
**Inauguraldissertation
zur Erlangung des akademischen Doktorgrades (Dr. phil.)
im Fach Psychologie
an der Fakultät für Verhaltens- und Empirische Kulturwissenschaften
der Ruprecht-Karls-Universität Heidelberg**

Titel der Dissertation
Entwicklung und Validierung eines Situational Judgement Tests (SJT)
zur Erfassung sozialer Kompetenzen von Studienplatzbewerbern
und -interessenten der Human- und Zahnmedizin

vorgelegt von
Dipl.-Psych. Janine Kahmann

Jahr der Einreichung
2014

Dekan: Prof. Dr. Klaus Fiedler
Beraterinnen: apl.-Prof. Dr. med. Martina Kadmon, MME (Bern)
Prof. Dr. Birgit Spinath

Danksagung

An erster Stelle möchte ich Frau apl.-Prof. Dr. med. Martina Kadmon für ihre großartige Betreuung und ihr Vertrauen danken. Sie ermöglichte es mir, zu diesem Thema zu promovieren und gab viele wertvolle Hinweise und Ideen. Sie stärkte mir stets den Rücken und verhalf mir, die richtigen Ansprechpartner innerhalb der Fakultät zu finden. Für die Übernahme der Zweitbegutachtung möchte ich Frau Prof. Dr. Birgit Spinath danken, die mir als methodische Ratgeberin zur Seite stand und deren Doktoranden-seminare ich besuchen durfte. Für die wertvollen Diskussionen im Büro danke ich meinem Kollegen Gerald Wibbecke, der mir in Problemsituationen zu neuen Ansätzen und Lösungen verholfen hat. Ich möchte allen Teilnehmern¹ des Workshops danken, die ihre Ideen in den *SJT* einfließen ließen. Auch den Schülern- und Schülerinnen des sozialwissenschaftlichen Gymnasiums im Heidelberger Privatschul Centrum danke ich für das Vortesten der schriftlichen Items und ihren kreativen Vorschlägen zur Überarbeitung der Antwortoptionen. Des Weiteren möchte ich allen fakultätsinternen und -externen Ärzten und Psychologen danken, die sich an den Fragebogenuntersuchungen beteiligt haben. Der nächste Dank gilt dem Filmteam des Action House, die mir mit ihren Ideen zur filmischen Umsetzung des Drehbuches und mit ihrem fachlichen und technischen Knowhow zur Seite standen. Ein großer Dank geht an die 78 freiwilligen Schauspieler, die ihre Rolle mit viel Engagement gespielt und viel Zeit für die Verfilmung investiert haben. Danke auch an alle Personen, die es mir ermöglicht haben, die Dreharbeiten in ihren Räumlichkeiten durchzuführen. Für die Unterstützung im Hintergrund danke ich meiner Freundin Nadia El-Ali, die dafür sorgte, dass sich die Workshop-Teilnehmer und Schauspieler gut gepflegt gefühlt haben. Für die technische Umsetzung des *SJTs* danke ich vielmals unserem Informatiker Lukas Jurkowski und Herrn Goldecker von q-set, die das Online-System für die Studienplatzbewerber funktionsfähig gemacht haben. Vielen Dank auch an alle Korrekturleser dieses Manuskripts. Abschließend möchte ich meiner Familie, meinen Freunden und meinen WG-Mitbewohnern danken, die mich motivational und emotional über die Jahre der Doktorarbeit unterstützt haben und immer für mich da waren.

¹ Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung der weiblichen und männlichen Sprachform verzichtet. Alle Personenbezeichnungen gelten für beide Geschlechter.

Inhaltsverzeichnis

1.	Einleitung und Zielsetzung der Arbeit	9
2.	Die Auswahl von Medizinstudierenden	12
2.1.	Hintergründe der medizinischen Studierendenauswahl.....	12
2.2.	Forschungsergebnisse zur Medizinischen Studierendenauswahl.....	17
2.2.1.	Kognitive Auswahlkriterien	17
2.2.2.	Nicht-kognitive Auswahlkriterien.....	22
2.3.	Studierendenauswahl an der Medizinischen Fakultät Heidelberg.....	25
2.4.	Schlussfolgerungen für die Medizinische Studierendenauswahl	26
3.	Soziale Kompetenzen.....	29
3.1.	Definition sozialer Kompetenzen.....	29
3.2.	Modelle sozialer Kompetenzen	33
3.3.	Soziale Kompetenzen in der Medizin	38
3.4.	Arbeitsmodell sozialer Kompetenzen	40
3.5.	Messung sozialer Kompetenzen	42
4.	Der Situational Judgement Test.....	44
4.1.	Ursprung, Aufbau und Funktionsweise.....	44
4.2.	Was messen SJTs?	46
4.3.	Entwicklung eines SJTs	49
4.4.	Gütekriterien des SJT	57
4.4.1.	Objektivität	57
4.4.2.	Reliabilität	57
4.4.3.	Validität	59
4.5.	Vor- und Nachteile des SJT	62
4.6.	SJTs in der medizinischen Studierenden- und Facharztauswahl.....	63
5.	Schlussfolgerung für die SJT-Entwicklung an der Medizinischen Fakultät Heidelberg	65
6.	Die Entwicklung des SJTs der Medizinischen Fakultät Heidelberg	66
6.1.	Anforderungsanalyse	66
6.1.1.	Forschungsfragen	67
6.1.2.	Eingesetzte Verfahren und Untersuchungsvariablen.....	67
6.1.3.	Stichprobenrekrutierung und -beschreibung.....	68
6.1.4.	Statistische Auswertung	70
6.1.5.	Ergebnisse	72
6.1.6.	Diskussion.....	79
6.2.	Workshops zur Itementwicklung.....	82

6.2.1.	Generieren der Situationen	83
6.2.2.	Generieren der Verhaltensweisen	84
6.2.3.	Limitationen.....	86
6.3.	Vortestung der Items.....	87
6.3.1.	Stichprobenbeschreibung.....	87
6.3.2.	Methodik	87
6.3.3.	Ergebnisse	89
6.3.4.	Ergebnisdiskussion und Itemüberarbeitung.....	92
6.4.	Bildung des Auswertungsschlüssels	94
6.4.1.	Stichprobenrekrutierung und -beschreibung	94
6.4.2.	Untersuchungsablauf	96
6.4.3.	Statistische Auswertung	96
6.4.4.	Ergebnisse	97
6.4.5.	Diskussion der sozialen Angemessenheit der Antwortoptionen.....	99
6.4.6.	Punktbildung	100
6.4.7.	Limitationen.....	103
6.5.	Verfilmen der Items.....	104
7.	Validierung an medizinischen Studienplatzbewerbern und -interessenten.....	107
7.1.	Hypothesen.....	107
7.2.	Untersuchungsablauf und eingesetzte Verfahren.....	109
7.3.	Stichprobenrekrutierung und -beschreibung.....	112
7.4.	Statistische Auswertung.....	114
7.5.	Ergebnisse.....	117
7.5.1.	Ergebnisse auf Itemebene	117
7.5.2.	Ergebnisse auf Skalenebene.....	136
7.6.	Diskussion	145
8.	Validierung an Medizinstudierenden	152
8.1.	Hypothesen.....	152
8.2.	Untersuchungsablauf und eingesetzte Verfahren.....	156
8.3.	Stichprobenrekrutierung und -beschreibung.....	162
8.4.	Statistische Auswertung.....	163
8.5.	Ergebnisse.....	165
8.5.1.	Deskriptive Statistik und vorausgegangene Analysen	165
8.5.2.	konvergente und diskriminante Validität des SJTs	178
8.5.3.	Prädiktive und inkrementelle Validität des SJTs	185
8.6.	Diskussion	188
9.	Studienübergreifende Zusammenfassung der Ergebnisse.....	199

10. Implikationen für Forschung und Praxis	202
Literaturverzeichnis	206
Tabellenverzeichnis.....	233
Abbildungsverzeichnis.....	236
Abkürzungsverzeichnis	237
Anhang.....	241

1. Einleitung und Zielsetzung der Arbeit

Der Wunsch nicht-kognitive Faktoren in Selektionsprozessen für Studierende mit zu berücksichtigen, besteht bei den medizinischen Fakultäten und der Öffentlichkeit bereits seit den 1970er Jahren (Benbassat & Baupal, 2007; Bucksch-Beudt et al., 2006) und verliert auch heute nicht an Aktualität (Powis, 2010). Dies kann zum einen auf das siebente Gesetz zur Änderung des Hochschulrahmengesetzes (28.08.2004) zurückgeführt werden. Dieses fordert von den Hochschulen Baden-Württembergs, neben der Abiturdurchschnittsnote zusätzliche Auswahlkriterien für zulassungsbeschränkte Studiengänge zu berücksichtigen. Zum anderen tragen wissenschaftliche Erkenntnisse zur Integration nicht-kognitiver Auswahlverfahren bei. Die Vorhersageleistung kognitiver Kriterien für den Studienerfolg nimmt im Verlauf des Medizinstudiums ab, während sie für nicht-kognitive Kriterien ansteigt (Eva et al., 2009; Eva et al., 2012; Hojat et al., 1993; Kulatunga-Moruzi & Norman, 2002; Lievens, 2013). Zudem schränkt eine rein kognitiv orientierte Auswahl die sozioökonomische Vielfalt unter den Ausgewählten ein (Powis, 2010; Prideaux et al., 2011). Aus emotionaler Perspektive geht das Anliegen vermutlich auf die Befürchtung zurück, dass die alleinige Beschränkung auf kognitive Leistungskriterien bei der medizinischen Studierendenauswahl, wenig emotionale Persönlichkeitstypen bevorzugen könnte (Hampe et al., 2009). Darüber hinaus werden Persönlichkeitsmerkmale als aussagekräftig für den langfristigen Studien- und Berufserfolg angesehen (Hossiep, Paschen & Mühlhaus, 2000; Schmidt & Hunter, 1998). Die Konzentration auf kognitive Auswahlkriterien allein gewährleistet außerdem keine ausreichende Passung zwischen den Studiengangprofilen und den Bewerbern (Wenzel, 2008; Wissenschaftsrat, 2004). Auch stellen falsche und unklare Studierenerwartungen, unzureichende Information über das Studium, sowie mangelnde Fachidentifikation häufige Gründe für einen Studienabbruch dar (Gold & Souvignier, 2005; Heublein, Spangenberg & Sommer, 2003). Die Deutsche Gesellschaft für Psychologie e. V. schlägt daher in ihrer Stellungnahme vom 22. November 2004 vor, Studierfähigkeitstests zu entwickeln, die sowohl allgemeine als auch fachspezifische Komponenten beinhalten. Darüber hinaus sollten beratende *Self-Assessments* angeboten werden, die die Auseinandersetzung der Studienplatzinteressenten mit dem Studienangebot und den Anforderungen fördern, um eine systematische Studienwahlentscheidung zu unterstützen. Die Herausforderung für die medizinische Studierendenauswahl liegt nun darin, reliable und valide Verfahren zur Erhebung nicht-kognitiver Kriterien zu entwickeln und zu erforschen (Benbassat & Baupal, 2007). Gleiche Autoren empfehlen daher, nicht-kognitive Kriterien bislang lediglich für die Selbstselektion von Bewerbern einzusetzen.

Für eine adäquate Patientenversorgung und -sicherheit nehmen soziale Kompetenzen einen hohen Stellenwert in der Medizin ein (Crushing et al., 2014; Gaba, 2000; Kohn, Corrigan & Donaldson, 2000). Ihre Relevanz bildet sich in der zunehmenden Aufnahme in medizinische Lernzielkatalogen ab (z.B. Frank et al., Entwurf 2015; Kiessling et al., 2010; Makoul, 2001; von Fragstein et al., 2008). Auch das *Heidelberger Curriculum Medicinale (HeiCuMed)* zielt mit seinen aktiven Lehr-Lernmethoden wie dem Medizinischen Kommunikations- und Interaktionstraining (MediKIT), dem Problemorientierten Lernen (POL) und dem hohen Anteil an Patientenkontakt nicht nur auf die Erlangung von fachlichen Fähigkeiten, sondern auch auf die Förderung von sozialen Kompetenzen wie Kommunikationsfähigkeit und Empathie ab.

Ein Instrument zur Erfassung sozialer Kompetenzen mit Potenzial zur Vorhersage des klinischen Studienerfolgs ist der *Situational Judgement Test (SJT)*; Lievens, 2013; Patterson, Baron, Carr, Plint & Lane, 2009). Der *SJT* ist ein in der Personalpsychologie eingesetztes Auswahlverfahren, welches Bewerbern arbeitsbezogene Situationen in mündlicher, schriftlicher oder videobasierter Form präsentiert, um eine möglichst realistische Vorschau auf die Arbeitstätigkeit zu geben (Motowidlo, Dunnette & Carter, 1990; Weekley & Ployhart, 2006). Anschließend werden die Bewerber aufgefordert, anzugeben, wie sie in dieser Situation reagieren würden. Somit besitzt der *SJT* sowohl eine informierende als auch testende Funktion.

