

Universität Mannheim

**Aspekte der Konstruktvalidität und sozialen
Validität von strukturierten Interviews am
Beispiel des strukturierten Interviews Opus[®]**

Diplomarbeit

Juniorprofessur für Wirtschaftspsychologie

vorgelegt bei Prof. Dr. Karsten Müller

Zweitgutachten von Prof. Dr. Thorsten Meiser

betreut von Tammo Straatmann

eingereicht von Sarah Klomann

Sarah Klomann

14.06.11

Schorlenberger Straße 32

67677 Enkenbach-Alsenborn

Danksagung

An dieser Stelle möchte ich denjenigen Personen danken, ohne die diese Arbeit nicht möglich gewesen wäre.

Ich danke Herrn Prof. Dr. Karsten Müller und Herrn Prof. Dr. Thorsten Meiser, die sich als Gutachter zur Verfügung gestellt haben. Ebenso danke ich Herrn Tammo Straatmann für die engagierte Betreuung dieser Arbeit.

Mein Dank gilt auch der O&P Consult AG, die es mir ermöglicht hat, diese empirische Arbeit zu verfassen. Mein besonderer Dank gilt hierbei Frau Dr. Anja Schmitz für ihr Vertrauen und ihre Unterstützung.

Abschließend möchte ich mich ganz herzlich bei meiner Familie, meinem Freund und all meinen Freunden bedanken, die mir in dieser Zeit geduldig zur Seite standen.

Inhaltsverzeichnis

Tabellenverzeichnis.....	III
Zusammenfassung	IV
1 Einleitung.....	1
2 Theoretischer Hintergrund.....	3
2.1 Verfahren zur Personalauswahl	3
2.1.1 Überblick der Verfahren zur Personalauswahl	4
2.1.2 Strukturierte Interviews.....	8
2.2 Gütekriterien zur Beurteilung von Verfahren zur Personalauswahl	9
2.2.1 Überblick der Validitätsarten	9
2.2.1.1 Konstruktvalidität.....	10
2.2.1.2 Soziale Validität	14
2.3 Validität strukturierter Interviews	16
2.4 Ableitungen für die empirische Untersuchung.....	19
2.4.1 <i>Opus[®]</i> als Beispiel eines strukturierten Interviews	19
2.4.2 <i>Zentrale Fragestellungen und Hypothesen</i>	21
2.4.2.1 Fragestellung und Hypothesen zur Konstruktvalidität	21
2.4.2.2 Fragestellung und Hypothese zur sozialen Validität	22
3 Empirischer Teil.....	24
3.1 Methodisches Vorgehen.....	24
3.1.1 <i>Vorstellung des Untersuchungssettings</i>	24
3.1.2 <i>Ablauf der Studie</i>	25
3.1.3 <i>Beschreibung der Stichprobe</i>	26
3.1.4 <i>Beschreibung der Erhebungsinstrumente</i>	26
3.1.4.1 <i>Selbsteinschätzung</i>	26
3.1.4.2 <i>Wilde Intelligenztest-2</i>	27
3.1.4.3 <i>ASK</i>	29
3.1.4.4 <i>Fragebogen zur sozialen Validität</i>	31
3.1.5 <i>Beschreibung der Analyseverfahren</i>	35
3.1.5.1 <i>Analyse des Fragebogens zur sozialen Validität</i>	36

3.1.5.2 Analyse der Konstruktvalidität.....	38
3.1.5.3 Analyse der sozialen Validität	40
3.1.5.4 Umgang mit Missings.....	40
3.1.5.5 Poweranalyse	42
3.2 Darstellung der Ergebnisse	42
3.2.1 <i>Soziodemografische Zusammensetzung der Stichprobe</i>	43
3.2.2 <i>Gütekriterien des Fragebogens zur sozialen Validität</i>	43
3.2.3 <i>Ergebnisse zur Konstruktvalidität</i>	51
3.2.4 <i>Ergebnisse zur sozialen Validität</i>	57
3.2.5 <i>Weiterführende Analysen</i>	60
4 Diskussion.....	61
4.1 Interpretation der empirischen Ergebnisse.....	61
4.2 Kritische Reflexion der Studie	66
4.3 Implikationen für Theorie und Praxis	68
5 Literaturverzeichnis.....	71
Anhang	V

Tabellenverzeichnis

Tabelle 2-1: Veranschaulichung einer fiktiven Multitrait-Multimethod-Matrix (nach Campbell & Fiske, 1959).....	13
Tabelle 3-1: Rotierte Faktorladungsmatrix und Varianzaufklärung der fünf extrahierten Faktoren des Fragebogens zur sozialen Validität	47
Tabelle 3-2: Cronbachs Alpha und Trennschärfe der verwendeten Items und Skalen der fünf extrahierten Faktoren des Fragebogens zur sozialen Validität	50
Tabelle 3-3: Komplette Korrelationsmatrix zur Konstruktvalidität	53
Tabelle 3-4: Multitrait-Multimethod-Matrix zur Konstruktvalidität.....	56

Zusammenfassung

Trotz seines häufigen Einsatzes zur Personalauswahl sowie seiner guten Vorhersagegüte, herrscht über das strukturierte Interview bisweilen wenig Konsens. Aufgrund der aktuellen Nachfrage nach einer fundierten Personalauswahl in Anbetracht des steigenden Fachkräftemangels, stellt sich aus Forschungs- wie auch aus Anwendungssicht die Frage: Was genau messen strukturierte Interviews? Neben dieser Frage nach der Konstruktvalidität hat auch die Frage nach der sozialen Validität in den letzten Jahrzehnten immer mehr an Bedeutung gewonnen: Wie kommt das Verfahren bei den Bewerbern an? Die vorliegende Arbeit greift zur Beantwortung dieser Fragen exemplarisch ein strukturiertes Interview (Opus[®]) heraus und setzt es in Bezug zu anderen Testverfahren (WIT-2, ASK) sowie einer Selbsteinschätzung. Dazu wurden die Daten von 57 Personen erfasst. Die Konstruktvalidität des Opus[®]-Interviews wurde anhand einer Multitrait-Multimethod-Matrix (Campbell & Fiske, 1959) untersucht. Es zeigten sich zwar signifikante Korrelationen hinsichtlich der konvergenten Validität, allerdings konnte keine diskriminante Validität nachgewiesen werden. Am höchsten fielen die Korrelationen zwischen den verschiedenen Opus[®]-Kriterien aus, was auf einen Methodeneffekt hindeutet. Dieser zeigte sich jedoch teilweise auch bei den anderen Verfahren. Um die Frage nach der sozialen Validität des Opus[®]-Interviews zu beantworten, wurde eigens ein Fragebogen entwickelt, der fünf Faktoren der sozialen Validität enthält. In den Faktoren „Akzeptanz und resultierende Motivation“ ($t(30) = 3.34$, $p = .001$), „Testerleben“ ($t(30) = 3.92$, $p < .001$) und „Jobbezug“ ($t(30) = 3.00$, $p = .003$) sowie im direkten Vergleich der Verfahren schnitt das Opus[®]-Interview durchweg besser ab als die Testverfahren. Der Faktor „Transparenz“ wurde nicht signifikant, während die Signifikanz beim Faktor „Achtung der Privatsphäre“ von der Reihenfolge der durchlaufenen Verfahren abhing. Die Studie zeigt, dass die Befunde zur Konstruktvalidität von Opus[®] durchmischt sind und die entsprechenden Hypothesen nicht bestätigt werden konnten. Allerdings zeigte sich auch, dass das Opus[®]-Interview wie erwartet von den Teilnehmern zum Großteil besser bewertet wurde als die Testverfahren.

1 Einleitung

Zurzeit gewinnt das Thema Personalauswahl nicht zuletzt aufgrund der momentanen wirtschaftlichen Lage wieder an Brisanz. In Anbetracht des steigenden Mangels an Fachkräften, steigt auch die Nachfrage nach einer fundierten Personalauswahl. Dabei sind Unternehmen stets bemüht den passenden Bewerber für die entsprechende Position auszuwählen. Dies soll nicht nur ein einwandfreies Funktionieren der Unternehmung garantieren, sondern ist auch unter wirtschaftlichen Aspekten unerlässlich, wie Brodgen (1949) schon herausarbeitete. In seiner viel zitierten Arbeit „When testing pays off“ entwickelte er bereits 1949 eine Formel, die den monetären Nutzen einer guten Personalauswahl erfasst. Somit wird deutlich, dass eine wissenschaftlich fundierte Personalauswahl einen direkten Wettbewerbsvorteil darstellen kann.

Auf dem Markt gibt es heute eine nahezu unendlich erscheinende Auswahl an Verfahren, die zur Personalauswahl eingesetzt werden. Neben den klassischen Intelligenztests entwickelt zurzeit fast jedes (Beratungs-)Unternehmen sein eigenes Auswahlinstrument. Doch nicht alle Verfahren erscheinen für diesen Zweck gleich geeignet. Umso wichtiger ist eine Beurteilung dieser Verfahren. Diese erfolgt zumeist anhand klassischer Gütekriterien, wie der Validität des Verfahrens. Darüber hinaus wurde z. B. auch im Sinne eines Verbraucherschutzes die DIN 33430 veröffentlicht, die „Anforderungen an Verfahren und ihren Einsatz bei berufsbezogenen Eignungsbeurteilungen“ enthält (Hornke & Winterfeld, 2004, S. 5). Dabei verlagert sich der Blick langsam in Richtung Bewerber. So ist es für das Unternehmen nicht nur wichtig, zur Personalauswahl ein fundiertes Verfahren mit hohen klassischen Gütekriterien einzusetzen, sondern dabei auch die Akzeptanz dieses Verfahrens durch die Bewerber zu berücksichtigen: die soziale Validität des Verfahrens. Da mit dieser Akzeptanz eine Reihe von weiteren Verhaltensweisen verknüpft ist (Smither, Reilly, Millsap, Pearlman & Stoffey, 1993; Hausknecht, Day & Thomas, 2004), sollte es im Sinne eines jeden Unternehmens sein, die Wirkung des eingesetzten Verfahrens auf den Bewerber zu untersuchen.

Somit wird deutlich, dass verschiedene Gütekriterien bei der Personalauswahl sowohl aus Sicht der Anwender als auch aus Bewerbersicht eine zentrale Rolle spielen. Deshalb wurde in der vorliegenden Arbeit exemplarisch ein konkretes Verfahren zur Personalauswahl herangezogen, um daran verschiedene Gütekriterien zu untersuchen. Dabei wurde der Fokus auf die Konstruktvalidität und die soziale Validität strukturierter Interviewverfahren gerichtet.

Das übergeordnete Thema der vorliegenden Arbeit sind Gütekriterien von Verfahren zur Personalauswahl. Um zu einem besseren Verständnis der zugehörigen Grundlagen zu gelangen, werden im theoretischen Teil unterschiedliche Verfahren der Personalauswahl vorgestellt. Des Weiteren wird auf einzelne Gütekriterien als Beurteilungsmaßstäbe eingegangen. Danach wird zusammenführend die Validität von strukturierten Interviews betrachtet. Schließlich werden Ableitungen für die empirische Studie getroffen, um zum zentralen Thema dieser Arbeit hinzuführen: Konstruktvalidität und soziale Validität am Beispiel des strukturierten Interviews Opus[®]. Der empirische Teil beschreibt das methodische Vorgehen und die Ergebnisse der Untersuchung. In der abschließenden Diskussion werden die gefundenen Ergebnisse sowie die Studie kritisch beleuchtet und es werden Schlussfolgerungen für Theorie und Praxis gezogen.

2 Theoretischer Hintergrund

Der theoretische Hintergrund dieser Arbeit befasst sich zunächst mit Verfahren zur Personalauswahl, wobei insbesondere strukturierte Interviewverfahren im Vordergrund stehen. Danach werden Gütekriterien zur Beurteilung von Verfahren zur Personalauswahl erläutert, wobei besonders auf die Konstruktvalidität sowie die soziale Validität eingegangen wird. Anschließend wird die Validität strukturierter Interviewverfahren betrachtet. Schließlich werden Ableitungen für die empirische Studie getroffen, wobei das hier eingesetzte Interviewverfahren Opus[®] näher erläutert und konkrete Hypothesen zur Untersuchung der Konstruktvalidität und sozialen Validität des Opus[®]-Interviews aufgestellt werden.

2.1 Verfahren zur Personalauswahl

Verfahren zur Eignungsdiagnostik traten in der Geschichte erstmals im Zuge des Ersten Weltkrieges in den Vordergrund (Greif, 2007). Es galt damals in kurzer Zeit viele fähige Soldaten auszuwählen. Insbesondere in den USA wurden dazu Intelligenztests eingesetzt. Nach Kriegsende wurde die Idee zur Auswahl von Arbeitern anhand psychologischer Tests von der Industrie übernommen, woraus sich später der Zweig der Eignungsdiagnostik entwickelte. Die Berufseignungsdiagnostik bezeichnet im weiteren Sinne „die Methodologie der Entwicklung, Prüfung und Anwendung psychologischer Verfahren zum Zwecke eignungsbezogener Erfolgsprognosen und Entscheidungshilfen im beruflichen Kontext“ (Schuler, 2007, S. 430). So wurden im Laufe der Zeit verschiedene Instrumente entwickelt, anhand derer eine Eignung des Bewerbers vorhergesagt werden soll. Im Folgenden werden die gebräuchlichsten dieser Verfahren kurz vorgestellt, wobei der Schwerpunkt auf den strukturierten Interviews liegen soll.

2.1.1 Überblick der Verfahren zur Personalauswahl

Allgemeine kognitive Fähigkeitstests

Psychologische Tests stellen die klassische Methode zur Erfassung von Eigenschaften schlechthin dar. Sie sind die am häufigsten eingesetzten Instrumente in der Eignungsdiagnostik und werden in den unterschiedlichsten Formen z. B. als Leistungstests, Persönlichkeitstests oder Interessentests eingesetzt (Schuler & Höft, 2001). Allgemeine kognitive Fähigkeiten werden häufig auch als „Intelligenz“ bezeichnet. In der Forschung gibt es die verschiedensten Ansätze, wie viele Intelligenzfacetten zu unterscheiden sind und ob man von einem g-Faktor (Spearman, 1904), also einem übergeordneten allgemeinen Intelligenzfaktor ausgehen kann. Oft wird dabei eine Unterscheidung in Generalfaktortheorien, nichthierarchische Intelligenztheorien, hierarchische Intelligenztheorien und Integrationsmodelle getroffen (Hell, 2007). Unabhängig von der jeweiligen zugrunde liegenden Theorie zielen Intelligenztests auf die Qualität und Geschwindigkeit der Lösung von neuen Aufgaben ab (Schuler & Höft, 2001). Ihnen kommt insbesondere deshalb eine hohe Aufmerksamkeit zu, weil sie weit verbreitet sind und eine hohe prognostische Validität besitzen (Schmidt & Hunter, 1998; Hülshager, Maier, Stumpp & Muck, 2006). Etwa ein Drittel der deutschen Unternehmen nutzt Intelligenztests zur Personalauswahl (Schuler, Hell, Trapmann, Schaar & Boramir, 2007). Seit ihrem Einsatz im Ersten Weltkrieg bis heute zeigen Intelligenztests, dass sich das allgemeine intellektuelle Leistungsvermögen zwischen verschiedenen Berufsgruppen unterschiedlich verteilt (Schuler & Höft, 2001). Schmidt und Hunter ermittelten in ihrer viel zitierten Metaanalyse von 1998 eine Vorhersagegüte von .51 für allgemeine kognitive Fähigkeitstest, was nur durch Arbeitsproben noch übertroffen wurde. Es gibt wohl keinen Beruf für dessen Erfolgsprognose sich Intelligenztests nicht eignen (Schuler & Höft, 2007). Somit gelten Intelligenztests unabhängig von der Art des Berufes als bester Prädiktor für Ausbildungs- und Berufserfolg.

Arbeitsproben

Arbeitsproben gehören zu den ältesten Methoden der Personaldiagnostik und wurden schon 1912 von Münsterberg eingesetzt. Das Vorgehen erscheint auf den ersten Blick recht simpel. Es werden Arbeitsaufgaben aus dem beruflichen Alltag gewählt, die vom Bewerber bewältigt werden müssen. Aufgrund dieses engen Bezugs zur Arbeitstätigkeit sind hierbei Anforderungsanalysen besonders wichtig (Görlich, 2007). In der Metaanalyse von Schmidt und Hunter (1998) zeigten Arbeitsproben die höchste Vorhersagegüte mit .54. Allerdings ist die Konstruktion einer Arbeitsprobe aufgrund des engen Tätigkeitsbezuges in Relation zum Einsatz eines allgemeineren Intelligenztests ungleich aufwendiger. Zudem beansprucht ihre Durchführung viele zeitliche und personelle Ressourcen. Ein weiteres Problem besteht darin, dass die erfolgreiche Bewältigung einer Arbeitsprobe eine gewisse Fachkompetenz voraussetzt und somit nur bedingt bei Berufsanfängern einsetzbar ist (Kanning, 2004). Man versucht diesen Umstand durch sogenannte „Trainability-Tests“ zu umgehen. Dabei wird dem Bewerber eine vorgeschaltete Trainingsphase gewährt, bevor er die eigentliche Arbeitsprobe bewältigen muss. Somit wird die Lernfähigkeit des Bewerbers miteinbezogen. Arbeitsproben kommen bei Bewerbern meist gut an, weil ihre Funktionsweise einfach und intuitiv verständlich ist.

Assessment-Center

Ein Assessment-Center kombiniert verschiedene eignungsdiagnostische Verfahren. Es handelt sich dabei um eine multiple Verfahrenstechnik, bei der „simultan mehrere Teilnehmer von mehreren Beobachtern in mehreren Verfahren hinsichtlich mehrerer definierter Anforderungen“ beurteilt werden (Höft & Funke, 2001, S. 150). Den Kern eines Assessment-Centers stellt die Anforderungs-Verfahrens-Matrix dar. In ihr zeigt sich, wie der Bewerber in verschiedenen Übungen hinsichtlich der untersuchten Kriterien abgeschnitten hat. Ein Gesamtergebnis für jedes Kriterium ergibt sich entweder aus der Aggregation der einzelnen Werte oder innerhalb einer Diskussion in der

abschließenden Beobachterkonferenz. Assessment-Center liegen mit einer Vorhersagegüte von .37 im mittleren Bereich (Schmidt & Hunter, 1998), wobei ihre Einsatzhäufigkeit in den vergangenen Jahren stetig zugenommen hat (Schuler et al., 2007).

Auswertung der Bewerbungsunterlagen und Referenzen

Etwa 99% der deutschen Unternehmen nutzen die Auswertung der Bewerbungsunterlagen zur Auswahl von Bewerbern (Schuler et al., 2007). Allerdings werden Bewerbungsunterlagen nur selten als alleinige Entscheidungsgrundlage zur Besetzung einer Stelle genutzt. Vielmehr bilden sie den Anfang eines längeren Auswahlprozesses. Die Analyse der Bewerbungsunterlagen erfolgt dabei aus zeitökonomischen Gründen häufig anhand formaler Kriterien, wie beispielsweise der Vollständigkeit und Fehlerfreiheit (Moser, 2007). Inhaltlich wird vor allem auf die Übereinstimmung von Lebenslauf und Belegen, die Plausibilität eines Stellenwechsels und die Qualität von Arbeitszeugnissen geachtet. Anhand von Bewerbungsunterlagen bilden sich die Verantwortlichen einen ersten Eindruck vom Bewerber. Dabei wird die Information meist intuitiv kombiniert. Das Prüfen von Referenzen in Form von Einholen zusätzlicher Beurteilungsinformation von einem früheren Arbeitgeber wird häufiger im angloamerikanischen Raum praktiziert (Schuler & Höft, 2007). Im Gegensatz zur eher geringen Validität der intuitiven Urteilsbildung aufgrund von Bewerbungsunterlagen zeigt das Prüfen von Referenzen eine Vorhersagegüte von .26 (Schmidt & Hunter, 1998).

Biografische Fragebogen

Biografische Fragebogen können die unterschiedlichsten Fragen enthalten. Es ist ihnen jedoch gemein, dass vergangene und gegenwärtige Fakten eines Bewerbers die Grundlage zur Prognose zukünftiger Arbeitsleistung bilden. In der geschichtlichen Entwicklung wurden formal überprüfbare Fragen immer mehr durch Items ersetzt, die etwas über die Persönlichkeit des Bewerbers aussagen (Schuler & Höft, 2007). Die Items unterscheiden sich dabei von anderen Fragebogen, weil sie statt abstrakteren Verhaltensweisen und

Eigenschaften konkrete Fakten aus der Biografie des Bewerbers abfragen. Häufig werden dabei Items anhand ihrer Validität bezüglich eines Erfolgskriteriums ausgewählt (Schuler & Marcus, 2001). Ein Nachteil von biografischen Fragebogen ist, dass sie meist recht spezifisch sind und damit für jede Auswahl-situation neu konstruiert werden müssen. Dafür weisen sie eine recht hohe Validität auf. Für biografische Daten ermittelten Schmidt und Hunter (1998) eine Vorhersagegüte von .35.

Einstellungsinterviews

Einstellungsinterviews werden neben der Auswertung der Bewerbungsunterlagen am häufigsten in Personalauswahlprozessen genutzt. In der Studie von Schuler et al. (2007) setzten sogar fast alle befragten Unternehmen Interviews zur Auswahl neuer Mitarbeiter ein. Es geht im Einstellungsinterview um ein persönliches Gespräch zwischen dem Bewerber und einem Verantwortlichen des Unternehmens. Im Laufe des Gesprächs tauschen die beiden Informationen aus, wobei beurteilt werden soll, ob Bewerber und Zielposition zusammenpassen. Kennzeichnend für das Interview ist die damit einhergehende soziale Situation (Kanning, 2004). Da die Entscheidung am Ende des Gesprächs über komplexe Prozesse der sozialen Urteilsbildung zustande kommt, ist dieser Prozess in mancherlei Hinsicht störanfällig. Insbesondere problematisch ist, dass der Interviewer meist einen intuitiven Gesamteindruck bildet, womit die Objektivität des Verfahrens stark eingeschränkt ist (Schuler & Höft, 2007). Um typischen Problemen des Einstellungsinterviews entgegen zu wirken hat Schuler (2002) das Multimodale Interview vorgeschlagen. Es besteht aus acht aufeinander folgenden Komponenten, die sukzessiv durchlaufen werden. Die Besonderheit liegt in der Kombination verschiedener Frageformate. Zu Beginn des Gespräches stellen sich alle Beteiligten vor und stimmen sich auf das bevorstehende Interview ein. So soll gleich zu Beginn eine angenehme Gesprächsatmosphäre hergestellt werden. Die folgende Selbstvorstellung entspricht meist einer Präsentationsübung, wie sie häufig auch in Assessment-Centern zur Anwendung kommt. Sie mündet in freie Fragen. Fragen zur Berufswahl oder zu Handlungswissen

dienen dazu, das Interesse des Bewerbers an der Organisation und der Tätigkeit zu erfassen. Die biografiebezogenen Fragen erfassen vergangenes Verhalten des Bewerbers, welches als Prädiktor für zukünftiges Verhalten dient. In der anschließenden realistischen Tätigkeitsinformation wird der Bewerber über Anforderungen und Arbeitsbedingungen informiert, so dass er sich ein konkretes Bild von seinem zukünftigen Arbeitsplatz machen kann. Die situativen Fragen zielen darauf ab, wie der Bewerber in verschiedenen Situationen handeln würde und können im weitesten Sinne als hypothetische Arbeitsprobe betrachtet werden. Der Gesprächsabschluss bietet beiden Seiten die Möglichkeit, offene Fragen zu klären sowie das weitere Vorgehen zu besprechen. Das Multimodale Interview wird immer häufiger auch in verschiedenen Kontexten eingesetzt und zeigt gute Validitäten zwischen .30 und .50 (Schuler & Höft, 2007).

