work group. *Small Group Behavior*, 20, 551-567.
- Drasgow, F., Olson, J. B., Keenan, P. A., Moberg, P. J. & Mead, A. D. (1993). Computerized assessment. In G. R. Ferris & K. M. Rowland (Eds.), *Research in personnel and human resources management*, (pp. 163-206). Greenwich: JAI Press.
- DuBois, P. H. (1970). *A history of psychological testing*. Boston: Allyn & Bacon.
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E. & Cortina, J. C. (2006). A Meta-Analytic Investigation of Conscientiousness in the Prediction of Job Performance: Examining the Intercorrelations and the Incremental Validity of Narrow Traits. *Journal of Applied Psychology*, 91, 40-57.
- Ellingson, J. E., Sackett, P. R. & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, 84, 155-166.
- Endres, J. & Putz-Osterloh, W. (1994). Komplexes Problemlösen in Kleingruppen: Effekte des Vorwissens, der Gruppenstruktur und der Gruppeninteraktion. *Zeitschrift für Sozialpsychologie*, 25, 54-70.

- Eggert, A. (2006). Die zwei Perspektiven des Kundenwerts: Darstellung und Versuch einer Integration. In B. Günter & S. Helm (Hrsg.), *Kundenwert: Grundlagen - Innovative Konzepte - Praktische Umsetzungen* (S. 41-59). Wiesbaden: Gabler.
- Erpenbeck, J. (1997). Selbstgesteuertes, selbstorganisiertes Lernen. In A. Qualifikations-Entwicklungs-Management (Hrsg.), *Kompetenzentwicklung '97 : Berufliche Weiterbildung in der Transformation. Fakten und Visionen* (S. 309-316). Münster: Waxmann.
- Evans, J. R. & Laskin, R. L. (1994). The relationship marketing process: A conceptualization and application. *Industrial Marketing Management*, 23, 439-452.
- Farmer, S. M. & Roth, J. (1998). Conflict-handling behavior in work groups: Effects of group structure, decision processes, and time. *Small group research*, 29, 669-713.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- File, Q. W. (1945). The Measurement of Supervisory Quality in Industry. *Journal of Applied Psychology*, 29, 323-337.
- Fisseni, H.- J. (2004). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Foscht, T. (2001). *Kundenloyalität - Integrative Konzeption und Analyse der Verhaltens- und Profitabilitätsauswirkungen*. Wiesbaden: Gabler.
- Fournier, S., Dobscha, S. & Mick, G. D. (1998). Beziehungsmarketing: Des Guten zuviel für die Stammkäufer. *Harvard Business Manager*, 3, 101-108.

- Frei, R. L. & McDaniel, M. A. (1998). Validity of customer service measures in personell selection: A review of criterion and construct evidence. *Human Performance*, 11, 1-27.
- Frey, A. (2007). Adaptives Testen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 261-278). Berlin: Springer.
- 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, 115-123.
- Gagne', M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26, 331-362.
- Georgi, D. (2001). Einfluss der normativen Erwartungen auf die Transaktionsqualität - Bedeutung der Beziehungsqualität. In M. Bruhn & B. Stauss (Hrsg.), *Dienstleistungsmanagement Jahrbuch 2001. Interaktionen im Dienstleistungsbereich* (S. 91-113). Wiesbaden: Gabler.
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *The Academy of Management Review*, 18, 694-734.
- Gilliland, D. I. & Bello, D. C (2002). Two sides to attitudinal commitment: the effect of calculative and loyalty commitment on enforcement mechanisms in distribution channels. *Journal of Academic Marketing Science*, 30, 24-43.
- Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59, 1216-1229.
- Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, 54, 493-503.

- Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment*, 5, 141-154.
- Grant, K., Cravens, D. W., Low, G. S. & Moncrief, W. C. (2001). The Role of Satisfaction with Territory Design on the Motivation, Attitudes, and Work Outcomes of Salespeople. *Journal of the Academy of Marketing Science*, 29, 165-178.
- Guion R M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah: Erlbaum.
- Hair, J. F., Tatham, R. L., Black, W. C., Anderson, R. E., & Black, B. (2008). *Multivariate Data Analysis*. New Jersey: Prentice Hall.
- Hanson, M. A. & Ramos, R. A. (1996). Situational judgment tests. In R. S. Barrett (Eds.), *Fair employment strategies in human resource management* (pp. 118-124). Westport: Quorum Books.
- Heneman H. G. & Judge T. (2003). *Staffing organizations*. New York: Irwin/McGraw-Hill.
- Herzberg, F., Mausner, B. & Synderman, B. B. (2004). *The motivation to work*. New Brunswick: Wiley & Sons.
- Heskett, J., Sasser, W. E. & Schlesinger, L. A. (1997). *The Service Profit Chain*. New York.
- Hess, R. L., Ganesan, S. & Klein, N. M. (2003). Service failure and recovery: the impact of relationship factors on customer satisfaction. *Journal of Academic Marketing Science*, 31, 127-145.
- Höft, S. & Funke, U. (2006). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 145-188). Göttingen: Hogrefe
- Hogadone, E. & Smith, L. F. (1947). Some Evidence on the Validity of the Cardall Test of Practical Judgment. *Journal of Applied Psychology*, 31, 54-56.

- Hogan, J. & Hogan, R. (1992). *Hogan personality inventory manual*. Tulsa: Hogan Assessment Systems.
- Hogan, J., & Roberts, W. (1996). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior*, 17, 627-637.
- Homburg, C. (2001). *Kundenzufriedenheit. Konzepte - Methoden - Erfahrungen*. Wiesbaden: Gabler.
- Homburg, C. & Bucerius, M. (2003). Kundenzufriedenheit als Managementherausforderung In C. Homburg (Hrsg.), *Kundenzufriedenheit. Konzepte – Methoden – Erfahrungen* (S. 53-86). Wiesbaden: Gabler.
- Homburg, C., Becker, A. & Hentschel, F. (2003). Der Zusammenhang zwischen Kundenzufriedenheit und Kundenbindung. In M. Bruhn & C. Homburg (Hrsg.), *Handbuch Kundenbindungsmanagement: Strategien und Instrumente für ein erfolgreiches Customer-Relationship-Management* (S. 91-121). Wiesbaden: Gabler.
- Homburg, C. & Bruhn, M. (2008): Kundenbindungsmanagement – Eine Einführung. In M. Bruhn & C. Homburg (Hrsg.), *Handbuch Kundenbindungsmanagement. Strategien und Instrumente für ein erfolgreiches Customer Relationship Management* (S. 3-37). Wiesbaden: Gabler.
- Homburg, C. & Giering, A. (2001). Personal characteristics as moderators of the relationship between customer satisfaction and loyalty – An empirical analysis. *Psychology & Marketing*, 18, 43-66.
- Homburg, C. & Stock, R. (2001). Der Zusammenhang zwischen Mitarbeiter- und Kundenzufriedenheit: Eine dyadische Analyse. *Zeitschrift für Betriebswirtschaft*, 71, 789-806.
- Homburg, C. & Werner, H. (1998). *Kundenorientierung mit System: Mit Customer Orientation Management zu profitablen Wachstum*. Frankfurt: Campus.

- Hooper, A. C., Cullen, M. J. & Sackett, P. R. (2006). Operational Threats to the Use of SJTs: Faking, Coaching, and Retesting Issues. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 205-252). Mahwah: Lawrence Erlbaum Associates Publishers.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Hossiep, R. & Paschen, M. (2003). *Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP)*. Göttingen: Hogrefe.
- Hough, L. M. (1992). The “big five” personality variables – construct confusion: Description vs. Prediction. *Human Performance*, 11, 209-244.
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment*, 9, 152-194
- Huber, F., Herrmann, A. & Braunstein, C. (2009). Der Zusammenhang zwischen Produktqualität, Kundenzufriedenheit und Unternehmenserfolg. In H. H. Hinterhuber & K. Matzler (Hrsg.), *Kundenorientierte Unternehmensführung* (S. 69-85). Wiesbaden: Gabler.
- Huck-Schade, J. M. (2003). *Soft Skills auf der Spur. Soziale Kompetenzen: weiche Fähigkeiten – harte Fakten*. Weinheim: Beltz.
- Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Degroot, T. G. & Jones, C. (2001). Comparison of situational and behaviour description interview questions for higher-level positions. *Personnel Psychology*, 54, 619-644.
- Hunt, T. (1928). The measurement of social intelligence. *Journal of Applied Psychology*, 12, 317-334.

- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hurtz, G. M. & Alliger, G. M. (2002). Influence of Coaching on Integrity Test Performance and Unlikely Virtues Scale Scores. *Human Performance*, 15, 255-273.
- Ingenkamp, K. (1997). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Jacoby, J. & Chestnut, R. W. (1978). *Brand Loyalty*. New York: Wiley & Sons.
- Jackson D. N. (1994). *Jackson personality inventory – revised manual*. London: Sigma Assessment Systems.
- Jackson, D. N., Wroblewski, V. R. & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced-choice offer a solution? *Human Performance*, 13, 371-388.
- Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentkontrollierte Weiterentwicklung des deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 18, 195-225.
- Jöreskog, H. G. (1999). *How large can a standardized coefficient be?* [online]. Verfügbar unter: www.ssicentral.com/lisrel/techdocs/HowLargeCanaStandardizedCoefficientbe.pdf. [März 2010]
- Jones, C. & DeCotiis, T. A. (1986). Video-assisted selection of hospitality employees. *The Cornell Hotel and Restaurant Administration Quarterly*, 27, 67-73.
- Jones, T. O. & Sasser, W. E. (1995). Why satisfied costumers defect. *Harvard Business Review*, 73, 88-99.
- Kaiser, H. F. & Dickmann, K. (1959). Analytic determination for common factors. *American Psychologist*, 14, 425-439.

- Kandampully, J. & Suhartanto, D. (2000). Customer loyalty in the hotel industry: The role of customer satisfaction and image. *International Journal of Contemporary Hospitality Management*, 12, 346-351.
- Kanning, U. P. (2002). Soziale Kompetenz: Definition, Strukturen und Prozesse. *Zeitschrift für Psychologie*, 210, 154-163.
- 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. (2008). *Akzeptanz von Assessment Center Übungen bei AC-Teilnehmern*. Manuskript eingereicht zur Publikation. Westfälische-Wilhelms Universität Münster.
- Kanning, U. P. (2009). *Inventar Sozialer Kompetenzen (ISK)*. Göttingen: Hogrefe.
- Kanning, U. P., Grewe, K., Hollenberg, S. & Hadouch, M. (2006). From the Subject's Point of View: Reaction to Different Types of Situational Judgment Items. *European Journal of Psychological Assessment*, 22, 168-176.
- Kanning, U. P., Pöttker, J. & Klinge, K. (2008). *Personalauswahl: Ein Leitfaden für die Praxis*. Stuttgart: Schäffer-Poeschel.
- Kauffeld, S., & Grote, S. (2000). Untersuchungsdesign und Stichprobenbeschreibung. In E. Frieling, E. Kauffeld, S. Grote & H. Bernard (Hrsg.), *Flexibilität und Kompetenz: Schaffen flexible Unternehmen kompetente und flexible Mitarbeiter?* (S. 21-32). Münster: Waxmann.

- Kelley, S. (1992). Developing Customer Orientation among Service Employees. *Journal of the Academy of Marketing Science*, 20, 27-36.
- Kempf, W. & Meder, G. (1993). Theoretische und empirische Untersuchungen zur Restriktivität der Klassischen Testtheorie. *Diagnostica*, 1, 1-21.
- Kihlstrom, J. F. & Cantor, N. (2000). Social Intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 359-379). Cambridge: Cambridge University Press.
- King, C. A. & Garey, J. G. (1997). Relational quality in service encounters. *International Journal of Hospitality Management*, 16, 39-63.
- Klages, L. (1926). *Die Grundlagen der Charakterkunde*. Leipzig: Barth.
- Klapproth, H. E. (1981). Kritik und Weiterentwicklung der Testtheorie: Realisierungsmöglichkeiten moderner Psychodiagnostik. *Psychologie und Praxis*, 25, 139-150.
- Kleinbeck, U. (1991). Die Wirkung von Zielsetzungen auf die Leistung. In H. Schuler (Hrsg.), *Beurteilung und Förderung beruflicher Leistung* (S. 41-56). Göttingen: Hogrefe.
- Kline, P. (1997). *An Easy Guide to Factor Analysis*. London: Routledge.
- Kluger, A. N. & Rothstein, H. R. (1993). The influence of selection test type on applicant reactions to employment testing. *Journal of Business and Psychology*, 76, 698-707.
- Krämer, C. K. (2009). *Eine empirische Überprüfung der Selbst- und Fremdwahrnehmung sozialer Kompetenzen. Analyse und Validierung des Inventars Sozialer Kompetenzen 360 Grad*. Unveröffentlichte Diplomarbeit. Westfälische-Wilhelms Universität Münster.

- Krafft, M. & Marzian, S. (1997). Dem Kundenwert auf der Spur. *Ab-satzwirtschaft*, 40, 104-107.
- Krauth, J. (1995). *Testkonstruktion und Testtheorie*. Weinheim: Psy-chologieVerlagsUnion.
- Krause, A. & Dunckel, H. (2003). Arbeitsgestaltung und Kundenzu-friedenheit. Auswirkungen der Einführung teilautonomer Grup-penarbeit auf die Kundenzufriedenheit unter Berücksichtigung von Mitarbeiterzufriedenheit und Arbeitsleistung. *Zeitschrift für Arbeits- und Organisationspsychologie*, 47, 182-193.
- Kuncel, N. R. & Borneman, M. J. (2007). Toward a New Method of Detecting Deliberately Faked Personality Tests: The use of idio-syncratic item responses. *International Journal of Selection and Assessment*, 15, 220-231.
- Landry, F. J. (2007). The validation of personell decisions in the twenty-first century: back to the future. In S. M. McPhail (Ed.), *Alternate validation strategies: Developing and leveraging exist-ing validity evidence* (pp. 409-426). San Francisco: Josey-Bass.
- Latham, G. P., Saari, L. M., Pursell, E. D. & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.
- Latham G. P & Saari, L. M. (1984). Do people do what they say? Fur-ther studies on the situational interview. *Journal of Applied Psy-chology*, 69, 569-573.
- Latham G. P. & Skarlicki D. P. (1995). Criterion-related validity of the situational and patternedbehavior description interviews with or-ganizational citizenship behavior. *Human Performance*, 8, 67-80.
- Latham, G. P. & Sue-Chan, C. (1999). A Meta-Analysis of the Situ-ational Interview: An Enumerative Review of Reasons for its Va-lidity. *Canadian Psychology*, 40, 56-67.

- Legree, P. (1995). Evidence for an oblique social intelligence factor established with a likert-based testing procedure. *Intelligence*, 21, 247-266.
- Leventhal, G. S. (1980). What should be done with equity theory? In K. J. Gergen, M. S. Greenberg & R. H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27-55). New York: Plenum.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Psychologie Verlags Union.
- Lievens, F. & 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., Buyse, T., & Sackett P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, 90, 442-452.
- Lievens, F. & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, 91, 1181-1188.
- Lievens, F. & Sackett, P. R. (2007). Situational judgment tests in high-stakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology*, 92, 1043-1055.
- Lievens, F. & Peeters, H. (2008). Impact of Elaboration on Responding to Situational Judgment Test Items. *Journal of Selection and Assessment*, 16, 345-355.
- Lievens, F., Sackett, P. R. & Buyse, T. (2009). The Effects of Response Instructions on Situational Judgment Test Performance and Valid-

- ity in a High-Stakes Context. *Journal of Applied Psychology*, 94, 1095-1101.
- Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 3, 157-189.
- Locke, E. A. & Latham, G. P. (1990). *A theory of goal setting & task performance*. Englewood Cliffs: Prentice Hall.
- MacKenzie, S., Podsakoff, P. & Ahearne, M. (1998). Some Possible Antecedents and Consequences of In-Role and Extra-Role Salesperson Performance. *Journal of Marketing*, 62, 87-98.
- Marlowe, D. & Crowne D. P. (1961). Social desirability and response to perceived situational demands. *Journal of Consulting Psychology*, 25, 109-115.
- Martin, W. C., Ponder, N. & Lueg, J. E. (2009). Price fairness perceptions and customer loyalty in a retail context. *Journal of Business Research*, 62, 588-593.
- Mattila, A. S. (2001). The impact of relationship type on customer loyalty in a context of service failures. *Journal of Service Research*, 4, 91-101.
- Matzler, K., Pechlaner, H. & Siller, H. (2001). Die Ermittlung von Basis-, Leistungs- und Begeisterungsfaktoren der Gästezufriedenheit. *Tourismus Journal*, 4, 445-469.
- Matzler, K. & Sauerwein, E. (2002). The Factor Structure of Customer Satisfaction: An empirical Test of the Importance Grid and the Penalty-Reward-Contrast Analysis. *International Journal of Service Industry Management*, 13, 314-332
- Matzler, K., Sauerwein, E. & Stark, C. (2009). Identifikation von Basis-, Leistungs- und Begeisterungsfaktoren. In H. H. Hinterhuber & K. Matzler (Hrsg.), *Kundenorientierte Unternehmensführung* (S. 319-344). Wiesbaden: Gabler

- Maurer, S. D. (2002). A practitioner-based analysis of interviewer job expertise and scale format as contextual factors in situational interviews. *Personnel Psychology, 55*, 307-327.
- Maurer, S. D., Sue-Chan, C. & Latham, G. P. (1999). The situational interview. In R.W. Eder & M. M. Harris (Eds.), *The employment interview handbook* (pp. 159-177). Thousand Oaks: Sage.
- Maurer, T., Solamon, J. & Troxtel, D. (1998). Relationship of coaching with performance in situational employment interviews. *Journal of Applied Psychology, 83*, 128-136.
- McClelland, D. C. (1973). Testing for competence rather than for "intelligence". *American Psychologist, 28*, 1-14.
- McCrae, R. R. & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81-90.
- McDaniel, M. A., Finnegan, E. B., Morgeson, F. P., 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.
- McDaniel, M. A. & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103-113.
- McDaniel, M. A., Schmidt, F. L. & Hunter, J. E. (1988). Job experience correlates of job performance. *Journal of Applied Psychology, 73*, 327-330.
- McDaniel, M. A. & Whetzel, D. L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence, 33*, 515-525.
- McDaniel, M. A., Whetzel, D. L., Hartman, N. S., Nguyen, N. T. & Grubb, W. L. (2006). Situational Judgment Tests: Validity and an

- Integrative Model. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 183-204). Mahwah: Lawrence Erlbaum Associates Publishers.
- McDaniel, M. A., Whetzel, D. L., Hartman, N. S., Nguyen, N. T. & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63-91.
- McFarland, L. A. & Ryan, A. M. (2000). Variance in faking across non-cognitive measures. *Journal of Applied Psychology*, 85, 812-821.
- McFarland, L. A., Ryan, A. M. & Aleksander, E. (2002). Item placement on a personality measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment*, 78, 348-369.
- Meister, U. & Meister, H. (1996). *Kundenzufriedenheit im Dienstleistungsbereich*. München: Oldenbourg.
- Mesmer-Magnus, J. & Viswesvaran, C. (2006). Assessing Response Distortion in Personality Tests: A review of research designs and analytic strategies. In R. L. Griffith & M. H. Peterson (Eds.), *A Closer Examination of Applicant Faking Behavior* (pp. 85-114). Greenwich: Information Age Publishing.
- Messick, S. & Jungeblut, A. (1981). Time and method in coaching for the SAT. *Psychological Bulletin*, 89, 191-216.
- Meyer, A. & Kantsperger, R. & Schaffer, S. (2006). Die Kundenbeziehung als ein zentraler Unternehmenswert. Kundenorientierung als Werttreiber der Kundenbeziehung. In B. Günter & S. Helm (Hrsg.), *Kundenwert: Grundlagen - Innovative Konzepte - Praktische Umsetzungen* (S. 61-81).