Der Ansatz dieser Dissertation besteht nun darin, das kognitiv geprägte Auswahlverfahren der Medizinischen Fakultät Heidelberg durch ein ressourcensparendes nicht-kognitives Selbstselektionsinstrument zu ergänzen. Dafür soll ein videobasierter *SJT* für medizinische Studienplatzbewerber und -interessenten entwickelt und validiert werden. Dieser soll die Studienplatzbewerber und -interessenten über die Anforderungen an soziale Kompetenzen im Medizinstudium informieren und dafür sensibilisieren. Auch soll er zum Reflektieren des eigenen sozialen Verhaltens anregen und als systematische Entscheidungshilfe für oder gegen ein Medizinstudium dienen, um die Passung zwischen Individuum und Studium zu erhöhen. Zudem soll er in der Lage sein, soziale Kompetenzen der Studienplatzbewerber und -interessenten zu erfassen. Neben diesem praktisch-medizinischen Handlungsbezug soll die Dissertation einen methodischen Beitrag zur *SJT*-Forschung leisten, indem die Faktorenstruktur, die Testfairness, die Konstruktvalidität, sowie prädiktive und inkrementelle Validität an Bewerbern und Stelleninhabern untersucht werden. Ferner soll eine Instruktionsform und ein Auswertungsschlüssel für *SJTs* mit mehreren Verhaltensoptionen gleicher Effektivität entwickelt werden.

Die Arbeit unterteilt sich in drei Studien, die sich der Entwicklung des *SJTs* widmen und in zwei, die die Validierung des *SJTs* beschreiben. Die erste Studie berichtet von

der Anforderungsanalyse zur Festlegung der wichtigsten sozialen Kompetenzen in der Medizin. Die zweite Studie schildert die Vortestung der *SJT*-Items an Gymnasiasten. Die dritte Studie beschreibt die Befragung von Medizinerinnen und Psychologen zur Bildung des Auswertungsschlüssels. Die erste Validierungsstudie geht anhand zweier Stichproben von Studienplatzbewerbern und -interessenten auf die Faktorenstruktur, den Einfluss von Geschlecht und Alter, sowie die Nützlichkeit und Akzeptanz des *SJT*s ein. Die zweite Validierungsstudie an Medizinstudierenden fokussiert stärker auf die Konstruktvalidierung des *SJT*s. Sie untersucht den Zusammenhang zwischen dem *SJT* und sozialen Kompetenzkriterien, der Intelligenz, der Konzentrationsfähigkeit und der Persönlichkeit. Darüber hinaus wird die Testfairness hinsichtlich des Geschlechtes, des Alters, Deutsch als Erstsprache, dem Textverständnis und einer abgeschlossenen medizinischen Ausbildung erforscht. Abschließend wird die prädiktive Validität des *SJT*s für den subjektiven und objektiven Studienerfolg analysiert.

Inhaltlich gliedert sich die Arbeit in acht Kapitel. Der Einleitung folgen Kapitel 2-5, die die theoretischen Säulen der Arbeit bilden. Kapitel 2 geht auf die rechtlichen, politischen, organisatorischen und wissenschaftlichen Hintergründe der medizinischen Studierendenauswahl ein. Es stellt den aktuellen Forschungsstand zu kognitiven und nicht-kognitiven Auswahlkriterien der Medizin dar und erläutert die Entwicklung des hochschulinternen Auswahlverfahrens in Heidelberg. Kapitel 3 führt in Definitionen, Modelle und Messmethoden sozialer Kompetenzen ein. Zudem geht es auf soziale Kompetenzen ein, die in der Medizin zum Tragen kommen. Die allgemeinen, sowie medizinspezifischen Ansätze werden anschließend in ein Arbeitsmodell sozialer Kompetenzen integriert. Das vierte Kapitel ist der Methode des *SJT*s gewidmet. Es beschreibt dessen Funktionsweise, messmethodische Ansätze, Gütekriterien sowie Vor- und Nachteile. Auch wird auf die Entwicklung eines *SJT*s eingegangen und dessen derzeitige Verwendung in der medizinischen Studierenden- und Facharztauswahl aufgezeigt. Im Kapitel 5 werden Schlussfolgerungen für die *SJT*-Entwicklung an der Medizinischen Fakultät Heidelberg aufgrund der theoretischen Kapitel gezogen. Das sechste Kapitel beschreibt die Entwicklung des eigenen *SJT*s, die sich von einer Anforderungsanalyse und Entwicklungs-Workshops, über eine Vortestung und Bildung des Auswertungsschlüssels bis hin zum Verfilmen der Items erstreckt. Kapitel 7 und 8 erläutern das Vorgehen, die Ergebnisse und Schlussfolgerungen der zwei durchgeführten Validierungsstudien. Kapitel 9 rundet die Arbeit mit Implikationen für Forschung und Praxis ab.

2. Die Auswahl von Medizinstudierenden

Die folgenden Kapitel beschreiben den theoretischen Hintergrund der Studierendenauswahl, auf den die Dissertation aufbaut. Kapitel 2.1 geht zunächst auf die rechtlichen und politischen Hintergründe der medizinischen Studierendenauswahl ein, bevor die theoretische Basis für die Auswahl von Prädiktoren und Kriterien des Studienerfolgs erläutert wird. Kapitel 2.2 beschreibt anschließend den Forschungsstand zu kognitiven und nicht-kognitiven Auswahlkriterien des Medizinstudiums. In Kapitel 2.3 wird die Studierendenauswahl der Medizinischen Fakultät Heidelberg vorgestellt. Kapitel 2.4 stellt abschließend generelle und heidelbergsspezifische Schlussfolgerungen für die medizinische Studierendenauswahl dar. Die Darlegungen zur medizinischen Studierendenauswahl bilden sich in der Vorgehensweise der Arbeit (Kapitel 5) und in der Studienerfolgsmessung der Validierungsstudie an Medizinstudierenden ab (Kapitel 8).

2.1. Hintergründe der medizinischen Studierendenauswahl

Rechtliche und politische Hintergründe. Laut Grundgesetz ist jeder Deutsche berechtigt ein von ihm gewähltes Hochschulstudium absolvieren zu dürfen, sofern die erforderlichen Qualifikationen zum Studium vorliegen (Art. 116). Überschreitet jedoch die Nachfrage die verfügbaren Studienplätze, darf eine Zulassungsbeschränkung für den Studiengang (Numerus-Clausus) beantragt werden (BVerfGE 33,303 vom 18. Juli 1972). Seit 1973 werden die Studienplätze der Human- und Zahnmedizin zentral über die Zentralstelle für Vergabe von Studienplätzen vergeben (ZVS), deren heutiger Name Stiftung für Hochschulzulassung ist. Sechzig Prozent der Studienplätze wurden über die Hochschulzulassungsberechtigung (HZB) und 40% Prozent über die Wartezeitquote besetzt. Anfang der 1970er Jahre stieg die Studienplatzbewerberanzahl stark an, sodass Bewerber mit Abiturnotendurchschnitten bis 1.7 noch einen Platz erhielten, während Bewerber mit einem Abiturnotendurchschnitt von 1.8 bis zu sieben Jahre auf einen Studienplatz warten mussten (Trost et al., 1998). Dies veranlasste dazu, zusätzliche Kriterien in die medizinische Studierendenauswahl zu integrieren. Seit 1976 wurden neben der Abiturnote, gewichteten Einzelnoten und einer Berufsausbildung/Berufstätigkeit, auch fachspezifische Studierfähigkeitstests, Auswahlgespräche oder eine Kombination obiger Kriterien für die Auswahl als zulässig erklärt (HRG vom 26.01.1976, §32). Daraufhin wurde der Test für Medizinische Studiengänge (TMS) vom Institut für Test- und Begabungsforschung (ITB) bundesweit eingeführt. Der TMS ist ein spezifischer Studierfähigkeitstest für medizinische Studiengänge, der fachnah konzipiert ist, aber kein spezifisches medizinisches Wissen voraussetzt. Er wurde zum ersten Mal in den 1980er und 1990er Jahren und wird aktuell erneut seit 2007 als Auswahlinstrument eingesetzt und einmal pro Jahr deutschlandweit unter standardisierten

Bedingungen durchgeführt. Er besteht aus neun Aufgabengruppen von insgesamt 307 Minuten Dauer. Eine kurze Beschreibung der Aufgabengruppen enthält Tabelle 1.

Tabelle 1: *Aufgabengruppen des TMS*

Aufgabengruppe	Testung von:	Itemanzahl
Muster zuordnen	Fähigkeit, Ausschnitte in einem komplexen Bild wieder zu erkennen	24
Medizinisch-naturwissenschaftliches Grundverständnis	Verständnis für Fragen der Medizin und Naturwissenschaften	24
Schlauchfiguren	Fähigkeit, räumlich zu denken	24
Quantitative und formale Probleme	Fähigkeit, im Rahmen medizinischer und naturwissenschaftlicher Fragestellungen mit Zahlen, Größen, Einheiten und Formeln richtig umzugehen	24
Konzentriertes und sorgfältiges Arbeiten	Fähigkeit, rasch, konzentriert und sorgfältig zu arbeiten	
Merkfähigkeit (Einprägen)	Fähigkeit, sich visuelle Einzelheiten von Gegenständen bzw. schriftliche Fakten einzuprägen	20
Figuren lernen		15
Fakten lernen		15
Textverständnis	Fähigkeit, umfangreiches und komplexes Textmaterial zu bearbeiten	24
Merkfähigkeit (Reproduktion)	Fähigkeit, visuelle Einzelheiten von Gegenständen bzw. schriftliche Fakten zu reproduzieren	5
Figuren lernen		7
Fakten lernen		7
Diagramme und Tabellen	Fähigkeit, Diagramme und Tabellen richtig zu analysieren und interpretieren	24

Anmerkung. Quelle: <http://www.tms-info.org>. Zugriff am 26.05.2014.

Die Studienplätze wurden nach der initialen Einführung des TMS zu 45% nach der Kombination von Abiturnotendurchschnitt und TMS, zu 10% allein nach dem TMS, zu 20% über die Wartezeit und von einigen Hochschulen zu 15% über Auswahlgespräche vergeben (Trost et al., 1998). Wegen sinkender Bewerberzahlen wurde der TMS 1996 eingestellt. In der Folge konnten die Hochschulen 24% der Studienplätze selbst vergeben, worauf viele Hochschulen allein die Abiturnote als Auswahlkriterium verwendeten

(Hampe et al., 2009). Schließlich forderte der Wissenschaftsrat (2004) die Hochschulen auf, sich wieder aktiver in die Auswahl einzubringen, um die Studienabbruchquoten zu reduzieren und die Profilbildung der Hochschulen zu stärken. Durch das 7. Gesetz zur Änderung des Hochschulrahmengesetzes (28.08.2004) wurde das Auswahlrecht der Hochschulen weiter gestärkt. Seit dem Wintersemester 2005/2006 werden die medizinischen Studienplätze nach einer Vorabquote (Ausländer, Härtefälle, Zweit-Studienplatzbewerber, etc.) zu jeweils 20% über die Abiturbesten- und Wartezeitquote zentral und zu 60% durch die Hochschule selbst im sogenannten Auswahlverfahren der Hochschule (AdH) vergeben.

Theoretische Hintergründe. Die Festlegung der Auswahlkriterien orientiert sich im Optimalfall nach ihrem Vorhersagewert für den Studienerfolg. Für den Begriff Studienerfolg existiert keine allgemein akzeptierte Definition. „Vielmehr liegen Kriterien vor, an denen der Studienerfolg gemessen werden kann“ (Freyer, 2013, S. 35). Meist wird darunter ein Studienabschluss mit guten Noten und Zwischennoten verstanden (Wissenschaftsrat, 2004). Auch international umfasst die Mehrzahl der Definitionen quantitative statt qualitative Studienerfolgskriterien (Venezia, Callan, Finney, Kirst & Usdan, 2005). Das Problem scheint jedoch weniger an dem Verständnis des Studienerfolgs an sich, sondern in der praktischen Durchführbarkeit von Studienerfolgsmessungen zu liegen. Modelle des Studienerfolgs (z.B. Kuh, Kinzie, Buckley, Bridges & Hayek, 2007; Rindermann & Oubaid, 1999; Thiel, Veith, Blüthmann, Lepa & Ficzk, 2008) enthalten neben den quantitativen sehr wohl qualitative Erfolgskriterien. Beispielhaft soll das Bedingungsmodell des Studienerfolgs (Rindermann & Oubaid, 1999) als einflussreiches Modell im deutschsprachigen Raum erläutert werden. Im Zentrum stehen die Studienanforderungen und -bedingungen, welche durch Personenmerkmale, Rahmenbedingungen der Hochschule und Gesetzmäßigkeiten beeinflusst werden und sich gemeinsam auf den Studienerfolg auswirken (Abbildung 1). Die Personenmerkmale bilden die Prädiktoren und Auswahlkriterien für den Studienerfolg ab, während die Erfolgskriterien zur Messung des Studienerfolgs herangezogen werden. Die Auswahl- und Erfolgskriterien werden im folgenden Abschnitt beschrieben.