2.1.2 Strukturierte Interviews

Interviews kommen in den verschiedensten Bereichen der Psychologie zum Einsatz. Unabhängig davon, ob sie nun in einem klinischen Kontext oder zur Personalauswahl eingesetzt werden, unterscheiden sie sich prinzipiell hinsichtlich ihres Grades an Strukturierung. In einem unstrukturierten Interview gibt es keine festgelegten Fragen. Somit kann es sein, dass jedem Bewerber unterschiedliche Fragen gestellt werden, worunter die Vergleichbarkeit stark leidet (Weinert, 2004). Auch gibt es für die Auswertung kein festgelegtes Verfahren, weshalb meist nicht die einzelnen Antworten des Bewerbers gewertet werden, sondern nur ein Gesamturteil gebildet wird, was zu weniger validen Ergebnissen führt (Ganzach, Kluger & Klayman, 2000). Das Ausmaß der Strukturierung lässt sich anhand eines Kontinuums erfassen. Den Gegenpol zum unstrukturierten Interview bilden vollständig strukturierte Interviews, bei denen allen Bewerbern „mit demselben Wortlaut dieselben Fragen in einer zuvor festgelegten Reihenfolge gestellt werden“ (Kanning, 2004, S.418). Die Strukturierung kann sich dabei auf verschiedene Variablen des Interviews beziehen, vornehmlich auf den Inhalt wie beispielsweise der Einsatz einer Anforderungsanalyse oder auf die Auswertung wie beispielsweise die

Verwendung verankerter Rating-Skalen (Campion, Palmer & Campion, 1997). Wenn also von strukturierten Interviews gesprochen wird, handelt es sich dabei nicht um eine eindeutig definierte Kategorie (Chapman & Zweig, 2005). Wie sich der Grad der Strukturierung auf die Validität des Interviews auswirkt, soll in Kapitel 2.3 näher beleuchtet werden.

2.2 Gütekriterien zur Beurteilung von Verfahren zur Personalauswahl

Verfahren zur Personalauswahl benötigen wie andere Testverfahren auch Bewertungsmaßstäbe, die eine Beurteilung ihrer Güte erlauben. In der Literatur sind hierzu zentrale Kriterien definiert, anhand derer eine Aussage über die psychometrischen Eigenschaften solcher Verfahren getroffen werden kann (Amelang & Schmidt-Atzert, 2006). Diese Gütekriterien orientieren sich hauptsächlich an der klassischen Testtheorie und lauten Objektivität, Reliabilität und Validität (Amelang & Zielinski, 1994). Diese sogenannten Hauptgütekriterien sind in strengem Maße anzulegen, um die Qualität eines Verfahrens zu beurteilen (Amelang & Schmidt-Atzert, 2006). Darüber hinaus gibt es noch eine Reihe weiterer Gütekriterien, die als Nebengütekriterien bezeichnet werden und die drei Hauptgütekriterien ergänzen. Dabei wird insbesondere auf die Normierung und Fairness eines Verfahrens verwiesen. Die Fairness wird laut Görlich und Schuler (2006) nochmals in fünf Teilaspekte untergliedert, unter anderem in die prozedurale Fairness, also die soziale Validität.

2.2.1 Überblick der Validitätsarten

Einen zentralen Stellenwert innerhalb der Hauptgütekriterien nimmt die Validität ein. Im traditionellen Drei-Validitäten-Modell (Cronbach & Meehl, 1955) werden dabei Inhalts-, Kriteriums- und Konstruktvalidität unterschieden. Bei der Inhaltsvalidität handelt es sich um die Repräsentativität des Verfahrens hinsichtlich des zu erfassenden Merkmals. Sie basiert meist auf logischen und fachlichen Überlegungen (Fisseni, 2004). Die Kriteriumsvalidität beschreibt den Zusammenhang zwischen einem Verfahren und einem weiteren Kriterium. Sie lässt sich untergliedern in eine konkurrente und eine prädiktive Gültigkeit, je nachdem, ob Verfahren und Kriterium nahezu zeitgleich oder in zeitlichem

Abstand erhoben wurden (Amelang & Schmidt-Atzert, 2006). Die Konstruktvalidität zielt auf die theoretische Einbettung eines Verfahrens in Bezug auf andere Verfahren und Konzepte ab und wird in Abschnitt 2.2.1.1 näher erläutert.

Im Gegensatz zu den bisher genannten Validitätsarten ist die soziale Validität nicht mit den klassischen Validitätskonzepten verwandt und von ihrer Bedeutung her in die Klasse der Nebengütekriterien einzuordnen. Darin ist wohl auch begründet, dass sie weniger scharf definiert ist und nicht wie die Hauptgütekriterien bei Verfahren konventionell angegeben wird. Allerdings ist ihre Bedeutung im Kontext der Personalauswahl nicht zu vernachlässigen (Smither et al., 1993; Hausknecht et al., 2004), weswegen sie in Abschnitt 2.2.1.2 genauer betrachtet wird.

2.2.1.1 Konstruktvalidität

Bedeutung

Die Konstruktvalidität eines Verfahrens ist kein Konzept, das sich einfach in Zahlen ausdrücken lässt. Im Gegensatz zur Kriteriumsvalidität, die eine klare Vorgehensweise und Quantifizierung zulässt, gibt es zur Konstruktvalidität keine Handlungsanweisung (Fisseni, 2004). Die Vorgehensweise der Ermittlung von Konstruktvalidität besteht grundsätzlich darin, das einem Test zugrunde liegende Konstrukt zu erfassen und mit anderen Tests und deren zugrunde liegenden Konstrukten zu vergleichen und vor allen Dingen auch voneinander abzugrenzen, um eine Einbettung in ein größeres nomologisches Netzwerk zu erzielen (Amelang & Schmidt-Atzert, 2006). Diese Vorgehensweise ist eher als ein andauernder Validierungsprozess zu sehen. Da die damit einhergehenden Konzepte, Überlegungen und Maßnahmen aus verschiedenen Bereichen der Psychologie hierbei verknüpft werden, wird die Konstruktvalidität zum übergeordneten Ziel. In diesem Sinne lässt sich auch behaupten: „Die Konstruktvalidität schließt inhaltliche und kriteriumsbezogene Validität ein“ (Fisseni, 2004, S.68). Ein Beispiel dafür, wie sich in diesem Sinne verschiedene

Validierungsarten miteinander verknüpfen lassen, stellt die Multitrait-Multimethod-Validierung (Campbell & Fiske, 1959) dar.

Multitrait-Multimethod-Validierung

Die Multitrait-Multimethod-Validierung wurde erstmals 1959 von Campbell und Fiske veröffentlicht und beeinflusste im Laufe der Jahre grundlegend die Psychodiagnostik (Eid & Diener, 2006). Sie wird inzwischen als „vergleichsweise sophistiziertes Validierungskonzept“ (Amelang & Schmidt-Atzert, 2006, S.159) angesehen.

Der Multitrait-Multimethod-Ansatz folgt einer Reihe von Grundsätzen zur Validierung (Campbell & Fiske, 1959): (1) Um Validität nachzuweisen, betrachtet man in erster Linie die Konvergenz. Es wird versucht, ein Konstrukt mittels verschiedener Messungen zu belegen. (2) Um ein Konstrukt zu validieren, ist zusätzlich diskriminante Validität erforderlich. Das bedeutet, dass sich das getestete Konstrukt deutlich von anderen abheben muss. (3) Jedes eingesetzte Verfahren um ein Konstrukt zu messen, bildet eine Trait-Method-Unit. Das heißt, dass jedes Messergebnis sowohl Anteile des zu messenden Konstrukts als auch Anteile der gewählten Methode in sich vereint und somit nicht ein Konstrukt an sich erfasst werden kann. Diese Konfundierung von Konstrukt und Methode führt dahin, dass man versucht, den Methodenanteil aus dem Messergebnis herauszufiltern. (4) Dies geschieht, indem man mehr als ein Konstrukt mit mehr als einer Methode erfasst. Die Trennung von Methode und Konstrukt gelingt dann am besten, wenn die untersuchten Konstrukte und eingesetzten Methoden stark divergieren (Fisseni, 2004). Das Ziel dieser Form von Validierung ist zum einen der Nachweis konvergenter Validität: Die Ergebnisse für ein und dasselbe Konstrukt sollen über verschiedene Methoden hinweg hoch korrelieren. Der Nachweis diskriminanter Validität hingegen ist gelungen, wenn die verschiedenen Konstrukte, welche mit ein und derselben Methode erfasst wurden, niedrig korrelieren (Bortz & Döring, 2009). Auf den Punkt gebracht heißt das: Die Multitrait-Multimethod-Analyse untersucht

mehrere Konstrukte mit jeweils mehreren Methoden, wobei sie das Ziel verfolgt, konvergente und diskriminante Validität nachzuweisen.

Um diesen Nachweis zu bringen, wird eine Matrix aus Konstrukten und Methoden erstellt, die eigentliche Multitrait-Multimethod-Matrix. In ihr finden sich die verschiedenen Korrelationen zwischen den Konstrukten und Methoden wieder. Um die Bedeutung der einzelnen Matrix-Komponenten anschaulicher darlegen zu können, findet sich in Tabelle 2-1 exemplarisch eine fiktive Multitrait-Multimethod-Matrix nach Campbell und Fiske (1959). In ihr sind die Konstrukte A, B und C jeweils erfasst durch die Methoden 1, 2 und 3 enthalten. Aus der Kombination ergeben sich neun Variablen von A_1 bis C_3 . Die Multitrait-Multimethod-Matrix umfasst vier Hauptkomponenten: die Reliabilitätsdiagonale, die Validitätsdiagonalen, die Heterotrait-Monomethod-Dreiecke und die Heterotrait-Heteromethod-Dreiecke. Um diese vier Komponenten systematisch beschreiben zu können, ist eine Einteilung in 2 Blöcke sinnvoll: Monomethod-Blöcke und Heteromethod-Blöcke. Die Monomethod-Blöcke enthalten alle Werte, die zu einer Methode gehören. Sie setzen sich aus der Reliabilitätsdiagonalen und den Heterotrait-Monomethod-Dreiecken zusammen. Die Reliabilitätsdiagonale ist die Hauptdiagonale der Multitrait-Multimethod-Matrix. Sie enthält die Reliabilitäten einer Methode, sozusagen die Monotrait-Monomethod-Werte. Die Heterotrait-Monomethod-Dreiecke beinhalten die Korrelationen der verschiedenen Konstrukte, die jeweils mit derselben Methode erhoben wurden. Die Heteromethod-Blöcke enthalten alle Korrelationen zwischen verschiedenen Konstrukten, die mit verschiedenen Methoden erhoben wurden. Sie setzen sich dabei aus den Validitätsdiagonalen und den Heterotrait-Heteromethod-Dreiecken zusammen. Die Validitätsdiagonalen beinhalten die Korrelationen zwischen den jeweils selben Konstrukten, die mit unterschiedlichen Methoden erhoben wurden. Diese Monotrait-Heteromethod-Werte zeigen konvergente Validität an. Die Heterotrait-Heteromethod-Dreiecke schließlich beinhalten die Korrelationen zwischen verschiedenen Konstrukten, die mit verschiedenen Methoden erhoben wurden.

Tabelle 2-1: Veranschaulichung einer fiktiven Multitrait-Multimethod-Matrix (nach Campbell & Fiske, 1959)

		Methode 1			Methode 2			Methode 3					
Konstrukt		A ₁	B ₁	C ₁	A ₂	B ₂	C ₂	A ₃	B ₃	C ₃			
Methode 1	A ₁	(.89)											
	B ₁	.51	(.89)										
	C ₁	.38	.37	(.76)									
Methode 2	A ₂	[.57]			(.93)								
	B ₂	.22	[.57]		.68		(.94)						
	C ₃	.11	.11	[.46]		.59		.58		(.84)			
Methode 3	A ₃	[.56]			[.67]			(.94)					
	B ₃	.23	[.58]		.43		[.66]		.67		(.92)		
	C ₃	.11	.11	[.45]		.34		.32		[.58]		.58	.60

Anmerkung: Reliabilitätswerte stehen in (), Validitätswerte stehen in []; ----- Monomethod-Blöcke, - - - Heteromethod-Blöcke.

Campbell und Fiske (1959) haben Hinweise geliefert, wie diese Matrix zu bewerten ist. Dabei stellen sie vier Forderungen: (1) Die Werte auf der Validitätsdiagonalen sollten signifikant von null verschieden und genügend hoch sein. Hierin manifestiert sich konvergente Validität. In den übrigen drei Forderungen zeigt sich diskriminante Validität. (2) Die Werte auf der Validitätsdiagonalen sollten höher sein als alle anderen Werte in derselben Spalte und Zeile der zugehörigen Heterotrait-Heteromethod-Dreiecke. Das bedeutet, dass der Validitätswert einer Variablen immer höher sein sollte als alle Korrelationen, die mit dieser Variablen weder Konstrukt noch Methode gemein haben. Wenn ein und dasselbe Konstrukt mit zwei verschiedenen Methoden erfasst wurde, sollte diese Korrelation immer höher sein, als wenn zwei verschiedene Konstrukte mit zwei Methoden erfasst wurden. (3) Die Werte auf der Validitätsdiagonalen sollten höher sein als die Korrelationen der entsprechenden Variablen in den Heterotrait-Monomethod-Dreiecken. Das bedeutet, dass die Maße der konvergenten Validität höher sein sollten als die Korrelationen zwischen verschiedenen Konstrukten, die mit derselben Methode erfasst wurden. (4) Das Korrelationsmuster in den Heterotrait-Monomethod-

Dreiecken und den Heterotrait-Heteromethod-Dreiecken sollte annähernd äquivalent sein. Das bedeutet, dass der Zusammenhang zwischen allen Konstrukten unter allen Methoden ähnlich sein sollte. Anhand dieser vier Forderungen, lassen sich konvergente und diskriminante Validität bewerten.

Am häufigsten kritisiert wurde am Multitrait-Multimethod-Ansatz, dass eine solche Vorgehensweise sehr komplex und kaum zu realisieren sei. Zudem wurden die vier Forderungen als zu vage bezeichnet: Was nun als genügend hoch oder äquivalent zu bezeichnen sei, lasse sich numerisch nicht einheitlich bestimmen (Fisseni, 2004). Da es sich dennoch um ein weit verbreitetes und renommiertes Verfahren handelt, soll die Multitrait-Multimethod-Validierung im empirischen Teil dieser Arbeit zur Anwendung kommen.

2.2.1.2 Soziale Validität

Die soziale Validität ist den Nebengütekriterien zuzuordnen. Sie kann als Ergänzung der testtheoretischen Kriterien von Auswahlverfahren dienen und sagt etwas über die Sicht der Bewerber aus. Schuler und Stehle (1983, 1985) verstehen unter dem Begriff „soziale Validität“ konkret eine Sammelbezeichnung von Situationsmerkmalen, die darüber entscheiden, ob eine Personalauswahlsituation vom Bewerber als akzeptable soziale Situation bewertet wird. Deswegen soll mit dem Begriff der sozialen Validität in der vorliegenden Arbeit die Akzeptanz eines Verfahrens durch den Bewerber gemeint sein. Als wichtigste Parameter, die aus einer Auswahlssituation eine akzeptable soziale Situation machen, nannten Schuler und Stehle (1983, 1985): Information, Partizipation oder Kontrolle, Transparenz und Urteils-kommunikation oder Feedback. Der Aspekt der Information bezieht sich auf die Aufgabenbereiche und Anforderungen, aber auch auf Organisationsmerkmale wie Kultur und Ziele der Organisation. Anhand dieser Information soll auch die Selbstselektion der Bewerber erleichtert werden. Der Aspekt der Partizipation oder Kontrolle bezieht sich im engeren Sinne auf die Beteiligung des Bewerbers an der Gestaltung der Auswahlssituation, der Instrumente oder der Entscheidung. Im weiteren Sinne, und das ist weit häufiger der Fall, bezieht sich

dieser Aspekt auf die Möglichkeit, Kontrolle in der Situation auszuüben, z. B. über das eigene Verhalten oder die Entscheidung anderer. Von der anderen Seite betrachtet geht es hierbei um die Freiheit von der Machtausübung anderer. Die Transparenz bezieht sich auf die Auswahl-situation mit den beteiligten Personen, die eingesetzten Instrumente und den gesamten Bewertungsprozess. Bei der Urteilskommunikation geht es um die diagnostische Information durch die Verfahren und die Beurteiler. Das Feedback schließlich sollte inhaltlich offen sein und auf Entwicklungsmöglichkeiten hinweisen. Formal sollte es verständlich und rücksichtsvoll gestaltet sein sowie die Selbsteinsicht des Bewerbers erleichtern.

Neben diesem Modell befasst sich auch die englischsprachige Literatur mit Forschung zur sozialen Validität, die dort unter dem Begriff „applicant reactions“ subsumiert wird. „Applicant reactions“ beziehen sich auf „attitudes, affect and cognitions an individual might have during the hiring process“ (Ryan & Ployhart, 2000, S. 566). Gilliland (1993) beispielsweise hat ein Modell entwickelt, das auf Theorien der „organizational justice“ basiert. Er postuliert ein umfassendes Modell, das die Gesamtbewertung der Fairness auf die Erfüllung oder Verletzung gewisser Regeln zurückführt. Dabei bezieht er auch die Folgen für das Individuum und die Organisation mit ein. Dieses Modell diente vielen nachfolgenden Studien als Basis. Es wurde z. B. von Ryan und Ployhart (2000) in Teilen aufgegriffen und erweitert. Nichtsdestotrotz fordern sie dazu auf, Gillilands Modell umfassender zu überprüfen. Hausknecht et al. (2004) haben sowohl die Ansätze von Gilliland (1993) als auch die Erweiterungen von Ryan und Ployhart (2000) aufgegriffen und in ein umfassenderes Modell integriert. Dabei lassen sich die vielen verschiedenen Variablen in vier Teilbereiche gliedern. Der erste Teilbereich umfasst diejenigen Variablen, die der Teilnehmerreaktion vorausgehen, wie Charakteristika der Person, des Auswahlprozesses, des Jobs und der Organisation. Hieraus ergibt sich dann der Bereich der Teilnehmerreaktion in seinen verschiedenen Facetten. Daraus wiederum resultiert ein dritter Teilbereich, der sämtliche Folgen der Teilnehmerreaktion umfasst. Er beinhaltet die Leistung und Selbstwahrnehmung des Teilnehmers oder seine Haltung und Verhaltensweisen

bezüglich der Organisation und der Arbeit. Ein letzter Teilbereich umfasst diejenigen Variablen, die innerhalb des Modells als Moderatoren fungieren. Die meta-analytischen Ergebnisse der Autoren stützen dieses Modell weitgehend.

Unabhängig vom zugrunde gelegten Modell steht fest, dass soziale Validität nicht nur im Sinne des Bewerbers ist, sondern auch im Interesse des jeweiligen Unternehmens liegt. Generell sind Bewerber eher geneigt ein Jobangebot anzunehmen, wenn sie das durchlaufene Verfahren positiv bewerteten (Macan, Avedon, Paese & Smith, 1994). Smither et al. (1993) konnten zudem zeigen, dass einzelne Aspekte wie z. B. der Jobbezug des Verfahrens dazu führen können, dass das Unternehmen in den Augen der Bewerber an Attraktivität gewinnt und sie dadurch auch bereit sind, es an andere Bewerber weiterzuempfehlen. Es ist jedoch zu beachten, dass nicht alle Auswahlverfahren von den Bewerbern gleich gut beurteilt werden. Bewerber scheinen Interviews psychologischen Teste vorzuziehen und schätzen dabei besonders den hohen Jobbezug sowie die Möglichkeit zur Einflussnahme (Carless, 2006).

2.3 Validität strukturierter Interviews

Prädiktive Validität

Über Jahrzehnte hinweg divergierten die Meinungen über das Interview zwischen Forschern und Anwendern (Weinert, 2004). Während sich die Anwender reichlich des beliebten Interviews bedienten, warnten die Forscher vor der mangelnden Validität von Interviews. Die prognostische Validität wurde auf .05 bis .25 geschätzt (Arvey & Campion, 1982). Einige Ursachen dieser geringen Validität waren z. B. der mangelnde Anforderungsbezug oder die emotionalen Einflüsse auf die Urteilsbildung. So ergab sich mit der Zeit eine Reihe von Verbesserungsvorschlägen, um die Güte von Interviews zu steigern. Dazu zählt unter anderem eine anforderungsbezogene Gestaltung, eine strukturierte Durchführung, die Verwendung verschiedener Fragetypen, die Nutzung verankerter Bewertungsskalen und auch ein Training der Interviewer. Im Zuge des wissenschaftlichen Fortschritts war es schließlich auch das Interview, das von allen eignungsdiagnostischen Verfahren am meisten

profitieren konnte (Schuler, 2001). Die Steigerung seiner prognostischen Validität hängt dabei unmittelbar mit dem Grad der Strukturierung zusammen, weswegen Campion et al. (1997) empfehlen, jegliches Verbesserungspotenzial hinsichtlich der Strukturierung voll auszunutzen. Schmidt und Hunter (1998) ermittelten für unstrukturierte Interviews lediglich eine prognostische Validität von .38, während die für strukturierte Interviews bei .51 lag. Somit zeigte das strukturierte Interview eine gleich hohe Validität wie allgemeine kognitive Fähigkeitstests. Auch McDaniel, Whetzel, Schmidt und Maurer (1994) zeigten eine Validität von .33 für unstrukturierte, aber eine Validität von .44 für strukturierte Verfahren. In diesem Zusammenhang soll auch die Studie von Huffcut und Arthur (1994) erwähnt werden. Sie unterschieden vier Stufen der Strukturierung und zeigten, dass mit steigender Strukturierung die Validität von .20 auf .57 stieg. Dazu passen auch die Ergebnisse von Marchese und Muchinsky (1993). Sie konnten nämlich eine Korrelation von .45 zwischen Strukturierung und Validität nachweisen. Aktuelle Überblicksarbeiten (Moscoso, 2000) kommen zu dem Ergebnis, dass fundiert entwickelte Interviews zu den besten Instrumenten für die Prognose von Berufserfolg zählen. Huffcut und Arthur (1994) ziehen bezüglich der Validität von strukturierten Interviews folgenden Schluss: Interviews weisen eine hohe prädiktive Validität auf, die mit der von allgemeinen kognitiven Fähigkeitstests vergleichbar ist. Der Grad der Strukturierung ist dabei der größte Moderator. Allerdings zeigt sich ab einem gewissen Strukturierungsgrad ein Deckeneffekt.

Konstruktvalidität

Die Frage, was ein strukturiertes Interview überhaupt misst, lässt sich nicht so einfach beantworten. Die Befunde zur Konstruktvalidität sind dementsprechend gemischt (Huffcutt, Conway, Roth & Stone, 2001). Man sollte theoretisch erwarten, dass Interviews genau das messen, worauf sie ausgerichtet sind. So konnten z. B. hohe Korrelationen zwischen einem Interview, welches auf soziale Kompetenz ausgerichtet ist und weiteren Werten der sozialen Kompetenz ermittelt werden (Schuler & Funke, 1989). Andererseits zeigt sich, dass intelligente Personen sowie Personen mit hoher sozialer Kompetenz in

Interviews generell besser abschneiden, was möglicherweise auf die zwischenmenschlich geprägte Begegnung in der Interviewsituation zurückzuführen ist (Mussel, 2007). So fanden Huffcutt, Roth und McDaniel (1996) meta-analytische Hinweise, dass die Intelligenz mit dem Interview zu .40 korreliert. Betrachtet man die oben zitierten divergierenden Befunde zur prognostischen Validität von strukturierten und unstrukturierten Interviews, stellt sich die Frage, worin dieser Unterschied begründet sein könnte und ob strukturierten Interviews eventuell andere Konstrukte zugrunde liegen als unstrukturierten. Huffcutt et al. (2001) kommen zu dem Schluss, dass strukturierte Interviews eher diejenigen Merkmale erfassen, die für den Berufserfolg verantwortlich sind. Konventionelle Interviews hingegen erfassen eher Intelligenz, globale Persönlichkeitsmerkmale und soziale Kompetenz (Salgado & Moscoso, 2002). Generell zeigt sich, dass die Konstruktvalidität strukturierter Interviews eher schwach ist, auch wenn die prädiktive Validität hoch ist. Strukturierte Interviews weisen sowohl mangelnde konvergente als auch mangelnde diskriminante Validität auf, wobei sich dieser Aspekt durch eine Erhöhung der Transparenz verbessern lässt (Klehe, König, Richter, Kleinmann & Melchers, 2008).