- Millard, K. A. (1952). Is How Supervise? an intelligence test? *Journal of Applied Psychology*, 36, 221-224.
- Mimouni-Chaabane, A. & Volle, P. (2010). Perceived benefits of loyalty programs: Scale development and implications for relational strategies. *Journal of Business Research*, 63, 32-37.
- Moosbrugger, H. & Schermelleh-Engel, K. (2007). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (KFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 307-324). Berlin: Springer.
- Morrison Coulthard, L. J. (2004). Measuring service quality. A review and critique of research using SERVQUAL. *International Journal of Market Research*, 46, 479-499.
- Moser, K. (2004). Planung und Durchführung organisationspsychologischer Forschung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 89-120). Bern: Huber.
- Moss, F. A. (1926). Do you know how to get along with people? Why some people get ahead in the world while others do not. *Scientific American*, 135, 26-27.
- Motowidlo, S. J., Borman, W. C. & Schmitt, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, 10, 71-83.
- Motowidlo, S. J., Crook, A. E., Kell, H. J. & Naemi, B. (2009). Measuring Procedural Knowledge More Simply with a Single-Response Situational Judgment Test. *Journal of Business Psychology*, 24, 281-288.
- Motowidlo, S., Dunnette, M. D. & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75, 640-647.

- Motowidlo, S. J., Hanson, M. A. & Crafts, J. L. (1997). Low-fidelity simulations. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement methods in industrial psychology* (pp. 241-260). Palo Alto: Davis Black.
- Mount, M. K. & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in Human Resources Management. *Research in Personnel and Human Resource Management*, 13, 152-200.
- Mumford, T. V., van Iddekinge, C. V., Morgeson, F. P. & Campion, M. A. (2008). The Team Role Test: Development and Validation of a Team Role Knowledge Situational Judgment Test. *Journal of Applied Psychology*, 93, 250-267.
- Nathan, B. R. & Alexander, R. A. (1988). A comparison of criteria for test validation: A meta-analytic investigation. *Personell Psychology*, 41, 517-535.
- Naumann, E. (1995). *Customer Satisfaction Measurement and Management: Using the Voice of the Customer*. Cincinnati: Thomson Executive Press
- Nguyen, N. T. & McDaniel, M. A. (2003). Response instructions and ethnic differences in a situational judgment test. *Applied H. R. M. Research*, 8, 33-44.
- Nerdinger, F. W. (2003). *Kundenorientierung*. Göttingen: Hogrefe.
- Northrop, L. C. (1989). *The psychometric history of selected ability constructs*. Washington: U.S. Office of Personell Management.
- O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. E. & Lawrence, A. (2007). Incremental Validity of Situational Judgment Test for Task and Contextual Job Performance. *International Journal of Selection and Assessment*, 15, 19-30.

- O`Conner, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer`s MAP test. *Behavioral Research Methods, Instruments and Computers*, 32, 396-402.
- Oderkerken-Schröder, G., van Birgelen, M., Lemmink, J. de Ruyter, K. & Wetzels, M. (2000). Moments of sorrow and joy. An empirical assessment of the complementary value of critical incidents in understanding customer service evaluations. *European Journal of Marketing*, 34, 107-125.
- Oliver, R. L. (1980). A cognitive model of the antecedences and consequences of satisfaction decisions. *Journal of Marketing Research*, 17, 46-90.
- Oliver, R. L. (1997). *Satisfaction: A Behavioral Perspective on the Customer*. New York: McGraw-Hill.
- Oliver, R. L. (1999). Whence customer loyalty. *Journal of Marketing*, 63, 33-44.
- Olshavsky, R. & Miller, J. A. (1972). Consumer expectations, product performance, and perceived product quality. *Journal of Marketing Research*, 9, 9-21.
- Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A. & Donovan, M. (1998). The conflict resolution skills assessment: Model-based, multi-media measurement. *Personnel Psychology*, 51, 1-24.
- Olson-Buchanan, J. B. & Drasgow, F. (2006). Multimedia Situational Judgment Tests: The Medium Creates the Message. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 253-278). Mahwah: Lawrence Erlbaum Associates Publishers.
- Ones, D. S. (1993). *The construct validity of integrity tests*. Doctoral dissertation, University of Iowa.

- Ones, D. S., Viswesvaran, C. & Schmidt, F. L. (1993). Meta-analysis of integrity test validities: Findings and implications for personell selection and theories of job performance. *Journal of Applied Psychology, 78*, 679-693.
- Ones, D. S. & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17*, 609-626.
- Ones, D. S. & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personell selection. *Human Performance, 11*, 145-169.
- Ones, D. & Viswesvaran, C. (2007). A research note on the incremental validity of job knowledge and integrity tests for predicting maximal performance. *Human Performance, 20*, 293-303.
- Ostendorf, F. (1990). *Sprache und Persönlichkeitsstruktur: Zur Validität des Fünf-Faktoren-Modells der Persönlichkeit*. Regensburg: Roderer.
- Ostrowsky, P. L., O'Brien, T. V. & Gordon, G. L. (1993). Service quality and customer satisfaction in the commercial airline industry. *Journal of Travel Research, 3*, 16-24.
- 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.
- Parasuraman, A., Zeithaml, V. A. & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing, 49*, 41-50.
- Parasuraman, A., Zeithaml, V. A. & Berry, L. L. (1988). SERVQUAL: A Multiple-Item-Scale for Measuring Customer Perceptions of Service Quality. *Journal of Retailing, 64*, 12-40.

- Parasuraman, A., Zeithaml, V. A. & Berry, L. L. (1991). Refinement and Reassessment of the SERVQUAL scale. *Journal of Retailing*, 67, 420-451.
- Peeters, H. & Lievens, F. (2005). Situational Judgment Tests and their Predictiveness of College Student Success: The influence of faking. *Educational and Psychological Measurement*, 65, 70-89.
- Pepels, W. (2003). Darstellung und Bedeutung des Kundenlebenszeitwerts im Business to Business Marketing. In S. Helmke, M. Uebel & W. Dangelmaier (Hrsg.), *Effektives Customer Relationship Management* (S. 233-268). Wiesbaden: Gabler.
- Phillips, J. F. (1992). Predicting sales skills. *Journal of Business and Psychology*, 7, 151-160.
- Phillips, J. F. (1993). Predicting negotiation skills. *Journal of Business and Psychology*, 8, 403-411.
- Pieters, R., Bottenschen, G. & Thelen, E. (1998). Customer Desire Expectations about Service Employees: An Analysis of Hierarchical Relations. *Psychology & Marketing*, 15, 755-773.
- Pizam, A. & Ellis, T. (1999). Customer satisfaction and its measurement in hospitality enterprises. *International Journal of Contemporary Hospitality Management*, 11, 326-339.
- Ployhart, R. E. & Ehrhart, 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.
- 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., Schneider B. & Schmitt N. (2006). *Staffing organizations: Contemporary practice and research*. Hillsdale/Mahwah: Erlbaum.
- Ployhart, R. E., Weekley, J. A., Holtz, B. C. & Kemp, C. (2003). Web-Based and Paper-and-Pencil Testing of Applicants in a Proctored Setting: Are personality, biodata, and situational judgment comparable? *Personnel Psychology*, 56, 733-752.
- Ployhart, R. E. & Ryan, A. M. (1998). Applicants' Reactions to the Fairness of Selection Procedures: The Effects of Positive Rule Violations and Time of Measurement. *Journal of Applied Psychology*, 83, 3-16.
- Pöttker, J. (2009). *Personaldiagnostik in Ausbildungsberufen des Handwerks: Entwicklung eines Situational Judgment Tests und Überprüfung verschiedener Scoringmethoden*. Münster: Monsenstein & Vannerdat.
- Prakash, V. (1984). Validity and reliability of the confirmation of expectations paradigm as a determinant of consumer satisfaction. *Journal of the Academy of Marketing Science*, 12, 63-76.
- Pulakos, E. D. & Schmitt, N. & Keenan, P. A. (1994). *Validation and implementation of the FBI special agent entry-level selection system*. Alexandria: Human Resources Research Organization.
- Quinn, R. W. (2005). Flow in knowledge work: High performance experience in the design of national security technology. *Administrative Science Quarterly*, 50, 610-641.
- Quinones, M. A., Ford, J. K. & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personell Psychology*, 49, 1-8.
- Ramsey, R. & Sohi, R. (1997). Listening to Your Customers. The Impact of Perceived Salesperson Listening Behavior on Relationship

- Outcomes. *Journal of the Academy of Marketing Science*, 25, 127-137.
- Reichheld, F. F. (1996). *The loyalty effect*. Cambridge: Cambridge University Press.
- Reichheld, F. F. & Kenny, D. W. (1990). The hidden advantages of customer retention. *Journal of Retail Banking*, 12, 19-23.
- Reichheld, F. F. & Sasser, W. (1990). Zero defections: Quality comes to service. *Harvard Business Review*, 68, 105-111.
- 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.
- Rosen, N. A. (1961). How Supervise? – 1943-1960. *Personell Psychology*, 14, 87-99.
- Rosenthal, R. (1994). On being one's own case study: Experimenter effects in behavioural research – 30 years later. In R. William & S. Fuller (Eds.), *The social psychology of science. Conduct of science series* (pp. 214-229). New York: Guilford Press.
- Rosse, J. G., Stecher, M. D., Levin, R. A. & Miller, J. L. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634-644.
- Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Göttingen: Hogrefe.
- Roth, P. L., Bobko, P. & McFarland, L. A. (2005). A Meta-Analysis of Work Sample Test Validity: Updating and integrating some classic literature. *Personell Psychology*, 58, 1009-1037.

- Rust, R. T. & Zahorik, A. J. (1993). Customer satisfaction, customer retention, and market share. *Journal of Retailing*, 69, 193-215.
- Rust, R. T., Zahorik, A. J. & Keiningham, T. L. (1996). *Service Marketing*. New York: HarperCollins.
- 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.
- Ryan, A. M., & Ployhart, R. E. (2003). Customer service behavior. In W. C. Borman, D. R. Ilgen & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (pp. 377-397). London: Wiley.
- Salgado, J. F. (1997). The FiveFactor Model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30-43.
- Salgado, J. F. (2002). The Big Five Personality Dimensions and Counterproductive Behaviors. *International Journal of Selection and Assessment*, 10, 109- 125.
- Salgado, J. F., Viswesvaran, C. & Ones, D. S. (2001). Predictors used in personell selection;an overview of constructs, methods and techniques. In N. Anderson, D. S. Ones , H. K. Sinangil & C. Viswesvaran (Eds.), *International Handbook of Work and Organizational Psychology* (pp. 165-199). London: Sage.
- Sartain, A. Q. (1946). Relation Between Scores on Certain Standard Tests and Supervisory Success in an Aircraft Factory. *Journal of Applied Psychology*, 30, 328-339.
- Saxe, R. & Weitz, B. A. (1982). The SOCO Scale: A Measure of the Customer Orientation of Salespeople. *Journal of Marketing Research*, 19, 343-351.

- Schmidt, F. L. & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability and job knowledge. *Current Directions in Psychological Science*, 2, 8-9.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmidt, F. L., Hunter, J. E. & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, 71, 432-439.
- Schmitt, N. & Chan, D. (2006). Situational Judgment Tests: Method or Construct? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 135-156). Mahwah: Lawrence Erlbaum Associates Publishers.
- Schmitt, N. & Gilliland, S. W. (1992). Beyond differential prediction: Fairness in selection. In D. M. Saunders (Ed.), *New approaches to employee management: Fairness in employee selection* (pp. 21-46). Greenwich: JAI Press.
- Schmitt, N., Gooding R. Z., Noe, R. A. & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422.
- Schmitt, N. & Mills, A. E. (2001). Traditional test and job simulations: Minority and majority performance and test validities. *Journal of Applied Psychology*, 86, 451-458.
- Schmit, M. J. & Ryan, A. M. (1992). Test-taking dispositions: A missing link? *Journal of Applied Psychology*, 77, 629-637.

- Schneider, B., Ehrhart, M. G., Mayer, D. M. & Saltz, J. L. (2005). Understanding organizational – customer linkages in service settings. *Academy of Management Journal*, 48, 1017-1032.
- Schubert, S., Ortwein, H., Dumitsch, A., Schwantes, U., Wilhelm, O. & Kiessling, C. (2008). A situational judgment test of professional behavior: Development and validation. *Medical Teacher*, 30, 528-533.
- Schuler, H. (1993). Social validity of selection situations: A concept and some empirical results. In H. Schuler, J. L. Farr & M. Smith (Eds.), *Personell selection and assessment: Individual and organizational perspectives* (pp. 11-26). Hillsdale: Erlbaum.
- Schuler, H. (2000). *Psychologische Personalauswahl*. Göttingen: Hogrefe.
- Schuler, H. (2004). *Lehrbuch Organisationspsychologie*. Bern: Verlag Hans Huber.
- Schuler, H. & Höft, S. (2004). Diagnose beruflicher Eignung und Leistung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 289-343). Bern: Verlag Hans Huber.
- Schuler, H. & Höft, S. (2006). Konstruktorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 101-144). Göttingen: Hogrefe.
- Schuler, H. & Marcus, B. (2006). Biographieorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 189-226). Göttingen: Hogrefe.
- Schuler, H. & Stehle, W. (1983). Neuere Entwicklungen des Assessment-Center-Ansatzes – beurteilt unter dem Aspekt der sozialen Validität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 27, 33-44.

- Snell, A. F., Sydell, E. J. & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, 9, 219-242.
- Smiderle, D., Perry, B. A. & Cronshaw, S. F. (1994). Evaluation of video-based assessment in transit operator selection. *Journal of Business and Psychology*, 1, 3-22.
- Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K. & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, 46, 49-76.
- Stauss, B. (1999). Kundenzufriedenheit. *Marketing-Zeitschrift für Theorie und Praxis*, 21, 5-24.
- Sternberg, R. J., Wagner, R. K. & Okagaki, L. (1993). *Practical intelligence: The nature and role of tacit knowledge in work and at school*. New Haven: Yale University Press.
- Stevens, M. A. & Campion, M. J. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management*, 25, 207-228.
- Stitt-Gohdes, W. L., Lambrecht, J. J. & Redmann, D. H. (2000). The Critical-Incident Technique in Job Behavior Research. *Journal of Vocational Education Research*, 25, 59-84.
- Strauß, B. & Kleinmann, M. (1995). *Computersimulierte Szenarien in der Personalarbeit*. Göttingen: Hogrefe.
- Streiner, D. L. (2006). Sample Size in Clinical Research: When is Enough Enough? *Journal of Personality Assessment*, 87, 259-260.
- Tabachnik, B. G. & Fidell, L. S. (2004). *Using Multivariate Statistics*. New York: HarperCollins.

- Taylor, P. J. & Small, B. (2002). Asking applicants what they would do versus what they did do: a meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology*, 75, 277-294.
- Tesluk, P. E. & Jacobs, R. R. (1998). Toward an integrated model of work experience. *Personell Psychology*, 51, 321-355.
- Tett, R. P., Jackson, D. N. & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44, 703-742,
- Thorndike, E. L. (1920). The measurement of social intelligence. *Journal of Applied Psychology*, 12, 317-334.
- Thorndike, E. L. (1949). *Personell selection; test and measurement techniques*. Oxford: Wiley.
- Thorndike, E. L. & Stein, S. (1937). An evaluation of the attempts to measure social intelligence. *Psychological Bulletin*, 34, 275-289.
- Thornton, G. C. & Byham, W. C. (1982). *Assessment centers and managerial performance*. New York: Academic Press.
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, 1. Chicago: University of Chicago Press.
- Tupes, E. C. & Christal, R. C. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality*, 60, 225-252.
- Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.
- Uncles, M. D., Dowling, G. D. & Hammond, K. (2003). Customer loyalty and customer loyalty programs. *Journal of Consumer Marketing*, 20, 294-316.

- Van Iddekinge, C. H., Raymark, P. H., Eidson, C. E. & Attenweiler, W. J. (2004). What do structured selection interviews really measure? The construct validity of behavior description interviews. *Human Performance*, *17*, 71-93.
- VanYperen, N. W. & Van de Vliert, E. (2006). Sozialpsychologie in Organisationen. In W. Stroebe, K. Jonas & M. Hewstone (Hrsg.), *Sozialpsychologie – Eine Einführung* (S. 623-655). Berlin: Springer.
- Vasilopoulos, N. L., Reilly, R. R. & Leaman, J. A. (2000). The Influence of Job Familiarity and Impression Management on Self-Report Measure Scale Scores and Response Latencies. *Journal of Applied Psychology*, *85*, 50-64.
- Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L. & Reilly, R. R. (2006). Forced-Choice Personality Tests: A Measure of Personality and Cognitive Ability? *Human Performance*, *19*, 175-199.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321-327.
- Viswesvaran, C. & Ones, D. S. (2000). Measurement error in “Big Five Factors” personality measurement. *Educational and Psychological Measurement*, *60*, 197-210.
- Weekley, J. A. & Jones, C. (1997). Video-based situational testing. *Personell Psychology*, *50*, 25-49.
- Weekley, J. A. & Jones, C. (1999). Further studies of situational tests. *Personell Psychology*, *52*, 679-700.
- Weekley, J. A. & Ployhart, R. E. (2005). Situational judgment: Antecedents and relationships with performance. *Human Performance*, *18*, 81-104.

- Weekley, J. A., Ployhart, R. E. & Harold, C. (2004). Personality and situational judgment tests across applicant and incumbent contexts: An examination of validity, measurement, and subgroup differences. *Human Performance*, *17*, 433-461.
- Weekley, J. A., Ployhart, R. E. & Holtz, B. C. (2006). On the Development of Situational Judgment Tests: Issues in Item Development, Scaling, and Scoring. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 157-182). Mahwah: Lawrence Erlbaum Associates Publishers.
- Weinert, A. B. (2004). *Organisationspsychologie: Ein Lehrbuch*. Weinheim: Psychologie Verlags Union.
- Wernimont, P. & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, *52*, 372-376.
- Whetzel, D. L., McDaniel, M. A. & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, *21*, 291-309.
- Whetzel, D. L. & McDaniel, M. A. (2008). Situational judgment tests: An overview of current research. *Human Resource Management Review*, *19*, 188-202.
- Whyte, W. H. (1975). How to cheat on personality tests. In A. Liebers (Ed.), *How to pass employment tests* (pp. 96-100). New York: ARCO.
- Wong Ooi Mei, A., Dean, A. M. & White, C. J. (1999). Analysing service quality in the hospitality industry. *Managing Service Quality*, *9*, 136-143.
- Yousfi, S. (2005). Mythen und Paradoxien der klassischen Testtheorie (I): Testlänge und Gütekriterien. *Diagnostica*, *51*, 1-11.

Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. *Journal of Marketing*, 52, 2-22.

Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

Abbildungsverzeichnis

<i>Abbildung 2-1</i>	Zentrale Punkte in der Historie von SJTs	S. 10
<i>Abbildung 2-2</i>	Leitfaden zur Erstellung von SJTs in der Praxis	S. 32
<i>Abbildung 2-3</i>	Einzelne Verfahren der Personalauswahl und ihr Grad an fidelity	S. 34
<i>Abbildung 2-4</i>	Ein Item aus dem Videotest von Jones und DeCotiis (1986)	S. 37
<i>Abbildung 2-5</i>	Validitäten von Verfahren der Personalauswahl im Überblick	S. 49
<i>Abbildung 2-6</i>	Heuristisches Modell von McDaniel et al. (2006)	S. 64
<i>Abbildung 2-7</i>	Erfolgskette der Kundenorientierung	S. 78
<i>Abbildung 3-1</i>	Methodisches Vorgehen bei der Konstruktion des SJT	S. 93
<i>Abbildung 3-2</i>	Vorlage zur Instruktion der Interviewpartner nach Bitner et al. (1994)	S. 97
<i>Abbildung 3-3</i>	Beispiele für Situationen, welche bei Hotelgästen beziehungsweise Unzufriedenheit ausgelöst haben	S. 99
<i>Abbildung 3-4</i>	Hierarchisches Modell der Kundenorientierung	S. 100
<i>Abbildung 3-5</i>	Beispiele für geschlechtsspezifische Situationen	S. 101
<i>Abbildung 3-6</i>	Beispiel für ein Drehskript	S. 106

<i>Abbildung 3-7</i>	Interaktionsmuster I (Hotelmitarbeiter steht, Gast steht)	S. 108
<i>Abbildung 3-8</i>	Interaktionsmuster II (Hotelmitarbeiter steht, Gast sitzt)	S. 109
<i>Abbildung 3-9</i>	Interaktionsmuster III (Hotelmitarbeiter steht, zwei Gäste sitzen)	S. 109
<i>Abbildung 4-1</i>	Beispielitem aus dem ISK-K von Kanning (2009)	S. 115
<i>Abbildung 4-2</i>	Definitionen sozial kompetenten Verhaltens und sozialer Kompetenz nach Kanning (2002)	S. 116
<i>Abbildung 4-3</i>	Sekundärfaktoren des ISK-K und Itembeispiele	S. 120
<i>Abbildung 4-4</i>	Beispielitem aus dem SJT	S. 120
<i>Abbildung 4-5</i>	Scoring key in Anlehnung am Moto-widlo et al. (1990)	S. 122
<i>Abbildung 4-6</i>	Mess- und Strukturmodell der konfirmatorischen Faktorenanalysen	S. 130
<i>Abbildung 4-7</i>	Scree-Plot	S. 135
<i>Abbildung 4-8</i>	Grafische Gegenüberstellung eines empirischen mit einem zufällig generierten Eigenwerteverlauf	S. 136
<i>Abbildung 5-1</i>	Umrechnung der Rohwerte im SJT (I)	S. 162
<i>Abbildung 5-2</i>	Umrechnung der Rohwerte im SJT (II)	S. 163

Tabellenverzeichnis

<i>Tabelle 2-1</i>	Mittlere Korrelationen zwischen SJTs und Persönlichkeitsfaktoren	S. 62
<i>Tabelle 3-1</i>	Charakteristika der befragten SME	S. 96
<i>Tabelle 3-2</i>	Zuordnung der 25 SJT-Items zu den fünf Anforderungsdimensionen	S. 105
<i>Tabelle 4-1</i>	Aufbau des Fragebogens im Rahmen der zweiten Studie	S. 114
<i>Tabelle 4-2</i>	Erreichte Schulabschlüsse der Berufsschüler	S. 126
<i>Tabelle 4-3</i>	Deskriptive Statistiken der vier Hauptskalen des ISK-K	S. 126
<i>Tabelle 4-4</i>	Ergebnisse der Parallelanalyse	S. 136
<i>Tabelle 4-5</i>	Ergebnisse des MAP-Tests	S. 137
<i>Tabelle 4-6</i>	Faktorladungen auf dem extrahierten Faktor	S. 138
<i>Tabelle 4-7</i>	Zentrale Statistiken der Skala SJT_kurz und deren Items	S. 139
<i>Tabelle 4-8</i>	Trennschärfen der Items aus der Skala SJT_kurz	S. 140
<i>Tabelle 4-9</i>	Schwierigkeitsindizes der Items aus der Skala SJT_kurz	S. 142
<i>Tabelle 4-10</i>	Komplett richtig- und Komplett falsch-Antworten je Item	S. 143
<i>Tabelle 4-11</i>	Veränderungen der internen Konsistenz bei Entfernung eines Items	S. 145

<i>Tabelle 4-12</i>	Korrelationen zwischen der Skala SJT_kurz und den vier Hauptskalen des ISK-K	S. 146
<i>Tabelle 4-13</i>	Ergebnisse der schrittweisen Regression	S. 149
<i>Tabelle 5-1</i>	Kurzdefinitionen der mit dem BIP erfassten Konstrukte nach Hossiep und Paschen (2003)	S. 173
<i>Tabelle 5-2</i>	Zeitplan am Tag der Datenerhebung	S. 178
<i>Tabelle 5-3</i>	Deskriptive Statistiken des CFT-20-R	S. 181
<i>Tabelle 5-4</i>	Mittelwerte der Dimensionen sozialer Validität	S. 183
<i>Tabelle 5-5</i>	Zentrale Skalenstatistiken des BIP	S. 185
<i>Tabelle 5-6</i>	Faktorladungen auf dem extrahierten Faktor	S. 188
<i>Tabelle 5-7</i>	Zentrale Statistiken der Skala SJT_Hotelfach	S. 189
<i>Tabelle 5-8</i>	Vergleich der Rangordnungen (Video/Paper-Pencil)	S. 192
<i>Tabelle 5-9</i>	Trennschärfen der Items aus der Skala SJT_Hotelfach	S. 193
<i>Tabelle 5-10</i>	Schwierigkeitsindizes der Items aus der Skala SJT_Hotelfach	S. 194
<i>Tabelle 5-11</i>	Veränderungen der Reliabilität bei Eliminierung eines Items	S. 195
<i>Tabelle 5-12</i>	Korrelationen zwischen dem SJT_Hotelfach und Maßen der Intelligenz	S. 197

<i>Tabelle 5-13</i>	Bivariate Korrelationen zwischen dem SJT_Hotelfach und den Eignungsvoraussetzungen des BIP	S. 199
<i>Tabelle 5-14</i>	Einzelkorrelationen zwischen Items des SJT und Dimensionen des BIP	S. 205
<i>Tabelle 6-1</i>	Fragen zur Messung der Berufserfahrung	S. 221
<i>Tabelle 6-2</i>	Kurzbeschreibungen der fünf Persönlichkeitsfaktoren	S. 223
<i>Tabelle 6-3</i>	Deskriptive Statistiken der Skala SJT_Hotelfach	S. 228
<i>Tabelle 6-4</i>	Vergleich der Rangordnungen	S. 229
<i>Tabelle 6-5</i>	Deskriptive Statistiken der fünf Faktoren aus dem NEO-FFI	S. 230
<i>Tabelle 6-6</i>	Veränderungen der Reliabilität bei Entfernung eines Items	S. 231
<i>Tabelle 6-7</i>	Trennschärfen der Items aus dem SJT_Hotelfach	S. 232
<i>Tabelle 6-8</i>	Aufgabenschwierigkeiten der Items aus dem SJT_Hotelfach	S. 233
<i>Tabelle 6-9</i>	Korrelationen zwischen der Skala SJT_Hotelfach und den Hauptskalen des NEO-FFI	S. 235
<i>Tabelle 6-10</i>	Schrittweise Regression zur Vorhersage der Leistung im SJT_Hotelfach	S. 236

Anhang A: Erste Studie

A.1 Beschreibung der Anforderungsdimensionen

1. Serviceorientierung

Serviceorientierte Mitarbeiter legen gegenüber dem Hotelgast eine transparente und nachvollziehbare Vorgehensweise an den Tag. Unklarheiten, Missverständnisse oder Probleme in der Mitarbeiter-Gast Interaktion werden frühzeitig erkannt und behoben. Der Mitarbeiter vermittelt dem Gast ein Gefühl der Geborgenheit und Sicherheit in dem Hotel und dient als jederzeit erreichbarer Ansprechpartner. Mitarbeiter mit einer weniger stark ausgeprägten Serviceorientierung erkennen viele Wünsche des Gastes nicht und beschränken die Kommunikation mit dem Gast auf das Nötigste.

2. Beschwerdemanagement

Mitarbeiter, welche das Beschwerdemanagement beherrschen, nehmen berechtigte Kritik und Verbesserungsvorschläge von Hotelgästen ernst. Dabei entschuldigen sie sich für aufgetretene Fehler und übernehmen die Verantwortung für das Ausmerzen von Irrtümern oder Versäumnissen seitens des Hotels. Außerdem nehmen sie sich die notwendige Zeit für die Anliegen von Hotelgästen und zeigen Verständnis für deren Beschwerden. In Diskussionen mit Gästen bleiben sie stets ruhig und sachlich. Selbst gegenüber sehr schwierigen Gästen behalten sie die Souveränität. Dagegen bringen Mitarbeiter, welchen das Beschwerdemanagement Probleme bereitet, viele Beschwerden von Kunden nicht ernst und bringen kein Verständnis für legitime Beanstandungen von Hotelgästen auf. Bei Meinungsverschiedenheiten mit Hotelgästen verlieren sie schnell die Contenance und vergreifen sich im Ton, so dass oftmals unbeteiligte Gäste unter dem Streit zu leiden haben.

3. Durchsetzungsfähigkeit

Durchsetzungsfähige Mitarbeiter agieren selbst gegenüber sehr schwierigen Gästen selbst- und pflichtbewusst. Sie treten in den erforderlichen Situationen für die Einhaltung der Hotelordnung oder die Wahrung der Rechte anderer Hotelgäste ein. Falls es notwendig ist, sprechen sie passende Sanktionen oder Verbote aus. Der Erfüllung unberechtigter oder

überzogener Gastwünsche treten sie energisch entgegen. Mitarbeiter mit geringerer Durchsetzungsfähigkeit gehen Auseinandersetzungen mit problematischen Hotelgästen aus dem Weg und ignorieren deren Fehlverhalten. Sie lassen sich von Gästen oftmals unrealistische oder unangemessene Zugeständnisse abringen.

4. Handlungsfähigkeit

Handlungsfähige Mitarbeiter wissen um ihren Handlungsspielraum in der Betreuung von Hotelgästen und nutzen diesen entsprechend aus. Auf Anfragen von Gästen wird flexibel reagiert und dem Gast können verschiedene Angebotsalternativen präsentiert werden. Weniger handlungsfähige Mitarbeiter handeln in starren Bahnen und nutzen ihre zugeteilten Kompetenzen nicht voll aus.

5. Gewissenhaftigkeit

Gewissenhafte Mitarbeiter zeigen in der Interaktion mit Kunden stets eine loyale Einstellung gegenüber ihrem Arbeitgeber und ihren Kollegen. Die Philosophie und Grundsätze des Hotels werden nach außen gelebt. Vor Gästen werden Arbeitskollegen, welche Fehler begangen haben, in Schutz genommen. Bei gewichtigen Problemen von Kunden werden notfalls Vorgesetzte hinzugezogen und konsultiert. Der vorgegebene Dienstweg wird gegenüber Kunden immer eingehalten. Allen Gästen, ob neu oder seit längerer Zeit dem Hotel verbunden, wird dieselbe Serviceleistung zuteil. Mitarbeiter mit geringer ausgeprägter Gewissenhaftigkeit fühlen sich nicht verantwortlich für Fehler oder Versäumnisse von Kollegen. Bei der Bewältigung von Kundenproblemen nehmen sie Vorschriften nicht allzu streng und bewegen sich oft abseits der Vorschriften.

A.2 Verhaltensalternativen je Item im SJT

Item 1:

- A. Sie begleiten den Gast zum Saal, damit Sie sich das Problem ansehen können und gegebenenfalls selbst beheben können.
- B. Sie bitten den Gast freundlich, den Personalbereich zu verlassen, da der Zutritt in diesen Bereich des Hotels für Gäste untersagt ist.
- C. Sie bitten den Gast um Geduld, da Sie einen Haustechniker informieren wollen, der sich des Problems annimmt.
- D. Sie verweisen den Gast an die Rezeption, um dort um Hilfe zu bitten, da Sie bereits Feierabend haben.

Item 2:

- A. Sie bitten den Gast um Verständnis, dass Sie in diesem Fall nichts für ihn tun können. Immerhin könnte der Gast den Wein ja selbst ausgetrunken haben.
- B. Sie entschuldigen sich beim Gast und rufen unverzüglich den Zimmerservice an und bitten um die sofortige Auffüllung des Weins in der Minibar auf Kosten des Hauses.
- C. Sie bitten den Gast um Geduld, da Sie versuchen wollen, die Vormieter des Zimmers anzurufen, um diese zu bitten, für den entstandenen Schaden aufzukommen.
- D. Sie bieten dem Gast als Wiedergutmachung einen Gutschein für einen Wein an der Hotelbar an.

Item 3:

- A. Sie ermuntern den Gast, sich beim nächsten Anlass direkt zu beschweren, da Sie sich gerne sofort um sein Wohl gekümmert hätten.
- B. Sie bedauern, dass es dem Gast nicht geschmeckt hat und versprechen ihm, dass sich die Küche für ihn besonders ins Zeug legt, wenn er wiederkommt.
- C. Sie bieten dem Gast ein weiteres Hauptgericht an, welches sie ihm nicht berechnen würden.

D. Sie erinnern den Gast freundlich daran, dass er aufgegegessen hat und dass es ihm dementsprechend geschmeckt haben muss, so dass Sie seine Beschwerde nicht nachvollziehen können.

Item 4:

- A. Sie verzichten auf eine weitere Bedienung der Gäste, da diese sich nicht an die vorgegebene Ordnung halten und übergeben den Tisch einem Kollegen.
- B. Sie lassen die Gäste ungestört weiterräumen und bedienen diese wie gehabt.
- C. Sie informieren den Veranstalter über die rauchenden Gäste und bitten ihn, das Rauchverbot selbst durchzusetzen.
- D. Sie bieten den rauchenden Gästen an, für sie einen Stehtisch im Flur außerhalb der Tagung herzurichten, an welchem sie in Ruhe rauchen können.

Item 5:

- A. Sie stellen dem Gast auf Kosten des Hauses eine Monatskarte für Ihre Garage aus, damit ihm beim nächsten Besuch der Ärger über die Garagegebühr erspart bleibt.
- B. Sie erlassen dem Gast wie von ihm gewünscht die Parkgebühr.
- C. Sie erklären dem Gast höflich, dass die Benutzung der hoteleigenen Garage der Inanspruchnahme einer Dienstleistung entspricht und dass alle Dienstleistungen bezahlt werden müssen.
- D. Sie bitten den Gast um die Herausgabe der Restaurantrechnung und überprüfen den Betrag. Die Parkgebühr fordern Sie in jedem Fall ein.

Item 6:

- A. Sie erklären den Gästen freundlich, dass das Menü leider die einzige Möglichkeit ist, die sich ihnen als Gruppe bietet, da sich ein Büffet für solch kleine Gruppen seitens Ihres Hotels nicht rentiert.
- B. Sie sprechen mit dem Küchenchef, ob die Zubereitung eines Büffets für die kleine Gruppe ausnahmsweise möglich ist.
- C. Sie erklären den Gästen, dass sie mit einem kleinen Büffet nicht zufrieden wären, da es nicht so vielfältig und reichhaltig wäre wie ein

normales Büffet. Stattdessen legen Sie ihnen ein individuelles Menü nahe.

- D. Sie weisen die Gäste auf alternative Angebote, z. B. spezielle Business-Gerichte, in Ihrem Restaurant hin, welche schnell zubereitet werden können.

Item 7:

- A. Sie entsprechen dem Wunsch des Gastes und nehmen die weiteren bestellten Getränke von der Rechnung.
- B. Sie klären den Gast höflich auf, dass Weine aus Flaschen mit Drehverschluss nicht korkig sein können und empfehlen ihm andere Getränke aus Ihrem Sortiment.
- C. Sie erklären dem Gast höflich, dass eine Stornierung der Getränke in solchen Fällen nicht möglich ist und bitten ihn, andere Getränke zu bestellen.
- D. Sie versuchen den Gast doch noch umzustimmen und für andere Weine zu begeistern, die Ihrer Meinung nach gut schmecken.

Item 8:

- A. Sie bitten den Gast auf höfliche Weise auf, zur Seite zu treten und zu warten, bis der Kirschkuchen aufgefüllt ist, damit Sie in der Zwischenzeit die anderen wartenden Gäste bedienen können.
- B. Sie bitten einen Kollegen, in der Küche nach dem Kirschkuchen zu schauen und dem Gast ein Stück Kuchen zum Sitzplatz zu bringen.
- C. Sie empfehlen dem Gast zur Überbrückung der Wartezeit weitere Obstkuchen, die Ihrer Meinung nach besonders schmackhaft sind.
- D. Sie bitten den Gast, in Ruhe und ohne Stress auf seinem Platz zu warten, bis der Kuchen beim Büffet wieder aufgefüllt ist.

Item 9:

- A. Sie rufen die zwei Gruppen an, die den Wintergarten gebucht haben und fragen jene, ob sie bereit wären zu tauschen.
- B. Sie führen den Gast durch das Restaurant und empfehlen ihm andere Plätze für seinen Besuch, die im Vergleich zum ausgebuchten Wintergarten ruhiger und intimer sind.

- C. Sie erklären dem Gast, dass Sie leider nichts für ihn tun können, da der Wintergarten besetzt ist und andere Plätze nicht dasselbe Ambiente bieten.
- D. Sie zeigen dem Gast andere Bereiche in Ihrem Restaurant, die noch nicht reserviert sind, so dass er sich einen Platz seiner Wahl aussuchen kann.

Item 10:

- A. Sie erklären dem Gast höflich, dass Sie nichts für ihn machen können, wenn die Garage besetzt ist, da andere Gäste auch ein Recht auf einen Parkplatz besitzen.
- B. Sie gestatten dem Gast, sein Auto ausnahmsweise auf den Personalparkplätzen hinter dem Hotel stehen zu lassen, damit er seinen wichtigen Termin wahrnehmen kann.
- C. Sie zeigen dem Gast auf einem Stadtplan alternative, gut erreichbare Parkplätze in der Umgebung.
- D. Sie empfehlen dem Gast, demnächst mit dem Bus oder dem Taxi zu kommen, damit er sich die nervenaufreibende Parkplatzsuche erspart.

Item 11:

- A. Sie bitten den Gast, diese Nacht durchzuhalten, damit der Hausmeister das Problem am nächsten Tag sofort beheben kann.
- B. Sie bitten den Gast um Geduld und versuchen denjenigen Haustechniker zu erreichen, welcher Bereitschaftsdienst hat.
- C. Sie bieten dem Gast für diese Nacht ein anderes freies Zimmer an, damit er ruhig schlafen kann.
- D. Sie bitten den Gast, an der Hotelbar Hilfe zu suchen, da Sie an der Rezeption ganz alleine sind und sich um andere Gäste kümmern müssen.

Item 12:

- A. Sie erklären den Gästen, dass die Größe ihrer Speisen der normalen Größe in Ihrem Hause entspricht und bieten ihnen weitere Beilagen an, falls sie später noch Hunger haben soll.

- B. Sie versichern den Gästen höflich, dass sie sich ausnahmsweise irren müssen, da in ihrem Hause stets gleichgroße Portionen ausgegeben werden.
- C. Sie nehmen den Fisch wieder mit und bestellen in der Küche einen kräftigen Nachschlag.
- D. Sie nehmen die Teller der Gäste wieder mit in die Küche und versuchen abzuklären, ob beim Auffüllen der Speisen etwas schief gelaufen ist.