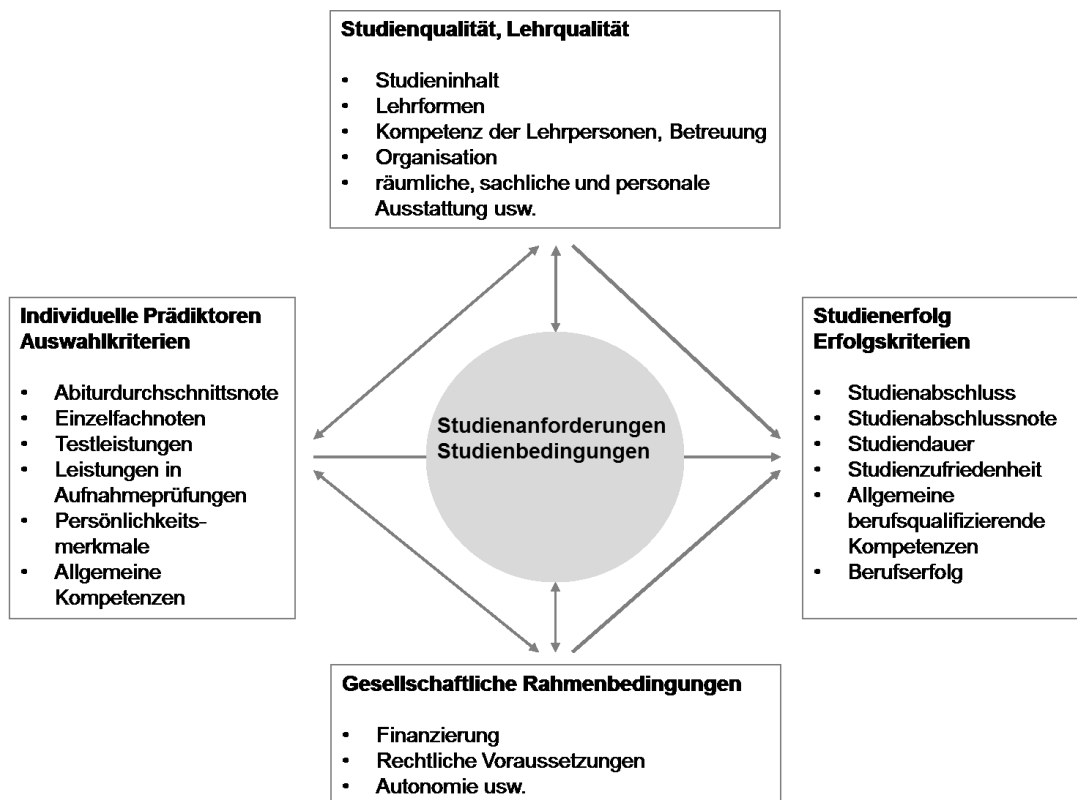


Abbildung 1. Bedingungsmodell des Studienerfolgs (Rindermann & Oubaid, 1999).

Als Prädiktoren werden aggregierte Schulnoten, Einzelfachnoten, Studierfähigkeits-tests, Leistungen in Aufnahmeprüfungen, Persönlichkeitsmerkmale, sowie allgemeine Kompetenzen verwendet. Die aggregierten Schulnoten wie der Abiturnotendurchschnitt werden als Indikator für die allgemeine Eignung zum Studium verwendet (Deidesheimer Kreis, 1993). Aggregierte Schulnoten sind leicht verfügbar und besitzen aufgrund der Leistungsorientierung eine höhere Augenscheinvalidität (Rindermann & Oubaid, 1999). Zudem zeigen sie aufgrund des höheren Aggregationsniveaus eine höhere Reliabilität und prognostische Validität als Einzelfachnoten (Baron-Boldt, Schuler & Funke, 1988). Einschränkungen weisen sie in ihrer Aussagekraft durch unterschiedliche Oberstufengestaltung, Fächerkombinationen, Lehrerurteile, Schwierigkeiten und Anzahl von Schuljahren auf (z.B. Althoff, 1986; Baumert & Watermann, 2000). Einzelfachnoten werden verwendet, wenn ein inhaltlicher Zusammenhang zwischen dem Schulfach und dem Studiengang besteht. Im Gegensatz zur Abiturnote können sie eher Stärken und Schwächen eines Bewerbers im Hinblick auf spezifische Studienvoraussetzung abbilden (Gold & Souvignier, 2005). Kenntnistests werden eingesetzt, um allgemeines oder fachspezifisches Wissen zu erfragen, wohingegen Studierfähigkeits-tests studienrelevante kognitive Fähigkeiten erheben (Deidesheimer Kreis, 1997). Sie können allgemeiner (z.B. Scholastic Aptitude Test und American College Test) oder

studienfachspezifischer Natur (Test für Medizinische Studiengänge, Eignungstest für Medizinstudium in der Schweiz) sein. Ihr Einsatz geht auf die Erkenntnis zurück, dass Fähigkeitstests eine prognostische Validität für den Berufserfolg haben (Hunter & Hunter, 1984). Um fachnahe Leistungsnachweise zu erhalten, werden Aufnahmeprüfungen durchgeführt. Diese werden jedoch weniger in den Auswahlverfahren zum Medizinstudium, als für die Zulassung zu Kunst-, Musik- und Sportstudiengängen durchgeführt. Seltener werden Persönlichkeitsmerkmale und allgemeine Kompetenzen wie soziale Kompetenz, Offenheit, Interessen und Studienmotivation über Auswahlgespräche, *Assessment-Center* und Selbstbeschreibungen berücksichtigt (Rindermann & Oubaid, 1999).

Den Prädiktoren stehen vielfältige Kriterien für den Studienerfolg gegenüber. Der Studienabschluss gilt im Gegensatz zum Studienabbruch als basales Erfolgskriterium für das Studium (Dlugosch, 2005). Eine genauere Differenzierung liefern Studienabschlussnoten und Zwischenprüfungsnoten. Sie bieten auch einen fachspezifischen Vorhersagewert für die Arbeitsmarktchancen der Absolventen (Gold & Souvignier, 1997). Als weiteres Studienerfolgskriterium gilt die Studiendauer. Eine Person, die ihren Abschluss in kürzerer Zeit erreicht, wird als erfolgreicher eingestuft. Der Nachteil in dieser Annahme liegt darin, dass die Studiendauer nicht nur die persönliche Eignung des Studierenden, sondern auch die Rahmenbedingungen der Universität widerspiegelt (Kersting, 2005). Zudem verzerren Fachwechsler, Studienortwechsler, Quereinsteiger und Studienunterbrechungen die Aussagekraft der Studiendauer. Die Studienzufriedenheit wird als Passung zwischen objektiven Studienbedingungen und individuellen Bedürfnissen verstanden (Gold & Souvignier, 1997). Die Studienzufriedenheit ist kein standardisiertes Kriterium, sondern variiert je nach den gewählten Items (Westermann, Heise, Spies & Trautwein, 1996). Die allgemeinen berufsqualifizierenden Kompetenzen erlangten als Studienerfolgskriterien durch die Umstellung auf Bachelor- und Masterstudiengänge eine höhere Bedeutung. So fordern die Programmakkreditierungsrichtlinien, dass die Studierenden neben fachlichen auch generische Kompetenzen erwerben sollen (Akkreditierungsrat, 2013, S.11). Diese gehen jedoch aufgrund von Definitions- und Messproblemen selten als Erfolgskriterium ein (Rindermann & Oubaid, 1999). Auch der Berufserfolg, operationalisierbar über Arbeitsmarktchancen, Position, Einkommen, Zufriedenheit, wird aus demselben Grund kaum herangezogen, zumal er nur mit einem zeitlichen Abstand zum Studienabschluss erhoben werden kann.

2.2. Forschungsergebnisse zur Medizinischen Studierendenauswahl

2.2.1. Kognitive Auswahlkriterien

Einen Forschungsausschnitt zur prädiktiven Validität der Hochschulzulassungsberechtigung als Auswahlkriterium in medizinischen Studiengängen stellt Tabelle 2 dar. Die durchschnittliche Korrelationshöhe des Abiturnotendurchschnitts mit den Studienleistungen liegt international um .40-.50 (Gold & Souvignier, 2005; Siu & Reiter, 2009). Die Durchschnittsnote hat einen höheren Vorhersagewert für die Leistung im humanmedizinischen als im zahnmedizinischen Studiengang, was auf den höheren Anteil mündlicher und praktischer Prüfungen in der Zahnmedizin zurückgeführt wird (Hell, Trapmann & Schuler, 2007; Trost, 1996). Untermuert wird die Annahme durch den höheren Zusammenhang der Abiturnote mit der schriftlichen als mit der mündlichen Physikumsprüfung ($r = .48$ vs. $.32$; Trost, 1994). Die Vorhersageleistung der Durchschnittsnote sinkt über die Studiendauer, wie die Längsschnittstudie von Mercer und Puddey (2011) zeigt. So können vorklinische Leistungen besser vorhergesagt werden als klinische (Stumpf & Nauels, 1990; Trapmann, Hell, Weigand & Schuler, 2007). Zum einen ist der vorklinische Studienabschnitt mehr theoretisch und der klinische mehr praktisch geprägt, zum anderen können Reifeprozesse zur sinkenden Validität beitragen (Trapmann et al., 2007). Selbige Autoren vermuten, dass die Abschlussnote über die Jahre an Validität gewonnen hat, da die mittleren minderungskorrigierten Korrelationen zwischen Schulnoten und Studienleistungen zwischen der Metaanalyse von Baron-Boldt et al. (1988) und der von Trapmann et al. (2007) angestiegen sind ($r = .46$ vs. $.53$). Andererseits kann dieser Effekt auch dadurch auftreten, dass die Messung der Studienleistung reliabler und valider erfolgt.

Tabelle 2: *Prädiktive Validität der Hochschulzulassungsberechtigung*

Autoren	Studienleistung	Stichprobe	Beta
McManus & Richards (1986)	vorklinisch	Humanmedizin N = 485	.23
Baron-Boldt, Schuler & Funke (1988)	Examen	Medizinische & andere Fächer	.46
Stumpf & Nauels (1990)	Erstes Staatsexamen	Humanmedizin N = 7000	.28
Trost (1994)	schriftliches Physikum	Humanmedizin N = 27 000	.48
	mündliches Physikum	Humanmedizin N = 27 000	.32

Fortsetzung Tabelle 2:

Autoren	Studienleistung	Stichprobe	Beta
Trost (1996)	Naturwissenschaftliche Vorprüfung	Zahnmedizin N = 21 000	.40
	Zahnärztliche Vorprüfung	Zahnmedizin N = 21 000	.31
Kulatunga-Moruzi & Norman (2002)	Examen	Humanmedizin N = 52-96	.33
Ferguson, James, O'Hehir & Sanders (2003)	vorklinisch und klinisch	Humanmedizin N = 118	.13
Gold & Souvignier (2005)	Erstes Staatsexamen	Human- und Zahnmedizin N = 86	.26
Julian (2005)	Studienjahr 1-2	Humanmedizin N = 65-148	.40
	Famulatur		.26
	Examen 1-3		.29-39
Trapmann, Hell, Weigand & Schuler (2007)	vorklinisch & klinisch	Medizinische Fächer	.33
	vorklinisch	k = 9-25	.39
	klinisch		.23
	vorklinisch & klinisch	Humanmedizin	.40
	vorklinisch	k = 3-7	.47
	vorklinisch & klinisch	Zahnmedizin	.25
	vorklinisch	k = 5-11	.31
Mercer & Puddey (2011)	klinisch		.20
	Studienjahr 1	Humanmedizin	.47
	Studienjahr 2	N = 547-1174	.40
	Studienjahr 3		.32
	Studienjahr 4		.23
	Studienjahr 5		.21
	Studienjahr 6		.21

Als Einzelfachnote zeigen die Französisch-, Latein- und die Mathematiknoten die beste prädiktive Eigenschaft für das Medizinstudium (Rindermann & Oubaid, 1999). In den in Tabelle 3 aufgeführten Studien erreicht die Mathematiknote Korrelationen von bis zu .20-.31 zu medizinischen Studienleistungen, wobei jedoch die Studie von Montague und Odds (1990) eine Nullkorrelation berichtet. Der Vorhersagewert von naturwissenschaftlichen Schulnoten schwankt zwischen .14 und .44. In der Studie von Gold und Souvignier (2005) schnitten die Chemienote ($r = .27$) und die Biologienote ($r = .25$) als beste Prädiktoren ab. Einzelfachnoten zeigen aufgrund der geringeren Aggregation eine niedrigere Reliabilität und prognostische Validität (Baron-Boldt et al., 1988).