Soziale Validität

Ein Grund für die Forschungsintensität in Richtung einer Verbesserung von Interviews ist neben ihrem häufigen Einsatz sicherlich ihrer Beliebtheit zu verdanken. Als Fruhner, Schuler, Funke und Moser (1991) Studierende unterschiedlicher Fachrichtungen baten, verschiedene Auswahlverfahren vergleichend zu bewerten, rangierte das Vorstellungsgespräch oder Interview an erster Stelle. Im Vergleich zu Testverfahren bietet das Interview laut Einschätzung der Befragten mehr Möglichkeiten das eigene Ergebnis zu beeinflussen, erfasst die eigenen Fähigkeiten besser, vermittelt mehr Information über den späteren Arbeitsplatz, ist transparenter und wird als weniger belastend erlebt. Auch Schuler, Frier & Kauffmann (1993) kommen zu dem Schluss, dass das Einstellungsinterview unter allen Personalauswahlverfahren am besten bewertet wird. Interviews sind wohl deshalb bei Bewerbern

so beliebt, weil sie alltäglichen Interaktionen deutlich näher kommen als Tests, Fragebogen oder Assessment-Center und dem Bewerber dadurch ein stärkeres Gefühl der Sicherheit vermitteln (Kanning, 2004). Das Interview wirkt somit auf gewisse Weise natürlich. Wie bereits erwähnt hat Carless (2006) gezeigt, dass Bewerber besonders den hohen Jobbezug sowie die Möglichkeit zur Einflussnahme bei Interviews schätzen. Andererseits zeigten verschiedene Studien (Conway & Peneno, 1999; Chapman & Zweig, 2005), dass diese Beliebtheit mit steigendem Strukturierungsgrad abnehmen kann. Dies ist wohl auf die steigende Schwierigkeit strukturierter Interviews zurückzuführen. Die wahrgenommene Fairness blieb von diesem Umstand unberührt. Zur sozialen Validität lässt sich trotz vieler offener Forschungsanliegen (Chan & Schmitt, 2004) sagen, dass das Interview mehr als andere eignungsdiagnostische Verfahren in der Lage ist, die Erwartungen der Bewerber zu erfüllen und damit eine höhere Wahrscheinlichkeit hat, als positiv erlebt zu werden (Schuler, 2002).

2.4 Ableitungen für die empirische Untersuchung

Aus den theoretischen Betrachtungen ergeben sich folgende Ableitungen für die empirische Untersuchung dieser Arbeit: Es gibt auf dem Markt eine Reihe von Verfahren zur Personalauswahl. Eines davon ist das strukturierte Interview. Da es wenig Konsens darüber gibt, was genau unter einem strukturierten Interview zu verstehen ist (Chapman & Zweig, 2005), wurde für die Studie ein spezielles Interviewverfahren herausgegriffen. Anhand dieses Interviews sollen nun exemplarisch die Gütekriterien der Konstruktvalidität und der sozialen Validität betrachtet werden. Das Untersuchungsziel der vorliegenden Studie ist somit die Überprüfung von Aspekten der Konstruktvalidität und sozialen Validität von strukturierten Interviewverfahren am Beispiel des strukturierten Interviews Opus[®].

2.4.1 Opus[®] als Beispiel eines strukturierten Interviews

Es gibt heutzutage eine große Auswahl an strukturierten Interviewverfahren, deren Inhalte sich zum Teil grundlegend unterscheiden. Das hier vorgestellte

Verfahren wird von einem Beratungsunternehmen im Kontext der Personalauswahl und -entwicklung im Zusammenspiel mit weiteren Instrumenten, wie z. B. einer darauf abgestimmten Anforderungsanalyse, eingesetzt. Es handelt sich um das strukturierte Interviewverfahren Opus[®] (Fischer, 2003), ein nach DIN 33430 (berufsbezogene Eignungsdiagnostik) TÜV-zertifiziertes Verfahren. Das Verfahren untergliedert sich in einen Teil zur Erfassung der Denkstrukturen und einen weiteren Teil zur Erfassung der Handlungsstrukturen der Teilnehmer. Da für die vorliegende Studie nur der Denkteil von Interesse ist, wird auf den Handlungsteil nicht weiter eingegangen.

Methode

Das Opus[®]-Interview zielt darauf ab, die Denkstrukturen des Teilnehmers zu erfassen. Dies geschieht in einem Interviewsetting mit einem geschulten Opus[®]-Interviewer. Es geht dabei nicht um das Überprüfen von (Fach-)Wissen, sondern ganz allgemein um kognitive Fähigkeiten. Folgende Strukturen werden überprüft: Analytisches Denken, Kreatives Denken und Entscheidungsvermögen. Diese Fähigkeiten werden nun anhand speziell entwickelter Fallstudien erfasst. Die Fallstudien sind für den Teilnehmer mit Allgemeinwissen zu bewältigen und so konstruiert, dass sie es dem Interviewer ermöglichen, die zugrunde liegenden Denkstrukturen zu erfassen, ohne inhaltlich in Details abzuschweifen.

Ablauf

Das Interview läuft folgendermaßen ab: Nach einem Bekanntmachen von Interviewer und Teilnehmer wird der Teilnehmer sowohl über den Ablauf als auch über das Opus[®]-Interview informiert. Danach hat der Teilnehmer 20 Minuten Zeit, die Fallstudie durchzulesen und sich auf das Interview vorzubereiten. Anschließend beginnt der Interviewer mit dem eigentlichen Interviewprozess. Dabei wird der Teilnehmer immer tiefer in die Fallstudie geführt. Es werden vorgegebene Fragen gestellt, anhand derer der Teilnehmer sein Analytisches und Kreatives Denken ebenso wie sein Entscheidungsvermögen unter Beweis stellen kann. Die Aufgabenstellung wird dabei

zunehmend komplexer, bis der Teilnehmer nicht mehr in der Lage ist, die gestellten Anforderungen zu bewältigen und der Interviewer sich ein Bild über die Denkstrukturen des Teilnehmers machen konnte. Abschließend wird der Teilnehmer anhand des gezeigten Analytischen Denkens, Kreativen Denkens und des Entscheidungsvermögens beurteilt. Der Teilnehmer erhält im Anschluss an das Interview ein ausführliches Feedback mit Hinweisen zur weiteren Entwicklung.

2.4.2 Zentrale Fragestellungen und Hypothesen

Aus dem theoretischen Hintergrund lassen sich für die empirische Untersuchung zentrale Fragestellungen und konkrete Hypothesen zur Konstruktvalidität und zur sozialen Validität des Opus[®]-Interviews ableiten.

2.4.2.1 Fragestellung und Hypothesen zur Konstruktvalidität

Wie im theoretischen Teil dieser Arbeit dargestellt, eignet sich eine Multitrait-Multimethod-Matrix, um sich der Konstruktvalidität eines Verfahrens zu nähern (Campbell & Fiske, 1959). Die vorliegende Studie wendet daher eine solche Matrix auf das strukturierte Interview Opus[®] an. Dabei werden anerkannte Testverfahren sowie eine Selbstauskunft als Validierungskriterien herangezogen. Damit wird der Frage nachgegangen, was das strukturierte Interviewverfahren überhaupt erfasst.

Um die Konstruktvalidität des Opus[®]-Interviews näher bestimmen zu können, wurden neben dem Interview selbst auch eine Testung mit Teilen des Wilde Intelligenztests-2 (WIT-2) von Kersting, Althoff und Jäger (2008) und mit Teilen der Analyse des Schlussfolgernden und Kreativen Denkens (ASK) von Schuler und Hell (2005) durchgeführt. Zudem füllten die Teilnehmer eine Selbsteinschätzung bezüglich der Opus[®]-Kriterien aus.

Hypothese 1 bezieht sich auf die Konstruktvalidität des Opus[®]-Interviews und gliedert sich in vier Unterpunkte a bis d. Die Grundannahme ist, dass das Opus[®]-Interview zuverlässig die Konstrukte Analytisches Denken, Kreatives

Denken und Entscheidungsvermögen erfasst. Es lassen sich in Anlehnung an Campbell und Fiske (1959) folgende Hypothesen ableiten:

H 1a: Es zeigen sich hohe Korrelationen zwischen dem Analytischen Denken, erfasst durch das Opus[®]-Interview und dem Schlussfolgerndem Denken, erfasst durch den Wilde Intelligenztest-2.

H 1b: Ebenso werden hohe Korrelationen zwischen dem Kreativen Denken, erfasst durch das Opus-Interview und dem Kreativen Denken, erfasst durch die ASK erwartet.

H 1c: Hohe Korrelationen zeigen sich zudem zwischen den Werten des Analytischen Denkens, Kreativen Denkens und Entscheidungsvermögens, erfasst durch das Opus[®]-Interview und erfasst mittels Selbsteinschätzung der Teilnehmer.

H 1d: Im Sinne einer Multitrait-Multimethod-Validierung wird erwartet, dass alle weiteren Korrelationen geringer ausfallen.

2.4.2.2 Fragestellung und Hypothese zur sozialen Validität

Viele Studien haben über die klassischen Gütekriterien (Amelang & Zielinski, 1994) hinaus bewiesen, dass für ein Verfahren der Personalauswahl die soziale Validität ebenfalls von großer Bedeutung ist (Hausknecht et al., 2004; Macan et al., 1994; Derous, Born & de Witte, 2004); wenngleich dieses Konzept noch nicht hinreichend erforscht ist (Chan & Schmitt, 2004). Insbesondere in Hinblick auf die hervorgerufenen Reaktionen beim Bewerber sollte der Aspekt der sozialen Validität bei der Entscheidung über den Einsatz eines Auswahlinstrumentes nicht außer Acht gelassen werden (Smither et al., 1993). Im Rahmen der empirischen Studie dieser Arbeit soll deshalb auch die soziale Validität des strukturierten Interviews Opus[®] untersucht werden. Die Fragestellung hierbei lautet, ob wie in der Literatur vermutet (Carless, 2006) das Interviewverfahren von den Teilnehmern besser bewertet und akzeptiert wird als typische Testverfahren.

Da die bisher publizierte Literatur für diesen Zweck kein geeignetes Instrument bietet, wird an dieser Stelle zudem ein eigener Fragebogen zur Erfassung der sozialen Validität aus Sicht der Teilnehmer entwickelt und beurteilt. Anhand dieses Fragebogens wird die soziale Validität der eingesetzten Verfahren bewertet.

Hypothese 2 bezieht sich also auf die soziale Validität des Opus[®]-Interviews. Die Grundannahme ist, dass Interviews von Teilnehmern generell als fair und jobbezogen erlebt werden (Hausknecht et al., 2004). Da es sich beim Opus[®]-Interview um ein interaktives Verfahren handelt, welches dem Teilnehmer Gestaltungsspielraum lässt und somit eher dem beliebten Einstellunginterview ähnelt als den bei Bewerbern eher unbeliebten psychologischen Testverfahren (Hausknecht et al., 2004), lässt sich daraus folgende Hypothese ableiten:

H 2: Das Opus[®]-Interview weist in allen untersuchten Aspekten eine höhere soziale Validität auf als die Testverfahren.

3 Empirischer Teil

Im empirischen Teil dieser Arbeit wird zunächst das methodische Vorgehen erläutert. Anschließend werden die unterschiedlichen Ergebnisse der Studie präsentiert.

3.1 Methodisches Vorgehen

Im folgenden Kapitel soll zuerst das Untersuchungssetting, der Ablauf der Studie und die Stichprobe beschrieben werden. Daraufhin wird ausführlich auf die eingesetzten Instrumente und Analyseverfahren eingegangen.

3.1.1 Vorstellung des Untersuchungssettings

Bei der hier vorliegenden Arbeit wurde eine empirische Studie durchgeführt, die in Kooperation mit einem Beratungsunternehmen entstand. In diesem Rahmen war es möglich, eine praxisnahe Untersuchung durchzuführen. Die Stichprobe basiert somit nicht nur auf einer studentischen Population, wie es bei Arbeiten dieser Art häufig der Fall ist, sondern es ergab sich aus der Kooperation die Möglichkeit, auch Arbeitnehmer verschiedener Herkunft und Arbeitsbereiche zu integrieren.

Die Studie wurde für die Teilnehmer als eine Art Standortbestimmung bzw. Training in Bewerbungstestverfahren konzipiert. Den Teilnehmern sollte die Möglichkeit geboten werden, verschiedene Instrumente, die in der Personalauswahl zum Einsatz kommen, kennen zu lernen und sich darin auszuprobieren. Im Gegenzug erklärten sie sich einverstanden, verschiedene Fragebogen auszufüllen. Eine ausführliche Rückmeldung der Ergebnisse sowie ein Ausblick hinsichtlich weiterer Karrierewege sollten den Teilnehmern als Ansporn dienen, an ihre tatsächliche Leistungsgrenze zu gehen und somit verwertbare Ergebnisse zu liefern.

Den Teilnehmern wurde garantiert, dass keinerlei personenbezogene Daten gespeichert wurden. Die Zuordnung der Einzelergebnisse erfolgte durch Codes. So ließ sich die Anonymität der Daten gewährleisten.

3.1.2 Ablauf der Studie

Die Daten wurden an jeweils zwei Terminen erhoben. Dies war einerseits von Vorteil für die Teilnehmer, da die Verfahren recht anstrengend und zeit-aufwendig sind und bei einer Gesamttagung eventuell Verzerrungen durch Ermüdungseffekte aufgetreten wären. Andererseits blieb so auch genügend Zeit, die Verfahren auszuwerten und das Feedback vorzubereiten.

Der Ablauf der Studie gestaltete sich folgendermaßen: Nachdem die Studie auf verschiedene Art beworben worden war (siehe Abschnitt 3.1.3), erhielten die Teilnehmer zunächst ein Informationsblatt zur Studie, auf dem die wichtigsten Inhalte und Informationen zusammengefasst waren. Mit den Teilnehmern wurden daraufhin zwei Termine, einen für das Interview und einen für die Testung, vereinbart. Zum ersten Termin wurden die Teilnehmer teils mündlich teils schriftlich über das Ziel der Studie, den Studienablauf und die eingesetzten Verfahren informiert. Es wurde auf die Anonymität der Ergebnisse hingewiesen und die Zuordnung über Codes erläutert. Danach füllten die Teilnehmer den Code sowie die Selbsteinschätzung aus. Anschließend absolvierte ein Teil der Teilnehmer am ersten Termin das Interview und am zweiten Termin die Testung. Um Reihenfolgeeffekte zu vermeiden, durchlief der andere Teil der Teilnehmer die beiden Verfahren in umgekehrter Reihenfolge, also zunächst die Testung und zum zweiten Termin das Interview. Das Interview wurde immer in einem Einzelsetting durchgeführt. Die Testung wurde aus organisatorischen Gründen entweder im Einzelsetting oder in kleinen Gruppen durchgeführt. Es erfolgte immer zuerst eine nähere Information zu dem folgenden Verfahren. Nach der Durchführung füllten die Teilnehmer sofort den Fragebogen zur sozialen Validität in der entsprechenden Version, „Interview“ und „Testung“, aus. Jeder Teilnehmer erhielt am Ende ein Feedback zu beiden Verfahren mit

Hinweisen zur Weiterentwicklung. Abschließend wurde ihm die Entscheidungsfrage gestellt, welches der beiden Verfahren er bevorzugt.

3.1.3 Beschreibung der Stichprobe

Die Stichprobe setzte sich aus verschiedenen Gruppen zusammen, wobei drei Gruppen dominierten. Ein Teil der Stichprobe stammte aus einer Outplacement-Gesellschaft, die ihre Mitglieder dabei unterstützt, einen neuen Arbeitsplatz zu finden. Ein weiterer Teil der Stichprobe befand sich aktuell in einem Karriereorientierungsprozess des Beratungsunternehmens. Ein letzter Teil der Stichprobe bestand aus Studenten, die sich im Endstadium ihres Studiums befanden. Die Rekrutierung der Teilnehmer erfolgte somit an drei Stellen: in Kooperation mit der Outplacement-Gesellschaft wurden die Mitglieder über die Möglichkeit einer Teilnahme informiert; in Kooperation mit dem Beratungsunternehmen wurden alle Personen, die sich aktuell in einer Karriereberatung befanden, darauf aufmerksam gemacht, dass es die Möglichkeit der Teilnahme an der vorliegenden Studie gibt und schließlich wurden über Aushänge an der Universität Studenten über die Teilnahmemöglichkeit informiert. Dabei wurde das Ziel verfolgt eine möglichst durchmischte Stichprobe zu rekrutieren, die aber gleichzeitig gemeinsam hat, dass sie eine große Nähe zum Thema „Bewerbung“ aufweist. Alle Teilnehmer befanden sich entweder gerade in einem Bewerbungsprozess oder waren in ihrer Ausbildung so weit, dass eine Bewerbung im kommenden Jahr anstand.

3.1.4 Beschreibung der Erhebungsinstrumente

In den folgenden Abschnitten sollen die einzelnen Erhebungsinstrumente vorgestellt werden. Ebenso wird die Entwicklung eines Fragebogens zur sozialen Validität beschrieben.

3.1.4.1 Selbsteinschätzung

Um das Selbstbild des Kandidaten zu erfassen und als weiteres Validierungskriterium heranzuziehen, wurde ein kurzer Fragebogen entworfen. Dieser sollte

die Kandidaten mit drei Items zu einer Selbsteinschätzung bewegen. Dabei wurden die drei Kriterien des Opus[®]-Interviews erfasst: Analytisches Denken, Kreatives Denken und Entscheidungsvermögen. Die Fragen lauteten: „Wie schätzen Sie insgesamt Ihre Analysefähigkeit ein?“, „Wie beurteilen Sie insgesamt Ihre Fähigkeit, kreativ zu denken?“ und „Wie schätzen Sie insgesamt Ihr Entscheidungsvermögen ein?“. Den Kandidaten wurde als Antwortmöglichkeit eine sechsstufige Likert-Skala mit folgenden Abstufungen geboten: deutlich unterdurchschnittlich, eher unterdurchschnittlich, leicht unterdurchschnittlich, leicht überdurchschnittlich, eher überdurchschnittlich, deutlich über-durchschnittlich. Diese Selbsteinschätzung sollte die Multitrait-Multimethod-Matrix erweitern, indem sie neben den objektiven Testverfahren (T-Daten) auch eine Selbstbeobachtung des Teilnehmers (Q-Daten) als weitere Datenquelle einbrachte (Cattell, 1978).

3.1.4.2 Wilde Intelligenztest-2

Beim Wilde Intelligenztest-2 (WIT-2) von Kersting et al. (2008) handelt es sich um ein Verfahren zur Erfassung der intellektuellen Leistungsfähigkeit. Dabei werden kognitive Fähigkeiten in Anlehnung an ein modifiziertes Modell der „primary mental abilities“ nach Thurstone (1938) erfasst. Das Verfahren eignet sich neben der Intelligenzdiagnostik und Forschung auch für die Eignungsdiagnostik. Insbesondere kommt es auch zur Personalauswahl und -entwicklung sowie bei der Berufswahl und Laufbahnberatung zum Einsatz. Der WIT-2 ist sowohl als Einzel- als auch als Gruppentestung durchführbar und für einen Altersbereich von 14 bis 42 Jahren geeignet. Das Verfahren wurde praxisnah normiert. Es stehen bildungs- und altersdifferenzierte Normdaten für jeden Subtest zur Verfügung. Der WIT-2 umfasst die acht Dimensionen (1) Schlussfolgerndes Denken, (2) Sprachliches Denken, (3) Rechnerisches Denken, (4) Räumliches Denken, (5) Merkfähigkeit, (6) Arbeitseffizienz, (7) Wissen „Wirtschaft“ und (8) Wissen „Informations-technologie“. Um den gesamten Test durchzuführen, ist eine Dauer von 143 Minuten vorgesehen. Da die Module sowohl einzeln als auch in beliebiger Kombination eingesetzt werden können, gestaltet sich die Anwendung sehr flexibel und zeitökonomisch.

Die Reliabilität zeigt sich in einer internen Konsistenz von $\alpha = .78$ bis $\alpha = .98$. Die Retestreliabilität variiert je nach Subtest zwischen $r_{tt} = .59$ und $r_{tt} = .93$. Die Validität des WIT-2 wurde auf vielfältige Weise belegt. Zur Bestimmung der Konstruktvalidität wurden die Dimensionen des WIT-2 mit 17 anderen Testverfahren abgeglichen, z. B. I-S-T 2000R (Amthauer, Brocke, Liepmann & Beauducel, 2001), CFT-3 (Weiß & Cattell, 1993) und NEO-PI-R (Ostendorf & Angleitner, 2004). Zur Bestimmung der Kriteriumsvalidität wurden zum einen die Ergebnisse bisheriger Studien zum WIT (Jäger & Althoff, 1994) meta-analytisch zusammengefasst. Zum anderen wurde die Kriteriumsvalidität des WIT-2 anhand mehrerer praxisnaher Studien näher bestimmt. Bedeutsame Zusammenhänge ergaben sich mit Selbsteinschätzungen und Schulnoten sowie mit Erfolg in Erstausbildung und Weiterbildung.

In der vorliegenden Studie wurde eine Auswahl an Modulen getroffen, so dass nur die Module Sprachliches Denken, Rechnerisches Denken, Räumliches Denken und Schlussfolgerndes Denken zur Anwendung kamen. Diese Auswahl garantierte ein angemessenes Maß an Information, ohne dabei die Teilnehmer durch lange Testzeiten zu sehr zu belasten. Beim Modul Sprachliches Denken geht es um die Fähigkeit, mit sprachlichen Konzepten umzugehen, wobei Wortschatz, Sprachverständnis und sprachlogisches Denken eine Rolle spielen. Dazu zählen die Aufgaben Analogien zu bilden und gleiche Wortbedeutungen zu erkennen. Das Modul Rechnerisches Denken überprüft die Fähigkeit, einfache Rechenoperationen der Addition, Subtraktion, Multiplikation und Division korrekt auszuführen. Dabei werden Grundrechenaufgaben und eingekleidete Rechenaufgaben gestellt. Durch das Modul Räumliches Denken wird die Fähigkeit zur Vorstellung räumlicher Relationen erfasst. Dies geschieht anhand der Abwicklungen von Figuren und Spiegelbildern. Das Modul Schlussfolgerndes Denken bestimmt die Fähigkeit, bestimmte logische Regeln und Gesetzmäßigkeiten zu erkennen und zweckgemäß anwenden zu können. Hier fließen zusätzlich zur Bildung von Zahlenreihen die oben genannten Wortanalogien und Abwicklungen ein. Um zwischen leistungsstarken und leistungsschwachen Personen differenzieren zu können, wurde bei der Konstruktion des WIT-2 auf eine ausreichende Bandbreite leichter bis schwerer

Items geachtet. Das Antwortformat ist eindeutig auf dem Antwortbogen vorgegeben und unterstützt somit eine objektive Auswertung sowie Interpretation der Ergebnisse. Ein ausführliches Instruktionsheft sorgt für eine standardisierte Durchführung. Das Verfahren entspricht zudem den Richtlinien der DIN 33430.

3.1.4.3 ASK

Die Analyse des Schlussfolgernden und Kreativen Denkens (ASK) von Schuler und Hell (2005) ist ein Verfahren zur Erfassung der beiden Fähigkeitsbereiche Schlussfolgerndes und Kreatives Denken. Generell können Schlussfolgerndes Denken und Kreativität als zwei komplementäre Facetten der allgemeinen Intelligenz aufgefasst werden. „Schlussfolgerndes Denken wird dabei als Fähigkeit definiert, Information verschiedenen Komplexitätsgrades angemessen zu interpretieren und logisch richtige Schlussfolgerungen abzuleiten“ (Schuler & Hell, 2005, S. 7). Kreatives Denken hingegen bezeichnet „die Fähigkeit, Verbindungen und Gemeinsamkeiten zwischen Sachverhalten zu erkennen sowie Ideen und Lösungsmöglichkeiten für offene Probleme zu finden“ (Schuler & Hell, 2005, S. 7).

Das Verfahren kommt in unterschiedlichen Bereichen zum Einsatz. Es können beispielsweise angewandte diagnostische Fragestellungen im Bereich der Berufs-, Bildungs- und Laufbahnberatung untersucht werden. Es kann außerdem zur Unterstützung von Entscheidungen in der Personalauswahl und -platzierung eingesetzt werden. Ebenso bietet sich das Verfahren für den Einsatz in der Forschung und in der experimentellen Psychologie, z. B. für Untersuchungen im Bereich der Fähigkeits- und Leistungsforschung, an. Es sind bei der ASK sowohl Einzel- als auch Gruppentestungen möglich. Die ASK wurde speziell für die Zielgruppe erwachsener Probanden zwischen 19 und 68 Jahren mit Bildungsvoraussetzung Abitur konstruiert und sollte dementsprechend auch nur auf diese Stichprobe angewendet werden. Der Test basiert auf der klassischen Testtheorie und ist so konzipiert, dass beide Testteile auch unabhängig voneinander eingesetzt werden können. Dabei ist für den

Gesamttest eine Dauer von ca. 70 Minuten angesetzt, 30 Minuten für den Testteil Schlussfolgerndes Denken und 40 Minuten für den Testteil Kreatives Denken.