Item 13:

- A. Sie erklären dem Gast, dass das Pay-TV-Angebot durch eine externe Firma abgerechnet wird und dass der Gast in jedem Fall vorab bezahlen muss. Sobald die Angelegenheit geklärt ist, wollen Sie den Gast telefonisch oder per Post benachrichtigen.
- B. Sie befragen einen Haustechniker oder den Hausmeister, ob tatsächlich ein technischer Fehler vorlag. Bis zur Klärung des Sachverhalts bitten Sie den Gast an der Rezeption zu warten.
- C. Sie ersparen dem Gast weitere unangenehme Ausführungen und stornieren die Pay-TV Gebühr.
- D. Sie lesen dem Gast die Namen der Filme vor, die er über Pay-TV angesehen hat, da er sich anscheinend nicht an diese erinnert und bitten ihn, zu bezahlen.

Item 14:

- A. Sie bringen den Gästen die Espressi, welche Sie nicht berechnen.
- B. Sie berechnen den Gästen jeweils nur eine Kugel Eis. Einen Espresso auf Kosten des Hauses erhalten sie jedoch nicht.
- C. Sie bitten die Gäste um Geduld und ziehen ihren Vorgesetzten hinzu, der die Sache klären soll.
- D. Sie bitten die Gäste, sich demnächst sofort bei Ihnen zu melden, wenn sie einen Fehler in der Bestellung bemerken, denn nur dann können Sie reagieren. Da die Gäste in diesem Fall aufgegessen haben, können Sie nichts mehr für sie tun.

Item 15:

- A. Sie verzichten auf die Begleichung der Garagengebühr und berechnen nur den Zimmerpreis für zwei Nächte.
- B. Sie bitten den Gast um Verständnis und erklären ihm höflich, dass die Benutzung der hoteleigenen Garage der Inanspruchnahme einer Dienstleistung entspricht und dass alle Dienstleistungen bezahlt werden müssen.
- C. Sie stellen dem Gast nur die Garagengebühr für einen Tag in Rechnung und erlassen ihm ausnahmsweise die Gebühr für den zweiten Tag.
- D. Sie weisen den Gast höflich auf die Preise der Garage, die für jeden Gast gut sichtbar an der Garageneinfahrt angebracht sind, hin und bitten ihn, die Parkgebühr so wie alle anderen Gäste zu bezahlen.

Item 16:

- A. Sie bitten die Gäste, an der Hotelbar zu warten und versprechen, sie schnellstmöglich zu informieren, wenn ein Platz frei wird.
- B. Sie bitten die Gäste, beim nächsten Mal vorher zu reservieren, um sich Ärger zu ersparen, und bieten ihnen einen schönen Platz für den nächsten Abend an.
- C. Sie fragen andere Gäste höflich, ob sich noch drei Personen zu ihnen an den Tisch setzen könnten.
- D. Sie bitten die Gäste, im Restaurant zu warten, um schnell einen Tisch zu ergattern, wenn etwas frei wird.

Item 17:

- A. Sie erklären dem Gast, dass Sie ihm leider kein Zimmer anbieten können, wenn er den Kreditkartenabzug ablehnt.
- B. Sie erklären dem Gast, dass die Geschäftsbedingungen Ihres Hotels einen Kreditkartenabzug bei Neukunden vorschreiben und dass dies ein einmaliger Vorgang ist.
- C. Sie verweisen den Gast auf die Tatsache, dass der Kreditkartenabzug bei Neukunden eine Anweisung Ihrer Vorgesetzten ist, an die Sie sich zu halten haben.
- D. Sie verzichten auf den Kreditkartenabzug, da Sie dem Gast nicht das Gefühl geben wollen, ihm zu misstrauen.

Item 18:

- A. Sie räumen den Teller ab und streichen das Essen von der Rechnung, da es fast nicht angerührt wurde.
- B. Sie bitten den Gast höflich, demnächst nur das bestellen, was er auch tatsächlich isst, damit ihm keine unnötigen Kosten entstehen.
- C. Sie versuchen, freundlich zu erfragen, warum der Gast nichts gegessen hat.
- D. Sie nehmen den Teller mit und bringen dem Gast wie gewünscht die Rechnung.

Item 19:

- A. Sie berechnen den Gästen ausnahmsweise den Angebotspreis mit dem nochmaligen Hinweis, dass dieses spezielle Angebot sonst erst ab 15 Uhr gilt.
- B. Sie bitten die Gäste um Geduld, da Sie den Sachverhalt zunächst mit Ihrem erwähnten Kollegen klären möchten.
- C. Sie bestehen auf den normalen Preis.
- D. Sie bitten die Gäste, das Restaurant zu verlassen, da Sie das Gefühl haben, diese Gäste wollen Sie über den Tisch ziehen.

Item 20:

- A. Sie bieten dem Gast als Alternative zum Frühstück eine Reservierung für ein Mittag- oder Abendessen an demselben Tag an.
- B. Sie erklären dem Gast, dass er den Aufenthalt in Ihrem überfüllten Restaurant ohnehin nicht genießen würde und bieten ihm eine Reservierung für den Abend oder Mittag desselben Tages an.
- C. Sie empfehlen dem Gast andere vorzügliche Frühstücksbüffets in der Stadt.
- D. Da der Gast die Beherrschung verliert, verzichten Sie auf eine weitere Unterhaltung und bitten ihn, das Haus zu verlassen.

Item 21:

- A. Sie rufen den Gast in Zimmer 111 an und fragen ihn höflich, ob er in ein anderes Zimmer gleicher Lage und Kategorie wechseln würde, damit Ihr Stammgast sein bevorzugtes Zimmer beziehen kann.

- B. Sie bieten Ihrem Stammgast an, das Zimmer 111 für seinen nächsten Besuch vorzureservieren und empfehlen ihm für diesen Besuch andere Zimmer mit ähnlicher Lage und Aussicht.
- C. Sie weisen Ihren Stammgast freundlich darauf hin, dass er wie alle anderen Gäste reservieren muss, da Sie das Zimmer 111 leider nicht die ganze Zeit für ihn freihalten können.
- D. Da Herr Schulze ein Stammgast ist und da sein Lieblingszimmer leider besetzt ist, bieten Sie ihm ein Zimmer einer höheren Kategorie zu einem Vorzugspreis an.

Item 22:

- A. Sie versprechen Herrn Meier, dass Sie und Ihr Team sich ganz besonders um ihn bemühen wollen, wenn er trotz des Fehlens seines Stammkellners bleibt.
- B. Sie bitten den Gast, sich beim nächsten Mal vorab zu erkundigen, ob sein Stammkellner Dienst hat, damit er nicht enttäuscht ist.
- C. Sie versprechen Herrn Meier, dass die Speisen und Getränke von gleicher Qualität sind, auch wenn sein Stammkellner heute nicht da ist.
- D. Sie bitten den Gast, nächste Woche wiederzukommen, wenn sein Stammkellner Herr Schneider wieder Dienst hat.

Item 23:

- A. Sie teilen die Gruppe auf und verteilen sie auf freie Plätze an verschiedenen Tischen, so dass alle wie gewünscht am Brunch teilnehmen können.
- B. Sie müssen die Gäste leider wieder nach Hause schicken, da diese nicht auf ihrer Liste stehen.
- C. Sie bitten die Gäste um Verzeihung, da dies scheinbar ein Fehler des Hotels ist. Zudem versichern Sie den Gästen, das Problem so schnell wie möglich zu beheben.
- D. Sie beruhigen den Gast und rufen an der Rezeption an, ob dort eine Reservierung für die Gäste vorliegt.

Item 24:

- A. Sie bitten den Gast um Geduld, da Sie ihren Vorgesetzten einschalten möchten, um diese Sache zu klären.

- B. Sie bitten den Gast höflich, erst zu Beginn der Veranstaltung wiederzukommen, damit Sie bis dahin in Ruhe aufbauen können.
- C. Sie erklären dem Gast, dass Sie mit dem Aufbau genau im Zeitplan liegen und dass er sich keine Sorgen machen braucht, etwas würde schief laufen.
- D. Sie bitten den Gast, beim Aufbau zu helfen, damit es schneller geht.

Item 25:

- A. Sie rufen den Zimmerservice an und bitten um eine möglichst schnelle Freigabe des Zimmers von Herrn Müller. Für die Wartezeit geben Sie Herrn Müller einen Gutschein für ein Getränk an der Bar.
- B. Sie erklären dem Gast, dass er leider warten muss, da die Zimmerreinigung in einem großen Hotel viel Zeit beansprucht und tadellos erfolgen muss, um die Gäste zufrieden zu stellen.
- C. Sie empfehlen Herrn Müller, selber den Zimmerservice aufzusuchen, um sein Anliegen zu erklären.
- D. Sie erinnern den Gast an seine verfrühte Anreise und an die Tatsache, dass andere Gäste ebenfalls warten müssen, wenn Sie zu früh ankommen.

A.3 Expertenrangordnungen bei den 25 Items

Tabelle A-1: Expertenrangordnungen für die SJT-Items

Item	Expertenrangordnung
1	CADB
2	BDAC
3	ABDC
4	DCBA
5	ACBD
6	CDBA
7	CBDA
8	BCAD
9	BDCA
10	BCAD
11	BCDA
12	ADCB
13	ABCD
14	CDBA
15	BDCA
16	ABDC
17	BCDA
18	CDAB
19	ABCD
20	BACD
21	BDCA
22	ABCD
23	CDBA
24	ADCB
25	ABDC

A.4 Kurzbeschreibungen der Inhalte in den 25 SJT-Situationen

Tabelle A-2: Inhalte in den 25 SJT-Situationen

Item	Inhalt
1	Ein Hotelgast bittet einen Mitarbeiter, welcher bereits Feierabend hat, um Unterstützung bei der Behebung eines technischen Problems.
2	An der Rezeption beschwert sich ein Gast darüber, dass der Wein in der Minibar seines Zimmers Wasser statt Wein enthalte.
3	Zunächst bejaht ein Gast die Frage eines Mitarbeiters, ob es ihm schmecke. Jedoch meint derselbe Gast beim Abräumen, dass er doch
4	Bei einer Veranstaltung verstoßen zwei Gäste gegen das geltende Rauchverbot.
5	Nach der Begleichung einer hohen Restaurantrechnung möchte ein Gast nicht auch noch die Garagengebühr bezahlen.
6	Zwei einzelne Gäste erbitten von einem Mitarbeiter die Zubereitung eines separaten Büffets.
7	Ein Gast beschwert sich über die Qualität des ihm empfohlenen Weins und erwartet, dass alle weiteren Getränke kostenlos sind.
8	Beim gut besuchten Küchenbüffet wartet ein Gast auf das Nachfüllen der von ihm gewünschten Torte.
9	Ein Gast möchte unbedingt einen Platz im Wintergarten reservieren, welcher für den betreffenden Zeitpunkt schon ausgebucht ist
10	Aufgrund zweier Tagungen ist auch die komplette Hotelgarage besetzt. Ein Gast findet deshalb keinen Parkplatz.
11	In der Nacht beklagt sich ein Gast an der Rezeption, dass er wegen eines nicht schließenden Fensters kaum Schlaf finden könne.
12	Im Restaurant hat ein Gast den Eindruck, der Mitarbeiter habe ihm eine viel zu kleine Portion gebracht.
13	Vor seiner Abreise weigert sich ein Gast an der Rezeption das von ihm wahrgenommene Pay-TV Angebot zu bezahlen.
14	Ein Gast moniert, er habe von einem Mitarbeiter mehr Eis bekommen, als er bestellt habe, und fordert ein Getränk als Wiedergutmachung.
15	Da ein Gast bei seiner Auskunft keinen anderen Parkplatz als in der Hotelgarage finden konnte, möchte er die anfallende Gebühr nicht zahlen.
16	Im Restaurant sucht ein Gast einen Tisch, um einen besonderen Anlass zu feiern. Leider ist alles besetzt.
17	Beim Check-In verweigert sich ein Gast dem im Hotel obligatorischen Kreditkartenabzug.

Anhang A: Erste Studie

18	Im Restaurant bittet ein Gast um das Abräumen seines Tellers, obwohl er sein Essen kaum angerührt hat.
19	Ein Gast besteht auf den Erhalt einer zeitlich begrenzten Offerte.
20	Da der Weihnachtsbrunch bereits ausgebucht ist, kann ein Gast keine Plätze mehr für diesen Anlass reservieren.
21	Das von einem Stammgast stets bezogene Zimmer ist belegt, als dieser ohne Vorankündigung an der Rezeption erscheint.
22	An einem Sonntag vermisst ein Gast einen Kellner, von dem er normalerweise bedient wird.
23	Für einen Gast, welcher angibt, vorab für seine Familie reserviert zu haben, findet sich kein Vermerk.
24	Im Rahmen der Vorbereitung einer Veranstaltung reagiert der Gastgeber nervös auf den Stand der Organisation.
25	Obwohl sein Zimmer noch nicht verfügbar ist, möchte ein zu früh erschienener Gast dieses bereits beziehen.

Anhang B: Zweite Studie

B.1 Profil zur Ergebnisrückmeldung an die 97 Teilnehmer

Ihr persönlicher Code:

--	--	--	--	--	--

1. Feld: Erster Buchstabe des Vornamens der Mutter, z. B. A bei „Aнна“
2. Feld: Zweiter Buchstabe des eigenen Vornamens, z. B. O bei „JOchen“
3. Feld: Geburtstag + Geburtsmonat, z. B. 2204, falls Sie am 22. April Geburtstag haben

A. Inventar sozialer Kompetenzen (33 Aussagen)

Soziale Orientierung

Menschen mit einem hohen Maß an sozialer Orientierung gelingt es, sich in die Perspektiven anderer Menschen hineinzusetzen. Sozial orientierte Menschen widmen auch unter Belastung anderen Menschen ihre Aufmerksamkeit und holen bei Unklarheiten unterschiedliche Meinungen ein. In Diskussionen bemühen sie sich um einen Konsens und die Berücksichtigung unterschiedlicher Sichtweisen.

Menschen mit einem geringen Maß an sozialer Orientierung achten vor allem auf die Durchsetzung ihrer eigenen Meinung und hegen kaum Interesse an den Sichtweisen anderer. In Diskussionen mit anderen Menschen sind sie gedanklich oft nicht bei der Sache.

Ihr persönlicher Punktwert: ___/___ von maximal erreichbaren 40 Punkten.

Offensivität

Menschen mit einem hohen Maß an Offensivität suchen gezielt die konstruktive Auseinandersetzung mit anderen Menschen und genießen die Teilnahme an lebhaften Diskussionen. In diesen Diskussionen übernehmen sie zumeist eine führende Rolle. Außerdem treffen Sie wichtige Entscheidungen schnell und gezielt.

Menschen mit einem geringen Maß an Offensivität scheuen den konstruktiven Austausch mit anderen Personen. Wichtige Entscheidungen schieben sie auf.

Ihr persönlicher Punktwert: ___/___ von maximal erreichbaren 32 Punkten.

Selbststeuerung

Menschen mit einem hohen Maß an Selbststeuerung halten ihre Emotionen in Begegnungen mit anderen Menschen unter Kontrolle. Sie sind davon überzeugt, dass sie ihr Leben selbst in der Hand haben und nicht auf die übermäßige Hilfe anderer angewiesen sind.

Menschen mit einem geringen Maß an Selbststeuerung erleben einen häufigen Stimmungswechsel und haben Probleme dabei, ihre Emotionen gegenüber anderen Menschen zu kontrollieren. Zudem fühlen sie sich in ihrem Leben oft hilflos und von anderen Menschen abhängig.

Ihr persönlicher Punktwert: ___/___ von maximal erreichbaren 32 Punkten.

Reflexibilität

Menschen mit einem hohen Maß an Reflexibilität sind stets um eine positive Außendarstellung bemüht und reflektieren ihre Wirkung auf andere Menschen. Sie sind darum bemüht, ihr eigenes Verhalten und Auftreten dem jeweiligen Gegenüber anzupassen.

Menschen mit einem geringen Maß an Reflexibilität denken wenig über ihre Wirkung auf andere Menschen nach. Zudem passen sie ihr Verhalten nicht an unterschiedliche soziale Umfelder an.

Ihr persönlicher Punktwert: ___/___ von maximal erreichbaren 28 Punkten.

B. Fragebogen zum Verhalten gegenüber Hotelgästen (25 Videos)

Bei der Entscheidung für bzw. gegen eine Verhaltensalternative war abzuwägen zwischen der Berücksichtigung berechtigter Interessen der Hotelgäste auf der einen Seite und der Beachtung der Interessen des Hotels auf der anderen Seite.

Ihr persönlicher Punktwert: __/__ von maximal erreichbaren 50 Punkten.

Ihr persönlicher Prozentrang: __/__ %

Ein Prozentrang von 78 % bedeutet z. B., dass 22 % Ihrer Mitschüler einen besseren Punktwert als Sie erreicht haben, während 78 % ihrer Mitschüler einen schlechteren Punktwert erlangt haben.

Ich möchte mich nochmals bei Ihnen für die Teilnahme an der Studie bedanken und wünsche Ihnen alles Gute für Ihre weitere Karriere im Hotelfach.

B.2 Zusätzliche Statistiken

B.2.1 Ergebnis der konfirmatorischen Faktorenanalyse

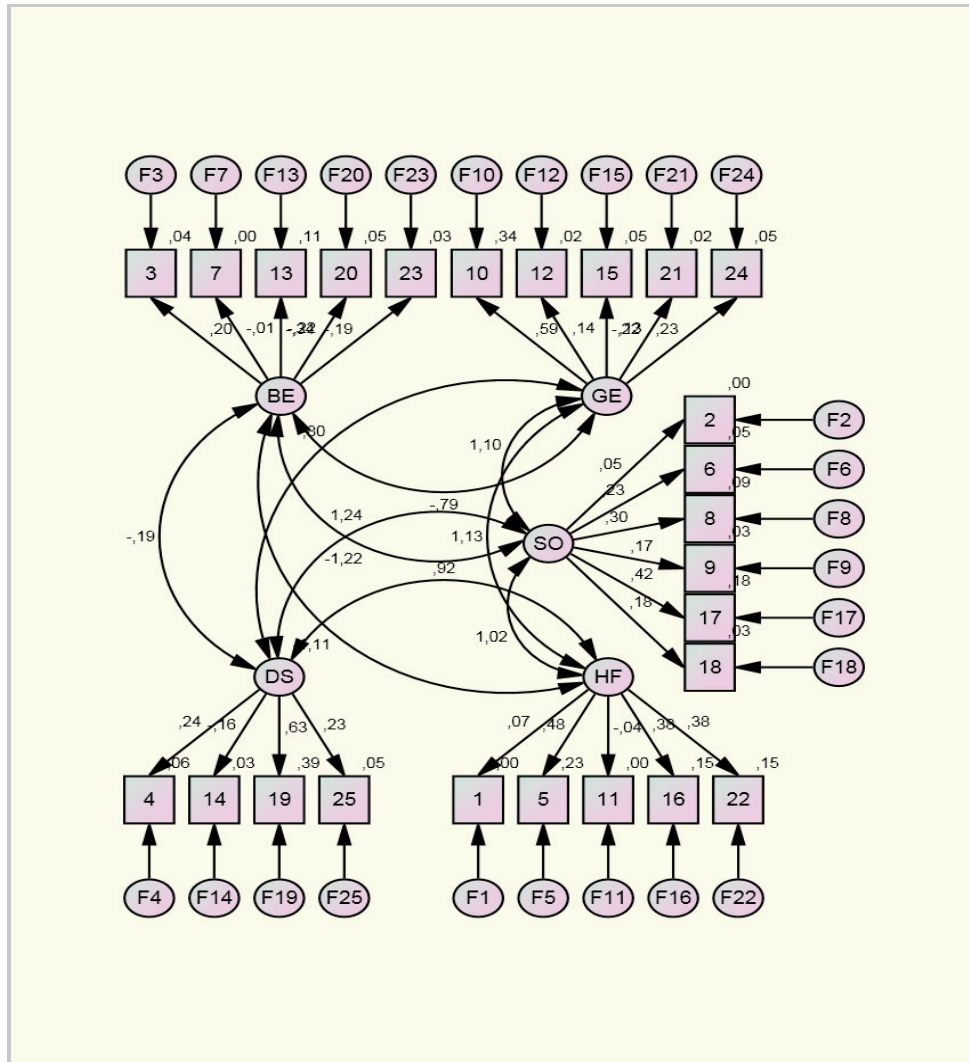


Abbildung B-1: Ergebnis der konfirmatorischen Faktorenanalyse⁶⁷

Anmerkung: Quadratische Variablen repräsentieren die 25 Items des SJT (1-25), ovale Variablen deuten die Faktoren 1. bzw. 2. Ordnung und relevante Fehleranteile (F1 - F26) an, Einfachpfeile (\rightarrow) symbolisieren Pfade und Zweifachpfeile (\leftrightarrow) Korrelationen bzw. Kovarianzen. KO = Kundenorientierung (Faktor 2. Ordnung)/ BE = Beschwerdemanagement/ GE = Gewissenhaftigkeit/ SO = Serviceorientierung/ HF = Handlungsfähigkeit/ DS = Durchsetzungsstärke (Faktoren 1. Ordnung).