Tabelle 3: *Prädiktive Validität Einzelfachnoten*

Autoren	Prädiktor	Studienleistung	Stichprobe	Beta
Giesen, Gold, Hummer & Jan- sen (1986)	Fachaffine Einzelnote	vorklinisch	Humanmedizin	.26
	Mathematiknote		N = 85	.31
Montague & Odds (1990)	Fachaffine Einzelnote	vorklinisch & klinisch	Humanmedizin	.16
	Mathematiknote		N = 137-207	-.03
Höschl & Kozeny (1997)	Mathematiknote	vorklinisch	Humanmedizin	.34
	Note Erstsprache	klinisch	N = 92	.24
Kulatunga-Moruzi & Norman (2002)	Grade Point Average	Examen	Humanmedizin	.45
	Science		N = 52-96	
Gold & Souvig- nier (2005)	Abiturnote	Erstes Staatsexamen	Human- und	.26
	Mathematiknote		Zahnmedizin	.14
	Noten Naturwissen- schaft		N = 86	.14-.27
	Noten Sprachen			.11-.17
Trapmann, Hell, Weigand & Schuler (2007)	Mathematiknote	vorklinisch & klinisch	Medizinische Fächer k = 3	.21

Ähnlich hohe Vorhersagewerte wie die Abiturnote erreichen Studierfähigkeitstests (Hell et al., 2007; Trapmann et al., 2007) wie Tabelle 4 veranschaulicht.

Tabelle 4: *Vorhersageleistung von Studierfähigkeitstests*

Autoren	Prädiktor	Studienleistung	Stichprobe	Beta
Stumpf & Nauels (1990)	TMS	1. Staatsexamen	Humanmedizin	.27
	Abitur & TMS		N = 7000	.35
Fisseni, Olbrich, Halsig, Mailahn & Iltner (1993)	Abitur & TMS	Physikum	Humanmedizin	
		1980	N = 149	.44
	1981		.39	
Trost (1994)	TMS	schriftliches Physikum	Humanmedizin	.53
	Abitur & TMS		N = 27 000	.58
	TMS	mündliches Physikum	Humanmedizin	.28
	Abitur & TMS		N = 27 000	.36

Fortsetzung Tabelle 4:

Autoren	Prädiktor	Studienleistung	Stichprobe	Beta
Trost (1996)	TMS	Naturwissenschaftliche	Zahnmedizin	.35
	Abitur & TMS	Vorprüfung	N = 21 000	.44
	TMS	Zahnärztliche Vorprü-	Zahnmedizin	.29
	Abitur & TMS	fung	N = 21 000	.35
Kulatunga-Moruzi & Norman (2002)	MCAT	Studienjahr 1-2	Humanmedizin	.44
		Famulatur	N = 65-148	.32
		Examen 1-3		.49-.61
	GPA & MCAT	Studienjahr 1-2		.53
		Famulatur		.29
		Examen 1-3		.52-.64
Bergman, Susar- la, Howell & Ka- rimbux (2006)	DAT- Wahrneh- mungsfähigkeit	Examen 1: Anatomie und Okklusion	Zahnmedizin N = 244	.37
		Examen 1: Anatomie und Okklusion		.37
	DAT- Leseverständnis	Examen 1: Mikrobiolo- gie & Pathologie		.32
Donnon, Paolucci & Violato (2007)	MCAT	vorklinisch	Humanmedizin	.39
		Examen 1	k = 23	.60
Hell, Trapmann & Schuler (2007)	TMS/EMS	gesamt	Humanmedizin	.51
		vorklinisch	k = 9	.52
		klinisch		.56
		gesamt	Zahnmedizin k = 8	.35
Hampe, Klus- mann, Buhk, Münch-Harrach & Harendza (2008)	HAM-NAT	Anzahl bestandener vorklinischer Leis- tungsnachweise	Humanmedizin N = 333	.26
Lievens (2013)	Naturwissen- schaftlicher & fluide Intelligenz- test	Kurse ohne Patienten- kontakt		.32
		Kurse mit Patienten- kontakt		.09
		OSCE Examen		.00

Anmerkung. DAT = Dental Admission Test; EMS = Eignungstest für Medizinstudium in der Schweiz; GPA = Grade Point Average; HAM-Nat = Hamburger Auswahlverfahren für Medizinische Studiengänge-Naturwissenschaftsteil; MCAT = Medical College Admission Test, OSCE = Objective structured clinical examination, TMS = Test für Medizinische Studiengänge.

In Nordamerika wird der *Medical College Admission Test (MCAT)* für naturwissenschaftliche Kenntnisse und Fähigkeiten seit 1991 an medizinischen Hochschulen eingesetzt. Die Metaanalyse von Donnon, Paolucci & Violato (2007) berichtet eine prädiktive Validität des MCAT von .39 für vorklinische Leistungen und .60 für die erste staatliche Prüfung, die klinisches Schlussfolgern erfasst. Die Vorhersageleistung für Studien- und Examensnoten erhöht sich bei der Kombination des Grade Point Averages (GPA) und des *MCATs* (Julian, 2005). Vergleichbare Vorhersagewerte zeigt der TMS, der derzeit an 17 humanmedizinischen und 11 zahnmedizinischen Fakultäten eingesetzt wird (www.hochschulstart.de/medizin, Zugriff am 27.03.2014). Der TMS korreliert höher mit schriftlichen als mit mündlichen Leistungen im Physikikum ($r = .53$ vs. $.28$; Trost, 1994). Er zeigt einen höheren Vorhersagewert für das humanmedizinische als für das zahnmedizinische Studium ($r = .51$ vs. $.31$; Hell et al., 2007). Die Kombination von Abiturnote und TMS erhöht die Vorhersageleistung auf $.58$ (Trost, 1994). Als Alternative setzen derzeit zwei medizinische Fakultäten in Deutschland das Hamburger Auswahlverfahren für Medizinische Studiengänge-Naturwissenschaftsteil (HAM-Nat) ein (www.hochschulstart.de/medizin, Zugriff am 27.03.2014). Das HAM-NAT besteht aus 52 naturwissenschaftlichen *Multiple Choice* Fragen auf Schulniveau. Eine bisherige Studie berichtet einen Zusammenhang von $.26$ mit der Anzahl bestandener vorklinischer Leistungsnachweise (Hampe, Klusmann, Buhk, Münch-Harrach & Harendza, 2008). In der Zahnmedizin wird der Dental Admission Test (DAT) eingesetzt, der naturwissenschaftliche Kenntnisse, Wahrnehmungsfähigkeit, Leseverständnis und logisches Schlussfolgern erfasst. Dieser zeigt Korrelationen um $.30$ zu Subtests der Examensleistung (Bergman, Susarla, Howell & Karimbu, 2006). In der Studie von Lievens (2013) fällt auf, dass die kognitiven Studierfähigkeitstests gute Prädiktoren für kognitive Studienkriterien sind, aber kaum Vorhersageleistung für praktische Prüfungen mit Patientenkontakt aufweisen.

Zusammenfassend liegt die Stärke der kognitiven Auswahlkriterien in der Vorhersage von kognitiv geprägten Studienleistungen, welche überwiegend in den vorklinischen Studienjahren erbracht werden. Schwäche zeigen sie indes in der Vorhersage von klinischen Studienleistungen, die einen höheren Anteil nicht-kognitiver Fähigkeiten beinhaltet.

2.2.2. Nicht-kognitive Auswahlkriterien

Wissenschaftliche Studien zur Bedeutung nicht-kognitiver Zulassungskriterien in medizinischen Studiengängen haben in den letzten Jahren an Bedeutung gewonnen (Hampe et al. 2009). Zum Einsatz kommen vor allem Auswahlgespräche von unstrukturierten Interviews (Kulatunga-Moruzi & Norman, 2002) bis hin zu hoch standardisierten *Multiple Mini-Interviews (MMI)* (Dodson et al., 2009; Eva, Reiter, Rosenfeld & Norman, 2004), in denen die Studienplatzbewerber eine festgelegte Anzahl von Interviewstationen durchlaufen. Auch *Assessment Center* (Hofmann, Rieger & Ostermann, 2007; Ziv et al., 2008) und *SJTs* (Lievens & Sackett, 2006) werden an medizinischen Hochschulen zur Studierendenauswahl eingesetzt. Der Schwerpunkt des Kapitels richtet sich gemäß dem übergeordneten Forschungsziel auf Studien, die psychische, soziale und emotionale Kriterien, sowie Persönlichkeitsfaktoren zum Gegenstand hatten.

Ein Ansatz den Einfluss nicht-kognitiver Auswahlkriterien auf den Studienerfolg zu untersuchen, geht auf die *Student Integration Theory* (Tinto, 1975, 1993) und die *Students Attrition Theory* (Bean, 1980, 1985) zurück. Diese nehmen an, dass sich die soziale Integration an der Hochschule direkt auf die Studienzufriedenheit und indirekt auf den Studienerfolg auswirkt. Als Beispiel sei die Studie von Heublein et al. (2003) angeführt. Sie verglichen die soziale Integration von Studienabbrechern und Weiterstudierenden verschiedener Fachrichtungen, darunter Medizin. Die Studienabbrecher wiesen deutlich weniger soziale Kontakte an der Hochschule auf, fühlten sich unter den Kommilitonen unwohl und von den Dozenten nicht anerkannt.

Der Einfluss von psychosozialen Faktoren auf den Studienerfolg in der Humanmedizin wurde von Hojat et al. (1993) untersucht. Als psychosoziale Faktoren gingen Angst, Prüfungsangst, Depression, externale Attribution, Einsamkeit, Neurotizismus, Selbstachtung, Extraversion und herausfordernde Lebensereignisse, erhoben über Fragebögen, ein. Die psychosozialen Faktoren konnten die Leistung in naturwissenschaftlichen und klinischen Kursen in vergleichbarem Ausmaß wie der *MCAT* vorhersagen (Abbildung 2). In der Vorhersage der klinischen Kompetenz waren jedoch die psychosozialen Faktoren den kognitiven Faktoren überlegen.

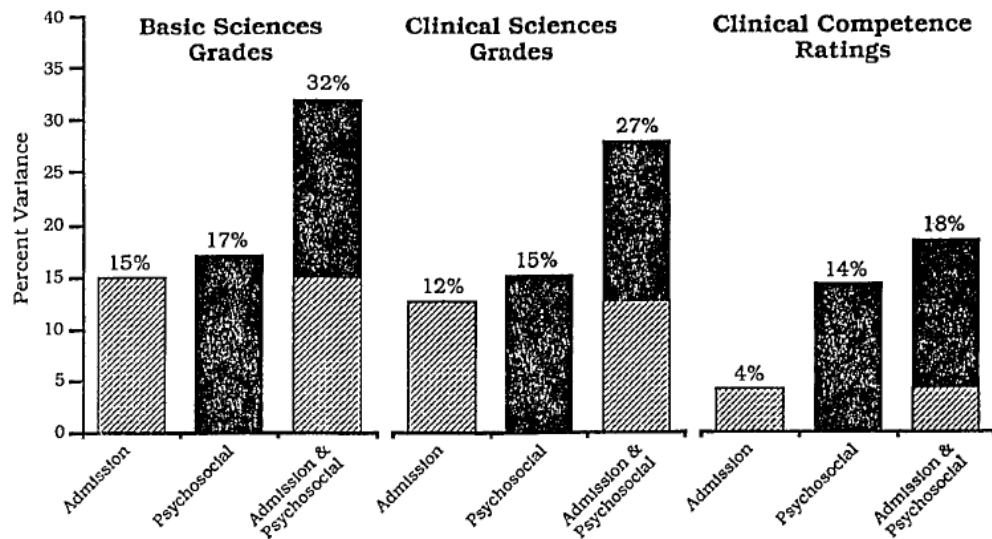


Abbildung 2. Varianzaufklärung des Studienerfolgs durch psychosoziale Faktoren (Hojat et al., 1993, S. 636).