Die Reliabilität der beiden Testteile zeigt sich in einer internen Konsistenz von $\alpha = .72$ für das Schlussfolgernde Denken und $\alpha = .70$ für das Kreative Denken. Die Retestreliabilitäten sind mit $r_{tt} = .78$ für das Schlussfolgernde Denken und $r_{tt} = .77$ für das Kreative Denken angegeben. Es besteht die Möglichkeit einen Gesamtwert zu bilden, dessen Retestreliabilität $r_{tt} = .86$ beträgt. Die Validität der ASK wurde durch explorative und konfirmatorische Faktorenanalysen belegt, die eindeutig die zweifaktorielle Struktur widerspiegeln. Zur Bestimmung der Konstruktvalidität der ASK wurde die konvergente und diskriminante Validität zu den folgenden Verfahren belegt: Berliner Intelligenzstruktur-Test BIS-Test (Jäger, Süß & Beauducel, 1997), Wonderlic Personnel Test WPT (Wonderlic, 1966) und Verbaler Kreativitätstest VKT (Schoppe, 1975). Für den Bereich Kreativität wurde die Validität zudem anhand von biografischen Daten und der Creative Personality Scale CPS (Gough, 1979) belegt. Auch hat sich eine hohe kriterienbezogene Validität hinsichtlich Abitur- und Studiennoten gezeigt. Im Folgenden wird nur der Testteil zum Kreativen Denken zum Einsatz kommen. Dieser umfasst vier Aufgabengruppen. Beim „Sätze kombinieren“ werden die Teilnehmer in zwei Aufgaben aufgefordert aus vorgegebenen Anfangsbuchstaben Wörter zu formen, die einen sinnvollen Satz ergeben. In den vier Aufgaben zum „Hypothesen generieren“ geht es darum, zur Erklärung vorgegebener Sachverhalte Hypothesen über mögliche Ursachen aufzustellen. Beim „Bedingungsgefüge definieren“ sollen die Teilnehmer in zwei Aufgaben Wirkzusammenhänge zu einem vorgegebenen Sachverhalt darstellen. In den beiden Aufgaben zum „Kategorien bilden“ geht es darum, vorgegebene Begriffe zu sinnvollen Kategorien zusammenzufassen. Es handelt sich also jeweils um ein freies Antwortformat. Um eine objektive Auswertung zu gewährleisten, sind dem Verfahren genaue Richtlinien zur Auswertung beigefügt, so dass eine Auswertungsobjektivität zwischen .87 und .99 erreicht werden kann. Das Verfahren entspricht ebenfalls den Richtlinien der DIN 33430.

3.1.4.4 Fragebogen zur sozialen Validität

Theoretische und methodische Überlegungen zum Einsatz von Fragebogen

Der Einsatz von Fragebogen ist in der psychologischen Forschung und Anwendung weit verbreitet (Mummendey, 2003). Bungard (2000) schätzt sogar, dass ca 90% aller Daten in der empirischen Sozialforschung auf diese Art erhoben werden. Für die Verwendung von Fragebogen spricht vor allem die gute Praktikabilität. Mit Fragebogen ist es möglich, die Einstellung eines Individuums ökonomisch zu erfassen, quantifizierbar zu machen und somit auch mit der Einstellung weiterer Individuen zu vergleichen. Trotz der theoretischen und methodischen Probleme, die Fragebogen und Rating-Skalen eigen sind (Mummendey, 2003), wurde in der vorliegenden Studie ein solches Instrument entworfen, da es der Fragestellung angemessen schien. Beim Einsatz von Rating-Skalen werden in der Literatur häufig zwei Probleme erörtert: die Anzahl der Skalenstufen und das Skalenniveau der Rating-Skalen (Bortz & Döring, 2009). Im vorliegenden Fragebogen fiel die Entscheidung zugunsten einer sechsfach abgestuften Likert-Skala. Die sechs Antwortstufen wurden gewählt, um eine hinreichend differenzierte Antwort abbilden zu können. Es wurde bewusst auf eine mittlere Antwortkategorie verzichtet, weil eine solche Kategorie oft verschiedene Bedeutungen in sich vereint und somit nicht eindeutig interpretierbar ist (Mummendey, 2003). Bezüglich des Skalenniveaus von Rating-Skalen wird oft diskutiert, ob sie intervallskaliert sind oder ob die Daten lediglich Ordinalskalenniveau erfüllen. In der vorliegenden Studie wird wie bei Bortz und Döring (2009) angenommen, dass Rating-Skalen intervallskaliert sind, wodurch sich auch der Einsatz der weiter unten aufgeführten Analyseverfahren rechtfertigt.

Zielsetzung

Um die bereits dargestellte Fragestellung zur sozialen Validität beantworten zu können, musste ein Fragebogen eingesetzt werden, der die soziale Validität eines Verfahrens aus der Sicht der Teilnehmer erfasst. Da weder in der deutsch- noch in der englischsprachigen Literatur ein solcher Fragebogen

gefunden werden konnte, der für die vorliegende Fragestellung hinreichend passend erschien, musste ein solches Instrument erst entwickelt werden.

Basierend auf Modellen zur sozialen Validität, z. B. Schuler und Stehle (1983, 1985), und englischsprachiger Literatur zu „applicant reactions“, wie Smither et al. (1993), sollte ein Fragebogen entworfen werden, mit dem die Teilnehmer ihre Einstellung hinsichtlich verschiedener Aspekte des Verfahrens zum Ausdruck bringen konnten. Aus diesem Fragebogen sollten dann zwei quasi identische Parallelversionen entstehen, die sich einmal auf das Interview und einmal auf die Testung mit dem WIT-2 und der ASK beziehen. Der Unterschied beider Fragebogen bestand lediglich in der Formulierung „Interview“ und „Testung“. So sollte in einem späteren Analyseschritt ein direkter Abgleich der Teilnehmereinstellung zum Interview und zur Testung ermöglicht werden.

Konstruktion

Generell zielt der Fragebogen auf die Prinzipien der Fairness in Auswahl-situationen ab. Er wurde basierend auf der Forschung zur sozialen Validität und „applicant reactions“ entwickelt. Sehr bekannt ist hierbei insbesondere in Deutschland das Konzept der sozialen Validität nach Schuler und Stehle (1983, 1985). Sie vertreten vier Aspekte der sozialen Validität: Information, Partizipation oder Kontrolle, Transparenz und Urteils-kommunikation oder Feedback. Darüber hinaus wird besonders in der englischsprachigen Literatur weitere Forschung zu „applicant reactions“ betrieben, die teilweise auf Ansätzen von Gilliland (1993) basiert. Dabei wird beispielsweise auch versucht, Bewerberreaktionen in einen größeren Zusammenhang zu bringen (vgl. Hausknecht et al., 2004). Nach eingehender Sichtung der entsprechenden Literatur wurden im vorliegenden Fragebogen nun verschiedene Ansätze kombiniert. Dies erschien insbesondere deshalb nötig, da es sich nicht um Auswahl-situationen vor Ort bei einem Arbeitgeber handelt, sondern um extern durchgeführte, jobunabhängige Verfahren. So können Schuler und Stehles Aspekte Information und Urteils-kommunikation z. B. nicht passend erfasst werden, da kein Kontakt mit einem konkreten

Unternehmen, welches Personal sucht, bestand. Außerdem sollte der Fragebogen weitere Bereiche als die von Schuler und Stehle genannten abdecken. Deshalb ergänzen weitere in der Forschung untersuchte Aspekte wie z. B. die Testmotivation und „organisational attractiveness“ Schuler und Stehles Aspekte. Es wurden insgesamt 48 geschlossene und 3 offene Items generiert. Diese wurden auf theoretischer Basis neun Faktoren zugeordnet und abschließend in einer Globalen Frage zusammengefasst. Da die meisten Items aus englischsprachigen Skalen entlehnt wurden, können Übersetzungsfehler nicht gänzlich ausgeschlossen werden. Es wurde jedoch mehr Wert darauf gelegt, eine sinngemäße, inhaltliche Übereinstimmung zu erzielen, als eine wortgetreue Übersetzung zu leisten. Wo es notwendig erschien, wurden zusätzliche Items generiert, die inhaltlich zu den gebildeten Faktoren passten. Der Fragebogen wurde mehrmals inhaltlich überarbeitet und abschließend zwei Experten zu einer letzten Revision, vor allem in sprachlicher Hinsicht, vorgelegt.

Inhalte

Im Folgenden sollen die neun theoretisch abgeleiteten Faktoren kurz näher beschrieben werden. Eine Übersicht über die den Faktoren zugehörigen Items findet sich in Anhang A. Da die theoretisch abgeleiteten Faktoren teilweise englischsprachiger Literatur entstammen und eine inhaltliche Verzerrung und Ungenauigkeit vermieden werden soll, werden sie im Folgenden nicht ins Deutsche übersetzt.

Die ersten beiden Faktoren entsprechen zwei der vier von Schuler und Stehle entwickelten Aspekte der sozialen Validität. Der erste Faktor lautet „Partizipation“ und beschreibt die Möglichkeit zur Einflussnahme auf das eigene Verhalten bzw. auf die Situation, die frei von Machtausübung ist. Diese Kontrollmöglichkeit spiegelt sich z. B. in Item 1 „Ich hatte das Gefühl, die Situation im Interview in angemessenem Maße mitsteuern zu können“ wider. Einen weiteren Aspekt der sozialen Validität nach Schuler und Stehle beschreibt Faktor 2 unter dem Begriff „Transparenz“. Diese Transparenz bezieht sich vor allem auf die Auswahl-situation und den damit einhergehenden

Verhaltenserwartungen an die Teilnehmer. Item 10 beschreibt dies so: „Mir wurde nachvollziehbar erläutert, was im Laufe des Interviews von mir erwartet wird.“ Faktor 3, „Affektive Bewertung“, wird von Smither et al. (1993) beschrieben als „affective reaction to taking the examination“ (S. 64). Ein Item, das diese affektive Bewertung ausdrückt wäre z. B. Item 16: „Im Interview habe ich mich als Person wertgeschätzt gefühlt.“ Faktor 4, „Job Acceptance Intention“, bezieht sich darauf, inwiefern ein Teilnehmer bereit ist einen angebotenen Job zu akzeptieren. Item 22 beschreibt dies folgendermaßen: „Wenn ich dieses Verfahren im Rahmen eines Bewerbungsprozesses durchlaufen hätte, würde ich ein im Anschluss erhaltenes Stellenangebot der Organisation annehmen.“ Faktor 5 versteht unter „Organisational Attractiveness“ die Anziehungskraft des Unternehmens, welches das Verfahren einsetzt. Dies äußert sich auch darin, ob ein Teilnehmer das Unternehmen an andere weiter empfehlen würde, wie es Item 24 beschreibt: „Eine Organisation, die dieses Verfahren zur Personalauswahl einsetzt, würde ich an andere interessierte Bewerber weiterempfehlen.“ „Perceived Predictive Validity“, Faktor 6, trifft eine Aussage darüber „how well the procedure predicts future job performance, regardless of how it looks“ (Smither et al., 1993, S. 54). Item 29 spiegelt diese Vorhersage wider: „Ich bin sicher, dass das Verfahren vorhersagen kann, wie gut ein Bewerber auf einer bestimmten Position sein wird.“ Der Faktor 7, „Perceived Job Relatedness“, beschreibt den wahrgenommenen Link zwischen dem Inhalt des Verfahrens und dem zukünftigen Jobinhalt und definiert sich folgendermaßen: „The extend to which a test appears to measure content relevant to the job situation or appears to be valid“ (Bauer, Truxillo, Sanchez, Craig, Ferrara & Campion, 2001, S. 391). Dieser Zusammenhang wird in Item 34 deutlich: „Ich erkenne einen klaren Zusammenhang zwischen den Anforderungen im Interview und den Anforderungen im Beruf.“ Faktor 8 thematisiert die „Achtung der Privatsphäre“ und bezieht sich auf „the extend to which questions avoid [...] invasion of privacy [...] and are deemed fair and appropriate“ (Bauer et al., 2001, S. 391). Item 41 macht dies deutlich: „Das Verfahren war diskriminierend.“ Ein letzter Faktor 9 bezieht sich auf die „Testmotivation“, wie Item 46 zeigt: „Die Art der

Aufgabenstellung war motivierend.“ Mit Hilfe der abschließenden Globalen Frage, „Zusammenfassend betrachtet: Welchen Eindruck hat das Durchlaufen des Interviews bei mir hinterlassen?“, kann der Teilnehmer ein übergreifendes Fazit ziehen.

Aufbau

Der Fragebogen beginnt mit einem Dank an den Teilnehmer für das Durchlaufen des Verfahrens und der Bitte, die während des vorangegangenen Verfahrens entstandenen Eindrücke wiederzugeben. Dem Teilnehmer wird hierbei absolute Anonymität zugesichert und er wird in die Verwendung von Codes zur Datenanonymisierung eingeführt. Nach der Generierung des Teilnehmercodes findet sich eine Erläuterung zum Ausfüllen des Fragebogens inklusive der verwendeten Antwortskala in Zahlen und Worten. Das geschlossene Antwortformat spiegelt sich in einer sechsstufigen Likert-Skala wider. Die Abstufungen lauten: trifft gar nicht zu, trifft überwiegend nicht zu, trifft eher nicht zu, trifft eher zu, trifft überwiegend zu und trifft voll zu. Sie werden durch die Zahlen 1 bis 6 ausgedrückt. Diese Antwortskala wird zur Erinnerung zu Beginn jeder Seite des Fragebogens über dem Itemblock wiederholt. Hinter den offenen Fragen befindet sich ausreichend Platz für eine geschriebene Antwort und bei der abschließenden Globalen Frage wird der Teilnehmer dazu aufgefordert, seinen Eindruck auf einer sechsstufigen Skala in Zahlen und Worten von sehr negativ bis sehr positiv anzukreuzen. Abschließend wird dem Teilnehmer gedankt. Der Fragebogen findet sich in der Version „Interview“ im Anhang B.

3.1.5 Beschreibung der Analyseverfahren

Im Folgenden werden die unterschiedlichen Analyseverfahren erläutert, anhand derer der Fragebogen, die Hypothesen zur Konstruktvalidität sowie die Hypothese zur sozialen Validität ausgewertet wurden. Abschließend wird sowohl der Umgang mit Missings als auch die Poweranalyse beschrieben. Soweit nicht anders erwähnt wurden alle Analysen mit PASW Statistics 17 (Kinnear & Gray, 2010) gerechnet.

3.1.5.1 Analyse des Fragebogens zur sozialen Validität

Da es sich beim Fragebogen zur sozialen Validität um ein neu entworfenes Instrument handelt, müssen seine Gütekriterien ausführlich bestimmt werden. Dazu wurde die folgende Vorgehensweise gewählt: Zunächst wurden die Gütekriterien der theoretisch abgeleiteten Skalen bestimmt. Anhand dieser Ergebnisse wurden erste Items ausgeschlossen. Anschließend wurde der Fragebogen einer faktorenanalytischen Überprüfung unterworfen. Es wurden faktoriell interpretierbare Skalen gebildet. Dabei wurden weitere Items eliminiert. Danach wurden die Gütekriterien dieser faktorenanalytisch gewonnenen Skalen bestimmt. Die daraus resultierenden Skalen wurden schließlich genutzt, um die soziale Validität der Testverfahren zu bestimmen. Hierzu wurden dann die Mittelwerte der Skalen gebildet. Im Folgenden werden die Begriffe Faktor und Skala synonym verwendet, auch wenn diese Bezeichnung etwas ungenau ist.

Gütekriterien

Zur Bestimmung der Reliabilität des Fragebogens wurden sowohl die Trennschärfen der Items als auch die interne Konsistenz der einzelnen Skalen durch den Alpha-Koeffizienten α von Cronbach (1951) berechnet. Die Trennschärfe r_{it} eines Items entspricht der Korrelation eines Items mit dem Gesamtergebnis und gibt damit an, wie gut ein einzelnes Item das Gesamtergebnis eines Tests, bzw. in diesem Falle einer Skala repräsentiert (Fisseni, 2004). Cronbachs Alpha hingegen ist eine Schätzung der Korrelation zwischen zwei beliebigen Itemstichproben aus allen Items des Tests (Cronbach, 1951) und zeigt somit die interne Konsistenz, also die Reliabilität des Tests, an. Anhand dieser Kennwerte wurden die Items jeder Skala beurteilt und gegebenenfalls eliminiert.

Hauptkomponentenanalyse

Bei der Konstruktion des Fragebogens wurde durch die sorgfältige Itemauswahl versucht, eine hinreichende Inhaltsvalidität zu erzielen. Da dieses Konzept umstritten und seine Güte schwer zu beurteilen ist (Lienert & Raatz, 1994),

wurde zur weiterführenden Beurteilung des Fragebogens eine Hauptkomponentenanalyse gerechnet. Insbesondere wenn ein neuer Fragebogen entworfen wird, ist nicht von vornherein klar, auf welche zugrunde liegenden Faktoren die einzelnen Items zurückzuführen sind (Lienert & Raatz, 1994). Eine Hauptkomponentenanalyse dient in erster Linie der Datenreduktion (Bühner 2006). Sie zielt darauf ab, die Datenmenge der Ausgangsvariablen nach Zusammenhängen zu durchforsten und damit auf wenige Faktoren zu reduzieren, welche aber immer noch die essentielle Ausgangsinformation enthalten. Die Grundlogik (Eid, Gollwitzer & Schmitt, 2010; Bortz & Schuster, 2010; Tabachnick & Fidell, 2004) lässt sich wie folgt zusammengefasst darlegen: Anhand der Interkorrelationsmatrix der Ausgangsvariablen werden nach dem Prinzip der maximalen Varianzabschöpfung Faktoren aus der ursprünglichen Datenmenge extrahiert. Um dem Ziel der Datenreduktion gerecht zu werden, muss der Anwender ein Abbruchkriterium festlegen, um nicht genauso viele Faktoren wie Ausgangsvariablen zu erhalten. Dazu eignen sich zwei Vorgehensweisen: Nach dem Kaiser-Guttman-Kriterium (Guttman, 1954; Kaiser, 1960) werden nur solche Faktoren mit einem Eigenwert $\lambda > 1$ extrahiert. Das bedeutet, dass ein Faktor mehr Varianz aufklären muss als eine Ausgangsvariable. Legt man die Anzahl der Faktoren anhand des Scree-Tests (Cattell, 1966) fest, extrahiert man hingegen so viele Faktoren, bis im Eigenwertverlauf ein deutlicher Knick eintritt. Hat man die Anzahl der Faktoren festgelegt, erhält man eine Faktorladungsmatrix. Um diese inhaltlich gut interpretieren zu können, ist meist eine Rotation der ursprünglich gefundenen Faktorenlösung notwendig. Ein häufig eingesetztes Rotationsverfahren ist die Varimax-Rotation (Kaiser, 1958); zum einen, da sie zu den orthogonalen Rotationsmethoden gehört, bei denen im Gegensatz zu einer obliquen Rotationsmethode die Faktoren unabhängig voneinander bleiben und zum anderen, weil sie im Allgemeinen gut interpretierbare Lösungen liefert. Dieser Schritt zielt auf eine Einfachstruktur der Faktoren nach Thurstone (1947) ab. Dabei ist es nötig, dass einzelne Items hoch auf einem Faktor laden, während sie nur gering auf anderen Faktoren laden. Die rotierte Faktorladungsmatrix dient schließlich als Basis der inhaltlichen Interpretation der Faktoren. Falls

benötigt, lassen sich für jede Versuchsperson entsprechende Faktorwerte berechnen.

Um eine Hauptkomponentenanalyse zu rechnen, empfiehlt sich eine Stichprobengröße von mindestens 50 Personen (Tabachnick & Fidell, 2004). Alternativ sollten dreimal so viele Personen wie Variablen vorhanden sein (Wirtz & Nachtigall, 2008). In der vorliegenden Studie konnte keine dieser Minimalannahmen erfüllt werden, weshalb die diesbezüglichen Ergebnisse mit äußerster Vorsicht zu betrachten sind. De Winter, Dodou und Wieringa (2009) haben allerdings gezeigt, dass eine kleine Stichprobe kein Hinderungsgrund für eine solche Analyse sein sollte. Auch wick die Hälfte der Items signifikant von einer Normalverteilung ab, was sich ungünstig auf die Hauptkomponentenanalyse auswirken kann. Da es sich hier allerdings um ein eher exploratives Setting mit einem komplett neu entworfenen Fragebogen handelt, wäre es trotz der schlechten Voraussetzungen unangebracht, eine solche Analyse von vornherein auszuschließen.

In der vorliegenden Studie wurde deshalb eine Hauptkomponentenanalyse mit Varimax-Rotation gerechnet, um die Struktur des Fragebogens zur sozialen Validität zu beurteilen. Die Analysen wurden am Fragebogen „Interview“ durchgeführt. Eine Berechnung anhand der Version „Testung“ brachte nur gering abweichende Ergebnisse, die deshalb im Folgenden vernachlässigt werden konnten.

3.1.5.2 Analyse der Konstruktvalidität

Zur Überprüfung der Konstruktvalidität des Opus[®]-Interviews wurde eine Multitrait-Multimethod-Matrix herangezogen. Diese beinhaltet sämtliche Korrelationen zwischen den eingesetzten Verfahren Opus[®]-Interview, WIT-2 und ASK. Zudem wurde die Selbsteinschätzung der Teilnehmer als weiteres Kriterium herangezogen. Die Analyse der Hypothese 1 mit ihren Unterpunkten basiert somit auf Produkt-Moment-Korrelationen. Zur Berechnung eines Korrelationskoeffizienten r müssen verschiedene Voraussetzungen erfüllt sein. Die zugrunde liegenden Daten müssen Intervallskalenniveau erfüllen. Da es

sich um aus psychologischen Testverfahren gewonnene Werte handelt und diese in der Praxis, wenn auch nicht unumstritten, meist als intervallskaliert angesehen werden, kann diese Voraussetzung als erfüllt gelten (Bortz, Lienert & Boehnke, 2000). Des Weiteren wurden die Korrelationen auf ihre Signifikanz getestet. Hierzu muss die Voraussetzung der bivariaten Normalverteilung der betrachteten Variablen erfüllt sein. Aus praktischen Gründen geht man allerdings dazu über, schlicht die Normalverteilung jeder einzelnen Variablen zu prüfen (Bortz & Schuster, 2010). Deshalb wurden alle relevanten Variablen einer Überprüfung mittels Kolmogorov-Smirnov-Einstichproben-Test unterzogen. Dieser Test schien angemessen, da er die Anpassung an eine Normalverteilung prüft und dabei auch für kleine Stichproben relativ robust ist (Eid et al., 2010). Der Kolmogorov-Smirnov-Einstichproben-Test zeigte bei drei von elf Variablen eine Signifikanz von $p < .05$. Das bedeutet, dass bei den Variablen Analytisches Denken (Selbsteinschätzung), Kreatives Denken (Selbsteinschätzung) und Entscheidungsvermögen (Selbsteinschätzung) signifikante Abweichungen von einer Normalverteilung gefunden wurden. Auffällig ist, dass nur die Variablen der Selbsteinschätzung von der Normalverteilung abwichen. Zudem näherten sich diese Variablen bei grafischer Betrachtung einer Normalverteilung. Deshalb fiel auch trotz der Testergebnisse die Entscheidung zugunsten der Berechnung von Produkt-Moment-Korrelationen. Der Produkt-Moment-Korrelationskoeffizient r zeigt sich robust hinsichtlich Verletzungen der Normalverteilungsannahme (Havlicek & Peterson, 1977) und die Intervallsaklenniveau-Annahme kann als hinreichend erfüllt angesehen werden (Bortz et al., 2000). Zudem hätte die Verwendung von dem auf Ordinalskalenniveau basierenden Rangkorrelationskoeffizienten ρ nach Spearman (1906) in diesem Fall zu einem Informationsverlust führen können (Eid et al. 2010). Aufgrund der zuvor postulierten Einzelhypothesen konnte auf eine Alpha-Fehler-Adjustierung verzichtet werden (Bortz et al., 2000). Da auch die Richtung des jeweils erwarteten Zusammenhangs bereits in den Hypothesen postuliert wurde, konnten einseitige Signifikanztests gerechnet werden (Bortz & Schuster, 2010).

3.1.5.3 Analyse der sozialen Validität

Um bestimmen zu können, welches Verfahren bei den Teilnehmern auf größere Resonanz stößt, wurden die Verfahren gegenübergestellt. Dies geschah, indem die Mittelwerte der faktorenanalytisch ermittelten Skalen genutzt wurden. Jede Skala aus dem Fragebogen „Interview“ wurde mit der entsprechenden Skala aus dem Fragebogen „Testung“ verglichen. Da jede Frage vom Teilnehmer zweimal beantwortet wurde, einmal in der Version „Interview“ und einmal in der Version „Testung“, resultierten abhängige Antworten. Das bedeutet, dass sich Messwertpaare ergaben, die systematische Ähnlichkeit aufwiesen, da ein und dieselbe Person unter zwei Bedingungen getestet wurde (Bortz & Schuster, 2010). In diesem Fall war also ein t -Test für abhängige Stichproben das adäquate Analysemittel (Eid et al., 2010). Dieser t -Test trifft eine Aussage darüber, ob sich die beiden in den abhängigen Stichproben ermittelten Mittelwerte bedeutsam unterscheiden. Eine Voraussetzung zur Anwendung des t -Tests für abhängige Stichproben ist, dass sich der Differenzwert, der aus dem Vergleich der beiden Mittelwerte resultiert, normalverteilt. Da es sich bei der vorliegenden Stichprobe um ein $N \geq 30$ handelt, kann diese Voraussetzung als annähernd erfüllt gelten (Eid et al., 2010). In der vorliegenden Studie wurden gerichtete Hypothesen formuliert, weshalb die Testung einseitig erfolgte (Bortz & Schuster, 2010).