⁶⁷ Eine Diskussion von Pfadkoeffizienten > 1 findet sich z. B. bei Jöreskog (1999).

B.2.2 MSA-Koeffizienten der 25 Items aus dem SJT*Tabelle B-1: MSA-Koeffizienten der 25 SJT-Items (Codierung nach Motowidlo)*

Item	MSA-Koeffizient	Einstufung
Situation 1	.40	Untragbar
Situation 2	.40	Untragbar
Situation 3	.44	Untragbar
Situation 4	.40	Untragbar
Situation 5	.60	Mittelmäßig
Situation 6	.44	Untragbar
Situation 7	.52	Klänglich
Situation 8	.53	Klänglich
Situation 9	.58	Klänglich
Situation 10	.61	Mittelmäßig
Situation 11	.40	Untragbar
Situation 12	.60	Mittelmäßig
Situation 13	.49	Untragbar
Situation 14	.55	Klänglich
Situation 15	.58	Klänglich
Situation 16	.61	Mittelmäßig
Situation 17	.73	Ziemlich gut
Situation 18	.50	Klänglich
Situation 19	.68	Mittelmäßig
Situation 20	.52	Klänglich
Situation 21	.48	Untragbar
Situation 22	.54	Klänglich
Situation 23	.45	Untragbar
Situation 24	.55	Klänglich
Situation 25	.61	Mittelmäßig

Anmerkung: Die Kriterien der Abstufung sind Bühner (2006) beziehungsweise Backhaus et al. (2008) entnommen.

B.2.3 Überprüfung der Normalverteilungsannahm*Tabelle B-2: Ergebnisse der Kolmogorov-Smirnov-Tests*

Item	Z	Signifikanz	Normalverteilung
Situation 1	3.72	.00	Nein
Situation 2	2.95	.00	Nein
Situation 3	3.10	.00	Nein
Situation 4	2.54	.00	Nein
Situation 5	2.77	.00	Nein
Situation 6	3.02	.00	Nein
Situation 7	3.10	.00	Nein
Situation 8	2.83	.00	Nein
Situation 9	2.77	.00	Nein
Situation 10	3.04	.00	Nein
Situation 11	2.93	.00	Nein
Situation 12	3.12	.00	Nein
Situation 13	2.65	.00	Nein
Situation 14	3.24	.00	Nein
Situation 15	3.82	.00	Nein
Situation 16	3.79	.00	Nein
Situation 17	2.91	.00	Nein
Situation 18	2.73	.00	Nein
Situation 19	3.06	.00	Nein
Situation 20	4.00	.00	Nein
Situation 21	3.28	.00	Nein
Situation 22	3.45	.00	Nein
Situation 23	3.37	.00	Nein
Situation 24	5.29	.00	Nein
Situation 25	4.08	.00	Nein

B.3 Analyse von Extremwerten über Boxplots

B.3.1 Extremwerte im SJT_kurz (n = 2)

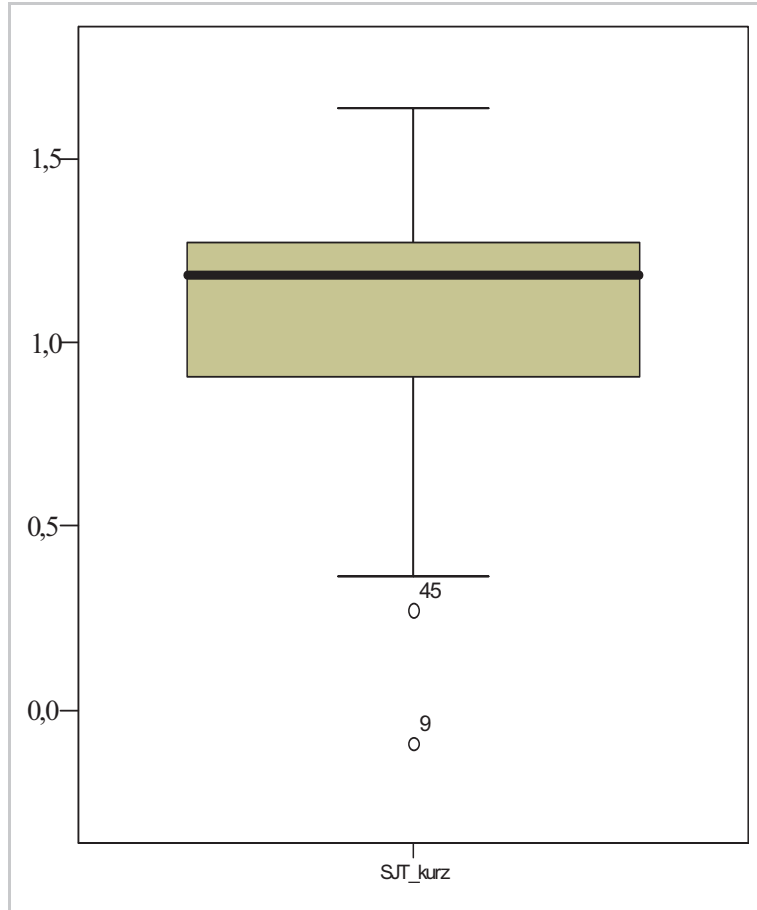


Abbildung B-2: Boxplot (SJT_kurz)

B.3.2 Extremwerte in der Skala Offensivität aus dem ISK-K (n = 4)

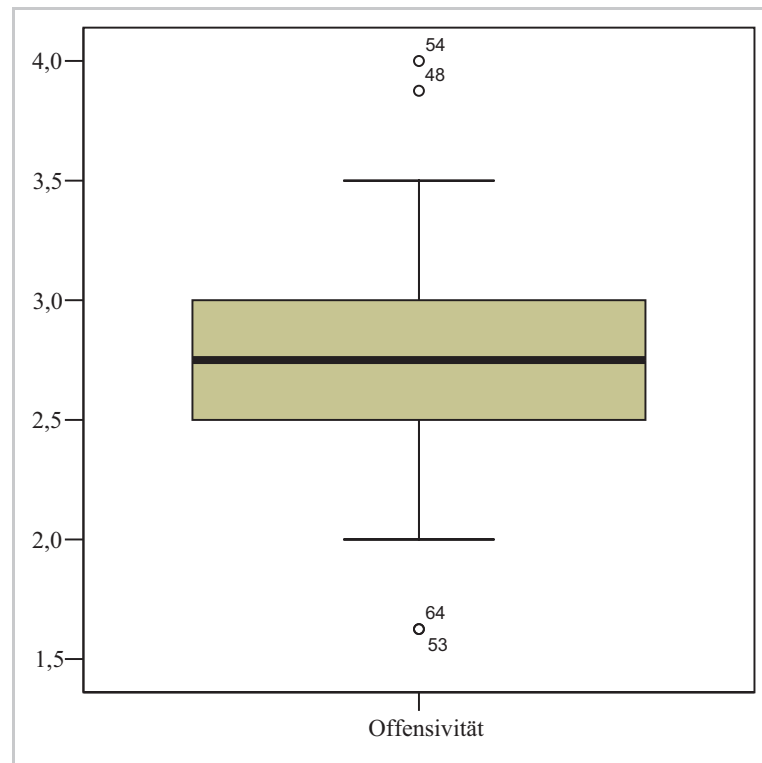


Abbildung B-3: Boxplot (Offensivität)

B.3.3 Extremwerte in der Skala Soziale Orientierung aus dem ISK-K (n = 7)

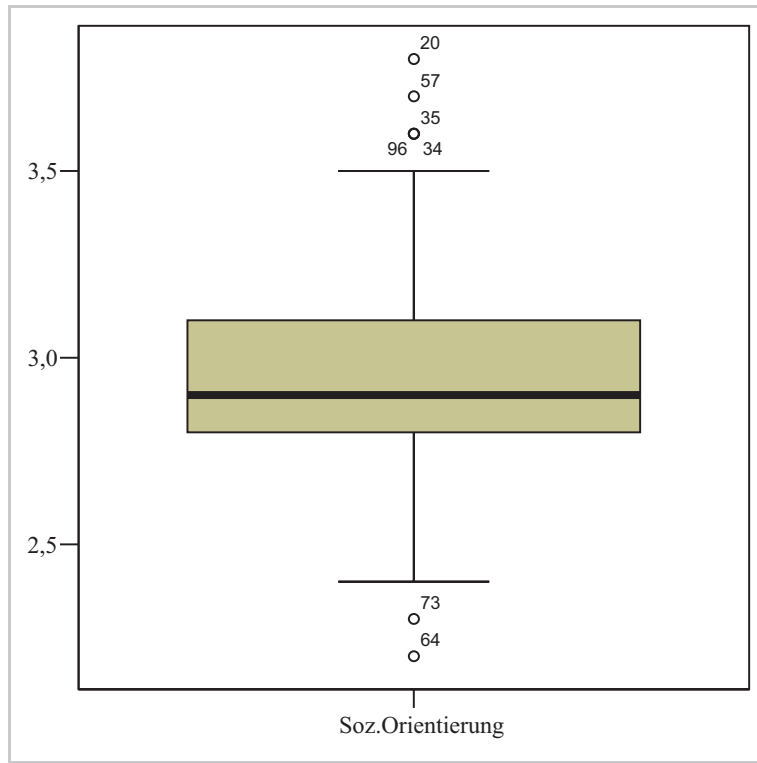


Abbildung B-4: Boxplot (Soziale Orientierung)

B.3.4 Extremwerte in der Skala Reflexibilität aus dem ISK-K (n = 7)

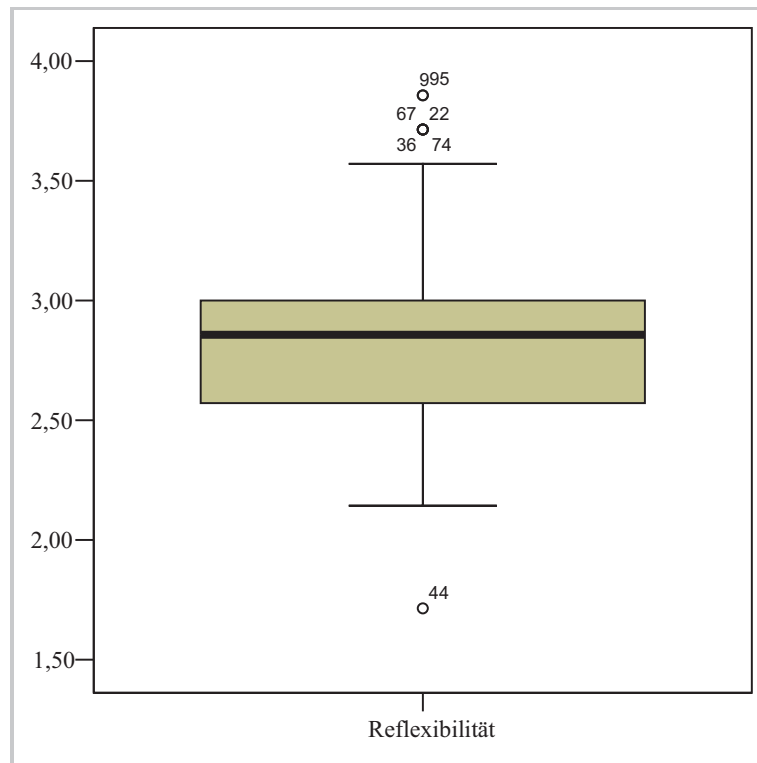


Abbildung B-5: Boxplot (Reflexibilität)

B.4 Korrelationen zwischen dem SJT_kurz und den Skalen des ISK-K nach einer einfachen Minderungskorrektur

Tabelle B-3: Korrelationen zwischen der Skala SJT_kurz und den Hauptskalen des ISK-K

	Soz. Orientierung	Offensivität	Selbststeuerung	Reflexibilität
SJT_kurz	.00	.26	.11	.01

Anmerkung: Aus den Korrelationen wurden Extremwerte (s. o.) entfernt.

Anhang C: Dritte Studie

C.1 Situationsbeschreibungen in der Paper-Pencil-Version des SJT

- **Situation 1** - Ihr Hotel verfügt über verschiedene Tagungsräume. Wenn es zu technischen Problemen in diesen Räumen kommt, wenden sich die Gäste an Mitarbeiter Ihres Hauses. In Ihrem Personalbereich, dessen Zutritt Gästen untersagt ist, spricht Sie ein besorgter Gast an, dass es zu technischen Problemen in einem Tagungsraum gekommen ist. Der Gast bittet Sie um Hilfe, da kein Arbeitskollege von Ihnen auffindbar war. Allerdings haben Sie bereits Feierabend und wollen gerade nach Hause.

- **Situation 2** - Jedes Ihrer Zimmer verfügt über eine kostenpflichtige Minibar, die verschiedene nicht-alkoholische und alkoholische Getränke enthält. Sie sind an der Rezeption beschäftigt, als Ihnen ein Gast entrüstet eine leere Weinflasche entgegenhält. Laut Aussage des Gastes entstammt diese Flasche aus seiner Minibar. In der Flasche des Gastes befand sich nach dessen Aussage jedoch kein Wein, sondern Wasser. Der Gast verdächtigt die Vormieter des Zimmers, den Wein ausgetauscht zu haben.

- **Situation 3** - Sie sind als Servicekraft in Ihrem Restaurant beschäftigt und begrüßen einen Gast an einem Einzeltisch, der bei Ihnen als Hauptgericht Kalbsleber bestellt. Nachdem Sie dem Gast das bestellte Gericht gebracht haben, erkundigen Sie sich zwischendurch, ob der Gast mit dem Gericht zufrieden ist. Der Gast bejaht dies. Nach einer Weile bittet Sie der Gast, den leeren Teller abzuräumen, und Sie fragen den Gast nochmals, ob es ihm geschmeckt hat. Nun verneint der Gast und weist enttäuscht darauf hin, dass ihm das Fleisch zu zäh war.

- **Situation 4** - Sie sind als Servicekraft bei einer Veranstaltung beschäftigt und werden zu zwei Gästen an einem Tisch im hinteren Bereich des Tagungsraumes gerufen. Als Sie die Bestellung der Gäste aufnehmen, bemerken Sie, dass beide Gäste trotz eines ausdrücklichen Verbotes durch den Veranstalter rauchen wollen. Sie weisen die Gäste nochmals auf das Rauchverbot und die Möglichkeit, in Ihrem Barbereich ungestört rauchen zu können, hin. Darauf erwidert einer der Gäste Ihnen jedoch nur barsch,

dass sich der Veranstalter nicht so anstellen sollte und dass Sie doch bitte schnellstens die Getränke bringen sollen.

- **Situation 5** - Viele Ihrer Restaurantgäste parken in der hoteleigenen Garage. Für die Benutzung der Garage wird eine Gebühr fällig. Sie sind an der Rezeption beschäftigt und begrüßen einen Gast, der Ihnen zornig seinen Parkschein über drei Euro aus Ihrer Garage zeigt. Der Gast weigert sich, diese Garagengebühr zu bezahlen, da er laut eigener Aussage bereits einen hohen dreistelligen Betrag in Ihrem Restaurant ausgegeben hat.

- **Situation 6** - Zwei Personen haben für den heutigen Tag einen kleinen Tagungsraum in Ihrem Hotel gemietet. Sie sind für die Betreuung dieser Gäste zuständig. Als Sie die Gäste fragen, was Sie zum Abendessen wünschen, bitten einer der Gäste Sie um die Zubereitung eines kleinen Büffets. Leider müssen Sie dem Gast mitteilen, dass Sie für Kleingruppen kein Büffet zubereiten können. Stattdessen bieten Sie den Gästen ein Drei-Gänge-Menü an. Der Gast reagiert auf diesen Vorschlag empört, da ihm ein Menü zu lange dauere.

- **Situation 7** - Sie sind als Servicekraft in Ihrem Restaurant beschäftigt. Zwei Gäste würden gerne Wein trinken und bitten um eine Empfehlung Ihrerseits. Sie empfehlen den Gästen Ihren Hauswein und bieten Ihnen eine Weinprobe, was die Gäste dankend annehmen. Als die Gäste Ihren Hauswein, welcher in Flaschen mit Drehverschluss serviert wird, probieren, reagiert ein Gast regelrecht angewidert und beschwert sich sofort, der Hauswein sei zu korkig. Ihr Angebot, weitere Weine zur Probe zu bringen, wird von dem Gast sofort mit der Forderung abgeschmettert, dass Sie ihm alle weiteren bestellten Getränke nicht berechnen sollen.

- **Situation 8** - Jeden Sonntag bieten Sie in Ihrem Restaurant ein Kuchenbüffet an, welches von den Gästen sehr gut angenommen wird, so dass sich oft lange Schlangen vor dem Büffet bilden. Einer der Gäste in der Schlange fragt Sie nach Ihrer Kirschtorte, welche im Moment leider vergriffen ist, was Sie dem Gast auch mitteilen. Der Gast würde nun gerne wissen, ob die Kirschtorte im Laufe des Tages nochmals aufgefüllt wird, was Sie bejahen. Der Gast freut sich und teilt Ihnen mit, dass er dann am

besten direkt vor dem Büffet wartet, damit er beim nächsten Mal auf jeden Fall ein Stück Kirschtorte bekommt.

- **Situation 9** - Der Wintergarten zählt zu den beliebtesten Sitzbereichen in Ihrem Restaurant, so dass die vorhandenen Plätze oftmals mehrere Tage im Voraus reserviert werden. Sie empfangen in Ihrem Restaurant eine Kundin, welcher für den kommenden Freitag einen größeren Tisch in Ihrem Wintergarten reservieren möchte. Sie müssen der Kundin leider mitteilen, dass an dem gewünschten Tag bereits der komplette Wintergarten durch zwei größere Gruppen belegt ist. Die Kundin reagiert besorgt und bittet Sie um alternative Vorschläge.

- **Situation 10** - Ihr Hotel verfügt über eine begrenzte Anzahl an kostenlosen Parkplätzen im Außenbereich sowie eine gebührenpflichtige Garage. Wenn Sie mehrere Tagungen in Ihrem Hause beherbergen, kommt es regelmäßig vor, dass sowohl die kostenlosen Parkplätze als auch Ihre Garage komplett besetzt sind. Sie sind an der Rezeption beschäftigt, als ein Gast auf Sie zukommt, der sich in großer Eile befindet. Der Gast teilt Ihnen mit, dass er keinen Parkplatz finden konnte, obwohl er in Ihrem Hotel einen wichtigen Termin hat. Als Sie ihm erklären, dass alle Parkplätze durch Tagungsteilnehmer belegt sind, reagiert der Gast gereizt.