Auch für die über ein Auswahlgespräch erhobenen psychosozialen Faktoren Sozialkontakte, Deutung der Prüfung als Herausforderung, Arbeitsstil, Relation zwischen Fremd- und Selbstbild, Antizipation von Prüfungsbelastung, Kenntnis von Schwächen und Besuch naturwissenschaftlicher Veranstaltungen ließen sich hohe Korrelate mit dem Studienerfolg in der Humanmedizin nachweisen (Fisseni, Olbrich, Halsig, Mailahn & Ittner, 1993). Die psychosozialen Faktoren korrelierten hoch mit der Leistung im Physikum (.66-.71, jeweils korrigiert für Rangrestriktionen und Reliabilitätseinschränkungen in Prädiktor- und Kriterium). In Kombination mit der Abiturnote und dem TMS erhöhte sich der Zusammenhang auf .73-.81. Interessanterweise zeigten Studierende, die sich weniger sozial kompetent einschätzten und weniger Kontakt zu Kommilitonen aufwiesen, höhere Leistungen im Physikum. Die Metaanalyse von Robbins, Lauer, Davis, Langley und Carlstrom (2004) zeigte eine Varianzaufklärung von 26% für die Kombination von psychosozialen und kognitiven Faktoren für den Studienerfolg, gemessen über die Durchschnittsnote im Studium. Die inkrementelle Validität der psychosozialen Faktoren betrug vier Prozent.

Bezüglich emotionaler Kompetenzen zeigte sich, dass Empathie besser geeignet ist, klinische statt kognitive Leistungsfähigkeit vorherzusagen (Gough & Hall, 1975). Dies steht im Einklang mit der Studie von Tutton (1996), der nachweisen konnte, dass Medizinstudierende mit hohen naturwissenschaftlichen Leistungen geringere Empathie-Werte aufwiesen und sich als zurückgezogener und unbeholfener in sozialen Kontakten einschätzten als Studierende mit weniger guten Noten. Auch aktuellere Studien belegen diesen Einfluss. Libbrecht, Lievens, Carette und Côté (2014) untersuchten den

Einfluss der emotionalen Intelligenz im Medizinstudium. Der Subfaktor Emotionsregulation konnte die Leistung in einem Arzt-Patienten-Kommunikationskurs über die ersten drei Studienjahre vorhersagen und besaß eine inkrementelle Validität von vier Prozent über kognitive Kriterien und Gewissenhaftigkeit.

Bezüglich sozialer Kompetenzen können ähnliche Ergebnisse festgehalten werden. Studierende, die über ein Auswahlgespräch zugelassen wurden, zeigten eine höhere Ausprägung sozialer Kompetenzen, aber schlechtere Studiennoten und längere Studiendauer als über Abiturdurchschnittsnote und TMS-Ergebnis zugelassene Studierende (Ittner & Halsig, 1993; Trost, Klieme & Nauels, 1997). Der in Belgien eingesetzte videobasierte *SJT* für zwischenmenschliche Fähigkeiten wies eine prädiktive Validität von .12-.21 und eine inkrementelle Validität von 1-4% für Kursleistungen mit Patientenkontakt über die gesamte Dauer des Studiums auf. Für die Vorhersage von Kursleistungen ohne Patientenkontakt waren hingegen die kognitiven Auswahlkriterien überlegen (Lievens, 2013). In der Studie von Kulatunga-Moruzi und Norman (2002) wies das Auswahlgespräch eine Vorhersageleistung von .24 für den zweiten Teil des staatlichen Examens auf, dessen Format ein *Objective Structured Clinical Examination (OSCE)* mit dem Schwerpunkt Kommunikationsfähigkeit und Problemexploration ist. Keine prädiktive Validität zeigte das Auswahlgespräch hingegen für den kognitiv geprägten ersten Examensteil. Auch das *MMI* sagt besser die Leistung in der *OSCE*-Prüfung am Ende einer klinischen Rotation vorher als die Leistung in einer *Multiple Choice* Klausur ($\beta = .23$ vs. $\beta = .09-.27$; Eva et al., 2004). Ebenfalls zeigte das *MMI* signifikante Zusammenhänge für die zeitlich spätere staatliche Prüfung im *OSCE*-Format (Eva et al., 2009; Eva et al., 2012).

Hinsichtlich Persönlichkeitsfaktoren scheint die Gewissenhaftigkeit am meisten Potenzial in der Vorhersage des Studienerfolgs zu haben. In der Studie von Lievens, Ones und Dilchert (2009) stieg der Vorhersagewert der Gewissenhaftigkeit im Verlauf des Medizinstudiums von .18 im ersten Jahr auf .45 im siebenten Studienjahr an (Rangrestriktionen und Reliabilitätseinschränkungen in Prädiktor und Kriterium korrigiert). Auch die Vorhersagekraft von Extraversion stieg im Verlauf des Studiums von .11 auf .31. Offenheit für neue Erfahrungen zeigte ab dem zweiten Studienjahr Korrelationen von .18 bis .30 mit den Studiennoten. Gleiche Korrelationshöhen wies die Verträglichkeit auf, jedoch nur im fünften und siebenten Studienjahr. Im zahnmedizinischen Bereich sind Studierende mit hohen Extraversionswerten in der klinischen Arbeit überlegen, während introvertierte Zahnmedizinische Studierende in kognitiv geprägten Studienabschnitten höhere Leistungen zeigen (Jones, Courts, Sandow & Watson, 1997). Auch unabhängig vom medizinischen Kontext werden Persönlichkeitsmerkmale als

aussagekräftig für den langfristigen Studium- und Berufserfolg angesehen (vgl. Hosiep et al., 2000; Schmidt & Hunter, 1998).

Zusammenfassend zeigen die nicht-kognitiven Kriterien Potenzial für die Vorhersage von klinischen Studienleistungen, in denen zwischenmenschliche und persönliche Kompetenzen eine höhere Rolle spielen. Wenig geeignet sind sie für die Vorhersage von kognitiven Studienleistungen, die gewöhnlich in den frühen Studienjahren überwiegen.

2.3. Studierendenauswahl an der Medizinischen Fakultät Heidelberg

Die in Kapitel 2.1 dargelegten rechtlichen und politischen Rahmenbedingungen spiegeln sich auch im AdH der Medizinischen Fakultät Heidelberg wieder, das seit 2005 dokumentiert ist (Tabelle 5). Die Hochschulzugangsberechtigung (HZB) wurde anfangs nur durch das Bonuskriterium einer medizinnahen Ausbildung oder Berufserfahrung ergänzt. Ab 2006 nahm die Fakultät weitere Bonuskriterien in das Zulassungsverfahren auf. Die Jahre 2007-2009 dienten der Implementierung des TMS. Seit 2009 werden die Abiturdurchschnittsnote und das TMS-Ergebnis kompensatorisch verrechnet, d.h. bei einer durchschnittlichen HZB-Note kann nur durch ein exzellentes TMS-Ergebnis eine Zulassung erreicht werden, während bei einer sehr guten HZB-Note eine überdurchschnittliche, aber geringere TMS-Leistung ausreicht.

Tabelle 5: *Entwicklung des AdH in Heidelberg*

Zulassung	1. Kriterium	2. Kriterium	Bonuskriterien
2005	HZB 1.0-4.0 (60%)		Medizinnahe Ausbildung/Berufserfahrung (40%)
2006	HZB 1.0-4.0 (60%)		Medizinnahe Ausbildung/Berufserfahrung, FSJ/Freiwilligendienst, Preise in bildungsbezo- genen Wettbewerben auf Bundesebene (40%)
2007-2012	HZB 1.0-2.3 (51%)	TMS > 100 (39%)	Bonuskriterien wie 2006 (10%)
2013	HZB 1.0-4.0 (46%)	TMS > 100 (44%)	Bonuskriterien wie 2006 (10%)

Anmerkung. HZB = Hochschulzugangsberechtigung; TMS = Test für Medizinische Studiengänge. Prozentwerte der formellen Gewichtung in Klammern.

Das kompensatorische Zulassungsverfahren senkte die Abbrecherzahlen in den AdH-Kohorten um 54-75% (Kadmon, Möltner, Kahmann, Kadmon, zur Publikation eingereicht). Die besten akademischen Leistungen erreichten Studierende entweder mit einer überdurchschnittlichen HZB-Note oder mit einem überdurchschnittlich guten TMS, aber mit einer durchschnittlichen HZB-Note (Kadmon et al., zur Publikation eingereicht). Das kompensatorische Verfahren ermöglichte somit, Bewerber mit einem breiteren Abiturnotenspektrum zuzulassen, die andererseits abgelehnt worden wären.

2.4. Schlussfolgerungen für die Medizinische Studierendenauswahl

Für den Status quo der medizinischen Studierendenauswahl stellen die Editoren des Lancet (2010) fest, dass seit dem Flexner Report (1910) der Schwerpunkt des Medizinstudiums und der Auswahl dafür stark auf naturwissenschaftliche Bestandteile gelenkt wurde und zu wenig humanitäre Aspekte berücksichtigt werden. Ein Problem der kognitiv-orientierten Auswahlverfahren liegt in der eingeschränkten Varianz in der Hochschulzulassungsberechtigung, die es kaum ermöglicht unter den Bewerbern mit sehr guten Leistungen genauer zu differenzieren (McManus, 2005; Mercer & Puddey, 2011; Rolf & Powis, 1997). Auch ließen die nicht-kognitiven Kompetenzen einiger Studierender im Studium und im Beruf zu wünschen übrig, was sich in Patientenbeschwerden äußert (Powis, 2010). Zudem sind die ausgewählten Studierenden nicht repräsentativ für die Bevölkerung, die später von den Absolventen versorgt werden (Powis, 2010; Prideaux et al., 2011). Die Ähnlichkeit von Arzt und Patient ist nicht nur politisch gewünscht, sondern führt auch zu einer besseren Arzt-Patienten-Kommunikation und -beziehung sowie höheren Therapieerfolgen (Magnus & Mick, 2000; Mathers, Sitch, Marsh & Parry, 2002). Die meisten Medizinstudierenden sind weiß, haben einen hohen sozioökonomischen Status und genossen eine gute Schulbildung (James & Nicholson, 2006). Die rein kognitive Studierendenauswahl wird auch als Mangelverwaltung des Bewerberüberhangs gegenüber den zur Verfügung stehenden Plätzen bezeichnet (Wottawa & Hossiep, 1987). Die Kritik bedeutet nicht, dass kognitive Kriterien aus der Studierendenauswahl zu verbannen sind, vielmehr sollten neben diesen Kriterien einbezogen werden, die eine größere sozioökonomische Vielfalt unter den Ausgewählten und eine Vorhersage für die klinischen Abschnitte des Studiums zulassen (Bore, Munro & Powis, 2009; Humphreys, 1960; Lievens et al., 2009a). Die kognitiven Fähigkeiten stellen eine Schwelle zum Medizinstudium dar, die genommen werden muss, aber die nicht alles sind, was einen guten Arzt ausmacht (Barr, 2010). Um eine hinreichende Passung zwischen dem Bewerber und dem Studiengangprofil zu gewährleisten, sollte demnach die Studierendenauswahl adaptiert

werden (Wenzel, 2008; Wissenschaftsrat, 2004). Das Wunschprofil eines guten Arztes beschreibt Powis (2010) wie folgt:

Any competency list for a generic medical practitioner should comprise excellent academic ability, good cognitive skills, and the ability to use academic knowledge appropriately in quantitative, verbal, and spatial domains. Practitioners should also have well developed decision making skills, professional integrity, highly developed moral standards, and excellent interpersonal skills, in addition to being accomplished and confident communicators who can empathize with patients. They must be able to function professionally when under stress, have a good measure of self-control, and not be prone to taking inappropriate risks in a practice setting. Their mental resilience and emotional stability must be high. We are not talking about a superhero, we are talking about the doctor we would all like to see when we are sick. (S. 708)

Ein Ansatz zur Adaptation der medizinischen Studierendenauswahl schlagen Bore et al. (2009) vor. Ihr *best-practice* Modell (Abbildung 3) kombiniert kognitive und nicht-kognitive Auswahlkriterien in einem zweistufigen Verfahren.

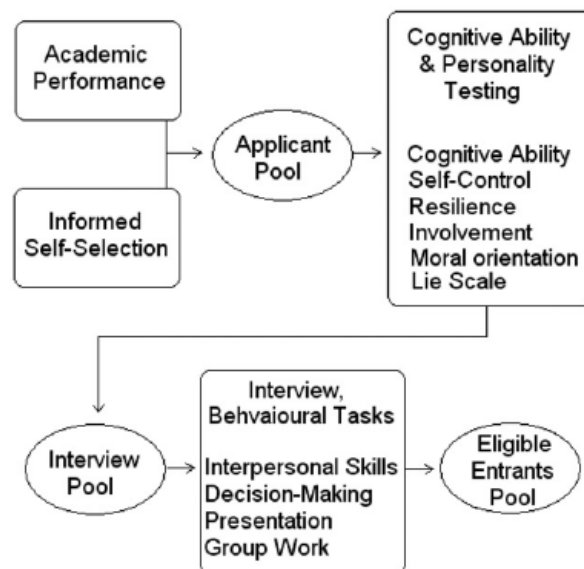


Abbildung 3. Studierendenauswahlmodell (Bore et al., 2009).