Die Globale Frage und die Entscheidungsfrage des Fragebogens wurden separat ausgewertet, da sie nicht Bestandteil der ermittelten Skalen waren. Zur Auswertung der Globalen Frage wurde ebenfalls der t -Test für abhängige Stichproben herangezogen. Die Entscheidungsfrage wurde anhand des Binomialtestes ausgewertet (Eid et al., 2010), um bestimmen zu können, ob die Teilnehmer eines der beiden eingesetzten Verfahren bevorzugten. Antworten auf offene Fragen werden exemplarisch berichtet.

3.1.5.4 Umgang mit Missings

Missings sind in empirischen Studien ein häufiges Problem. Sie entstehen z. B. dadurch, dass Teilnehmer nicht zu jedem Erhebungszeitpunkt kommen oder

nicht alle Items eines Fragebogens beantworten (Eid et al., 2010). Da Missings die Stichprobengröße dezimieren und methodische Schwierigkeiten mit sich bringen können, insbesondere wenn sie nicht rein zufälliger Natur sind (Raykov & Marcoulides, 2008), haben sich im Laufe der Zeit verschiedene Vorgehensweisen entwickelt mit ihnen umzugehen.

In der einschlägigen Literatur findet sich eine Reihe von Methoden zum Umgang mit Missings, die meistens auf den gleichen Grundprinzipien basieren (Cohen, Cohen, West & Aiken, 2003; Tabachnick & Fidell, 2004; Raykov & Marcoulides, 2008; Shadish, Cook & Campbell, 2002). Es gibt einerseits die Möglichkeit, fehlende Werte, z. B. durch den Mittelwert der betrachteten Variable oder komplexere Imputationsverfahren, zu ersetzen. Andererseits kann man Personen komplett oder teilweise ausschließen. Die erste Möglichkeit dabei ist, diejenigen Personen mit fehlenden Werten im Datensatz von der Analyse komplett auszuschließen, also listenweise. Dadurch befinden sich nur noch komplette Datensätze in der Auswertung. Allerdings verringert dieses Vorgehen die Größe der Gesamtstichprobe. Die zweite Ausschlussmöglichkeit besteht darin, Missings paarweise auszuschließen. Das bedeutet, dass z. B. bei Korrelationsanalysen die Stichprobengröße bei einzelnen Korrelationen unterschiedlich groß ist, je nachdem für wie viele Personen Werte auf den beiden korrelierten Variablen vorliegen. Neben diesen gängigen Möglichkeiten mit Missings umzugehen, gibt es zumindest theoretisch auch die Variante, diejenige Variable mit Missings komplett auszuschließen. Dieser Weg wird in der Forschung allerdings kaum gewählt. In einer Studie sollten nur Variablen erhoben werden, die von Interesse sind, so dass man sie im Nachhinein nicht ohne weiteres aufgrund von Missings komplett ausschließen kann (Cohen et al., 2003). Ein solches Vorgehen ist nur dann gerechtfertigt, wenn außergewöhnlich viele Missings vorliegen, die Berechnungen mit dieser Variablen unmöglich machen. Schließlich kann auch im Nicht-Beantworten einer Frage Information enthalten sein. So kann es sinnvoll sein, Missings als eigene Variable aufzunehmen oder verschiedene Methoden im Umgang mit Missings zu vergleichen (Tabachnick & Fidell, 2007).

In der vorliegenden Studie gab es weder außergewöhnlich viele Missings, noch ließ sich ein Muster in den fehlenden Werten erkennen. Da sich die Missings weder bei gewissen Variablen noch bei einzelnen Versuchspersonen häuften, wurde mit Missings folgendermaßen verfahren: Bei der Analyse der Konstruktvalidität wurde paarweise ausgeschlossen. Während der Faktorenanalyse wurden einzelne Missings durch den Mittelwert des jeweiligen Items ersetzt. Da bei der Analyse der sozialen Validität nur vereinzelte Missings vorlagen, wurden die Mittelwerte jeder Skala aus den jeweils vorhandenen Werten gebildet. Dieses Vorgehen schien den Zielen der Studie angemessen, da so die geringe Stichprobengröße nicht noch weiter dezimiert wurde.

3.1.5.5 Poweranalyse

Um die benötigte Stichprobe für die Analysen abschätzen zu können, wurde eine a priori-Poweranalyse mit G*Power 3 (Faul, Erdfelder, Lang & Buchner, 2007) gerechnet. Die Berechnung erfolgte zunächst für die Analyse der Konstruktvalidität. Da in Hypothese 1 von einem großen Effekt ausgegangen wird, wurde die Stichprobengröße bei einem großen Effekt, einem Signifikanzniveau von .05 und einer Teststärke von .80 berechnet. Hierbei ergab sich eine benötigte Stichprobengröße von 23 Teilnehmern. Bei einem lediglich mittleren Effekt wäre eine Stichprobengröße von 67 Teilnehmern erforderlich gewesen. Für die Analyse der sozialen Validität zeigte sich, dass bei einem großen Effekt, einem Signifikanzniveau von .05 und einer Teststärke von .80 eine Stichprobengröße von 21 Teilnehmern erforderlich ist. Bei einem lediglich mittleren Effekt wäre eine Stichprobe von 52 Teilnehmern erforderlich gewesen.

3.2 Darstellung der Ergebnisse

Nachdem in den vergangenen Kapiteln auf die methodischen Aspekte der Studie eingegangen wurde, werden im folgenden Kapitel die Ergebnisse der durchgeführten Analysen präsentiert.

3.2.1 Soziodemografische Zusammensetzung der Stichprobe

Insgesamt nahmen 57 Personen an der Studie teil. Von diesen 57 Teilnehmern waren 51% weiblichen und 49% männlichen Geschlechts. Das Alter der Teilnehmer variierte zwischen 16 und 56 Jahren, mit einem Durchschnitt von 31,3 Jahren und einer Standardabweichung von 10,7 Jahren. Wie sich schon aus der Rekrutierung der Stichprobe erkennen lässt, hatten die Teilnehmer einen recht breit gefächerten beruflichen Hintergrund. Ein Großteil der Teilnehmer war zum Zeitpunkt der Erhebung noch als Student einzustufen. Die teilnehmenden Berufstätigen deckten inhaltlich sämtliche Bereiche von Entwicklung, Produktion, Vertrieb, Personal, IT sowie weitere Tätigkeiten in Wirtschaft und Forschung ab. Auffällig ist, dass 86% der Teilnehmer im Laufe ihrer Schulbildung die (Fach-)Hochschulreife erreicht haben. Daraus ergibt sich, dass der Einsatz der unterschiedlichen Verfahren gerechtfertigt ist und der Zielgruppe entspricht. Zu beachten ist allerdings, dass das Abschneiden der Teilnehmer durch diesen Bildungshintergrund tendenziell als überdurchschnittlich zu werten ist. Aus dem Ablauf der Studie ergibt sich, dass die Teilnehmer nicht zufällig ausgewählt werden konnten. Deshalb besteht ein erhöhtes Risiko, dass ein Selektionseffekt vorliegen könnte. Es ist nicht auszuschließen, dass sich diejenigen Personen, die der Einladung zur Teilnahme folgten, hinsichtlich verschiedener Merkmale signifikant von den nicht teilnehmenden Personen unterscheiden. Da nicht alle beteiligten Unternehmen zustimmten, dass ihre Teilnehmer den Fragebogen zur sozialen Validität ausfüllen sollten sowie aus organisatorischen Gründen, gehen in die diesbezüglichen Analysen nur 31 Personen ein.

3.2.2 Gütekriterien des Fragebogens zur sozialen Validität

Die Bestimmung der Gütekriterien des Fragebogens zur sozialen Validität erfolgte in drei Schritten.

Erster Schritt

Zunächst wurden die Mittelwerte und Standardabweichungen der einzelnen Items berechnet (siehe Anhang C). Auffällig hierbei waren die durchgängig hohen Mittelwerte. In einem ersten Analyseschritt wurde dann jede einzelne Skala hinsichtlich ihrer internen Konsistenz sowie der Trennschärfe der jeweiligen Items untersucht (siehe Anhang C). Um die gefundenen Ergebnisse bewerten zu können, wurde in Anlehnung an Bühner (2006) eine Trennschärfe r_{it} von .30 bis .50 als mittelmäßig gewertet. Ein Wert von $r_{it} > .50$ gilt als hoch, wobei die Trennschärfe generell möglichst hoch sein sollte (Moosbrugger & Kelava, 2007). Die Interpretation von Cronbachs Alpha wurde in der entsprechenden Literatur kritisch diskutiert (Lance, Butts & Michels, 2006). In Anlehnung an Nunnally und Bernstein (2006) wird folgende Einteilung vorgenommen: Alpha-Koeffizienten $> .90$ werden als sehr gut, $> .80$ als gut und $> .70$ als akzeptabel gewertet. Bei Werten $< .70$ ist die interne Konsistenz fragwürdig und ab Werten $< .50$ inakzeptabel. Die Itemselektion ist allerdings kein rein technisches Vorgehen. So sind verschiedene Itemkennwerte und -inhalte immer gegeneinander abzuwägen (Moosbrugger & Kelava, 2007). Auch die Anzahl der Items pro Skala ist nicht zu vernachlässigen. Im Folgenden wurde deshalb ein Kompromiss gefunden: Als unterer Grenzwert der Trennschärfe wurde ein Wert von $r_{it} > .30$ festgelegt. Weiterhin wurden Items entfernt, die zu einer Steigerung von Cronbachs Alpha führten. Dabei wurde jedoch darauf geachtet, dass immer mindestens drei Items pro Skala übrig blieben. Zudem wurden Itemmittelwert und -streuung betrachtet. Im Zweifel führte die inhaltliche Betrachtung des jeweiligen Items zu einer Entscheidung über den eventuellen Ausschluss.

Die Skala des Faktors 1 „Partizipation“ wies mit ihren sieben Items zunächst ein Cronbachs Alpha von .64 auf und wäre somit als fragwürdig aber nicht inakzeptabel einzustufen. Da die Trennschärfen von Item 1 und Item 6 zu gering und tendenziell negativ waren, wurden diese beiden Items entfernt. Dadurch stieg Cronbachs Alpha stark und zwar auf .85, was als gut zu werten ist. Die Skala des Faktors 2 „Transparenz“ zeigte mit ihren sieben Items

zunächst ein Cronbachs Alpha von .69 an und lag somit knapp unter dem akzeptablen Wertebereich. Die Items 7 und 9 wurden aufgrund geringer Trennschärfe entfernt. Dabei stieg Cronbachs Alpha auf .76 an und erreichte somit einen akzeptablen Wert. Die Skala des Faktors 3 „Affektive Bewertung“ wies mit ihren sieben Items zunächst ein Cronbachs Alpha von .78 auf und erreichte somit einen akzeptablen Wert. Aufgrund der geringen Trennschärfe wurde Item 16 aussortiert, wodurch Cronbachs Alpha auf .80 stieg und somit als gut zu werten ist. Die Skala des Faktors 4 „Job Acceptance Intention“ zeigte mit ihren drei Items ein Cronbachs Alpha von .82 und ist ebenfalls als gut zu werten. Da die Skala nur drei Items enthielt und die Trennschärfen in einem sehr guten Bereich lagen, wurden keine Items entfernt. Gleiches galt für die Skala des Faktors 5 „Organisational Attractiveness“ bei einem sehr guten Cronbachs Alpha von .90. Die Skala des Faktors 6 „Perceived Predictive Validity“ wies mit ihren sieben Items ein Cronbachs Alpha von .84 bei sehr guten Trennschärfen auf und ist somit als gut anzusehen. Die Skala des Faktors 7 „Perceived Job Relatedness“ zeigte mit ihren fünf Items zunächst ein Cronbachs Alpha von .85 und erreichte somit einen guten Wert. Da einzig Item 34 eine lediglich mittelmäßige Trennschärfe aufwies, wurde dieses Item entfernt. Dadurch stieg Cronbachs Alpha weiter auf .89. Die Skala des Faktors 8 „Achtung der Privatsphäre“ erreichte mit ihren vier Items ein Cronbachs Alpha von .74 und damit einen akzeptablen Wert. Item 40 wies als einziges Item der Skala eine lediglich mittelmäßige Trennschärfe auf und wurde aussortiert, wodurch Cronbachs Alpha leicht auf .75 stieg. Die Skala des Faktors 9 „Testmotivation“ wies mit ihren vier Items ein Cronbachs Alpha von .84 bei sehr guten Trennschärfen auf und erreichte somit einen guten Wert. Schließlich gingen alle Items außer Item 1, 6, 7, 9, 16, 34 und 40 in die weiteren Analysen ein.

Zweiter Schritt

In einem zweiten Schritt wurde mit den verbliebenen Items eine Hauptkomponentenanalyse mit Varimax-Rotation gerechnet. Die Faktorextraktion anhand des Kaiser-Guttman-Kriteriums ergab elf Faktoren mit einem

Eigenwert $\lambda > 1$. Allerdings führte die Rotation zu keinem interpretierbaren Ergebnis. Da anhand des Kaiser-Guttman-Kriteriums oftmals zu viele Faktoren extrahiert werden (Russell, 2002), wurde der Scree-Test hinzugezogen. Bei Betrachtung des Scree-Plots zeichnete sich eine Zwei-Faktor-Lösung ab, die jedoch hinsichtlich der Ziele der Untersuchung, dem Auffinden mehrerer Fragebogenskalen, nicht ausreichend differenziert erschien. Bei näherer Betrachtung der Varianzaufklärung der gefundenen Faktoren zeigte sich eine weitere mögliche Lösung. Es wurden schließlich fünf Faktoren extrahiert, die alle einen Eigenwert $\lambda > 2$ aufwiesen und zusammen 64,76% der Varianz aufklärten. Im Scree-Plot zeichnet sich an dieser Stelle eine weitere Absenkung ab.

Zur Interpretation der einzelnen Faktoren lässt sich sagen, dass insbesondere Faktoren mit vier oder mehr Ladungen $> .60$ gut interpretierbar sind und zwar unabhängig von der Stichprobengröße (Guadagnoli & Velicer, 1988). Laut Comrey und Lee (1992, zitiert nach Tabachnick & Fidell, 2004) sind Faktorladungen von $.71$ als ausgezeichnet, $.63$ als sehr gut, $.55$ als gut, $.45$ als akzeptabel und $.32$ als schwach zu bezeichnen. In Tabelle 3-1, die die rotierte Faktorladungsmatrix der fünf Faktoren darstellt, wurden deshalb Faktorladungen $< .32$ ausgelassen. Items, die besonders hoch auf einem Faktor laden, sind als Markiervariablen anzusehen und beschreiben den Inhalt dieses Faktors besonders treffend.

Tabelle 3-1: Rotierte Faktorladungsmatrix und Varianzaufklärung der fünf extrahierten Faktoren des Fragebogens zur sozialen Validität

Item	Faktor 1: Akzeptanz und resultierende Motivation	Faktor 2: Testerleben	Faktor 3: Transparenz	Faktor 4: Jobbezug	Faktor 5: Achtung der Privatsphäre
25	.86				
24	.83	.36			
22	.75				
21	.75				
32	.71				
23	.66	.39			-.37
26	.66				
43	.64		.51		.39
30	.64			.36	
27	.61		-.45		
46	.60		.50		
33	.60			.51	
29	.58	.37			
44	.57		.35		
45	.50				
28	.48			.40	
2	.39	.76			
4	.33	.74			
47	.45	.73			
3		.63		.43	
19		.61			
18		.60	.46		
14		.56			.32
15		.51			.37
5	.33	.44			
8			.72		
12			.66		
11			.61		
13			.61		
10		.54	.59		
20			.56		.52
17	.33	.36	.49		
36				.81	
35				.79	
37				.78	
38		.40		.67	
31	.45			.64	
39					.84
42					.79
41				-.41	.69
Varianz- aufklärung	20,94%	13,25%	11,45%	10,85%	8,27%

Der erste Faktor klärt 20,94% Varianz auf. Auf ihm laden fast alle Items der theoretischen angenommenen Faktoren 4 („Job Acceptance Intention“), 5 („Organisational Attractiveness“), 6 („Perceived Predictive Validity“), und 9 („Testmotivation“). Dabei laden Item 25 und 24 (ehemals Faktor 5) sowie Item 21 und Item 22 (ehemals Faktor 4) ausgezeichnet auf diesem Faktor. Da ebenfalls Items akzeptabel bis sehr gut auf diesem Faktor laden, die ehemals den Faktoren 6 und 9 zugeordnet waren, ergibt sich inhaltlich ein Faktor, der die Haltung gegenüber dem Verfahren abbildet und die damit einhergehende Motivation an dem Verfahren teilzunehmen, ein folgendes Jobangebot anzunehmen und die zugehörige Organisation weiterzuempfehlen. Dieser Faktor wird künftig als „Akzeptanz und resultierende Motivation“ bezeichnet.

Der zweite Faktor klärt 13,25% Varianz auf. Er setzt sich aus sämtlichen Items des ehemaligen Faktors 1 („Partizipation“) sowie aus den meisten Items des ehemaligen Faktors 3 („Affektive Bewertung“) zusammen. Ausgezeichnet laden Item 2, 4 und 47, welche sich auf die Möglichkeit, Stärken zu zeigen beziehen. Da auch Items gut laden, die sich auf Anstrengung und Stress beziehen, wird dieser Faktor künftig als „Testerleben“ bezeichnet.

Der dritte Faktor klärt 11,45% Varianz auf. Auf ihm laden besonders hoch die Items des ehemaligen Faktors 2 („Transparenz“), bei guten bis ausgezeichneten Werten. Insbesondere die Items 8 und 12 laden ausgezeichnet und beschreiben den Inhalt des Faktors. Zusätzlich laden noch Item 20 und 17 von ehemals Faktor 3. Da diese Ladungen geringer sind als die der Items von ehemals Faktor 2 und inhaltlich nicht eindeutig in Einklang zu bringen sind, wurde entschieden diese beiden Items auszusortieren und den Faktor weiterhin als „Transparenz“ zu bezeichnen.

Faktor 4 klärt 10,85% Varianz auf. Er beinhaltet alle Items des ehemaligen Faktors 7 („Perceived Job Relatedness“) bei sehr guten bis ausgezeichneten Ladungen. Zusätzlich lädt Item 31 vom ehemaligen Faktor 6 auf diesem Faktor. Da dieses Item allerdings auch akzeptabel auf dem Faktor „Akzeptanz und resultierende Motivation“, wo es theoretisch auch vermutete wurde lädt, wurde

entschieden, das Item auszusortieren. Faktor 4 wurde dementsprechend unter der vereinfachten Bezeichnung „Jobbezug“ beibehalten.

Faktor 5 klärt schließlich 8,27% Varianz auf. Er ist der einzige Faktor, der die Forderung von vier oder mehr Ladungen $> .60$ nicht erfüllt, weil auf ihm insgesamt nur drei Items laden. Es handelt sich dabei allerdings um die drei einzigen Items zum ehemaligen Faktor 8 („Achtung der Privatsphäre“) bei sehr hohen bis ausgezeichneten Ladungen, weshalb der Faktor in dieser Form beibehalten wurde. Für den Fragebogen ergeben sich somit fünf faktoriell interpretierbare Skalen: „Akzeptanz und resultierende Motivation“, „Testerleben“, „Transparenz“, „Jobbezug“ und „Achtung der Privatsphäre“. Die Items 17, 20 und 31 wurden für die weitere Analyse ausgeschlossen.

Dritter Schritt

In einem dritten Schritt wurden nochmals die Trennschärfen und Cronbachs Alpha der faktoriell interpretierbaren Skalen untersucht. Die Ergebnisse finden sich in Tabelle 3-2 wieder. Für die erste Skala „Akzeptanz und resultierende Motivation“ mit 16 Items zeigte sich ein Cronbachs Alpha von .95 bei hohen Trennschärfen, was als sehr gut einzustufen ist. Für die zweite Skala „Testerleben“ mit neun Items ergab sich ein gutes Cronbachs Alpha von .85, wobei die Trennschärfen von Item 5 und 15 nur als mittelmäßig einzustufen sind. Die dritte Skala „Transparenz“ mit fünf Items wurde durch die Faktorenanalyse nicht verändert und behielt das akzeptable Cronbachs Alpha von .76. Ebenso behielt die vierte Skala „Jobbezug“ mit vier Items ihr hohes Cronbachs Alpha von .89 bei hohen Trennschärfen. Auch für die fünfte Skala „Achtung der Privatsphäre“ mit drei Items ergab sich keine Änderung und ihr Cronbachs Alpha von .75 ist als akzeptabel zu bewerten. Die fünf faktoriell interpretierbaren Skalen stellen somit die für die vorliegende Studie endgültige Version des Fragebogens dar und werden weiterhin zur Bestimmung der sozialen Validität genutzt.

Tabelle 3-2: Cronbachs Alpha und Trennschärfe der verwendeten Items und Skalen der fünf extrahierten Faktoren des Fragebogens zur sozialen Validität

Skala	Item	Cronbachs α der Skala	Trennschärfe r_{it}
Faktor 1: Akzeptanz und resultierende Motivation	Item 21	.95	.84
	Item 22		.77
	Item 23		.69
	Item 24		.87
	Item 25		.91
	Item 26		.67
	Item 27		.60
	Item 28		.56
	Item 29		.61
	Item 30		.69
	Item 32		.64
	Item 33		.80
	Item 43		.76
	Item 44		.69
Item 45	.64		
Item 46	.76		
Faktor 2: Testerleben	Item 2	.85	.68
	Item 3		.54
	Item 4		.74
	Item 5		.36
	Item 14		.67
	Item 15		.38
	Item 18		.62
	Item 19		.56
Item 47	.82		
Faktor 3: Transparenz	Item 8	.76	.73
	Item 10		.56
	Item 11		.50
	Item 12		.49
	Item 13		.59
Faktor 4: Jobbezug	Item 35	.89	.76
	Item 36		.78
	Item 37		.78
	Item 38		.74
Faktor 5: Achtung der Privatsphäre	Item 39	.75	.63
	Item 41		.65
	Item 42		.64

3.2.3 Ergebnisse zur Konstruktvalidität

Im Folgenden sollen die Ergebnisse der Analyse der Konstruktvalidität hinsichtlich der Hypothese 1 wiedergegeben werden. Es wird zwischen nicht signifikanten ($p \geq .05$), signifikanten ($p < .05$) und hochsignifikanten ($p < .01$) Korrelationen unterschieden. Um die Größe der auftretenden Korrelationen bewerten zu können, sei auf Cohen (1992) verwiesen, der als Konvention für die Effektstärken von Korrelationen folgende Einteilung vornimmt: Werte von .10 weisen auf einen kleinen, Werte von .30 auf einen mittleren und Werte von .50 auf einen großen Effekt hin.

Zur Übersicht sind sämtliche Korrelationen zwischen den Verfahren Opus[®], WIT-2, ASK und der Selbsteinschätzung tabellarisch in einer Korrelationsmatrix in Tabelle 3-3 wiedergegeben. In Bezug auf das Analytische Denken aus dem Opus[®]-Interview zeigten sich hochsignifikante Korrelationen mit Kreativem Denken (Opus[®]), Entscheidungsvermögen (Opus[®]), Kreativem Denken (ASK), Schlussfolgerndem Denken (WIT-2), Analytischem Denken (Selbsteinschätzung) und Sprachlichem Denken (WIT-2). Die Korrelationen mit Kreativem Denken (Opus[®]) und Entscheidungsvermögen (Opus[®]) sind als starke Effekte einzustufen. Alle weiteren hochsignifikanten Korrelationen entsprechen mittleren Effekten. Signifikante Korrelationen ergaben sich zwischen Analytischem Denken (Opus[®]) und Räumlichem Denken (WIT-2) sowie Rechnerischem Denken (WIT-2). Die restlichen Variablen erreichten korreliert mit Analytischem Denken (Opus[®]) keine Signifikanz. In Bezug auf das Kreative Denken aus dem Opus[®]-Interview zeigten sich hochsignifikante Zusammenhänge mit Entscheidungsvermögen (Opus[®]), Analytischem Denken (Opus[®]), Kreativem Denken (ASK) und Sprachlichem Denken (WIT-2). Die Korrelationen mit Entscheidungsvermögen (Opus[®]) und Analytischem Denken (Opus[®]) sind als starke Effekte zu werten. Signifikante Korrelationen ergaben sich zwischen Kreativem Denken (Opus[®]) und Rechnerischem Denken (WIT-2), Analytischem Denken (Selbsteinschätzung) und Schlussfolgerndem Denken (WIT-2). Hinsichtlich der restlichen Variablen ergaben sich mit Kreativem Denken (Opus[®]) keine signifikanten Korrelationen. In Bezug auf das

Entscheidungsvermögen aus dem Opus[®]-Interview wurden die Korrelationen mit Kreativem Denken (Opus[®]), Analytischem Denken (Opus[®]), Kreativem Denken (ASK), Analytischem Denken (Selbsteinschätzung) und Rechnerischem Denken (WIT-2) hochsignifikant, wobei nur die beiden ersten einen starken Effekt anzeigen.