- **Situation 11** - An der Rezeption beklagt sich ein Gast bei Ihnen, dass ein Zimmerfenster zur Straßenseite nicht richtig geschlossen werden kann, so dass der Gast wegen des Lärms nicht richtig schlafen kann. Um das Fenster zu reparieren, fragt der Gast nach dem Hausmeister, der jedoch leider schon Feierabend hat, was Sie dem Gast auch mitteilen. Der Gast reagiert besorgt und fürchtet, dass er dann überhaupt nicht schlafen kann.

- **Situation 12** - Sie sind als Servicekraft im Restaurant beschäftigt und servieren zwei Gästen Fisch als Hauptgericht. Als die Gäste einen Blick auf Ihre Gerichte werfen, reagiert einer der beiden Gäste empört und beschwert sich sofort, dass die Portionen kleiner als gewohnt seien.

- **Situation 13** - Jedes Ihrer Zimmer verfügt über ein gebührenpflichtiges Pay-TV-Angebot (Bezahlfernsehen). Dieses Angebot wird durch eine externe Firma betreut. Sie sind an der Rezeption beschäftigt und betreuen einen Gast beim Check-out. Als Sie ihm seine Pay-TV-Rechnung zeigen, reagiert der Gast verärgert und versichert Ihnen, dieses Angebot überhaupt nicht genutzt zu haben. Er ist der festen Überzeugung, dass ein technischer Fehler Ihrerseits vorliegen muss, so dass er die Pay-TV-Rechnung keinesfalls ohne Klärung des Sachverhalts begleichen möchte.

- **Situation 14** - Sie sind als Servicekraft im Restaurant beschäftigt und nehmen die Bestellung von zwei Gästen auf, die jeweils drei Kugeln gemischtes Eis haben möchten. Nachdem diese Gäste das Eis verzehrt haben und Sie bitten, den Tisch abzuräumen, erkundigen Sie sich nach der Zufriedenheit der Gäste. Einer der beiden Gäste wirft Ihnen nun vor, die Bestellung wohl missverstanden zu haben, da sie sich jeweils nur eine Kugel Eis gewünscht hätten. Als Wiedergutmachung wünscht sich der Gast jeweils einen Espresso.

- **Situation 15** - Sie sind an der Rezeption beschäftigt und betreuen die Gäste beim Check-out. Als Sie einem Ihrer Gäste mitteilen, dass er für die zweitägige Benutzung Ihrer Garage eine Gebühr von 30 Euro zu entrichten hat, reagiert dieser wütend. Der Gast schildert Ihnen, dass er in der Garage parken musste, da kein kostenfreier Parkplatz vor Ihrem Hotel frei war, und weigert sich die Gebühr zu bezahlen.

- **Situation 16** - An bestimmten Tagen in der Woche ist Ihr Restaurant so gut besucht, dass ohne vorherige Reservierung kaum ein Tisch zu ergattern ist. Sie sind als Servicekraft in Ihrem Restaurant beschäftigt, als Sie ein weiblicher Gast nach einem freien Tisch für sich und seine Gruppe fragt. Da die Dame vorher nicht reserviert hat, müssen Sie ihr leider mitteilen, dass Ihr Restaurant bereits komplett ausgebucht ist. Die Dame bittet Sie daraufhin nochmals um einen freien Tisch, da sie heute einen besonderen Abend mit ihren Freuden verbringen möchte.

- **Situation 17** - In Ihrem Hotel ist es Vorschrift, dass Sie bei Gästen, die Ihr Haus zum ersten Mal zu besuchen, einen Kreditkartenabzug vorneh-

men. Im Rahmen dieses Kreditkartenabzuges hinterlegen Sie sowohl die Kreditkartennummer als auch die Laufzeit der Kreditkarte in einem elektronischen Kundenprofil. Sie sind an der Rezeption beschäftigt. Einer Ihrer neuen Gäste empört sich nun über den Vorgang des Kreditkartenabzuges bei seinem Check-in. Der Gast äußert die Sorge, dass Sie ihm wohl misstrauen, wenn Sie zu solchen „Maßnahmen“ greifen müssen.

- **Situation 18** - Sie sind als Servicekraft in Ihrem Spezialitätenrestaurant beschäftigt. Einer Ihrer Gäste bitte Sie zu seinem Tisch, da er möchte, dass Sie seinen Tisch abräumen und ihm die Rechnung bringen. Als Sie am Tisch des Gastes angelangen, stellen Sie fest, dass der Gast sein Essen fast nicht angerührt hat.

- **Situation 19** - Sie sind als Servicekraft in Ihrem Restaurant beschäftigt. Sie bieten jeden Tag ab 15 Uhr ein spezielles Kuchenangebot an, welches von vielen Gästen wahrgenommen wird. Ein Gast bestellt bei Ihnen nun um 14 Uhr dieses Kuchenangebot. Sie erinnern den Gast daran, dass dieses Angebot leider erst ab 15 Uhr gilt. Der Gast besteht jedoch darauf, dass Sie ihm dieses Angebot bringen. Laut Aussage des Gastes hat ihm einer Ihrer Arbeitskollegen versichert, dieses Angebot gelte heute ausnahmsweise den ganzen Tag.

- **Situation 20** - Sie empfangen Anfang November in Ihrem Restaurant einen Gast, welcher Plätze für das Frühstücksbüffet am 1. Advent reservieren möchte. Bei einem Blick in Ihr Reservierungsbuch stellen Sie rasch fest, dass Sie an diesem Tag komplett durch Ihre Hotelgäste ausgebucht sind. Als Sie dies dem Gast mitteilen, reagiert dieser erbost und fragt Sie, ob er denn jetzt schon für das Jahr 2050 reservieren müsse.

- **Situation 21** - Ihr Hotel verfügt über viele Stammgäste, die mehrmals im Jahr in Ihrem Hause übernachten. Viele Hotelgäste nehmen immer wieder dasselbe Zimmer. Leider reisen die Stammgäste oft an, ohne vorab zu reservieren. Sie sind an der Rezeption beschäftigt und begrüßen einen Ihrer langjährigen Stammgäste beim Check-in. Dieser Gast besucht Sie dieses Jahr zum zehnten Mal. Der Stammgast möchte gerne sein gewohntes Zimmer belegen. Da der Stammgast ohne Reservierung angereist ist,

wurde das gewünschte Zimmer bereits durch andere Gäste belegt. Als Sie dies dem Stammgast mitteilen, reagiert dieser gekränkt und fragt Sie, wo er denn jetzt bleiben solle.

- **Situation 22** - Ihr Restaurant verfügt über viele Stammgäste, die sich im Laufe der Zeit an bestimmte Kellner gewöhnt haben, von denen sie nun am liebsten bedient werden. Sie sind als Servicekraft in Ihrem Restaurant beschäftigt und begrüßen einen Ihrer langjährigen Stammgäste. Der Stammgast erkundigt sich bei Ihnen, ob sein Stammkellner Dienst hat. Als Sie ihm mitteilen, dass dies nicht der Fall ist, möchte der Stammgast das Restaurant am liebsten wieder verlassen.

- **Situation 23** - Sie empfangen in Ihrem Restaurant einen Gast, welcher Ihnen mitteilt, für den heutigen Abend einen Tisch für sechs Personen reserviert zu haben. Bei einem Blick in Ihr Reservierungsbuch stellen Sie jedoch fest, dass die Dame nicht eingetragen ist. Als Sie dies der Dame mitteilen, reagiert diese aufgebracht und versichert Ihnen, vorab reserviert zu haben und dass ihr Mann dies bezeugen könne.

- **Situation 24** - Sie beteiligen sich am Aufbau einer größeren Veranstaltung für den heutigen Abend. Ihr Team hat Veranstaltungen dieser Größenordnung schon oft vorbereitet und Sie liegen genau im Zeitplan. Am Nachmittag wendet sich der Veranstalter, welcher bereits mehrere Stunden vor dem Beginn eingetroffen ist, an Sie. Sie teilen dem Veranstalter auf Nachfrage mit, dass Sie noch mit dem Aufbau beschäftigt sind. Der Veranstalter reagiert entsetzt und ermahnt Sie, sich zu beeilen. Zudem bemängelt er die Sauberkeit der Räumlichkeiten sowie fehlendes Geschirr auf den Tischen.

- **Situation 25** - Ihre neu anreisenden Gäste können immer erst nachmittags in ihre Zimmer, da die Zimmer vorab gereinigt werden müssen. Viele Gäste reisen jedoch schon morgens oder mittags an. Sie sind an der Rezeption beschäftigt und begrüßen einen neuen Gast beim Check-in. Der Gast möchte gerne jetzt schon sein Zimmer beziehen, doch Sie müssen ihm leider mitteilen, dass das Zimmer noch nicht gereinigt wurde und somit noch nicht bezugsfertig ist. Der Gast bietet Sie dennoch um eine Frei-

gabe des Zimmers, da er sich von seiner langen Anreise sehr erschöpft fühlt.

C.2 Rechenbeispiele für die vier scoring keys bei Rangordnungen

C.2.1: Komplettschlüssel

Expertenrangordnung	Teilnehmerrangordnung	Itemscore
ABCD	ABCD	1
ABCD	ABDC	0

C.2.2: Motowidlo-Schlüssel

Expertenrangordnung	Teilnehmerrangordnung	Itemscore
ABCD	ABCD	+2
ABCD	DBCA	-2
ABCD	ADBC	+1
ABCD	DACB	-1

C.2.3: Einzelvergleich-Schlüssel

Expertenrangordnung	Teilnehmerrangordnung	Itemscore
ABCD	ABCD	4
ABCD	ACBD	2
ABCD	DBAC	1

C.2.4: Distanz-Schlüssel

Expertenrangordnung	Teilnehmerrangordnung	Itemscore
ABCD	ABCD	+4
ABCD	ACBD	-2
ABCD	DBAC	-6
ABCD	DCBA	-8

C.3 Profil zur Ergebnismeldung an die 62 Teilnehmer

Dein persönlicher Code

--	--	--	--	--	--

1. Feld: 1. Buchstabe des Vornamens der Mutter, z. B. A bei „Anna“
2. Feld: 2. Buchstabe des eigenen Vornamens, z. B. O bei „Jochen“
3. Feld: Geburtstag + Geburtsmonat, z. B. 2204, falls Sie am 22. April Geburtstag

Am 20.1.09 hast Du an einem mehrstündigen Workshop im Hotel Münster teilgenommen. Im Rahmen dieses Workshops hattest Du die Gelegenheit, verschiedene Aufgaben und Situationen kennen zu lernen, welche Dir nach Abschluss Deiner Schullaufbahn in dieser oder ähnlicher Form begegnet könnten, wenn Du Dich zum Beispiel für einen Ausbildungsplatz bewirbst.

Der Workshop bestand aus drei verschiedenen Übungen⁶⁸:

1. Fragebogen zum Verhalten gegenüber Hotelgästen
2. Test zur allgemeinen kognitiven Leistungsfähigkeit
3. Wortschatztest

Im Folgenden werden Dir kurze Beschreibungen der drei Übungen samt Deiner persönlichen Ergebnisse vorgestellt.

⁶⁸ Über den BIP und die Leistungen der Schüler in diesem Test referierte der Autor in der PGR.

1. Fragebogen zum Verhalten gegenüber Hotelgästen

Bei der Bearbeitung des *Fragebogens zum Verhalten gegenüber Hotelgästen* wurden Dir 25 typische Situationen aus dem Berufsalltag einer Hotelfachfrau/- eines Hotelfachmannes präsentiert. In den Situationen wurden Interaktionen zwischen Hotelgästen und Mitarbeitern eines Hotels aufgegriffen. Für Dich war es wichtig, dass Du versuchst Dich in die Rolle des Hotelmitarbeiters hineinzusetzen und zu überlegen, wie Du in der Situation handeln würdest. Die 25 Situationen wurden Dir entweder als Video oder in schriftlicher Form präsentiert.

Im Anschluss an die Darstellung der Situationen war es Deine Aufgabe, einerseits vier verschiedene Verhaltensalternativen hinsichtlich der Wahrscheinlichkeit, mit der Du das beschriebene Verhalten tatsächlich zeigen würdest, einzuschätzen und andererseits die vier aufgelisteten Verhaltensalternativen in Form einer Rangordnung einzustufen.

Bei der Bearbeitung dieser Aufgaben war es wichtig, dass Du abwägst zwischen der Berücksichtigung von Wünschen und Interessen, welche Hotelgäste hegen und äußern, und Deiner Aufgabe als (fiktiver) Hotelmitarbeiter, die Interessen des Hotels zu wahren und den Hotelgästen nicht jeden unrealistischen Wunsch zu erfüllen.

Um zu beurteilen, wie gut Du abgeschnitten hast, wurden Deine Einschätzungen und Rangordnungen mit den Angaben von erfahrenen Hotelmitarbeitern verglichen, welche die relevanten Situationen schon oft in ihrer beruflichen Laufbahn bewältigen mussten. Dein Ergebnis fällt umso besser aus, je stärker Deine Angaben mit denen der Berufsexperten übereinstimmen.

a) Dein persönlicher Punktwert: _____ von maximal erreichbaren 20 Punkten. (20 Punkte = komplette Übereinstimmung mit langjährigen Hotelmitarbeitern)

Wenn Du wissen möchtest, wie Dein persönlicher Wert einzuordnen ist: Der durchschnittlich von Dir und Deinen MitschülerInnen erreichte Punktwert ist 13. Punktwerte > 13 sind überdurchschnittlich gut, Punktwerte < 13 eher unterdurchschnittlich.

b) Dein persönlicher Prozentrang: ____ %

Ein Prozentrang von 78% bedeutet zum Beispiel, dass 22% Deiner Mitschüler einen besseren Punktwert als Du erreicht haben, während 78% Deiner Mitschüler einen schlechteren oder gleich guten Punktwert erlangt haben. In einer Rangreihe gibt es immer einen ersten und einen letzten Platz. Wenn gerade Du den letzten Platz belegt hast, ist dies kein Grund, den Kopf hängen zu lassen, da dies ja noch keine reale Bewerbungssituation war und Du in den anderen Tests vielleicht deutlich besser abgeschnitten hast.

2. Test zur allgemeinen kognitiven Leistungsfähigkeit

Die allgemeine kognitive Leistungsfähigkeit zählt zu denjenigen Eigenschaften, welche besonders häufig in Bewerbungssituationen erfasst und beurteilt werden, da die allgemeine kognitive Leistungsfähigkeit eines Bewerbers Arbeitgebern wichtige Rückschlüsse geben kann über dessen (zukünftige) Arbeitsleistung. Unter dem Begriff *allgemeine kognitive Leistungsfähigkeit* werden gänzlich verschiedene Facetten zusammengefasst, welche bei unterschiedlichen Problemen und Aufgaben vonnöten sind. Stell Dir doch bitte eine Person vor, die Du als sportlich bezeichnen würdest: Eine sportliche Person muss ganz verschiedene Bewegungsabläufe beherrschen, zum Beispiel eine kraftvolle Schusstechnik beim Fußball, eine effektive Koordination beim Weitsprung oder eine gute Kondition beim Marathon. Ähnlich verhält es sich auch mit der allgemeinen kognitiven Leistungsfähigkeit: Die allgemeine kognitive Leistungsfähigkeit setzt sich wie die Eigenschaft *Sportlichkeit* aus verschiedenen „Bausteinen“ zusammen.

In dem von Dir bearbeiteten Leistungstest standen vier Facetten der allgemeinen kognitiven Leistungsfähigkeit im Blickpunkt, welche jeweils über zwei Aufgabengruppen erfasst wurden, so dass Du insgesamt acht Aufgabengruppen bearbeitet hast. Für jede Aufgabengruppe stand Dir immer nur eine bestimmte Bearbeitungszeit zur Verfügung, was womöglich dazu geführt hat, dass Du nicht jede einzelne Aufgabe bearbeiten konntest. Die einzelnen Aufgabengruppen enthielten jeweils zeichnerische Problemstellungen, welche von Dir zum Beispiel das Ergänzen einer

Objektreihe, das Vervollständigen eines Objektes oder das Feststellen von Gemeinsamkeiten zwischen Objekten erforderten.

Dein Gesamtwert für die allgemeine kognitive Leistungsfähigkeit: _____

Um Dein Ergebnis einzuordnen, sollte Dir folgende Kategorisierung hilfreich sein. Werte zwischen 85 und 115 sind durchschnittlich im Vergleich zu anderen Personen Deiner Altersstufe. Werte > 115 sind überdurchschnittlich zu nennen, während Werte < 85 als unterdurchschnittlich zu bezeichnen sind.

3. Wortschatztest

Im Gegensatz zu dem unter Punkt 2 erläuterten Verfahren ist für das erfolgreiche Abschneiden im Wortschatztest ein andere Komponente Deiner allgemeinen kognitiven Leistungsfähigkeit von Belang. Bei diesem Test hat die *sprachliche* Facette der allgemeinen kognitiven Leistungsfähigkeit eine dominante Rolle inne. Unser sprachliches Ausdrucksvermögen und damit einhergehend den Umfang unseres Wortschatzes bauen wir vor allem im Austausch mit unseren Mitmenschen auf, so dass dieser Teil unserer allgemeinen kognitiven Leistungsfähigkeit zu einem großen Teil auf Lernerlebnisse zurückzuführen ist.

Der Hauptbestandteil des Wortschatztests waren 30 Mehrfachwahlaufgaben, für deren Bewältigung Du einem Zielwort ein Wort gleicher oder ähnlicher Bedeutung aus einer Liste von fünf Alternativen zuordnen musstest. In Analogie zum dritten Test war auch diese Aufgabe innerhalb eines vorgegebenen Zeitrahmens durchzuführen.

Du hast _____ von 30 Wortpaaren korrekt zugeordnet.

In dieser Aufgabe habt Ihr alle gut abgeschnitten.