Die Bewerber sollten sich zunächst selbstständig über das Studium informieren, um entscheiden zu können, ob sie für das Studium geeignet sind (Selbst-Selektion). Die schulische Leistung wird über Durchschnittsnoten und wahlweise Fachnoten erfasst, wobei die Hürde unterschiedlich hoch gesetzt werden kann, um Personen mit unterschiedlichem sozio-ökonomischem Hintergrund einzubeziehen. Die kognitive Fähigkeit wird mit Kenntnis- oder Fähigkeitstests geprüft. Ebenfalls empfehlen sie den Einbezug von Persönlichkeitsfaktoren. Die in dieser Stufe geeigneten Kandidaten werden anschließend zu einem Auswahlgespräch eingeladen. Dieses sollte soziale Fertigkeiten, Entscheidungsfähigkeit, Belastbarkeit und Präsentationsfähigkeit testen. Die Herausforderung für die medizinische Studierendenauswahl liegt nun darin, reliable und valide Verfahren zur Erhebung nicht-kognitiver Kriterien zu entwickeln bzw. weiter zu erforschen (vgl. Benbassat & Baumal, 2007).

Für das kognitiv orientierte Auswahlverfahren der Medizinischen Fakultät Heidelberg bedeutet dies ebenfalls, dass es durch nicht-kognitive Auswahlkriterien ergänzt werden sollten, um stärker zwischen den kognitiven Leistungen der Bewerber zu differenzieren und den Studienerfolg im klinischen Studienabschnitt genauer vorherzusagen. Der Schwerpunkt kann dabei auf soziale Kompetenzen gelegt werden, da sie in der Arbeit mit Patienten, Kommilitonen und Lehrenden unabdingbar sind (Kapitel 3). Auswahlgespräche werden aufgrund der eingeschränkten Ressourcen jedoch als Instrument eher ausgeschlossen. Anlehnend an das Modell von Bore et al. (2009) und nach Empfehlungen der Deutschen Gesellschaft für Psychologie e. V. (Stellungnahme, 2004) kann in der Phase der Selbstselektion angesetzt werden. Der Vorteil ist, dass Studieninteressierte und Bewerber online Informationen erhalten und beraten werden können, ohne große personelle Ressourcen zu beanspruchen. Ziel der *Self-Assessments* ist die Auseinandersetzung der Studieninteressenten mit der eigenen Person, dem Studienangebot und den Studienanforderungen, um eine systematische Studienwahlentscheidung zu fördern (Organisation for Economic Cooperation and Development, 2004). Ein mögliches Instrument stellt der SJT dar, der in Kapitel 4 erläutert wird.

3. Soziale Kompetenzen

Ein Ansatzpunkt zur Ergänzung der medizinischen Studierendenauswahl um nicht-kognitive Kriterien sind soziale Kompetenzen (Kapitel 2.2.2). Auf diese gehen die folgenden Abschnitte ein. Kapitel 3.1 und 3.2 setzen sich mit allgemeinen Definitionen und Modellen sozialer Kompetenzen auseinander, während Kapitel 3.3 spezifische soziale Kompetenzen in der Medizin aufzeigt. Auf Basis der drei Kapitel wird anschließend ein Arbeitsmodell sozialer Kompetenzen als Grundlage für die Entwicklung eines *SJT*s für soziale Kompetenzen erstellt (Kapitel 3.4). Kapitel 3.5 beschreibt abschließend Messmethoden sozialer Kompetenzen, von denen der *SJT* im Bereich der Simulationsverfahren einzuordnen ist.

3.1. Definition sozialer Kompetenzen

Seit den 1970er Jahren gewinnt der Begriff der Sozialkompetenzen in vielen Bereichen an Bedeutung (Bastians & Runde; 2002; Lindgren & Heikkinen, 2004; Petermann, 2002). Einigkeit besteht bezüglich der Wichtigkeit sozialer Kompetenzen für das Individuum und die Gesellschaft. Soziale Kompetenzen ermöglichen dem Individuum die Teilhabe an der Gesellschaft und verhelfen der Gesellschaft zum wirtschaftlichen Wachstum und zu funktionierenden demokratischen Prozessen und sichern den sozialen Zusammenhalt (Rychen & Salganik, 2003). Des Weiteren stehen soziale Kompetenzen im positiven Zusammenhang mit der akademischen Leistungsfähigkeit (Wentzel, 1991), der Arbeitsleistung (Law, Wong & Song, 2004), der Problemlösefähigkeit (Simmons-Morton & Crump, 2003) und der Lebenszufriedenheit (Renk & Phares, 2004). Unklarheit herrscht indes hinsichtlich der Definition und des Konstrukts allgemeiner sozialer Kompetenzen (Döpfner, Schlüter & Rey, 1981; Kauffeld, Frieling & Grote, 2002; Mathews, Zeidner & Roberts, 2002). Oft werden die Begrifflichkeiten und Definitionen überlappend und widersprüchlich verwendet (Darden & Gintner, 1996; Spitzberg & Cupach, 1989). Neben der Multidimensionalität des Konstrukts liegen die Definitionsschwierigkeiten darin begründet, dass die Anzahl relevanter Dimensionen und deren Abstraktionsniveaus schwanken (Kanning, 2003; Schneider, Ackerman & Kanfer, 1996). Als gemeinsamen Nenner sehen Spitzberg und Cupach (1989) die Kontrolle und Anpassungsfähigkeit an die Umwelt und die Kollaboration mit den Interaktionspartnern.

Die umfangreichste Literatur zum sozialen Kompetenzbegriff bietet die klinische Psychologie mit dem Schwerpunkt der Behandlung sozial ängstlicher Personen, denen es schwer fällt, ihre eigenen Interessen in sozialen Interaktionen durchzusetzen (Fleischer, 2007; Kanning, 2003). Ein Beispiel für die Begriffsdefinition, die die klinische

Perspektive der Durchsetzungsfähigkeit eines Individuums widerspiegelt, ist die Definition von Hinsch und Pfingsten (2007). Sie beschreiben soziale Kompetenz als „die Verfügbarkeit und Anwendung von kognitiven, emotionalen und motorischen Verhaltensweisen, die in bestimmten sozialen Situationen zu einem langfristig günstigen Verhältnis von positiven und negativen Konsequenzen für den Handelnden führen“ (Hinsch & Pfingsten, 2007, S. 82). Derartige Kompetenzdefinitionen stellen den Handelnden in den Mittelpunkt, während die Interaktionspartner keine Berücksichtigung finden (Fleischer, 2007; Kanning, 2003). Im Gegensatz zur klinischen Psychologie bezieht die Entwicklungspsychologie die Interaktionspartner des Handelnden mit ein und versteht sozial kompetentes Verhalten eher als Anpassung und Sozialisationsprozess des Individuums an eine soziale Gruppe und an die Gesellschaft (DeRossier, Patterson, & Kuperschmidt, 1994; Hargie, Saunders & Dickson, 1981; Waters & Soufre, 1983). Durch die Verbindung beider Perspektiven kann sozial kompetentes Verhalten als Kompromiss zwischen Anpassung und Durchsetzung verstanden werden (Asendorpf & Neyer, 2012; Döpfner et al., 1981; Euler, 1997; Petermann & Petermann, 2003). In diesen Verständnisbereich fällt die Kompetenzdefinition von Kanning (2003, Abbildung 4), der zwischen sozial kompetentem Verhalten und sozialer Kompetenz differenziert.

Aus den Definitionen geht hervor, dass die Verfügbarkeit von sozialen Kompetenzen eine Voraussetzung, aber keine Garantie für sozial kompetentes Verhalten ist. Vielmehr ist die Umsetzung der Kompetenz in die Performanz entscheidend, um die Kompetenz beobachtbar zu machen. Umgekehrt ist ein Rückschluss auf die soziale Kompetenzausprägung eines Individuums erst durch die Beobachtung von spezifischen Situationen möglich.

Ob ein Verhalten als sozial kompetent gilt, ist abhängig vom situativen Kontext (Albrecht, 1997; Holtz, 1994; Reschke, 1995), dem Alter der Interagierenden (Holtz, 1994), gesellschaftlicher Werte und den kurz- oder langfristigen Zielen der Interaktionspartner (Kanning, 2003).

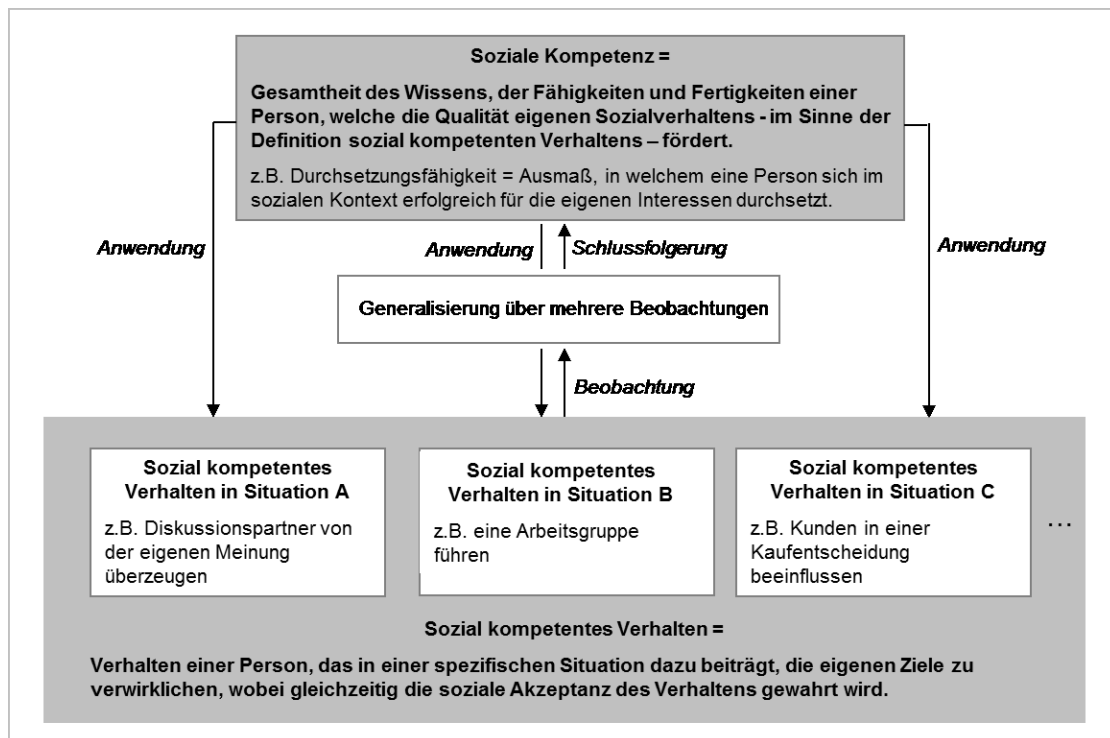


Abbildung 4. Definition von sozialer Kompetenz und sozial kompetentem Verhalten (Kanning, 2003, 2009)².

Der Begriff soziale Kompetenz kann als Oberbegriff gesehen werden, der viele verschiedene soziale Kompetenzen beinhaltet (Kanning, 2003). Analog dazu sprechen Hinsch & Pfingsten (2007) vom Begriff der sozialen Kompetenz lieber in der Pluralform, da in spezifischen Situationen immer ein Zusammenspiel von unterschiedlichen sozialen Kompetenzen sichtbar wird. Soziale Kompetenzen lassen sich in allgemeine und spezifische soziale Kompetenzen unterteilen (Reschke, 1995). Erstere können situationsübergreifend bei allen Menschen beobachtet werden (z.B. Extraversion, Perspektivenübernahme), während letztere nur bei denjenigen Personen beobachtbar sind, die entsprechende Lernerfahrung im spezifischen Kontext gesammelt haben (z.B. Gesprächsführung einer Kindergärtnerin mit den Kindern vs. Gesprächsführung einer Ärztin mit schwerkranken Patienten). „Dabei können die spezifischen Kompetenzen auch als eine auf Erfahrung basierende Ausdifferenzierung allgemeiner sozialer Kompetenzen verstanden werden“ (Kanning, 2003, S.19).

² Die ursprüngliche Abbildung (Kanning, 2003, S. 13) wurde durch die sozialen Kompetenzdefinitionen (Kanning, 2003, 2009), die Beispiele für Durchsetzungsfähigkeit und die nach unten gerichteten Pfeile ergänzt.