Bezüglich der Selbsteinschätzung zeigten Kreatives Denken (Selbsteinschätzung) und Entscheidungsvermögen (Selbsteinschätzung) keinerlei signifikante Korrelationen. Analytisches Denken (Selbsteinschätzung) korrelierte hochsignifikant mit Räumlichem Denken (WIT-2), Analytischem Denken (Opus[®]), Sprachlichem Denken (WIT-2), Entscheidungsvermögen (Opus[®]) und Schlussfolgerndem Denken (WIT-2) bei mittleren Effekten. Signifikante Korrelationen ergaben sich noch mit Rechnerischem Denken (WIT-2), Kreativem Denken (Opus[®]) und Kreativem Denken (ASK).

In Bezug auf das Kreative Denken aus der ASK zeigten sich hochsignifikante Korrelationen mit Sprachlichem Denken (WIT-2), Entscheidungsvermögen (Opus[®]), Schlussfolgerndem Denken (WIT-2), Kreativem Denken (Opus[®]) und Analytischem Denken (Opus[®]) bei mittleren Effekten. Signifikant korrelierte Kreatives Denken (ASK) mit Analytischem Denken (Selbsteinschätzung).

Bezüglich des Schlussfolgernde Denkens aus dem WIT-2 zeigten sich hochsignifikante Korrelationen mit Räumlichem Denken (WIT-2), Sprachlichem Denken (WIT-2), Rechnerischem Denken (WIT-2), Kreativem Denken (ASK), Analytischem Denken (Opus[®]) und Analytischem Denken (Selbsteinschätzung). Die Korrelationen zu Räumlichem Denken (WIT-2), Sprachlichem Denken (WIT-2) und Rechnerischem Denken (WIT-2) sind hierbei als starke Effekte zu bewerten. Signifikant korrelierte das Schlussfolgernde Denken (WIT-2) mit Kreativem Denken (Opus[®]) und Entscheidungsvermögen (Opus[®]).

Tabelle 3-3: Komplette Korrelationsmatrix zur Konstruktvalidität

		Opus [®]			Selbsteinschätzung			ASK	WIT-2			
		AD	KD	EV	AD	KD	EV	KD	Sf D	Sp D	Re D	Rä D
Opus [®]	Analytisches Denken (AD)	1										
	Kreatives Denken (KD)	.81**	1									
	Entscheidungsvermögen (EV)	.69**	.85**	1								
Selbsteinschätzung	Analytisches Denken (AD)	.40**	.28*	.37**	1							
	Kreatives Denken (KD)	-.11	.06	-.05	.16	1						
	Entscheidungsvermögen (EV)	.06	.16	.19	.08	-.11	1					
ASK	Kreatives Denken (KD)	.40**	.43**	.48**	.23*	-.07	-.02	1				
WIT-2	Schlussfolgerndes Denken (Sf D)	.40**	.26*	.23*	.36*	.05	-.12	.43**	1			
	Sprachliches Denken (Sp D)	.37**	.33**	.28*	.38**	.02	-.14	.50**	.64**	1		
	Rechnerisches Denken (Re D)	.30*	.29*	.35**	.30*	-.09	-.03	.16	.51**	.39**	1	
	Räumliches Denken (Rä D)	.32*	.07	.12	.44**	-.05	.01	.19	.70**	.38**	.45**	1

Anmerkung: mit * gekennzeichnete Korrelationen sind einseitig mit $p < .05$ signifikant, mit ** gekennzeichnete Korrelationen sind einseitig mit $p < .01$ signifikant.

Um die Ergebnisse im Sinne einer Multitrait-Multimethod-Validierung beschreiben zu können, wurden die relevanten Werte in eine Multitrait-Multimethod-Matrix übertragen (siehe Tabelle 3-4). Um das grundlegende Muster beizubehalten, wurde eine komplette Multitrait-Multimethod-Matrix erstellt und die nicht besetzten Zellen wurden leer gelassen. In der Hauptdiagonalen finden sich die Reliabilitätswerte wieder. Da für das Opus[®]-Interview zurzeit keine Reliabilitätskennwerte vorliegen, bleiben die entsprechenden Zellen unbesetzt. Gleiches gilt für die Selbsteinschätzungswerte. Für die ASK und den WIT-2 finden sich hier die Werte der Retestreliabilität laut Testmanual. In der Multitrait-Multimethod-Matrix findet sich nur eine Validitätsdiagonale wieder und zwar zwischen dem Opus[®]-Interview und der Selbsteinschätzung. Da mit der ASK und dem WIT-2 jeweils nur ein relevantes Kriterium erhoben wurde, ergeben sich hier nur einzelne Validitätswerte und keine komplette Diagonale.

Die erste Forderung von Campbell und Fiske (1959), dass die Werte in der Validitätsdiagonalen signifikant von null verschieden und genügend hoch sein sollen, kann für die Korrelation zwischen Analytischem Denken (Opus[®]) und Analytischem Denken (Selbsteinschätzung) bestätigt werden ($r = .40^{**}$). Die Werte des Kreativen Denkens und Entscheidungsvermögens in der Validitätsdiagonalen zwischen Opus[®]-Interview und Selbsteinschätzung erfüllen die Forderung nicht. Hinsichtlich der weiteren Validitätswerte lässt sich in Bezug auf das Kreative Denken sagen, dass nur die Korrelation zwischen Kreativem Denken (Opus[®]) und Kreativem Denken (ASK) mit $r = .43^{**}$ einen zufriedenstellenden Wert erreicht, die Korrelation zwischen Kreativem Denken (Selbsteinschätzung) und Kreativem Denken (ASK) dagegen nicht. Hinsichtlich des Analytischen Denkens erreichen die Korrelationen zwischen Analytischem Denken (Opus[®]) und Schlussfolgerndem Denken (WIT-2) mit $r = .40^{**}$ und zwischen Analytischem Denken (Selbsteinschätzung) und Schlussfolgerndem Denken (WIT-2) mit $r = .36^{**}$ zufriedenstellende Werte. Die zweite Forderung, dass die Werte auf der Validitätsdiagonalen höher sein sollten als alle anderen Werte in derselben Spalte und Zeile der zugehörigen Heterotrait-Heteromethod-

Dreiecke, kann nur für die Korrelationen zwischen den drei Variablen Analytisches Denken (Opus[®]), Analytisches Denken (Selbsteinschätzung) und Schlussfolgerndes Denken (WIT-2) als erfüllt gelten. Die Korrelationen in Zusammenhang mit Kreativem Denken erfüllen diese Forderung nicht. Die dritte Forderung, dass die Werte auf der Validitätsdiagonalen höher sein sollten als die Korrelationen der entsprechenden Variablen in den Heterotrait-Monomethod-Dreiecken, wird nicht erfüllt. Lediglich die Korrelation zwischen Analytischem Denken (Opus[®]) und Analytischem Denken (Selbsteinschätzung) ist höher, als die Korrelationen im Heterotrait-Monomethod-Dreieck der Selbsteinschätzungsmethode. Auch die vierte Forderung, dass das Korrelationsmuster in den Heterotrait-Monomethod-Dreiecken und den Heterotrait-Heteromethod-Dreiecken annähernd äquivalent sein sollte, kann nicht als erfüllt gelten. Da sich bei der ASK und beim WIT-2 keine Heterotrait-Monomethod-Dreiecke ergeben, lassen sich die beiden letzten Forderungen auch nur ungenügend beurteilen.

Tabelle 3-4: Multitrait-Multimethod-Matrix zur Konstruktvalidität

		Opus®			Selbsteinschätzung			ASK	WIT-2				
		AD	KD	EV	AD	KD	EV	-	KD	-	Sf D	-	-
Opus®	Analytisches Denken (AD)	(-)											
	Kreatives Denken (KD)	.81**	(-)										
	Entscheidungsvermögen (EV)	.69**	.85**	(-)									
Selbsteinschätzung	Analytisches Denken (AD)	[.40**]	.28*	.37**	(-)								
	Kreatives Denken (KD)	-.11	[.06]	-.05	.16	(-)							
	Entscheidungsvermögen (EV)	.06	.16	[.19]	.08	-.11	(-)						
ASK	-	-	-	-	-	-	-	-					
	Kreatives Denken (KD)	.40**	[.43**]	.48**	.23*	[-.07]	-.02	-	(.77)				
WIT-2	Schlussfolgerndes Denken (Sf D)	[.40**]	.26*	.23*	[.36**]	.05	-.12	-	.43**	-	(.86)		
	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-

Anmerkung: mit * gekennzeichnete Korrelationen sind einseitig mit $p < .05$ signifikant, mit ** gekennzeichnete Korrelationen sind einseitig mit $p < .01$ signifikant; Reliabilitätswerte stehen in (), Validitätswerte stehen in []; ----- Monomethod-Blöcke, - - - Heteromethod-Blöcke.

Sämtliche Hypothesen zur Konstruktvalidität konnten nicht bestätigt werden. Es zeigt sich zwar eine hochsignifikante Korrelation zwischen Analytischem Denken (Opus[®]) und Schlussfolgerndem Denken (WIT-2), allerdings lässt sich bei einem Wert von $r = .40^{**}$ nur von einem mittleren Effekt sprechen. Zudem korreliert das Analytische Denken (Opus[®]) deutlich höher mit den anderen Opus[®]-Kriterien und dem Kreativen Denken (ASK). Das Schlussfolgernde Denken (WIT-2) hingegen korreliert höher mit den anderen WIT-2-Kriterien und dem Kreativen Denken (ASK). Hypothese 1a konnte demzufolge nicht bestätigt werden. Zwischen dem Kreativen Denken (Opus[®]) und dem Kreativen Denken (ASK) zeigte sich eine Korrelation von $r = .43^{**}$, die zwar hochsignifikant ist, wobei der Effekt allerdings nur mittelgroß ist. Das Kreative Denken (Opus[®]) korreliert höher mit den beiden anderen Opus[®]-Kriterien und das Kreative Denken (ASK) zeigt höhere Korrelationen mit Sprachlichem Denken (WIT-2), Entscheidungsvermögen (Opus[®]) und Schlussfolgerndem Denken (WIT-2). Hypothese 1b konnte somit auch nicht bestätigt werden. Hypothese 1c bezieht sich auf die Korrelation zwischen den drei Kriterien des Opus[®]-Interviews und der Selbsteinschätzung. Das Analytische Denken (Selbsteinschätzung) korreliert zwar hochsignifikant mit dem Analytischen Denken (Opus[®]), allerdings noch stärker mit Räumlichem Denken (WIT-2) und weiterhin mit Sprachlichem Denken (WIT-2), Entscheidungsvermögen (Opus[®]) und Schlussfolgerndem Denken (WIT-2). Die Selbsteinschätzungen des Kreativen Denkens und Entscheidungsvermögens zeigen überhaupt keine signifikanten Korrelationen mit einer der erhobenen Variablen. Hypothese 1c konnte demzufolge auch nicht bestätigt werden. Hypothese 1d bezieht sich darauf, dass alle weiteren Korrelationen geringer ausfallen als die postulierten Korrelationen. Aus der Darstellung aller bisherigen Ergebnisse ergibt sich, dass Hypothese 1d ebenfalls nicht bestätigt werden konnte.

3.2.4 Ergebnisse zur sozialen Validität

Im Folgenden werden die Ergebnisse zur sozialen Validität (Hypothese 2) wiedergegeben. Hierzu wurden die Mittelwerte der fünf faktoriell interpretierbaren Skalen für die Fragebogenversionen „Interview“ und „Testung“

bestimmt und paarweise verglichen. Zusätzlich wurde die Effektstärke $d_z = \bar{x}_d / \delta_d$ (Bortz & Döring, 2009) berechnet und in Anlehnung an Cohen (1988) interpretiert. Demnach kann man bei einem d_z von 0.57 von einem großen Effekt sprechen. Ein d_z von 0.35 entspricht einem mittleren und ein d_z von 0.14 einem kleinen Effekt.

Für die erste Skala „Akzeptanz und resultierende Motivation“ zeigte sich, dass das Opus[®]-Interview signifikant besser bewertet wurde als die Testverfahren ($t(30) = 3.34, p = .001$) bei einem großen Effekt von $d_z = 0.60$. Auch hinsichtlich der zweiten, Skala „Testerleben“ schnitt das Opus[®]-Interview signifikant besser ab ($t(30) = 3.92, p < .001$) ebenfalls bei einem großen Effekt von $d_z = 0.70$. Bei der dritten Skala „Transparenz“ zeigten sich keine signifikanten Unterschiede zwischen den Verfahren ($t(30) = -1.67, p = .053$), wobei die Tendenz eher für die Testverfahren spricht. Für die vierte Skala „Jobbezug“ zeigte sich wieder, dass das Opus[®]-Interview signifikant besser bewertet wurde ($t(30) = 3.00, p = .003$) bei einem mittleren Effekt von $d_z = 0.54$. Auch hinsichtlich der fünften Skala „Achtung der Privatsphäre“ schnitt das Opus[®]-Interview signifikant besser ab ($t(30) = 1.84, p = .038$) bei einem kleinen Effekt von $d_z = 0.33$. Die Testung der Globalen Frage zeigte ebenfalls, dass das Opus[®]-Interview signifikant besser bewertet wurde als die Testverfahren ($t(29) = 3.39, p = .001$) bei einem großen Effekt von $d_z = 0.62$. Bei der Entscheidungsfrage gaben 73% der Teilnehmer an, dass sie das Opus[®]-Interview bevorzugten. Nur 27% entschieden sich für die Testverfahren. Der entsprechende zweiseitige Binomialtest zeigte, dass diese Antworten signifikant von einer Gleichverteilung abweichen ($p = .001$).

Die offenen Fragen wurden von den Teilnehmern nur teilweise beantwortet. Für die Testverfahren zeigten sich folgende Tendenzen: Auf die Frage, welche Faktoren es dem Teilnehmer ermöglicht haben seine Stärken zu zeigen, antworteten fast alle in Richtung Vielfalt der Aufgaben und Klarheit der Struktur. Manche Teilnehmer erwähnten ihre Fähigkeit hinsichtlich einzelner Aufgaben. Auf die Gegenfrage, warum sie ihre Fähigkeiten nicht zeigen konnten und auf die Frage, was ihnen geholfen hätte, ihre Fähigkeiten besser zeigen zu können,

antwortete ein Großteil, dass sie zu sehr unter Zeitdruck und Stress gerieten und dass sie sich vorher ein Training gewünscht hätten, weil die Aufgaben der Intelligenz- und Kreativitätstestung keinen beruflichen Bezug gehabt hätten. Zudem bemängelte ein Großteil der Teilnehmer, dass es kaum persönlichen Austausch gegeben hätte und sie sich ein Gespräch gewünscht hätten, in dem sie ihre sozialen und rhetorischen Fähigkeiten unter Beweis hätten stellen können.

Für das Opus[®]-Interview zeigten sich folgende Tendenzen: Auf die Frage, welche Faktoren es dem Teilnehmer ermöglicht haben seine Stärken zu zeigen, antworteten fast alle Teilnehmer, dass sie die angenehme Gesprächsatmosphäre schätzten, in der mit offenen Fragen das komplexe Thema beleuchtet wurde. Dadurch sank der Leistungsdruck. Ein Teil erwähnte die Zeit zur Vorbereitung und die Zeit zum Nachdenken während des Interviews als positiv. Auf die Gegenfrage, warum die Teilnehmer ihre Fähigkeiten nicht zeigen konnten und auf die Frage, was ihnen geholfen hätte, ihre Fähigkeiten besser zeigen zu können, antworteten nur wenige Teilnehmer. Dabei drehten sich die Antworten hauptsächlich um das Thema der Fallstudie, wobei manche Teilnehmer gern persönlich Stellung bezogen hätten, andere allerdings befürchteten, vor dem Unternehmen ihre persönliche Einstellung preisgeben zu müssen. Ein Teil hätte sich lieber mit einem geläufigeren Thema befasst oder zwischen Themen gewählt. Einzelne Teilnehmer hätten sich eine klarere Zielsetzung gewünscht oder schlugen Kombinationen von Verfahren vor.

Es zeigt sich, dass Hypothese 2 größtenteils bestätigt werden konnte. Das Opus[®]-Interview schnitt bei vier von fünf Skalen sowie der Globalen Frage und der Entscheidungsfrage signifikant besser ab als die Testverfahren. In den offenen Fragen bemängelten die Teilnehmer den Zeit- und Leistungsdruck der Intelligenz- und Kreativitätstestung sowie den mangelnden Berufsbezug, während wenn sie die Vielfalt der Aufgaben und die Klarheit der Struktur schätzten. Ihre Forderung nach einem persönlichen Gespräch in entspannter Atmosphäre ohne Zeitdruck kam das Opus[®]-Interview nach, auch wenn die Meinungen über die Inhalte der Fallstudie uneinheitlich waren.

3.2.5 Weiterführende Analysen

In einer weiteren Analyse wurden die Ergebnisse zur sozialen Validität nochmals hinsichtlich eventueller Gruppenunterschiede betrachtet. Dabei wurde die Gesamtstichprobe in zwei Gruppen geteilt, um zu überprüfen, ob die Reihenfolge der Verfahren einen Einfluss auf die Bewertung der sozialen Validität hatte. Für die Teilstichprobe, die zuerst die Testverfahren und danach das Opus[®]-Interview durchlaufen hatte, zeigte sich folgendes Ergebnis: Auch hier wurden die Skalen „Akzeptanz und resultierende Motivation“ ($t(15) = 1.90$, $p = .039$), „Testerleben“ ($t(15) = 1.95$, $p = .036$), „Jobbezug“ ($t(15) = 2.29$, $p = .019$) und „Achtung der Privatsphäre“ ($t(15) = 1.85$, $p = .042$) für das Opus[®]-Interview signifikant besser bewertet. Lediglich bei der dritten Skala „Transparenz“ zeigten sich keine signifikanten Unterschiede zwischen den Verfahren ($t(15) = -0.86$, $p = .201$). Auch wurde die Globale Frage in dieser Gruppe nicht signifikant ($t(14) = 1.61$, $p = .066$). Für die Teilstichprobe, die zuerst das Opus[®]-Interview und danach die Testverfahren durchlaufen hatte, zeigte sich folgendes Ergebnis: Die drei Skalen „Akzeptanz und resultierende Motivation“ ($t(14) = 2.95$, $p = .005$), „Testerleben“ ($t(14) = 4.00$, $p = .001$) und „Jobbezug“ ($t(14) = 1.88$, $p = .041$) wurden für das Opus[®]-Interview signifikant besser bewertet. Bei der dritten Skala „Transparenz“ ($t(14) = -1.46$, $p = .084$) und bei der fünften Skala „Achtung der Privatsphäre“ ($t(14) = 0.80$, $p = .219$) zeigten sich keine signifikanten Unterschiede zwischen den Verfahren. Die Testung der Globalen Frage zeigte wieder, dass das Opus[®]-Interview signifikant besser bewertet wurde als die Testverfahren ($t(14) = 3.56$, $p = .002$). Somit zeigte sich, dass ein Teil der gefundenen Ergebnisse zur sozialen Validität unabhängig von der Reihenfolge ist, in der die verschiedenen Verfahren durchlaufen wurden. Trotzdem ergaben sich je nach Reihenfolge geringfügige Unterschiede.

4 Diskussion

Im Diskussionsteil dieser Arbeit werden die gewonnenen Ergebnisse zusammenfassend dargestellt und interpretiert. Ebenso soll die vorliegende Studie einer kritischen Betrachtung unterzogen werden. Abschließend werden Implikationen für die Praxis sowie weitere Forschungsmöglichkeiten aufgezeigt.

4.1 Interpretation der empirischen Ergebnisse

Die Nachfrage nach fundierten Verfahren der Personalauswahl nimmt aufgrund der aktuellen Arbeitsmarktentwicklung stetig zu. Interviews als Personalauswahlverfahren sind sowohl beim Anwender als auch beim Bewerber besonders beliebt (Schuler et al., 2007; Fruhner et al., 1991). Da insbesondere strukturierte Interviews eine sehr gute prognostische Validität aufweisen (Schmidt & Hunter, 1998), wurde in der vorliegenden Arbeit beispielhaft ein strukturiertes Interview (Opus[®]) herausgegriffen, an dem verschiedene Gütekriterien näher untersucht wurden. Der Fokus lag dabei auf der noch wenig untersuchten Konstruktvalidität und auf der seit einigen Jahren an Bedeutung gewinnenden sozialen Validität. Die Studie verfolgte zwei zentrale Fragestellungen: zum einen, was das strukturierte Interview erfasst und zum anderen, ob es beim Bewerber besser ankommt als herkömmliche Testverfahren. Im Folgenden werden die Ergebnisse zur Konstruktvalidität und zur sozialen Validität des Opus[®]-Interviews diskutiert.

Qualität des Fragebogens zur sozialen Validität

Die ursprünglichen neun Faktoren der sozialen Validität, wie sie aus der Literatur abgeleitet wurden, konnten mittels der Hauptkomponentenanalyse nicht repliziert werden. Es wurden schließlich fünf Faktoren extrahiert, die einzelne theoretisch angenommene Skalen klar widerspiegeln, andere jedoch unter einem gemeinsamen Faktor zusammenfassen. So blieben die Skalen „Transparenz“, „Jobbezug“ und „Achtung der Privatsphäre“ erhalten. Die Skala „Transparenz“ entspricht einem der vier Aspekte der sozialen Validität nach

Schuler und Stehle (1985), während die Skala „Jobbezug“ sich bei Bauer et al. (2001) zeigte. Die ursprünglichen Skalen „Partizipation“ (Schuler & Stehle, 1985) und „Affektive Bewertung“ (Smither et al., 1993) wurden zur Skala „Testerleben“ zusammengefasst. Alle übrigen ursprünglichen Skalen resultierten in der Skala „Akzeptanz und resultierende Motivation“. Die zugehörigen Gütekriterien variierten von akzeptabel bis sehr gut und wurden für die nachfolgenden Analysen der Studie als hinreichend betrachtet.

Ergebnisse zur Konstruktvalidität

Wie im Ergebnisteil dargestellt, konnte keine der vier Hypothesen zur Konstruktvalidität bestätigt werden. Die Ergebnisse lassen nur zum Teil auf konvergente Validität des Opus[®]-Interviews schließen. Dafür spricht, dass sich hochsignifikante Korrelationen zwischen Analytischem Denken (Opus[®]) und Analytischem Denken (Selbsteinschätzung) sowie Schlussfolgerndem Denken (WIT-2) zeigen. Ebenso dafür spricht die hochsignifikante Korrelation zwischen Kreativem Denken (Opus[®]) und Kreativem Denken (ASK). Dagegen spricht die mangelnde Korrelation zwischen Kreativem Denken (Opus[®]) und Kreativem Denken (Selbsteinschätzung) sowie die mangelnde Korrelation zwischen Entscheidungsvermögen (Opus[®]) und Entscheidungsvermögen (Selbsteinschätzung). Selbst wenn man diese mangelnden Korrelationen einer schlechten Selbsteinschätzung der Teilnehmer zuschreiben würde, erklärt sich dadurch nicht, warum andere Korrelationen höher ausfallen als zu erwarten war. Betrachtet man nämlich die diskriminante Validität des Opus[®]-Interviews, zeigen sich alle drei Forderungen der Multitrait-Multimethod-Validierung nicht hinreichend erfüllt. Die höchsten Korrelationen des Opus[®]-Interviews zeigen sich auffälligerweise zwischen den drei Opus[®]-Kriterien. Das bedeutet, dass das Opus[®]-Interview nicht klar zwischen den untersuchten Kriterien differenziert. Auch wenn die Konstruktvalidität des Opus[®]-Interviews nicht eindeutig belegt werden konnte, zeigen sich hinsichtlich der eingesetzten Methoden und untersuchten Kriterien verschiedene Auffälligkeiten, die im Folgenden näher beleuchtet werden sollen.