C.4 Zusätzliche Statistiken

C.4.1 Soziale Validität und Akzeptanz des SJT (25 Situationen)

Tabelle C-1: Unterschiede in Bezug auf Dimensionen der sozialen Validität und Akzeptanz (Video-Bedingung vs. Paper-Pencil-Bedingung)- t-Tests für unabhängige Stichproben

		Levenes Test auf Varianzhomogenität		T	Signifikanz (2-seitig)
		F	Signifikanz		
Informationsgrad	Annahme gleicher Varianzen	1.64	.21	.97	.34
	Annahme ungleicher Varianzen			.97	.34
Beeinflussbarkeit	Annahme gleicher Varianzen	.08	.77	-.78	.44
	Annahme ungleicher Varianzen			-.78	.44
Anforderungsbezug	Annahme gleicher Varianzen	.00	.94	1.50*	.14
	Annahme ungleicher Varianzen			1.50	.14
Allg._Akzeptanz	Annahme gleicher Varianzen	2.80	.10	1.54	.31
	Annahme ungleicher Varianzen			1.54	.31

Anmerkung: * = $p < .20$

C.4.2 Selbsteinschätzung der eigenen Leistung im SJT (25 Situationen)

Tabelle C-2: Korrelationen zwischen der Einschätzung der eigenen Leistung im SJT und Dimensionen der sozialen Validität und Akzeptanz (Video-Bedingung)

		Selbsteinschätzung der eigenen Leistung	Informationsgrad	Beeinflussbarkeit	Anforderungsbezug	Allg. Akzeptanz
Selbsteinschätzung der eigenen Leistung	R	1	.37*	.24	.34	.00
	Signifikanz (2-seitig)	.	.04	.19	.07	.99

Anmerkung: * $p < .05$

Tabelle C-3: Korrelationen zwischen der Einschätzung der eigenen Leistung im SJT und Dimensionen der sozialen Validität und Akzeptanz (Paper-Pencil-Bedingung)

	Informationsgrad	Beeinflussbarkeit	Anforderungs-Bezug	Allg. Akzeptanz
Selbsteinschätzung der eigenen Leistung	.25	.35	.07	-.13

C.4.3 Überprüfung der Normalverteilungsannahme (SJT mit 25 Situationen - Video-Bedingung)

Tabelle C-4: Ergebnisse der Kolmogorov-Smirnov-Tests

Item	Z	Signifikanz	Normalverteilung
Situation 1	1.29	.07	Ja
Situation 2	.69	.72	Ja
Situation 3	.63	.82	Ja
Situation 4	.75	.63	Ja
Situation 5	.55	.92	Ja
Situation 6	.68	.75	Ja
Situation 7	.90	.40	Ja
Situation 8	.77	.60	Ja
Situation 9	.98	.30	Ja
Situation 10	.93	.36	Ja
Situation 11	.64	.80	Ja
Situation 12	.58	.89	Ja
Situation 13	.50	.97	Ja
Situation 14	.94	.34	Ja
Situation 15	.99	.28	Ja
Situation 16	.54	.93	Ja
Situation 17	1.04	.23	Ja
Situation 18	.76	.63	Ja
Situation 19	.91	.38	Ja
Situation 20	.74	.64	Ja
Situation 21	.89	.41	Ja
Situation 22	.88	.42	Ja
Situation 23	.87	.43	Ja
Situation 24	.52	.95	Ja
Situation 25	1.05	.22	Ja

C.4.4 MSA-Koeffizienten der 25 Items aus dem SJT (Video-Bedingung)

Tabelle C-5: MSA-Koeffizienten der 25 SJT-Items (Codierung mit Ratingschlüssel)

Item	MSA-Koeffizient
Situation 1	.42
Situation 2	.24
Situation 3	.17
Situation 4	.16
Situation 5	.14
Situation 6	.24
Situation 7	.19
Situation 8	.23
Situation 9	.25
Situation 10	.23
Situation 11	.17
Situation 12	.19
Situation 13	.22
Situation 14	.29
Situation 15	.55
Situation 16	.13
Situation 17	.20
Situation 18	.22
Situation 19	.21
Situation 20	.33
Situation 21	.29
Situation 22	.30
Situation 23	.20
Situation 24	.25
Situation 25	.19

C.4.5 Scree-Plot (Codierung mit Ratingschlüssel/Video-Bedingung)

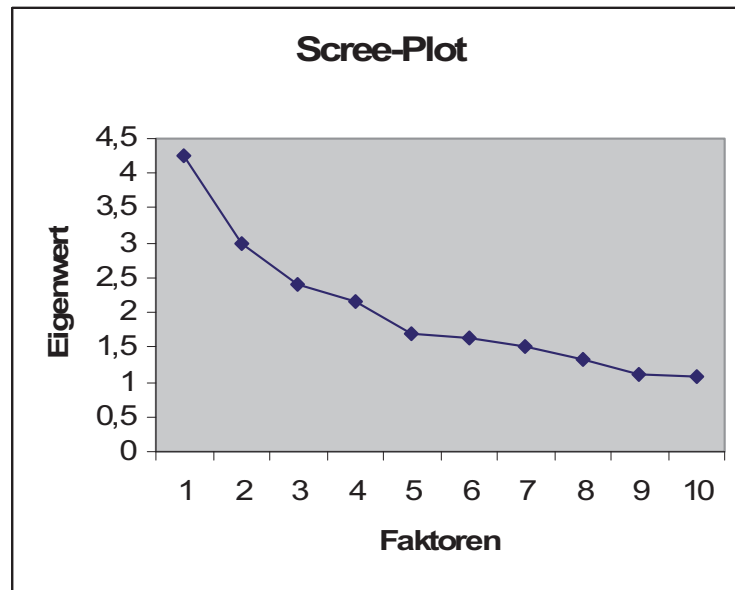


Abbildung C-1: Scree-Plot

C.4.6 Ergebnis des MAP-Test (Video-Bedingung)

Tabelle C-6: MAP-Test (Codierung mit Ratingschlüssel)

Faktor	Mittlere Partialkorrelation (r^2)
1	.039
2	.041
3	.043
4	.044
5	.047
6	.053
7	.055
8	.060
9	.066
10	.069

C.5 Analyse von Extremwerten über Boxplots (Video-Bedingung)

C.5.1 Wortschatztest (n = 1)

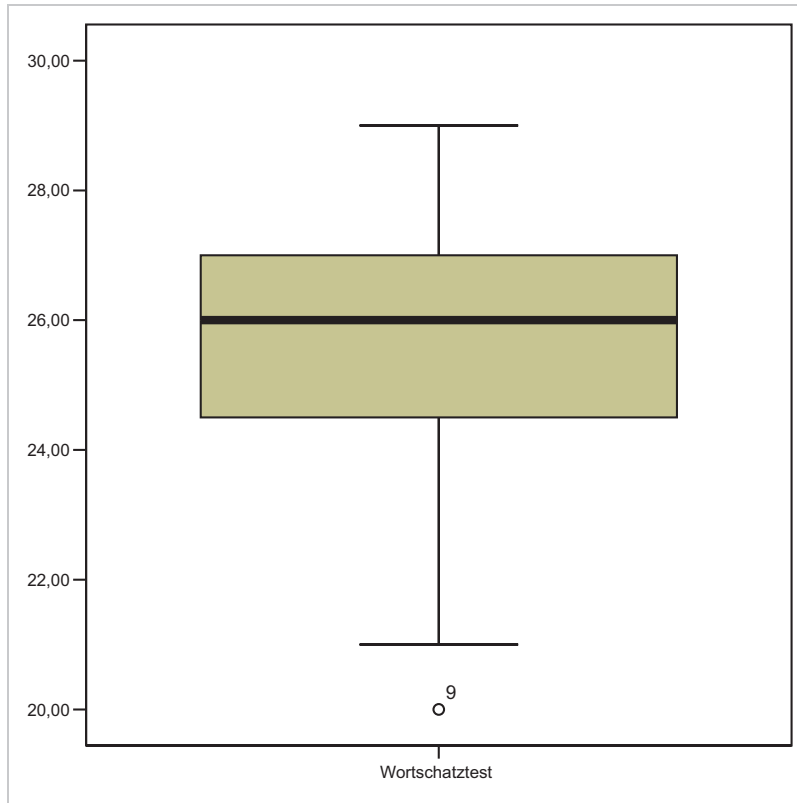


Abbildung C-2: Boxplot (WS)

C.5.2 BIP_Sensitivität (n = 1)

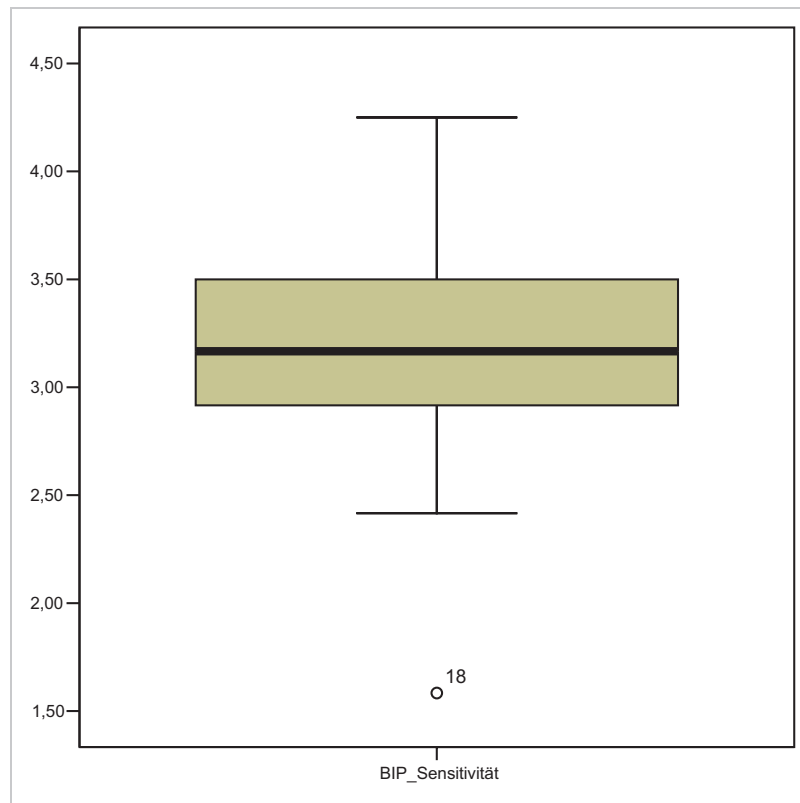


Abbildung C-3: Boxplot (Sensitivität)

C.5.3 BIP_Gestaltungsmotivation (n = 2)

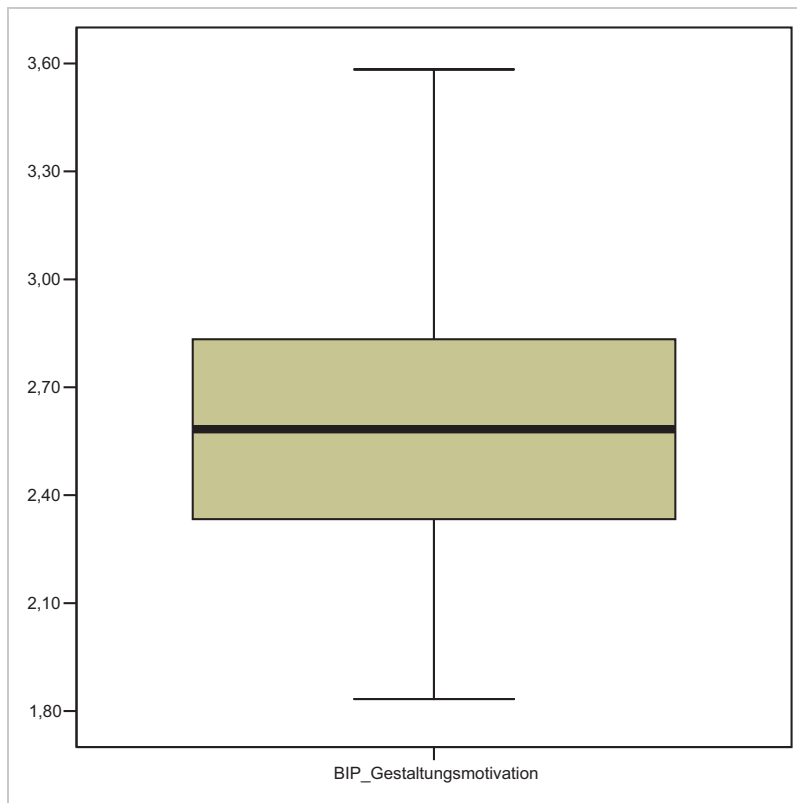


Abbildung C-4: Boxplot (Gestaltungsmotivation)

C.5.4 BIP_Gewissenhaftigkeit (n = 1)

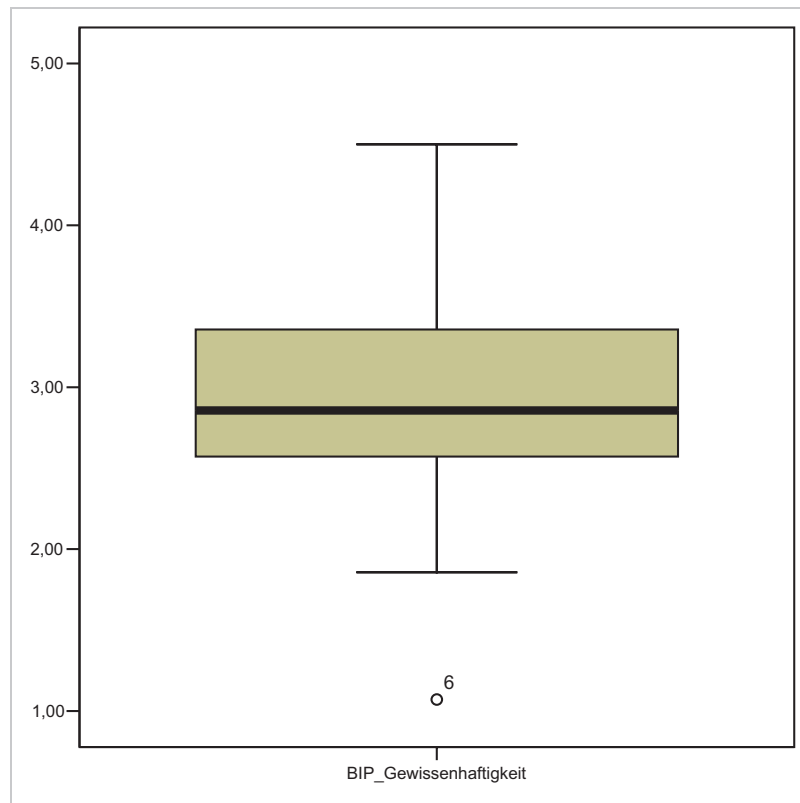


Abbildung C-5: Boxplot (Gewissenhaftigkeit)

C.5.5 BIP_Handlungsorientierung (n = 1)

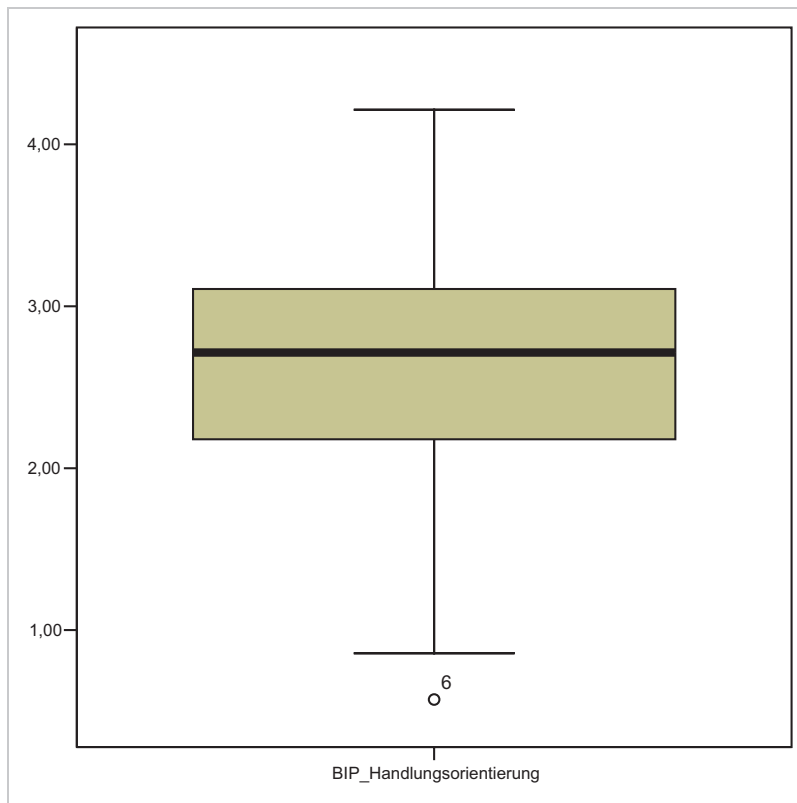


Abbildung C-6: Boxplot (Handlungsorientierung)

C.5.6 BIP_Selbstbewusstsein (n = 1)

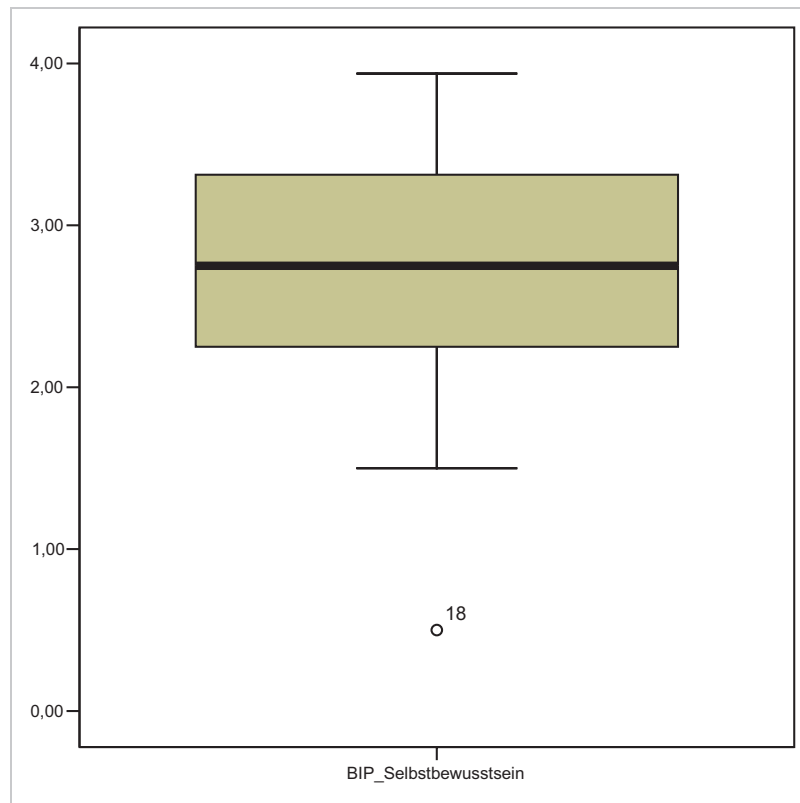


Abbildung C-7: Boxplot (Selbstbewusstsein)

C.6 Analyse von Extremwerten über Boxplots (Paper-Pencil-Bedingung)

C.6.1 CFT-20-R 1. Testteil (n = 1)

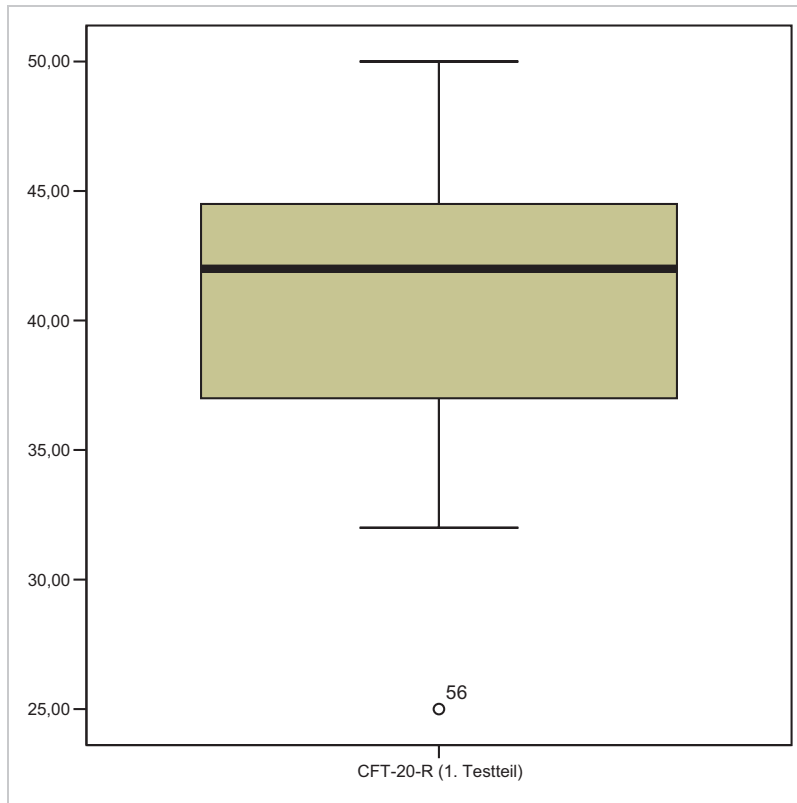


Abbildung C-8: Boxplot (CFT-20-R 1. Testteil)