Neben den sozialen Kompetenzen können die Begrifflichkeiten soziale Intelligenz, emotionale Intelligenz, soziale Fertigkeiten und interpersonale Kompetenz als Teilmenge der sozialen Kompetenzen verwendet werden (Abbildung 5; Kanning, 2003). Diese Subkonstrukte überlappen wiederum miteinander und sind nicht eindeutig voneinander abgrenzbar. Auf die Subkonstrukte wird im Folgenden kurz eingegangen.

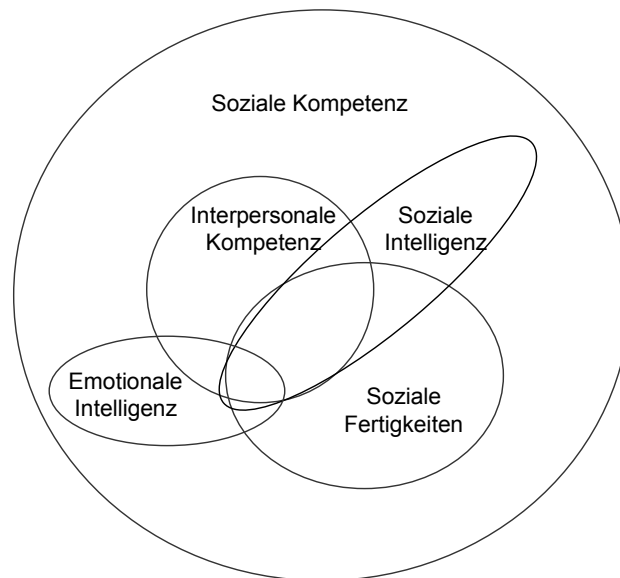


Abbildung 5. Verwandte Konstrukte sozialer Kompetenzen (Kanning, 2003).

Die soziale Intelligenz geht auf Thorndike (1920) zurück. Er definiert soziale Intelligenz als, „the ability to understand and manage men and women, boys and girls - to act wisely in human relations“ (S.228). Sternberg (1997, 1999) versteht die soziale Intelligenz neben der akademischen und praktischen Intelligenz als eine grundlegende Intelligenz des Menschen. Thorndike selbst (1936) kritisierte die große Überlappung der sozialen mit der allgemeinen Intelligenz. Auch korrelieren soziale Intelligenztests nicht mit Selbst- und Fremdbeobachtungen von sozial kompetentem Verhalten (Schmidt, 1995). Die soziale Intelligenz wird als die kognitive Voraussetzung für sozial kompetentes Verhalten angesehen, welche die soziale Informationsverarbeitung steuert (Cantor & Harlow, 1994; Law et al., 2004).

Das Konzept der emotionalen Intelligenz stammt von Salovey und Mayer (1990) und wurde vor allem durch Goleman (1995) populär. Unter der emotionalen Intelligenz wird die Fähigkeit verstanden, mit den eigenen und den Emotionen anderer umzugehen (Salovey & Mayer, 1990). Als Subfacetten werden meist das Erkennen und Verstehen von eigenen und fremden Emotionen, die Emotionsregulation und -expression beschrieben (Mayer, Salovey & Caruso, 2000; 2003; Rindermann, 2009). Mit der sozialen Kompetenz gemein hat die emotionale Intelligenz die vier Komponenten *self-*

awareness, empathy, self-management und *interpersonal relationship skills* (Law et al., 2004).

Soziale Fertigkeiten sind spezifische erlernbare Fertigkeiten, die eingesetzt werden, um sozial kompetent zu handeln (Becker & Heimberg, 1988; Buhrmester, Furman, Wittenberg & Reis, 1988; Gresham, Sugai & Horner, 2001). Diese können kommunikative, kognitive und behaviorale Aspekte umfassen (Schneider & Byrne, 1985). Darunter fallen z.B. jemandem ein Kompliment zu machen, Augenkontakt zu halten, Gestik adäquat einzusetzen oder gar Fragen stellen zu können (Segrin, 2000; Sheffer, Penn & Cassisi, 2001). Im Gegensatz zu sozialen Fertigkeiten beinhalten soziale Kompetenzen eine Wertung des Verhaltens (McFall, 1982).

Interpersonale Kompetenzen richten den Fokus stärker auf die Beziehung zwischen Menschen. So verstehen z.B. Buhrmester et al. (1988) unter der interpersonalen Kompetenz bei Freundschaften die Fähigkeit, Interaktionen zu initiieren, sich selbst anderen Menschen gegenüber zu öffnen, andere Menschen emotional zu unterstützen, eigene Interessen zu vertreten und Konflikte friedfertig zu lösen. Der Begriff wird oft synonym zu sozialer Kompetenz verwendet.

3.2. Modelle sozialer Kompetenzen

Modelle sozialer Kompetenzen lassen sich einteilen in Modelle, die eher die zugrunde liegenden Dimensionen sozialer Kompetenzen betrachten und Modelle, die eher Prozesse sozialer Interaktionen abbilden. Erstere bestehen meist aus einer Auflistung oder Beschreibung von sozialen Kompetenzdimensionen. Letztere befassen sich mit dem Prozess der sozialen Informationsverarbeitung (z.B. Dodge, 1993) oder der Entwicklung sozialer Kompetenzen (z.B. Malti & Perren, 2008). Beispielhaft werden jeweils zwei Vertreter vorgestellt.

Das Modell von Caldarella und Merell (1997) besteht aus fünf Dimensionen sozialer Kompetenzen (Tabelle 6). Diese wurden auf Basis einer Metaanalyse von 21 Studien über soziale Kompetenzen von Kindern und Jugendlichen extrahiert.

Tabelle 6: Dimensionen sozialer Kompetenzen (Caldarella & Merell, 1997)

Dimension	Facetten
Fähigkeiten zur Bildung positiver Beziehungen zu Gleichaltrigen:	Perspektivenübernahme Hilfsbereitschaft Komplimente machen soziale Verantwortung Führungsfähigkeit
Selbstmanagementkompetenzen:	Kontrollieren der eigenen Stimmungslage Konfliktfähigkeit Kompromissfähigkeit Kritikfähigkeit Kooperationsfähigkeit Organisationsfähigkeit
Akademische Kompetenzen:	Zuhören um Hilfe bitten selbstständiges Arbeiten
Kooperative Kompetenzen:	Anerkennen sozialer Regeln Kritikfähigkeit Teilen
Selbstsicherheit und Selbstbehauptung:	Gespräche initiieren Gefühle ausdrücken auf andere zugehen

Payton et al. (2000) entwickelten das Rahmenmodell *Social and Emotional Learning (SEL)* zur Förderung sozialer Schlüsselkompetenzen von Schülern. Die vier Dimensionen des *SEL* mit insgesamt 17 untergeordneten Facetten stellt Tabelle 7 dar.

Tabelle 7: Framework of Social and Emotional Learning (Payton et al., 2000, S. 180)

Dimension und Facetten	Definition
Awareness of Self and Others	
Awareness of feelings:	The capacity to accurately perceive and label one's feelings.
Management of feelings:	The capacity to regulate one's feelings.
Constructive sense of self:	The capacity to accurately perceive one's strengths and weaknesses and handle everyday challenges with confidence and optimism.
Perspective taking:	The capacity to accurately perceive the perspective of others.

Fortsetzung Tabelle 7:

Dimension und Facetten	Definition
Positive Attitudes and Values	
Personal responsibility:	The intention to engage in safe and healthy behaviors and be honest and fair in dealing with others.
Respect for Others:	The intention to accept and appreciate individual group differences and to value the rights of all people.
Social responsibility:	The intention to contribute to the community and protect the environment.
Responsible Decision Making	
Problem identification:	The capacity to identify situations that require a decision or solution and assess the associated risks, barriers, and resources.
Social norm analysis:	The capacity to critically evaluate social, cultural, and media messages pertaining to social norms and personal behavior.
Adaptive goal setting:	The capacity to set positive and realistic goals.
Problem solving:	The capacity to develop, implement and evaluate positive and informed solutions to problems.
Social Interaction Skills	
Active listening:	The capacity to attend to others both verbally and non-verbally to demonstrate that they have been understood.
Expressive communication:	The capacity to initiate and maintain conversations and to clearly express one's thoughts and feelings both verbally and non-verbally.
Cooperation:	The capacity to take turns and share in both pairs and group situations.
Negotiation:	The capacity to consider all perspectives involved in a conflict in order to resolve the conflict peacefully and to the satisfaction of all involved.
Refusal:	The capacity to make and follow through with clear "No" statements, to avoid situations in which one might be pressured, and to delay acting in pressure situations until adequately prepared.
Help seeking:	The capacity to identify the need for support and assistance and to access available and appropriate resources.

Beiden Modelle weisen Überschneidungen in den Facetten auf. So enthalten sie z.B. die Fähigkeiten zur sozialen Verantwortungsübernahme, zur Kooperation und zur Perspektivenübernahme. Jedoch sind diese unterschiedlichen Dimensionen zugeordnet, und auch die Namen der Dimensionen unterscheiden sich. Daneben gibt es Facetten, die nur in einem der beiden Modelle vorkommen. So wird z.B. Hilfsbereitschaft nur im Modell von Caldarella und Merrell (1997) abgebildet, während die Fähigkeit, Nein sagen zu können nur im Modell von Payton et al. (2000) aufgenommen wurde.

Das prozessorientierte Modell von Holtz (1994) betrachtet soziale Kompetenzen als Entwicklungsaufgabe angelehnt an Havighurst (1982) sowie Eisenberg und Harris (1994). Soziale Kompetenzen bauen aufeinander auf und werden zunehmend komplexer (Abbildung 6).

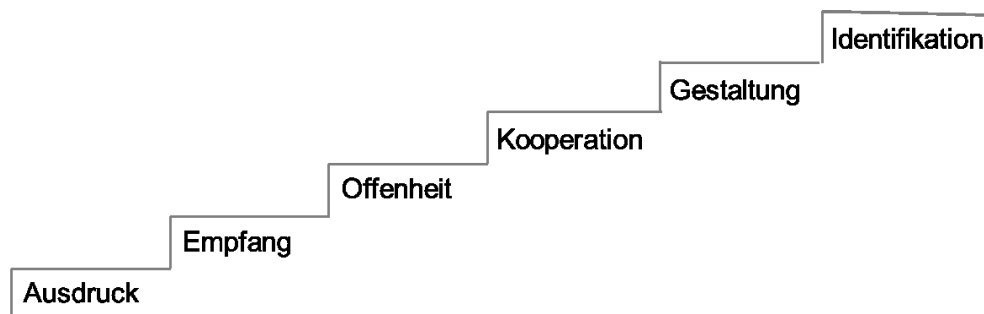


Abbildung 6. Modell sozialer Kompetenzen nach Holtz (1994)

Auf Basis der Ausdrucksfähigkeit zum Vermitteln der eigenen Bedürfnisse und Wünsche, folgt der Erwerb der Wahrnehmungsfähigkeit (Empfang). Hierunter fallen Fertigkeiten wie zuhören und beobachten, um soziale Situationen analysieren zu können. Danach folgt die Stufe der Offenheit, in der das Individuum lernt, sich mit anderen auseinander zu setzen und Kritik anzunehmen. Der Erwerb der Kooperationsfähigkeit stellt die nächste Stufe dar. Das Individuum lernt, eigene Handlungsmöglichkeiten und Verantwortlichkeiten wahrzunehmen und sich an die seiner Mitmenschen anzupassen. Darauf folgt der Erwerb der Gestaltung, in dem die Person Beziehungen initiiert, aufrecht erhält und sich in Gruppenprozessen zu Recht findet. Die komplexeste Stufe ist die Identifikation, in der die Person lernt, ein Gleichgewicht aus Durchsetzung und Anpassung zu finden, Möglichkeiten und Grenzen des eigenen Handelns einzuschätzen und Konflikte zu lösen. Die Entwicklung fokussiert sich zunächst auf das Individuum selbst und wendet sich nach und nach den Mitmenschen zu.

Das Drei-Ebenen Modell sozialer Kompetenzen (Abbildung 7; Malti & Perren, 2008) setzt intrapsychische Prozesse, soziales Verhalten und Verhaltenskonsequenzen miteinander in Beziehung. Es stellt ein Entwicklungs- und Prozessmodell sozialer

Kompetenz für die Kindheit und Adoleszenz dar und basiert auf konstruktivistischen und handlungstheoretischen Ansätzen. Im sozialen Setting laufen innerhalb einer Person (Ebene 1) sozial-kognitive (z.B. soziale Informationsverarbeitung), sozial-emotionale (z.B. Emotionskontrolle) und motivationale Prozesse (z.B. positive Selbstdarstellung) ab, bevor die Person handelt (Ebene 2). Die Person kann dabei primär selbstbezogen zur Befriedigung eigener Bedürfnisse oder primär fremdbezogen zur Befriedigung von Bedürfnissen anderer handeln. Die dritte Ebene bildet die Konsequenz der Handlung auf die Gesundheit und das Wohlbefinden des Individuums und die Qualität der sozialen Beziehung zu den Interaktionspartnern ab. Alle Ebenen wirken interaktiv, reziprok und sind dysmorph. Alle Ebenen werden durch Sozialisationsbedingungen und Persönlichkeitsmerkmale beeinflusst.