Die Korrelationen der drei Opus[®]-Kriterien lassen auf einen Methodeneffekt schließen. Die Interkorrelationen zwischen diesen drei Kriterien sind hochsignifikant bei starkem Effekt und übertreffen sämtliche andere Korrelationswerte. Das bedeutet, dass Opus[®] diese Kriterien nicht hinreichend differenziert abbildet. Diese Tatsache ist wohl hauptsächlich dem Umstand geschuldet, dass die drei Kriterien im Laufe des Opus[®]-Interviews immer der Reihe nach, Stufe für Stufe, erfasst werden und sich somit keine unabhängigen Werte ergeben können. Bezüglich der Selbsteinschätzung fällt auf, dass sich nur signifikante Korrelationen zwischen dem Analytischen Denken (Selbsteinschätzung) und weiteren Variablen ergeben, jedoch keine Korrelationen innerhalb der Selbsteinschätzungskriterien. Daraus lässt sich eventuell schließen, dass die Teilnehmer ein differenziertes Selbstbild aufweisen, da sie jedes Kriterium separat zu werten scheinen und sich nicht einfach generell in allen Kriterien eher gut oder schlecht einschätzen. Zudem legen die mangelnden Korrelationen zwischen dem Kreativen Denken und Entscheidungsvermögen erfasst durch die Selbsteinschätzung und erfasst durch das Opus[®]-Interview den Schluss nahe, dass die Teilnehmer hinter diesen beiden Kriterien etwas anderes vermutet hatten als das, was durch das Opus[®]-Interview erfasst wurde. Auffällig ist, dass dem Urteil der Teilnehmer nach ihr Analytisches Denken mit allen Interview- und Testvariablen signifikant bis hochsignifikant korreliert. Es liegt die Vermutung nahe, dass die Teilnehmer ein grobes Bild ihrer kognitiven Fähigkeit besitzen, das mit den tatsächlichen Testergebnissen tendenziell übereinstimmt und so etwas wie einen g-Faktor abbildet (Spearman, 1904). Die Korrelationen des Kreativen Denkens (ASK) spiegeln sehr gut die Art der Aufgabenstellung der ASK wider. Zum einen zeigt das Kreative Denken (ASK) den größten Zusammenhang zum Sprachlichen Denken (WIT-2), was zur sprachlichen Darbietung und Anforderung der Aufgaben der ASK passt. Zum anderen sind auch analytische bzw. schlussfolgernde Elemente klar in den Aufgaben der ASK verankert und zeigen dementsprechend hochsignifikante Korrelationen. Die Korrelation des Kreativen Denkens (ASK) mit dem Kreativen Denken (Opus[®]) fällt wohl deshalb nicht am größten aus, weil die ASK nur auf die Anzahl der gefundenen Lösungen abzielt,

während das Opus[®]-Interview auf der qualitativen Wertigkeit der Antworten basiert. Betrachtet man die starken Korrelationen des Schlussfolgernden Denkens (WIT-2) mit den anderen drei Kriterien des WIT-2, lässt sich auch hier ein starker Methodeneffekt, ähnlich dem des Opus[®]-Interviews erkennen. Auch der WIT-2 bildet die unterschiedlichen Kriterien nicht differenziert ab. Die mittleren Korrelationen des Schlussfolgernden Denkens (WIT-2) mit Analytischem Denken (Opus[®]) und Analytischem Denken (Selbsteinschätzung) weisen inhaltlich allerdings in die erwartete Richtung.

Abschließend lässt sich zur Konstruktvalidität sagen, dass die Korrelationen zwar teilweise in die erwartete Richtung weisen, die eingesetzten Verfahren aber letztendlich nicht klar genug differenzieren. Sowohl für das Opus[®]-Interview als auch für den WIT-2 zeigt sich ein starker Methodeneffekt. Die hohen Korrelationen zwischen den Opus[®]-Ergebnissen und den Werten der eingesetzten Testverfahren WIT-2 und ASK weisen darauf hin, dass das Opus[®]-Interview allgemeine kognitive Fähigkeiten misst.

Ergebnisse zur sozialen Validität

Wie im Ergebnisteil dargestellt, konnte Hypothese 2 zum Teil bestätigt werden. Wie bei Hausknecht et al. (2004) und Carless (2006) vermutet, ist das Interviewverfahren bei den Teilnehmern beliebter als die Testverfahren. Die Skalen „Akzeptanz und resultierende Motivation“, „Testerleben“, „Jobbezug“ und „Achtung der Privatsphäre“ wurden von den Teilnehmern für das Opus[®]-Interview signifikant höher bewertet als für die Testverfahren. Ebenso wurde die Globale Frage für das Opus[®]-Interview signifikant höherer beantwortet als für die Testverfahren. Schließlich entschieden sich auch signifikant mehr Teilnehmer bei der Entscheidungsfrage für das Opus[®]-Interview als für die Testverfahren. Lediglich für die dritte Skala „Transparenz“ zeigte sich kein signifikantes Ergebnis, nur eine leichte Tendenz in Richtung der Testverfahren.

Die Analysen für getrennte Gruppen zeigen, dass von den Teilnehmern, die zuerst das Opus[®]-Interview durchlaufen hatten, neben der Skala „Transparenz“ auch die Skala „Achtung der Privatsphäre“ nicht signifikant besser bewertet

wurde. Hinsichtlich der anderen drei Skalen sowie der Globalen Frage wurde aber auch hier das Opus[®]-Interview signifikant besser bewertet. Für die Teilnehmer, die zuerst die Testverfahren durchlaufen hatten, wurde wie bei der Gesamtgruppe die dritte Skala „Transparenz“ nicht signifikant. Auch die Globale Frage verfehlte knapp das Signifikanzlevel. Was sich deutlich abzeichnet ist, dass das Opus[®]-Interview im Bereich „Transparenz“ nicht besser abschneidet als die Testverfahren und sogar eine negative Tendenz aufweist. Eine mögliche Erklärung hierfür wäre, dass das Opus[®]-Interview ein wenig bekanntes Verfahren ist und die Bewertungsmaßstäbe für den Teilnehmer auf den ersten Blick nicht so klar erkennbar sind, wie bei den Testverfahren, die einfach nur auf die Anzahl richtiger Lösungen abzielen. Da die Bewertungsmaßstäbe recht komplex sind und dem Teilnehmer nicht explizit erläutert werden, fühlt dieser sich eventuell nicht umfassend aufgeklärt. Warum die Skala „Achtung der Privatsphäre“ bei der Gruppe, die zuerst das Opus[®]-Interview durchlaufen hat, nicht signifikant besser bewertet wurde, lässt sich nicht genau nachvollziehen. In diesem Fall sind die Mittelwerte der Skala „Achtung der Privatsphäre“ fast gleich hoch. Auch in der Gesamtgruppe zeigt sich hier nur ein kleiner Effekt. Denkbar wäre es, dass die Personen dafür verantwortlich sind, die in den offenen Antworten äußerten, sie hätten Sorge ihre persönliche Einstellung zu den Themen der Fallstudie vor einem Unternehmen preiszugeben.

Da die erste Skala „Akzeptanz und resultierende Motivation“ durchgängig für das Opus[®]-Interview besser bewertet wurde, lässt sich daraus schließen, dass die Akzeptanz unter den Teilnehmern für das Opus[®]-Interview hoch war. Damit einher gehen auch die Testmotivation sowie Verhaltensabsichten, wie ein Jobangebot bei diesem Unternehmen anzunehmen oder das Unternehmen an andere weiterzuempfehlen. Dass die zweite Skala „Testerleben“ durchgängig signifikant besser bewertet wurde zeigt, dass die Teilnehmer das Opus[®]-Interview angemessen fanden und ohne Stress ihre Fähigkeiten zeigen konnten. Auch für die vierte Skala „Jobbezug“ zeigen sich durchgängig signifikant bessere Ergebnisse für das Opus[®]-Interview. Daraus lässt sich schließen, dass die Teilnehmer dem Opus[®]-Interview einen größeren Bezug zum Berufsleben attestierten als den abstrakteren Testverfahren.

Zusammenfassend kann festgehalten werden, dass abgesehen von der dritten Skala „Transparenz“ alle, auch die nicht signifikanten Ergebnisse die erwartete Richtung aufweisen. Insofern lässt sich sagen, dass das Opus[®]-Interview tendenziell besser bewertet wurde als die Testverfahren, was mit den Befunden von Hausknecht et al. (2004) und Carless (2006) übereinstimmt.

4.2 Kritische Reflexion der Studie

Wie jede empirische Studie weist auch die vorliegende Arbeit Stärken und Schwächen auf, die nun kritisch beleuchtet werden sollen. In der vorliegenden Untersuchung wurden verschiedene Verfahren einer Multitrait-Multimethod-Validierung unterzogen, um die Konstruktvalidität des Opus[®]-Interviews zu untersuchen. Es handelt sich hierbei um ein anspruchsvolles Vorgehen. Leider zeigte sich, dass die hier eingesetzten Verfahren, sowohl das Opus[®]-Interview als auch die Intelligenz- und Kreativitätstestung, dem nicht gewachsen waren. Kritisch zu sehen ist, dass es sich um eine nicht vollständige Multitrait-Multimethod-Matrix handelt, da nicht alle Verfahren alle Konstrukte erfassten. Zudem lagen keine Reliabilitätswerte des Opus[®]-Interviews und der Selbsteinschätzung vor. Dementsprechend konnte die Multitrait-Multimethod-Matrix auch nicht ausschöpfend interpretiert werden. In einer weiteren Untersuchung sollte jedes Kriterium anhand mehrerer Verfahren geprüft werden. Dabei wäre z. B. auch zu überprüfen, ob die Anzahl der gefundenen kreativen Lösungen im Opus[®]-Interview mit dem Wert des Kreativen Denkens aus der ASK höher korreliert, als es beim Wert des Kreativen Denkens (Opus[®]) der Fall war. Positiv zu werten ist, die hinreichende Größe der Stichprobe für die Multitrait-Multimethod-Analyse. Es konnten signifikante und teils hochsignifikante Korrelationen beobachtet werden, die sogar einen großen Effekt anzeigen.

Die Stichprobengröße ist allerdings gleichzeitig das größte Manko der Studie, wenn man die soziale Validität betrachtet. Da hier nur 31 Personen in die Analysen eingingen, lassen sich sämtliche Befunde nicht gesichert deuten. Insbesondere muss dabei diskutiert werden, ob die Daten den Einsatz einer Hauptkomponentenanalyse überhaupt rechtfertigten. Die geforderte Mindest-

anzahl von 50 Personen (Tabachnick & Fidell, 2004) wurde ebenso wenig erreicht, wie ein Verhältnis von dreimal so vielen Personen wie Variablen (Wirtz & Nachtigall, 2008). Aufgrund der theoretisch abgeleiteten Skalen hätte man hier eventuell auch eine konfirmatorische Faktorenanalyse durchführen können, um das Neun-Faktoren-Modell zu überprüfen. Allerdings war das aufgrund der Datenlage nicht möglich. Wiederum positiv zu bemerken ist, dass sich bei der Hauptkomponentenanalyse recht klare und zum Teil auch erwartete Faktoren abzeichneten, auch wenn sich keine klare Einfachstruktur zeigte. Zudem lag das jeweilige Cronbachs Alpha der Skalen in einem akzeptablen bis sehr guten Bereich. Auch zeigen sich fast durchweg sehr gute Trennschärfen, was für die Qualität des Fragebogens zur sozialen Validität spricht. Nichtsdestotrotz sollte der Fragebogen dringend an einer größeren Stichprobe überprüft werden. Es gilt die Stabilität der gefundenen Faktorenlösung in weiteren Studien zu untersuchen, bevor man fundierte Schlüsse ziehen kann. Dies ist auch vor dem Hintergrund nötig, dass die Mittelwerte der Items auffallend hoch waren und somit der Fragebogen eventuell nicht über die gesamte Breite hinweg differenziert. Es liegt hier die Vermutung nahe, dass die Teilnehmer dem durchführenden Unternehmen einen Gefallen tun wollten. Wenn diese positive Bewertung allerdings auf „Impression Management“ beruht (Moosbrugger & Kelava, 2007), sind die hier gefundenen Ergebnisse generell in Frage zu stellen und somit nicht aussagekräftig. Trotz aller Kritik soll der vorliegenden Arbeit zugute gehalten werden, dass zumindest versucht wurde, verschiedene Aspekte der sozialen Validität anhand eines deutschsprachigen Fragebogens zu erfassen und damit die soziale Validität des Opus[®]-Interviews mit der sozialen Validität von Intelligenz- und Kreativitätstests zu vergleichen.

Zur Stichprobe lässt sich weiterhin sagen, dass sie im Gegensatz zu vielen vergleichbaren Studien nicht nur auf einer rein studentischen Population basiert. Neben Studenten enthielt die Stichprobe auch viele Arbeitnehmer aus den unterschiedlichsten Arbeitsfeldern. Da sich alle Teilnehmer in oder kurz vor einem Bewerbungsprozess befanden, lassen sich die gefundenen Ergebnisse recht gut auf Bewerber im Allgemeinen übertragen. An dieser Stelle soll auch erwähnt werden, dass der Drop-Out in der Stichprobe zwischen den Personen,

die den Fragebogen zur sozialen Validität ausfüllten und dem Rest hauptsächlich organisatorischer Natur war. Deshalb sollte dabei keine systematische Verzerrung der Stichprobe erfolgt sein. So ist es auch so gut wie auszuschließen, dass nur diejenigen Personen den Fragebogen beantworteten, die die Verfahren als positiv empfanden, auch wenn man das aufgrund der hohen Itemmittelwerte hätte vermuten können.

Zwei weitere organisatorische Komplikationen sollen hier erwähnt werden. Zum einen konnte nicht gewährleistet werden, dass alle Teilnehmer die Verfahren in einer ähnlichen Zeitspanne durchliefen. Während manche Teilnehmer die gesamte Untersuchung in nur einer Woche durchliefen, verteilte sie sich bei anderen über Monate hinweg. Es ist nicht auszuschließen, dass sich das auf die Motivation der Teilnehmer und somit auch auf die Ergebnisse ausgewirkt haben könnte. Zum anderen ergaben sich aus organisatorischen Gründen bei der Intelligenz- und Kreativitätstestung manchmal Einzel- und manchmal Gruppentestungen. Da dies im Vorhinein nicht beabsichtigt war und entsprechend auch nicht als Variable aufgenommen wurde, könnte sich dieser Umstand auf die einzelnen Teilnehmer unterschiedlich ausgewirkt haben. So könnte dieser Umstand im positiven wie im negativen Sinne Einfluss auf die Leistung der Teilnehmer gehabt haben. Zudem könnte er die Bewertung der Intelligenz- und Kreativitätstestung beeinflusst haben. In einer weiteren Untersuchung dieser Art sollte zur Überprüfung die Durchführung des Verfahrens, ob Einzel- oder Gruppentestung, als separate Variable aufgenommen werden.

4.3 Implikationen für Theorie und Praxis

Da sich hinter dem Begriff „strukturiertes Interview“ die unterschiedlichsten Verfahrensformen verbergen können (Chapman & Zweig, 2005), ist die Bezeichnung wenig aussagekräftig. So ist auch der Transfer der hier gefundenen Ergebnisse auf andere Verfahren nicht unproblematisch. Die vorliegenden Ergebnisse zum Opus[®]-Interview ermöglichen eigentlich nur dann Rückschlüsse auf andere strukturierte Interviewverfahren, wenn diese dem

Opus[®]-Interview grundsätzlich ähneln. Eine generelle Übertragung auf strukturierte Interviewverfahren müsste weiterer Überprüfung unterzogen werden. Für das Opus[®]-Interview zeigen sich durchmischte Befunde, die die Konstruktvalidität des Interviews nur bedingt untermauern. Einerseits zeigen sich bei der Multitrait-Multimethod-Validierung hohe Interkorrelationen zwischen den drei Opus[®]-Kriterien, die für einen Methodeneffekt sprechen. Somit bildet das Opus[®]-Verfahren das Analytische Denken, Kreative Denken und Entscheidungsvermögen nicht differenziert ab. Ähnliche Befunde zeigte allerdings auch der renommierte und häufig eingesetzte WIT-2. Andererseits zeigen sich hohe Korrelationen der Opus[®]-Kriterien mit den Testverfahren, woraus man schließen könnte, dass das Opus[®]-Interview so etwas wie „general mental abilities“ erfasst. Da diese der beste Prädiktor für Joberfolg sind (Schmidt & Hunter, 1998), sollte zukünftige Forschung die Kriteriumsvalidität des Opus[®]-Interviews fokussieren. Dies könnte insbesondere hilfreich sein, da die Daten der Studie nahe legen, dass das Opus[®]-Interview eine höhere soziale Validität besitzt als die Testverfahren. Zukünftige Forschung sollte auch die Faktoren der sozialen Validität inhaltlich erneut überprüfen. Wenn der Faktor „Akzeptanz und resultierende Motivation“ replizierbar wäre, könnte das ein wichtiger Anhaltspunkt für die Anwender sein, da mit diesem Faktor eng eine generelle Akzeptanz des Verfahrens einhergeht. Zudem scheint er die zukünftige Verhaltensabsicht des Bewerbers zu beeinflussen, da er unter anderem die Komponenten der „Job Acceptance Intention“, „Organisational Attractiveness“ und „Testmotivation“ enthält. Wie Ajzen (1985, 1991) schon diskutierte, muss eine Verhaltensabsicht nicht unmittelbar in einem entsprechenden Verhalten resultieren. Trotzdem wäre dieser Faktor besonders interessant für Unternehmen. Wenn dieser Faktor Einfluss auf Einstellung und Verhalten des Bewerbers hat, wird der Bewerbungsprozess zur Schnittstelle nach außen. Der Bewerber wird seinen Eindruck dazu nutzen, sein Bild von dem Unternehmen und eventuell auch von den Produkten des Unternehmens zu überdenken. Indem er seine Eindrücke an andere weitergibt und das Unternehmen und seine Produkte gegebenenfalls an andere weiterempfiehlt, wird er zu einem potenziellen Werbeträger. Das Unternehmen könnte

dementsprechend anhand des eingesetzten Auswahlverfahrens über den Bewerber ein positives Image nach außen transportieren. Wenn also über den Faktor „Akzeptanz und resultierende Motivation“ auf die Verhaltensabsicht des Bewerbers Einfluss genommen werden und dies eine Reihe von Folgen für das Unternehmen haben kann, sollte zukünftige Forschung sich der Frage widmen, welche Aspekte eines Verfahrens es genau sind, die diesen Faktor steigern lassen. In Zeiten, in denen der schleichende Fachkräftemangel auf Dauer dazu führen kann, dass fähige Arbeitskräfte zwischen den Unternehmen wählen können, die sich um sie reißen, könnte der Aspekt der sozialen Validität zum entscheidenden Wettbewerbsvorteil werden. Somit würde sich eine neue Dimension von „When testing pays off“ (Brodgen, 1949) eröffnen.

5 Literaturverzeichnis

- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckman (Hrsg.), *Action control: From cognition to behavior*. (S. 11-39). New York: Springer.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Heidelberg: Springer.
- Amelang, M. & Zielinski, W. (1994). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (1999). *I-S-T 2000. Intelligenz-Struktur-Test 2000*. Göttingen: Hogrefe.
- Arvey, R. D. & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology*, 35(2), 281-322.
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P. & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology*, 54(2), 388-420.
- Bortz, J. & Döring, N. (2009). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.

-
- Bortz, J., Lienert, G. A. & Boehnke, K. (2000). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Brodgen, H. E. (1949). When testing pays off. *Personnel Psychology*, 2, 171-183.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.
- Bungard, W. (2000). Mitarbeiterbefragungen als Feedbackinstrument im Rahmen eines systematischen Prozess-Controllings. *Wirtschaftspsychologie*, 7(3), 4-15.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Campion, M. A., Palmer, D. K. & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, 50(3), 655-702.
- Carless, S. A. (2006). Applicant reactions to multiple selection procedures for the police force. *Applied Psychology: An International Review*, 55(2), 145-167.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245-276.

- Cattell, R. B. (1978). *Die empirische Erforschung der Persönlichkeit*. Weinheim: Beltz.
- Chan, D. & Schmitt, N. (2004). An agenda for future research on applicant reactions to selection procedures: A construct-oriented approach. *International Journal of Selection & Assessment*, 12(1), 9-23.
- Chapman, D. S. & Zweig, D. I. (2005). Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. *Personnel Psychology*, 58(3), 673-702.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Conway, J. M. & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology*, 13(4), 485-506.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.

- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302.
- Derous, E., Born, M. P. & Witte, de K. (2004). How applicants want and expect to be treated: Applicants' selection treatment beliefs and the development of the social process questionnaire on selection. *International Journal of Selection & Assessment*, 12(1), 99-119.
- Eid, M. & Diener, E. (2006). *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz.
- Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Fischer, S. (2003). Opus Organisations- und Potenzial- Untersuchungssystem. In J. Erpenbeck & L. von Rosenstiel (Hrsg.), *Handbuch Kompetenzmessung: Erkennen, Verstehen und Bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* (S. 473-487). Stuttgart: Schäffer-Poeschel.
- Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.

- Fruhner, R., Schuler, H., Funke, U. & Moser, K. (1991). Einige Determinanten der Bewertung von Personalauswahlverfahren. *Zeitschrift Für Arbeits- und Organisationspsychologie*, 35, 170-178.
- Ganzach, Y., Kluger, A. N. & Klayman, N. (2000). Making decisions from an interview: Expert measurement and mechanical combination. *Personnel Psychology*, 53(1), 1-20.
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, 18(4), 694-734.
- Görlich, Y. (2007). Arbeitsproben. In H. Schuler & K. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 468-474). Göttingen: Hogrefe.
- Görlich, Y. & Schuler, H. (2006). Personalentscheidungen, Nutzen und Fairness. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 797-840). Göttingen: Hogrefe.
- Gough, H. G. (1979). A creative personality scale for the adjective check list. *Journal of Personality and Social Psychology*, 37(8), 1398-1405.
- Greif, S. (2007). Geschichte der Organisationspsychologie. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 21-57). Bern: Huber.
- Guadagnoli, E. & Velicer, W. F. (1988). Relation to sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265-275.

- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2), 149-161.
- Hausknecht, J. P., Day, D. V. & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57(3), 639-683.
- Havlicek, L. L. & Peterson, L.N. (1977). Effect of the violation of assumptions upon significance levels of the Pearson r . *Psychological Bulletin*, 84(2), 373-377.
- Hell, B. (2007). Messung kognitiver Fähigkeiten. In H. Schuler & K. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 441-458). Göttingen: Hogrefe.
- Höft, S. & Funke, U. (2001). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 135-173). Göttingen: Hogrefe.
- Hornke, L. & Winterfeld, U. (2004). *Eignungsbeurteilungen auf dem Prüfstand: DIN 33430 zur Qualitätssicherung*. Heidelberg: Spektrum.
- Huffcutt, A. I. & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79(2), 184-190.
- Huffcutt, A. I., Conway, J. M., Roth, P. L. & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86(5), 897-913.

- Huffcutt, A. I., Roth, P. L. & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*(5), 459-473.
- Hülshager, U. R., Maier, G. W., Stumpp, T. & Muck, P. M. (2006). Vergleich kriteriumsbezogener Validitäten verschiedener Intelligenztests zur Vorhersage von Ausbildungserfolg in Deutschland. *Zeitschrift für Personalpsychologie, 5*(4), 145-162.
- Jäger A. O. & Althoff K. (1994). *Der WILDE-Intelligenz Test (WIT)*. Göttingen: Hogrefe.
- Jäger, A. O., Süß, H. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test (BIS-Test)*. Göttingen: Hogrefe.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*(3), 187-200.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1), 141-151.
- Kanning, U. P. (2004). *Standards der Personaldiagnostik*. Berlin: Hogrefe.
- Kersting, M., Althoff, K. & Jäger, A. (2008). *Wilde-Intelligenz-Test 2*. Göttingen: Hogrefe.

- Kinnear, P. R. & Gray, C. D. (2010). *PASW statistics 17 made simple*. Hove: Psychology Press.
- Klehe, U., König, C. J., Richter, G. M., Kleinmann, M. & Melchers, K. G. (2008). Transparency in structured interviews: Consequences for construct and criterion-related validity. *Human Performance*, 21(2), 107-137.
- Lance, C. E., Butts, M. M. & Michels, L.C. (2006). The sources of four commonly reported cutoff criteria. *Organizational Research Methods*, 9(2), 202-220.
- Lienert, G. A. & Raatz, U. (1994). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Macan, T. H., Avedon, M. J., Paese, M. & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, 47(4), 715-738.
- Marchese, M. C. & Muchinsky, P. M. (1993). The validity of the employment interview: A meta-analysis. *International Journal of Selection and Assessment*, 1(1), 18-26.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L. & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79(4), 599-616.
- Moosbrugger, H. & Kelava, A. (2007). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.