C.6.2 BIP_Sensitivität (n = 1)

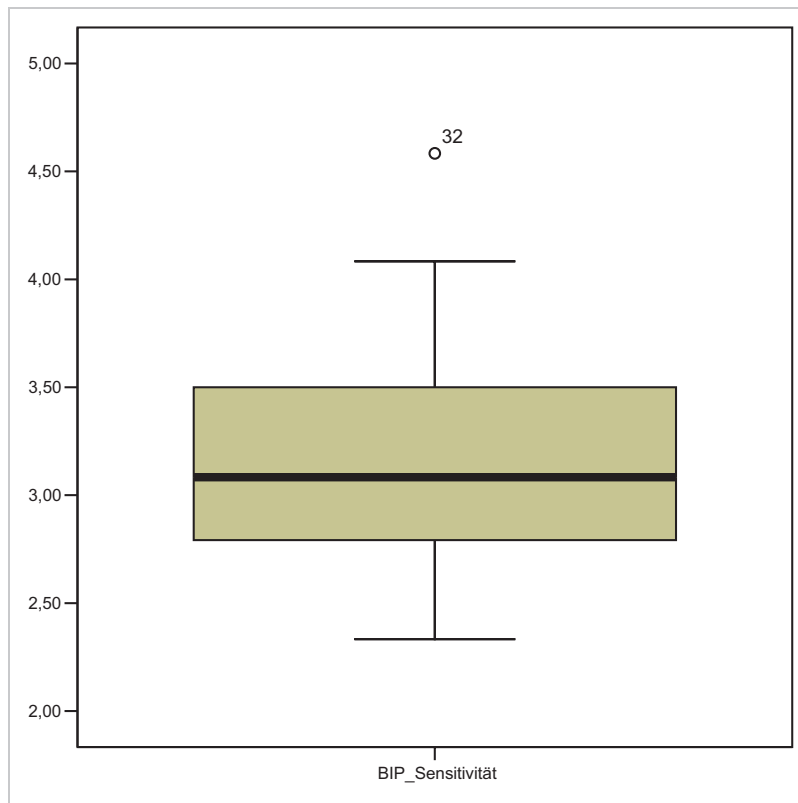


Abbildung C-9: Boxplot (Sensitivität)

C.6.3 BIP_Leistungsmotivation (n = 1)

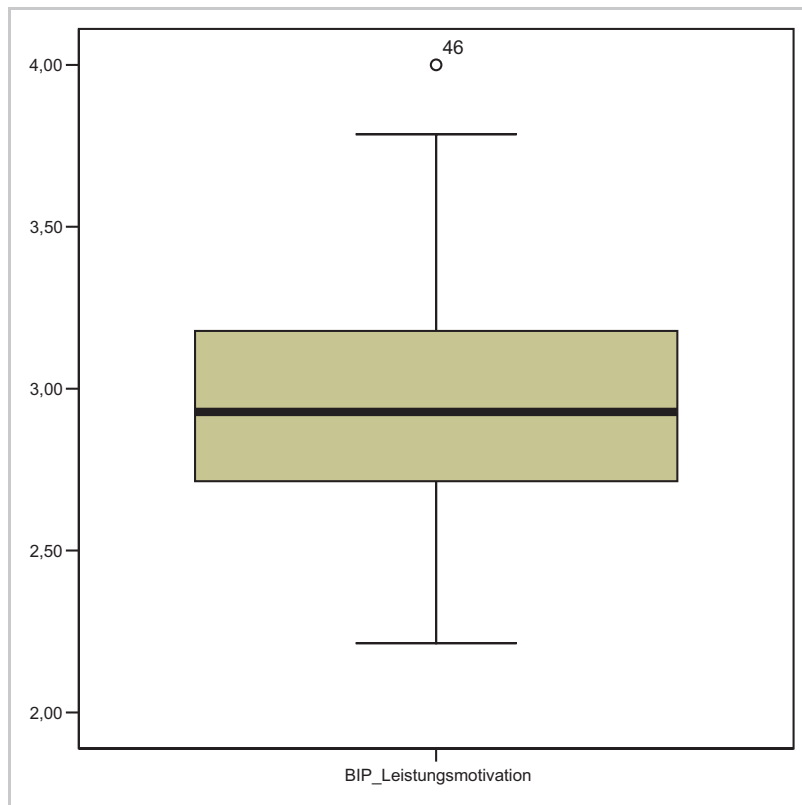


Abbildung C-10: Boxplot (Leistungsmotivation)

C.6.4 BIP_Gestaltungsmotivation (n = 1)

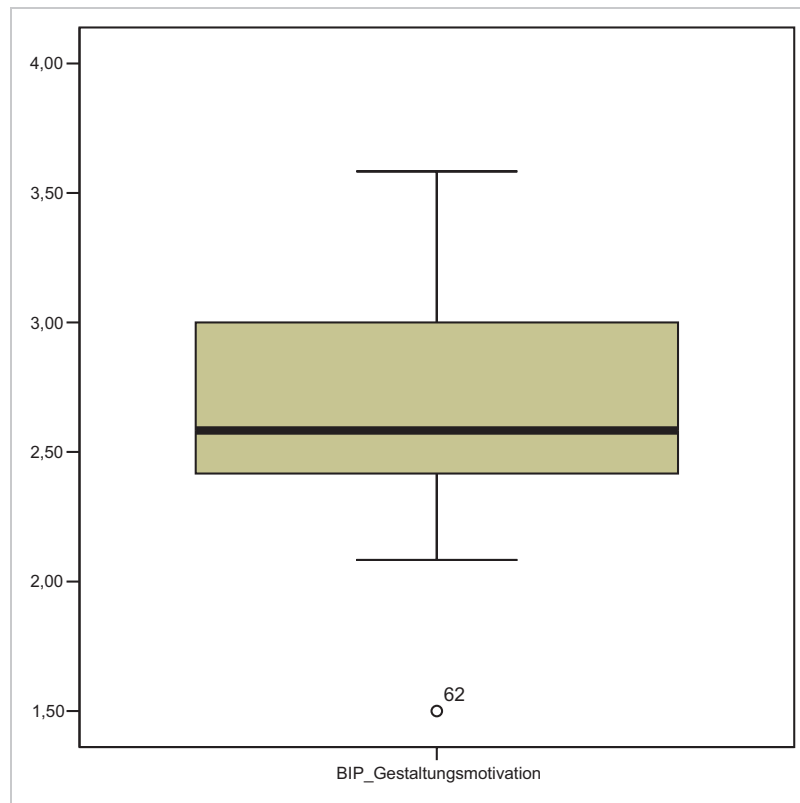


Abbildung C-11: Boxplot (Gestaltungsmotivation)

C.6.5 BIP_Handlungsorientierung (n = 2)

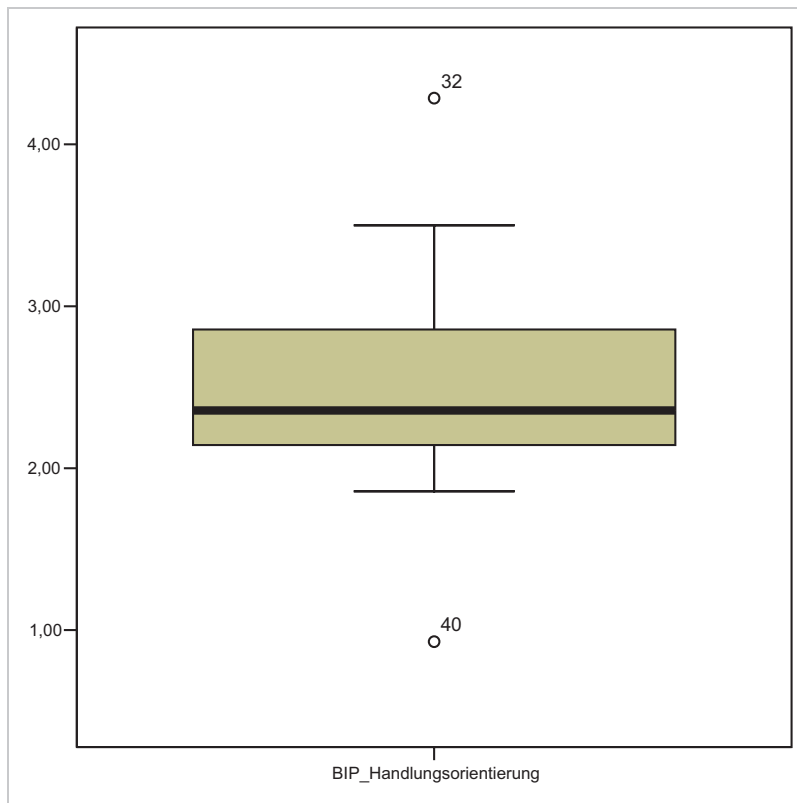


Abbildung C-12: Boxplot (Handlungsorientierung)

C.6.6 BIP_Belastbarkeit (n = 2)

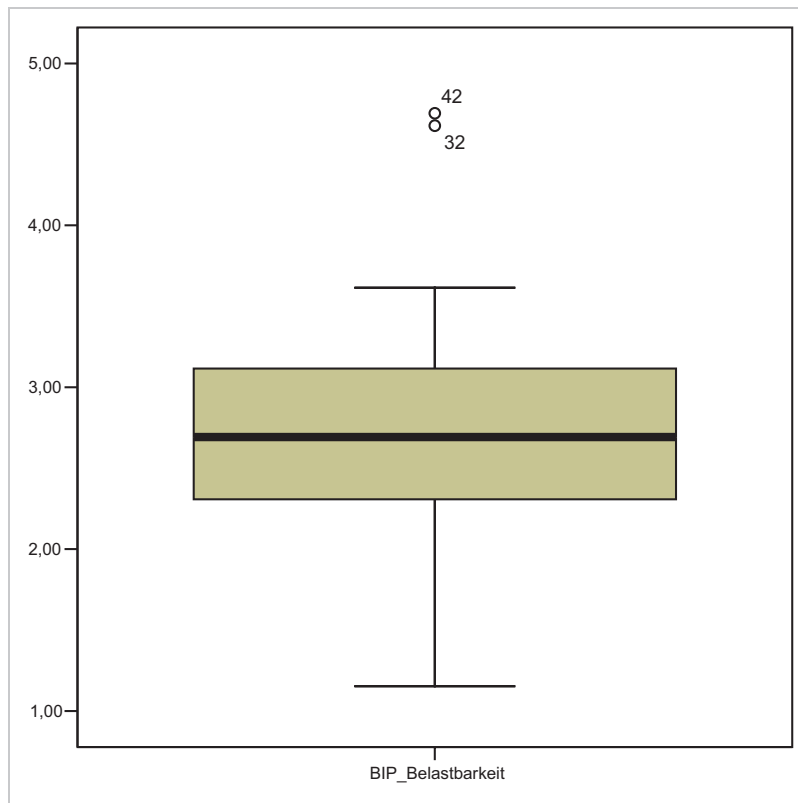


Abbildung C-13: Boxplot (Belastbarkeit)

C.6.7 BIP_Selbstbewusstsein (n = 1)

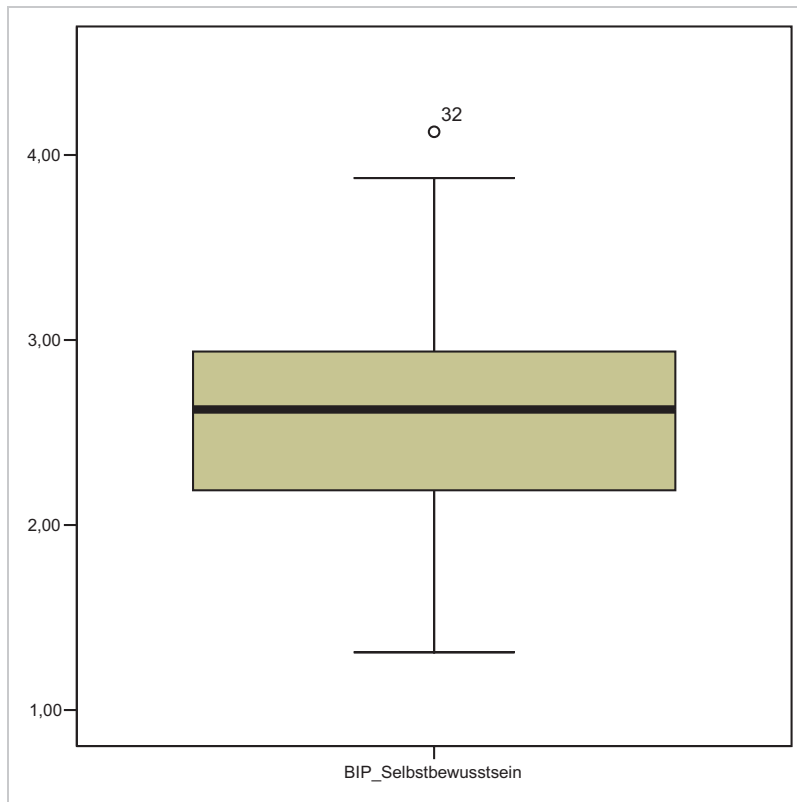


Abbildung C-14: Boxplot (Selbstbewusstsein)

C.6.8 BIP_Emotionale_Stabilität (n = 5)

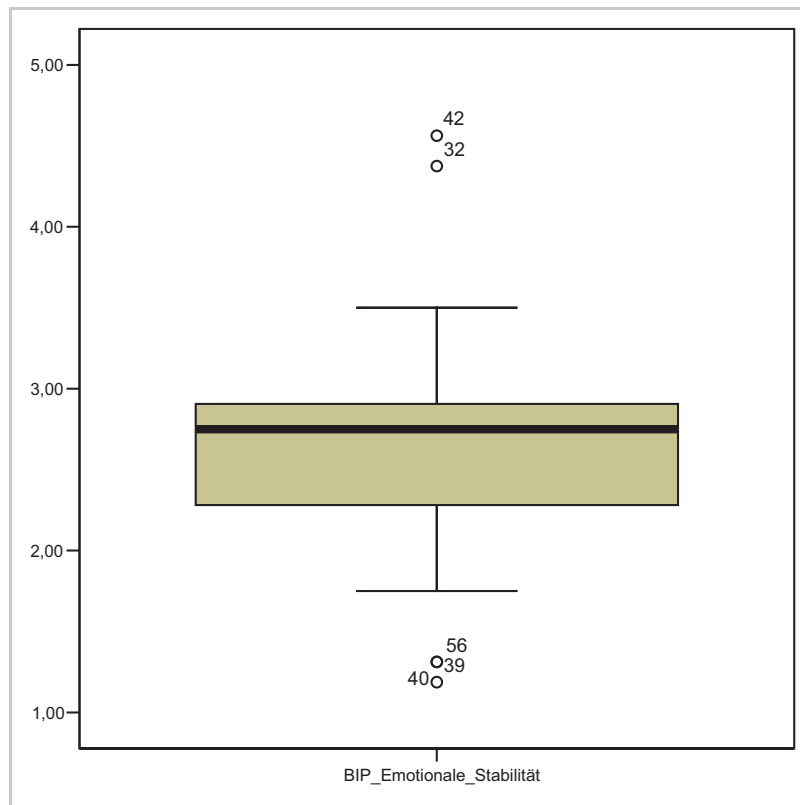


Abbildung C-15: Boxplot (Emotionale Stabilität)

C.7 Minderungskorrigierte Korrelationen zwischen dem SJT_Hotelfach und dem BIP

Tabelle C-8: Bivariate Korrelationen ($r_{cor.}$) zwischen dem SJT_Hotelfach und den Eignungsvoraussetzungen des BIP (nach einfachen Minderungskorrekturen)

Berufliche Orientierung		
	Video	Paper-Pencil
Gestaltungsmotivation	-.10	-.54
Leistungsmotivation	-.18	.05
Arbeitsverhalten		
	Video	Paper-Pencil
Handlungsorientierung	.53	.15
Flexibilität	.56	-.46
Gewissenhaftigkeit	.16	-.03
Psychische Konstitution		
	Video	Paper-Pencil
Emotionale Stabilität	.10	-.14
Belastbarkeit	.41	.23
Selbstbewusstsein	.18	-.09
Soziale Kompetenz		
	Video	Paper-Pencil
Sensitivität	.44	.43
Kontaktfähigkeit	-.17	-.02
Soziabilität	.12	-.25
Teamorientierung	.03	.37
Durchsetzungsstärke	-.06	-.26

Anhang D: Vierte Studie

D.1 Zusätzliche Statistiken

D.1.1 Überprüfung der Normalverteilungsannahme

Tabelle D-1: Ergebnisse der Kolmogorov-Smirnov-Tests

Item	Z	Signifikanz	Normalverteilung
Situation 1		1.35	.05 Ja
Situation 5		1.42	.04 Nein
Situation 10		.82	.51 Ja
Situation 11		1.03	.24 Ja
Situation 13		.99	.28 Ja
Situation 17		.83	.50 Ja
Situation 18		1.28	.08 Ja
Situation 19		1.21	.11 Ja
Situation 21		1.22	.10 Ja
Situation 22		1.23	.10 Ja
Situation 23		1.21	.11 Ja
Situation 24		.58	.88 Ja
Situation 25		1.28	.08 Ja

D.1.2 Der SJT_Hotelfach und verschiedene Aspekte der Berufserfahrung

Tabelle D-2: Korrelationen zwischen dem SJT_Hotelfach und Maßen der Berufserfahrung

	Erfahrung im Hotelfach insgesamt (in Jahren)	Erfahrung in jetziger Position (in Jahren)	Längster Zeitraum in einem Unternehmen	Anzahl an Unternehmen
SJT_Hotelfach	-.06	-.15	-.02	-.03

Tabelle D-3: Einfluss der Führungsverantwortung auf die Leistung im SJT_Hotelfach (t-Test für unabhängige Stichproben)

		Levene-Test auf Varianzhomogenität			
		F	Signifikanz	t	Signifikanz (2-seitig)
SJT_Hotelfach	Annahme gleicher Varianzen	.00	.97	.07	.95
	Annahme ungleicher Varianzen			.07	.95

ERKLÄRUNG

Hiermit versichere ich, die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet sowie Zitate kenntlich gemacht zu haben.

Münster, Juni 2010

Kundenorientierung im Hotelfach

Jan Henning Möller

Situational Judgment Tests (SJTs) gelten aufgrund ihrer überzeugenden psychometrischen Gütekriterien und ihres starken Anforderungsbezugs als nützliche Bereicherung klassischer Verfahren der Berufseignungsdiagnostik. Sie versetzen Bewerber in realistische Arbeitsplatzsituationen im Video- oder Paper-Pencil Format und konfrontieren diese mit einer Vielzahl unterschiedlich effektiver Verhaltensalternativen. In dem vorliegenden Buch wird die Entwicklung eines videobasierten SJT zur Messung von Kundenorientierung im Hotelfach beschrieben. Die Testentwicklung wurde in enger Zusammenarbeit mit einer international agierenden Hotelkette realisiert. An die Darstellung der Testentwicklung schließt sich die Beschreibung verschiedener Studien zur Untersuchung der Konstruktvalidität an. So werden u. a. Zusammenhänge mit Maßen wie der kognitiven Leistungsfähigkeit, der sozialen Kompetenz und der berufsbezogenen Persönlichkeit aufgeklärt sowie diskutiert. Neben den genannten Punkten enthält das Buch einen umfassenden Überblick über die aktuelle Forschungslage zu SJTs und eine kritische Auseinandersetzung mit eben jener.

ISBN 978-3-8405-0026-8 EUR 25,50

0 2 5 5 0



9 783840 500268