- Moscoso, S. (2000). Selection interview: A review of validity evidence, adverse impact, and applicant reactions. *International Journal of Selection & Assessment*, 8(4), 237-247.
- Moser, K. (2007). Bewerbungsunterlagen und Referenzen. In H. Schuler & K. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 490-496). Göttingen: Hogrefe.
- Mummendey, H. D. (2003). *Die Fragebogen-Methode: Grundlagen und Anwendungen in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung*. Göttingen: Hogrefe.
- Münsterberg, H. (1912). *Psychologie und Wirtschaftsleben: Ein Beitrag zur angewandten Experimental-Psychologie*. Leipzig: Barth.
- Mussel, P. (2007). Einstellungsinterviews. In H. Schuler & K. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 503-510). Göttingen: Hogrefe.
- Nunnally, J. C. & Bernstein, I. H. (2006). *Psychometric theory*. New York, NY: McGraw-Hill.
- Ostendorf, F. & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae: NEO-PI-R*. Göttingen: Hogrefe.
- Raykov, T. & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis*. New York, NY: Routledge.

- Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in personality and social psychology bulletin. *Personality Social Psychology Bulletin*, 28(12), 1629-1646.
- 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(3), 565-606.
- Salgado, J. F. & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work & Organizational Psychology*, 11(3), 299-324.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.
- Schoppe, K. J. (1975). *Verbaler Kreativitäts-Test*. Göttingen: Verlag für Psychologie.
- Schuler, H. (2001). Teil III: Personalentwicklung. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 213-214). Göttingen: Hogrefe.
- Schuler, H. (2002). *Das Einstellungsinterview*. Göttingen: Hogrefe.
- Schuler, H. (2007). Berufseignungstheorie. In H. Schuler & K. Sonntag (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 429-440). Göttingen: Hogrefe.

- Schuler, H., Frier, D. & Kauffmann, M. (1993). *Personalauswahl im Europäischen Vergleich*. Göttingen: Verlag für Angewandte Psychologie.
- Schuler, H. & Funke, U. (1989). The interview as a multimodal procedure. In R. W. Eder & G. R. Ferris (Hrsg.), *The employment interview: Theory, research and practice* (S. 183-192). Newbury Park, CA: Sage.
- Schuler, H. & Hell, B. (2005). *Analyse des Schlussfolgernden und Kreativen Denkens*. Göttingen: Hogrefe.
- Schuler, H., Hell, B., Trapmann, S., Schaar, H. & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen - ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie*, 6(2), 60-70.
- Schuler, H. & Höft, S. (2001). Konstruktorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 93-133). Göttingen: Hogrefe.
- Schuler, H. & Höft, S. (2007). Diagnose beruflicher Eignung und Leistung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 289-343). Bern: Huber.
- Schuler, H. & Marcus, B. (2001). Biographieorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 175-212). 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.
- Schuler, H. & Stehle, W. (1985). Soziale Validität eignungsdiagnostischer Verfahren: Anforderungen für die Zukunft. In W. Stehle (Hrsg.), *Organisationspsychologie und Unternehmenspraxis: Perspektiven der Kooperation* (S. 133-138). Stuttgart: Verlag für Angewandte Psychologie.
- Shadish, W. R., Cook, T. D. & Donald, T. C. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K. & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, 46(1), 49-76.
- Spearman, C. (1904). "General intelligence", objectively determined and measured. *American Journal of Psychology*, 15(2), 201-293.
- Spearman, C. (1906). 'Footrule' for measuring correlation. *British Journal of Psychology*, 2, 89-108.
- Tabachnick, B. G. & Fidell, L. S. (2004). *Using multivariate statistics*. Boston, MA: Allyn and Bacon.
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, 1.

-
- Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of the vectors of mind*. Chicago, IL: University of Chicago Press.
- Weinert, A. B. (2004). *Organisations- und Personalpsychologie*. Weinheim: Beltz.
- Weiß, R. H. & Cattell, R. B. (1993). *Grundintelligenztest CFT 3 - Skala 3*. Braunschweig: Westermann.
- Winter, de J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *44*(2), 147-181.
- Wirtz, M. A. & Nachtigall, C. (2008). *Statistische Methoden für Psychologen: 1. Deskriptive Statistik*. Weinheim: Juventa.

Anhang

Anhang A: Faktoren der sozialen Validität

Faktor	Item Nr.	Iteminhalt
1: Partizipation	1	Ich hatte das Gefühl, die Situation im Interview in angemessenem Maße mitsteuern zu können.
	2	Im Interview habe ich meine Fähigkeiten zeigen können.
	3	Meine Leistung im Interview war für mich schwer zu beeinflussen.
	4	In dem Verfahren kamen meine Stärken zum Tragen.
	5	Das, was meine Leistungsfähigkeit auszeichnet, konnte ich im Interview unter Beweis stellen.
	6	Das Ergebnis des Verfahrens ist durch den Teilnehmer leicht zu verfälschen.
	47	Das Interview ermöglichte es mir meine Stärken zu zeigen.
2: Transparenz	7	Ich war mir vor dem Verfahren im Klaren darüber, nach welchen Kriterien meine Leistung beurteilt werden würde.
	8	Mir wurden die Inhalte des Interviews verständlich dargelegt.
	9	Die im Vorfeld erhaltenen Informationen über das Interview waren ausreichend.
	10	Mir wurde nachvollziehbar erläutert, was im Laufe des Interviews von mir erwartet wird.
	11	Die Vorgehensweise des Interviews wurde mir verständlich dargelegt.
	12	Ich wurde angemessen über den Verfahrensablauf informiert.
	13	Ich konnte gut abschätzen, was beim Interview auf mich zukommen wird.

Faktor	Item Nr.	Iteminhalt
3: Affektive Bewertung	14	Ich habe das Interview als angenehm erlebt.
	15	Das Verfahren hat mich sehr angestrengt.
	16	Im Interview habe ich mich als Person wertgeschätzt gefühlt.
	17	Ich erlebte den Schwierigkeitsgrad des Interviews als angemessen.
	18	Ich habe mich im Interview überfordert gefühlt.
	19	Ich fühlte mich durch die Situation gestresst.
	20	Der Umgang mit dem Interviewer war angenehm.
4: Job Acceptance Intention	21	An einem nächsten Auswahlstadium würde ich nach diesem Interview weiterhin teilnehmen wollen.
	22	Wenn ich dieses Verfahren im Rahmen eines Bewerbungsprozesses durchlaufen hätte, würde ich ein im Anschluss erhaltenes Stellenangebot der Organisation annehmen.
	23	Das hier verwendete Auswahlverfahren würde meine Entscheidung, ein folgendes Jobangebot anzunehmen, positiv beeinflussen.
5: Organisational Attractiveness	24	Eine Organisation, die dieses Verfahren zur Personalauswahl einsetzt, würde ich an andere interessierte Bewerber weiterempfehlen.
	25	In einem Unternehmen, welches ein solches Auswahlverfahren einsetzt, möchte ich gerne arbeiten.
	26	Wenn ich höre, dass im Unternehmen eine Stelle mittels dieses Verfahrens besetzt werden soll, würde ich mich dafür bewerben.
6: Perceived Predictive Validity	27	Aus meiner Sicht ist das Verfahren ein gutes Instrument, um herauszufinden, ob eine Person geeignet ist für einen Job.
	28	Ich wäre bereit, an diesem Verfahren innerhalb eines Bewerbungsprozesses teilzunehmen.
	29	Ich bin sicher, dass das Verfahren vorhersagen kann, wie gut ein Bewerber auf einer bestimmten Position sein wird.
	30	Das Verfahren ist aus meiner Sicht eine geeignete Methode zur Unterstützung der Personalauswahl.
	31	Personen, die in dem Verfahren besser sind als andere, sind wahrscheinlich auch im Beruf erfolgreicher als andere.
	32	Ich glaube, dass mit diesem Verfahren die passende Person für eine offene Stelle ausgewählt werden kann.
	33	Ich halte den Einsatz dieses Verfahrens als ein Bestandteil in Personalauswahlprozessen für sinnvoll.

Faktor	Item Nr.	Iteminhalt
7: Perceived Job Relatedness	34	Ich erkenne einen klaren Zusammenhang zwischen den Anforderungen im Interview und den Anforderungen im Beruf.
	35	Das Interview erfasste Kompetenzen, die auch im Beruf erforderlich sind.
	36	Das Interview steht in engem Bezug zum wahren Berufsleben und den dort benötigten Fähigkeiten.
	37	Mir ist klar, was das Interview mit meiner Berufstätigkeit zu tun hat.
	38	Ich bin der Überzeugung, dass die den Aufgaben zugrundeliegenden Fähigkeiten im Interview etwas mit den in meinem Beruf benötigten Fähigkeiten zu tun haben.
8: Achtung der Privatsphäre	39	Durch das Interview bzw. die Interview-Ergebnisse musste ich mehr von mir preisgeben, als mir Recht war.
	40	Der Inhalt des Verfahrens war für eine Bewerbungssituation angemessen.
	41	Das Verfahren war diskriminierend.
	42	Insgesamt bin ich der Überzeugung, dass das Verfahren fair war.
9: Testmotivation	43	Ich habe versucht, in diesem Verfahren mein Bestes zu geben.
	44	Während des Interviews habe ich mich stark konzentriert.
	45	Ich war bemüht bei diesem Verfahren gut abzuschneiden.
	46	Die Art der Aufgabenstellung war motivierend.

Anhang B: Fragebogen zur sozialen Validität

Fragebogen Interview

1

Liebe(r) Teilnehmer(in),

Vielen Dank, dass Sie am Interview teilgenommen haben.

Wir möchten Sie darum bitten, Ihre Eindrücke über das Verfahren mit uns zu teilen. In diesem Fragebogen interessieren wir uns für Ihre Eindrücke zum Interview. Ihre Angaben in den Fragebogen werden wir ausschließlich dazu verwenden, das hier eingesetzte Instrument und den Prozess zu verbessern.

Wir versichern Ihnen hiermit, dass die Daten der Fragebogen vollkommen vertraulich behandelt und nicht an Dritte weitergegeben werden. Damit wir die Daten einander zuordnen können, bitten wir Sie, den nachfolgenden Code sorgfältig auszufüllen. Dieser Code ermöglicht es, die Daten zu anonymisieren.

Ihr Code	
Erster Buchstabe des Vornamens der Mutter	___
Zweiter Buchstabe des Vornamens des Vaters	___
Dritter Buchstabe des eigenen Geburtsmonats	___
Letzter Buchstabe des eigenen Geburtsortes	___

Ein Beispiel	
Die Tochter von Luise Mustermann	L
Die Tochter von Hans Mustermann	A
Geb. 14.März	R
in Raubach	H

Datum: __. __. ____

Bitte geben Sie das Ausmaß Ihrer Zustimmung zu den unten folgenden Aussagen an, die sich auf Ihre Eindrücke zum Interview beziehen.

Skala:	①	②	③	④	⑤	⑥
	Trifft gar nicht zu	Trifft überwiegend nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft überwiegend zu	Trifft voll zu

Aussagen zum Interview	Ausmaß ihrer Zustimmung
1. Ich hatte das Gefühl, die Situation im Interview in angemessenem Maße mitsteuern zu können.	① ② ③ ④ ⑤ ⑥
2. Im Interview habe ich meine Fähigkeiten zeigen können.	① ② ③ ④ ⑤ ⑥
3. Meine Leistung im Interview war für mich schwer zu beeinflussen.	① ② ③ ④ ⑤ ⑥
4. In dem Verfahren kamen meine Stärken zum Tragen.	① ② ③ ④ ⑤ ⑥
5. Das, was meine Leistungsfähigkeit auszeichnet, konnte ich im Interview unter Beweis stellen.	① ② ③ ④ ⑤ ⑥
6. Das Ergebnis des Verfahrens ist durch den Teilnehmer leicht zu verfälschen.	① ② ③ ④ ⑤ ⑥
7. Ich war mir vor dem Verfahren im Klaren darüber, nach welchen Kriterien meine Leistung beurteilt werden würde.	① ② ③ ④ ⑤ ⑥
8. Mir wurden die Inhalte des Interviews verständlich dargelegt.	① ② ③ ④ ⑤ ⑥
9. Die im Vorfeld erhaltenen Informationen über das Interview waren ausreichend.	① ② ③ ④ ⑤ ⑥

Skala:	①	②	③	④	⑤	⑥
	Trifft gar nicht zu	Trifft überwiegend nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft überwiegend zu	Trifft voll zu
Aussagen zum Interview	Ausmaß ihrer Zustimmung					
10. Mir wurde nachvollziehbar erläutert, was im Laufe des Interviews von mir erwartet wird.	①	②	③	④	⑤	⑥
11. Die Vorgehensweise des Interviews wurde mir verständlich dargelegt.	①	②	③	④	⑤	⑥
12. Ich wurde angemessen über den Verfahrensablauf informiert.	①	②	③	④	⑤	⑥
13. Ich konnte gut abschätzen, was beim Interview auf mich zukommen wird.	①	②	③	④	⑤	⑥
14. Ich habe das Interview als angenehm erlebt.	①	②	③	④	⑤	⑥
15. Das Verfahren hat mich sehr angestrengt.	①	②	③	④	⑤	⑥
16. Im Interview habe ich mich als Person wertgeschätzt gefühlt.	①	②	③	④	⑤	⑥
17. Ich erlebte den Schwierigkeitsgrad des Interviews als angemessen.	①	②	③	④	⑤	⑥
18. Ich habe mich im Interview überfordert gefühlt.	①	②	③	④	⑤	⑥
19. Ich fühlte mich durch die Situation gestresst.	①	②	③	④	⑤	⑥
20. Der Umgang mit dem Interviewer war angenehm.	①	②	③	④	⑤	⑥
21. An einem nächsten Auswahlschritt würde ich nach diesem Interview weiterhin teilnehmen wollen.	①	②	③	④	⑤	⑥
22. Wenn ich dieses Verfahren im Rahmen eines Bewerbungsprozesses durchlaufen hätte, würde ich ein im Anschluss erhaltenes Stellenangebot der Organisation annehmen.	①	②	③	④	⑤	⑥
23. Das hier verwendete Auswahlverfahren würde meine Entscheidung, ein folgendes Jobangebot anzunehmen, positiv beeinflussen.	①	②	③	④	⑤	⑥
24. Eine Organisation, die dieses Verfahren zur Personalauswahl einsetzt, würde ich an andere interessierte Bewerber weiterempfehlen.	①	②	③	④	⑤	⑥
25. In einem Unternehmen, welches ein solches Auswahlverfahren einsetzt, möchte ich gerne arbeiten.	①	②	③	④	⑤	⑥
26. Wenn ich höre, dass im Unternehmen eine Stelle mittels dieses Verfahrens besetzt werden soll, würde ich mich dafür bewerben.	①	②	③	④	⑤	⑥
27. Aus meiner Sicht ist das Verfahren ein gutes Instrument, um herauszufinden, ob eine Person geeignet ist für einen Job.	①	②	③	④	⑤	⑥
28. Ich wäre bereit, an diesem Verfahren innerhalb eines Bewerbungsprozesses teilzunehmen.	①	②	③	④	⑤	⑥
29. Ich bin sicher, dass das Verfahren vorhersagen kann, wie gut ein Bewerber auf einer bestimmten Position sein wird.	①	②	③	④	⑤	⑥

Skala:	①	②	③	④	⑤	⑥
	Trifft gar nicht zu	Trifft überwiegend nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft überwiegend zu	Trifft voll zu
Aussagen zum Interview					Ausmaß ihrer Zustimmung	
30. Das Verfahren ist aus meiner Sicht eine geeignete Methode zur Unterstützung der Personalauswahl.	①	②	③	④	⑤	⑥
31. Personen, die in dem Verfahren besser sind als andere, sind wahrscheinlich auch im Beruf erfolgreicher als andere.	①	②	③	④	⑤	⑥
32. Ich glaube, dass mit diesem Verfahren die passende Person für eine offene Stelle ausgewählt werden kann.	①	②	③	④	⑤	⑥
33. Ich halte den Einsatz dieses Verfahrens als ein Bestandteil in Personalauswahlauswahlprozessen für sinnvoll.	①	②	③	④	⑤	⑥
34. Ich erkenne einen klaren Zusammenhang zwischen den Anforderungen im Interview und den Anforderungen im Beruf.	①	②	③	④	⑤	⑥
35. Das Interview erfasste Kompetenzen, die auch im Beruf erforderlich sind.	①	②	③	④	⑤	⑥
36. Das Interview steht in engem Bezug zum wahren Berufsleben und den dort benötigten Fähigkeiten.	①	②	③	④	⑤	⑥
37. Mir ist klar, was das Interview mit meiner Berufstätigkeit zu tun hat.	①	②	③	④	⑤	⑥
38. Ich bin der Überzeugung, dass die den Aufgaben zugrundeliegenden Fähigkeiten im Interview etwas mit den in meinem Beruf benötigten Fähigkeiten zu tun haben.	①	②	③	④	⑤	⑥
39. Durch das Interview bzw. die Interview-Ergebnisse musste ich mehr von mir preisgeben, als mir Recht war.	①	②	③	④	⑤	⑥
40. Der Inhalt des Verfahrens war für eine Bewerbungssituation angemessen.	①	②	③	④	⑤	⑥
41. Das Verfahren war diskriminierend.	①	②	③	④	⑤	⑥
42. Insgesamt bin ich der Überzeugung, dass das Verfahren fair war.	①	②	③	④	⑤	⑥
43. Ich habe versucht, in diesem Verfahren mein Bestes zu geben.	①	②	③	④	⑤	⑥
44. Während des Interviews habe ich mich stark konzentriert.	①	②	③	④	⑤	⑥
45. Ich war bemüht bei diesem Verfahren gut abzuschneiden.	①	②	③	④	⑤	⑥
46. Die Art der Aufgabenstellung war motivierend.	①	②	③	④	⑤	⑥
47. Das Interview ermöglichte es mir meine Stärken zu zeigen.	①	②	③	④	⑤	⑥

Offene Fragen
48. „Welche Faktoren haben es mir ermöglicht, meine Fähigkeiten im Interview zu zeigen?“ _____
49. „Aus welchem Grund konnte ich meine Fähigkeiten im Interview nicht zeigen?“ _____
50. „Was hätte mir geholfen, meine Fähigkeiten besser zeigen zu können?“ _____

Aussagen zum Interview (bitte ankreuzen)						
51. Zusammenfassend betrachtet: Welchen Eindruck hat das Durchlaufen des Interviews bei mir hinterlassen?						
Skala:	①	②	③	④	⑤	⑥
	Sehr negativ	Überwiegend negativ	Eher negativ	Eher positiv	Überwiegend positiv	Sehr positiv

Vielen Dank!

Anhang C: Mittelwert und Standardabweichung der Items

Faktor 1: Partizipation

	<i>Item 1</i>	<i>Item 2</i>	<i>Item 3</i>	<i>Item 4</i>	<i>Item 5</i>	<i>Item 6</i>	<i>Item 7</i>
<i>MEAN</i>	5,13	4,50	4,58	4,26	4,16	2,20	4,58
<i>SD</i>	0,72	0,90	1,29	0,77	0,97	1,03	0,77

Faktor 2: Transparenz

	<i>Item 7</i>	<i>Item 8</i>	<i>Item 9</i>	<i>Item 10</i>	<i>Item 11</i>	<i>Item 12</i>	<i>Item 13</i>
<i>MEAN</i>	3,29	5,55	5,61	5,32	5,68	5,61	4,42
<i>SD</i>	1,47	0,62	0,62	0,87	0,54	0,80	1,34

Faktor 3: Affektive Bewertung

	<i>Item 14</i>	<i>Item 15</i>	<i>Item 16</i>	<i>Item 17</i>	<i>Item 18</i>	<i>Item 19</i>	<i>Item 20</i>
<i>MEAN</i>	5,29	4,26	5,39	5,23	5,03	4,74	5,74
<i>SD</i>	0,78	1,15	0,88	0,82	0,98	1,09	0,51

Faktor 4: Job Acceptance Intention

	<i>Item 21</i>	<i>Item 22</i>	<i>Item 23</i>
<i>MEAN</i>	5,48	5,20	4,74
<i>SD</i>	0,85	1,03	1,39

Faktor 5: Organisational Attractiveness

	<i>Item 24</i>	<i>Item 25</i>	<i>Item 26</i>
<i>MEAN</i>	4,87	4,97	4,86
<i>SD</i>	1,01	0,94	1,03

Faktor 6: Perceived Predictive Validity

	<i>Item 27</i>	<i>Item 28</i>	<i>Item 29</i>	<i>Item 30</i>	<i>Item 31</i>	<i>Item 32</i>	<i>Item 33</i>
<i>MEAN</i>	4,37	5,71	4,23	5,03	4,20	4,40	4,94
<i>SD</i>	0,85	0,59	0,90	0,85	1,16	0,86	0,89

Faktor 7: Perceived Job Relatedness

	<i>Item 34</i>	<i>Item 35</i>	<i>Item 35</i>	<i>Item 36</i>	<i>Item 37</i>	<i>Item 38</i>
<i>MEAN</i>	4,39	5,00	4,23	4,84	4,87	1,58
<i>SD</i>	1,23	0,93	0,99	1,00	0,76	0,92

Faktor 8: Achtung der Privatsphäre

	<i>Item 39</i>	<i>Item 40</i>	<i>Item 41</i>	<i>Item 42</i>
<i>MEAN</i>	5,42	4,97	5,84	5,33
<i>SD</i>	0,92	0,91	0,45	0,71

Faktor 9: Testmotivation

	<i>Item 43</i>	<i>Item 44</i>	<i>Item 45</i>	<i>Item 46</i>
<i>MEAN</i>	5,65	5,10	5,35	5,03
<i>SD</i>	0,49	0,87	0,88	0,91

Anhang D: Trennschärpen und Cronbachs Alpha der neun Faktoren

Skala	Item	Cronbachs α der Skala	Trennschärfe r_{it}
1: Partizipation	Item 1	.64	-.09
	Item 2		.63
	Item 3		.53
	Item 4		.77
	Item 5		.49
	Item 6		-.27
	Item 47		.78
2: Transparenz	Item 7	.69	.23
	Item 8		.68
	Item 9		.12
	Item 10		.53
	Item 11		.40
	Item 12		.44
	Item 13		.68
3: Affektive Bewertung	Item 14	.78	.62
	Item 15		.40
	Item 16		.21
	Item 17		.61
	Item 18		.68
	Item 19		.70
	Item 20		.41
4: Job Acceptance Intention	Item 21	.82	.58
	Item 22		.82
	Item 23		.71
5: Organisational Attractiveness	Item 24	.90	.80
	Item 25		.85
	Item 26		.74

Skala	Item	Cronbachs α der Skala	Trennschärfe r_{it}
6: Perceived Predictive Validity	Item 27	.84	.58
	Item 28		.56
	Item 29		.58
	Item 30		.65
	Item 31		.52
	Item 32		.71
	Item 33		.65
7: Perceived Job Relatedness	Item 34	.85	.46
	Item 35		.78
	Item 36		.76
	Item 37		.75
	Item 38		.67
8: Achtung vor Privatsphäre	Item 39	.74	.67
	Item 40		.43
	Item 41		.63
	Item 42		.56
9: Testmotivation	Item 43	.84	.70
	Item 44		.75
	Item 45		.66
	Item 46		.72

Eidesstattliche Erklärung

Hiermit erkläre ich ehrenwörtlich,

1. dass ich die vorliegende Arbeit ohne fremde Hilfe angefertigt habe,
2. dass ich die Übernahme wörtlicher Zitate aus der Literatur sowie die Verwendung der Gedanken anderer Autoren an den entsprechenden Stellen der Arbeit gekennzeichnet habe.

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen für mich haben kann.

Mannheim, den 14.06.11

Unterschrift

Freiwillige Erklärung

Ich stimme ausdrücklich zu, dass meine durch Tammo Straatmann betreute Diplomarbeit mit dem Titel:

„Aspekte der Konstruktvalidität und der sozialen Validität von strukturierten Interviews am Beispiel des strukturierten Interviews Opus®“

nach Beendigung der Diplom-Prüfung zu wissenschaftlichen Zwecken im Bereich der Fakultätsbibliothek PE aufgestellt und zugänglich gemacht wird (Veröffentlichung nach §§ 6 Abs. 1 UrhG), und hieraus im Namen des § 51 UrhG zitiert werden kann. Sämtliche Verwendungsrechte nach §15 UrhG verbleiben beim Verfasser der Diplomarbeit.

Unterschrift

Klomann, Sarah

Schorlenberger Straße 32

67677 Enkenbach-Alsenborn

Mannheim, den 14.06.11