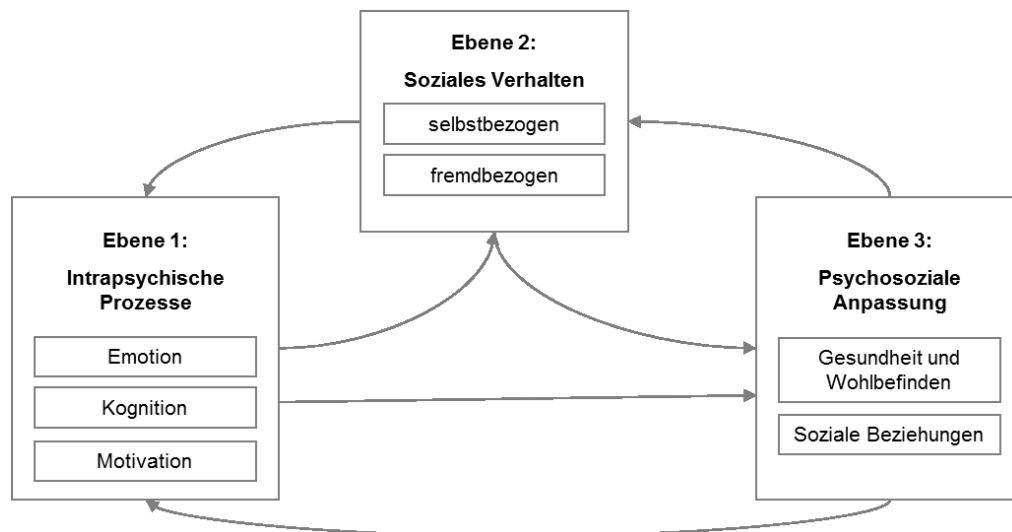


Abbildung 7. Drei-Ebenen Modell sozialer Kompetenzen (Malti & Perren, 2008).

Beiden prozessbezogenen Modellen ist gemein, dass sie sowohl Kompetenzen innerhalb der Person als auch zwischen Personen berücksichtigen. Während bei Holtz (1994) die Entwicklung sozialer Kompetenz beim Erreichen der letzten Stufe abgeschlossen scheint, entwickelt sie sich bei Malti und Perren (2008) durch Rückkopplungsprozesse kontinuierlich weiter.

Aus allen Modellen lässt sich ableiten, dass soziale Kompetenzen durch die Po-le Durchsetzungsfähigkeit vs. Anpassung (selbstbezogen vs. fremdbezogen) geprägt sind, wie es auch in den Definitionen sozialer Kompetenzen abgebildet ist. Die Modelle unterscheiden enthalten Dimensionen bzw. Prozesse innerhalb und zwischen Personen. Dabei stellen die Dimensionen bzw. Prozesse innerhalb einer Person meist eine Voraussetzung zum sozial kompetenten Handeln dar. Hierunter fallen u.a. die soziale

Wahrnehmungsfähigkeit und die Perspektivenübernahme. Petermann & Petermann (1989) fügen ein positives Selbstkonzept als Voraussetzung für sozial kompetentes Verhalten hinzu. Soziale Kompetenzen beinhalten kognitive, emotionale und motivationale Anteile. Sie werden mit verschiedenen Begrifflichkeiten und auf unterschiedlichen Abstraktionsniveaus beschrieben und je nach Modell unterschiedlichen Dimensionen zugeordnet. Soziale Kompetenzen scheinen zudem unterschiedlich komplex (Holtz, 1994). Diese Ableitungen sollen in einem Arbeitsmodell sozialer Kompetenzen als Grundlage für die Dissertation einfließen. Zuvor wird jedoch noch auf spezifische soziale Kompetenzen in der Medizin eingegangen.

3.3. Soziale Kompetenzen in der Medizin

Der einflussreichste Ansatz zur Integration nicht-kognitiver Kompetenzen in die medizinische Ausbildung stellt das *CanMed Framework* (Frank et al., Entwurf 2015; Abbildung 8) dar, welches seit 1996 kontinuierlich weiterentwickelt und adaptiert wird. Ziel des Frameworks ist es, eine hochqualitative Patientenversorgung und -sicherheit durch eine verbesserte Ausbildung zu gewährleisten. Es basiert auf Kompetenzen, die für alle medizinischen Fachrichtungen notwendig sind. Den Kompetenzen werden sieben Rollen zugeordnet, die nun in Hinblick auf soziale Kompetenzen erläutert werden.

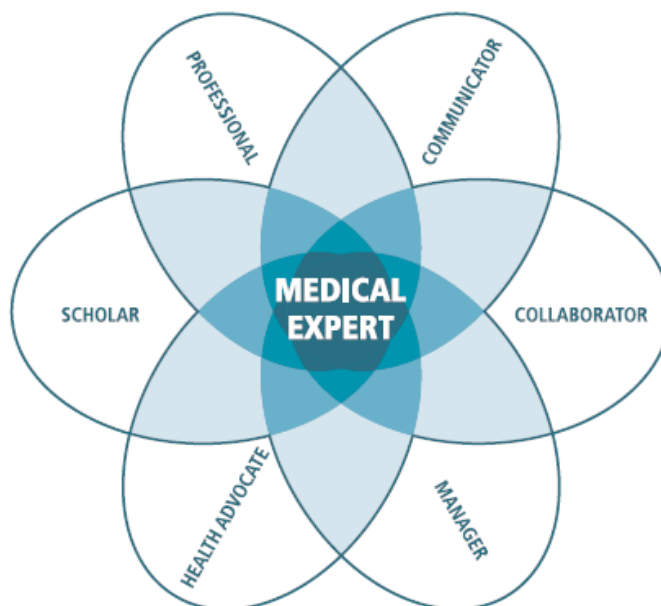


Abbildung 8. Draft CanMed Framework (Frank et al., Entwurf, 2015)

Die Kernaufgabe des *Communicators* liegt in der effektiven Kommunikation mit Patienten und deren Angehörigen. Die Ausführung der Rolle setzt eine verbale und nonverbale Kommunikationsfähigkeit, aktives Zuhören, Empathie, respektvolles Verhalten, ethisch-moralisches Handeln und den Umgang mit eigenen Emotionen voraus. Die Rolle des *Collaborators* liegt in der effektiven Zusammenarbeit in disziplinären und interdisziplinären Teams. Wichtige soziale Kompetenzen des *Collaborators* sind das Reflektieren der eigenen Rolle, der eigenen Grenzen und der eigenen Arbeitsweise. Zudem erfordert die Rolle, die Perspektive anderer Gruppenmitglieder einzunehmen, Verantwortung zu übernehmen und über ein konstruktives Konfliktmanagement zu verfügen. Die Rolle des *Managers (Leader)* setzt vor allem Führungsfähigkeiten und Verantwortungsübernahme voraus, um definierte Ziele zu erreichen. Die Rolle des *Health Advocate* bezieht sich auf die Förderung der Gesundheit in der Bevölkerung. Dies impliziert eine Kooperation mit Partnern und einen verantwortungsvollen Umgang mit Ratsuchenden. Die Rolle des *Scholars* beinhaltet lebenslanges Lernen der eigenen Person und die Weitergabe dieses Verhaltens an Auszubildende. Die Rolle erfordert ein hohes Maß an Selbstreflexion der eigenen Arbeitsweise und -ergebnisse, das Erkennen der eigenen Grenzen und Weiterbildungsmöglichkeiten, die Bitte um Unterstützung von kompetenten Personen und das Geben und Annehmen von Feedback. Die Rolle des *Professionals* bildet ein professionelles verantwortungsvolles Verhalten zwischen Arzt und Patienten ab, das durch moralisch-ethisches Handeln, Integrität, Ehrlichkeit, Altruismus und Respekt geprägt ist. Zur adäquaten Ausübung der Rolle sollten selbstregulatorische Fähigkeiten vorhanden sein, um die eigenen Gedanken, Gefühle und Verhaltensweisen zu beobachten und zu bewerten. Der *Medical Expert* integriert alle sechs Rollen in sein ärztliches Handeln.

Zahlreiche andere Rahmenwerke beschreiben ähnliche soziale Kompetenzen als essentiell für die medizinische Ausbildung und den Arztberuf (z.B. von Fragstein et al., 2008; Kiessling et al., 2010; Makoul, 2001; Makoul & Schofield, 1999; Simpson et al., 1999). Der Schwerpunkt liegt bei allen auf der Arzt-Patienten-Kommunikation. Die erarbeiteten Kompetenzen basieren auf Forschungsergebnissen zur Patientensicherheit und Versorgungsqualität und/oder wurden durch Konsensusmethoden definiert. Die dargestellten sozialen Kompetenzen werden über den gesamten Zeitraum der medizinischen Aus- und Weiterbildung bis zum Erlangen des Facharztes und darüber hinaus während der Berufstätigkeit erworben und sind bei Studienplatzbewerbern noch nicht beobachtbar. Sie sollen aber als Grundlage dienen, Studienplatzbewerbern und -interessenten ein Bewusstsein für die Wichtigkeit sozialer Kompetenzen in der Medizin zu vermitteln und ihre eigenen Kompetenzen und individuellen Weiterentwicklungsmöglichkeiten zu reflektieren.

3.4. Arbeitsmodell sozialer Kompetenzen

Als Basis für die Entwicklung eines SJTs für soziale Kompetenzen für die medizinische Studierendenauswahl (Kapitel 6) wurde ein eigenes Arbeitsmodell sozialer Kompetenzen entworfen (Abbildung 9). Ein Grund für dessen Entwicklung lag darin, dass bis heute keine allgemein akzeptierte Definition und keine allgemein akzeptierten Modelle sozialer Kompetenzen vorliegen (Döpfner et al., 1981; Kauffeld et al., 2002; Mathews et al., 2002). Außerdem sollte aufgrund der Kontextabhängigkeit (Albrecht, 1997; Holtz, 1994; Reschke, 1995) für die Medizin relevante Kompetenzen im Modell berücksichtigt werden.

Das Arbeitsmodell betrachtet soziale Kompetenzen als multimodales Konstrukt, das sich aus mehreren Abstraktionsebenen zunehmender Komplexität zusammensetzt. Im Zentrum des Arbeitsmodells stehen die intrapersonellen Kompetenzen (Ebene 0). Diese umfassen soziale Kompetenzen, die nur die eigene Person betreffen und eine Voraussetzung für die Kompetenzen höherer Ebenen darstellen. Auf die intrapersonellen Kompetenzen bauen die interpersonellen Kompetenzen der Ebenen eins bis vier auf. Diese sind notwendig, um mit anderen Personen sozial kompetent zu interagieren. Die Komplexität der Kompetenzen nimmt von Ebene zu Ebene zu.

Das Konzept für das Arbeitsmodell geht darauf zurück, dass Modelle sozialer Kompetenzen in den Abstraktionsniveaus der Dimensionen schwanken (Kanning, 2003; Schneider et al., 1996) und unterschiedlich komplex sind (Holtz, 1994). Die Komplexität sozialer Kompetenzen scheint zuzunehmen, wenn soziale Kompetenzen nicht nur innerhalb einer Person, sondern auch zwischen Personen Anwendung finden (Holtz, 1994). Folglich wurden die intrapersonellen Kompetenzen als Voraussetzung für die interpersonelle sozialen Kompetenzen in den Kern des Arbeitsmodells gerückt. Diese umfassen sozial-emotionale, -kognitive und -motivationale Kompetenzen (Malti & Perren, 2008). Je nach Autoren werden verschiedene intrapersonelle Kompetenzen aufgezählt. Payton et al. (2000) nennen z.B. die Perspektivenübernahme, Petermann und Petermann (1989) ein positives Selbstkonzept, Malti und Perren (2008) die soziale Wahrnehmungsfähigkeit und Emotionskontrolle. Aus der medizinischen Perspektive können selbstregulatorische Kompetenzen wie Selbstbeobachtung/Selbstreflexion und Selbstkontrolle/Selbstdisziplin hinzugefügt werden, die in den *CanMed*-Rollen des *Professionals*, *Collaborators* und *Scholars* abgebildet sind (Frank et al., Entwurf 2015). Das Arbeitsmodell versucht die verschiedenen Begrifflichkeiten zu integrieren.

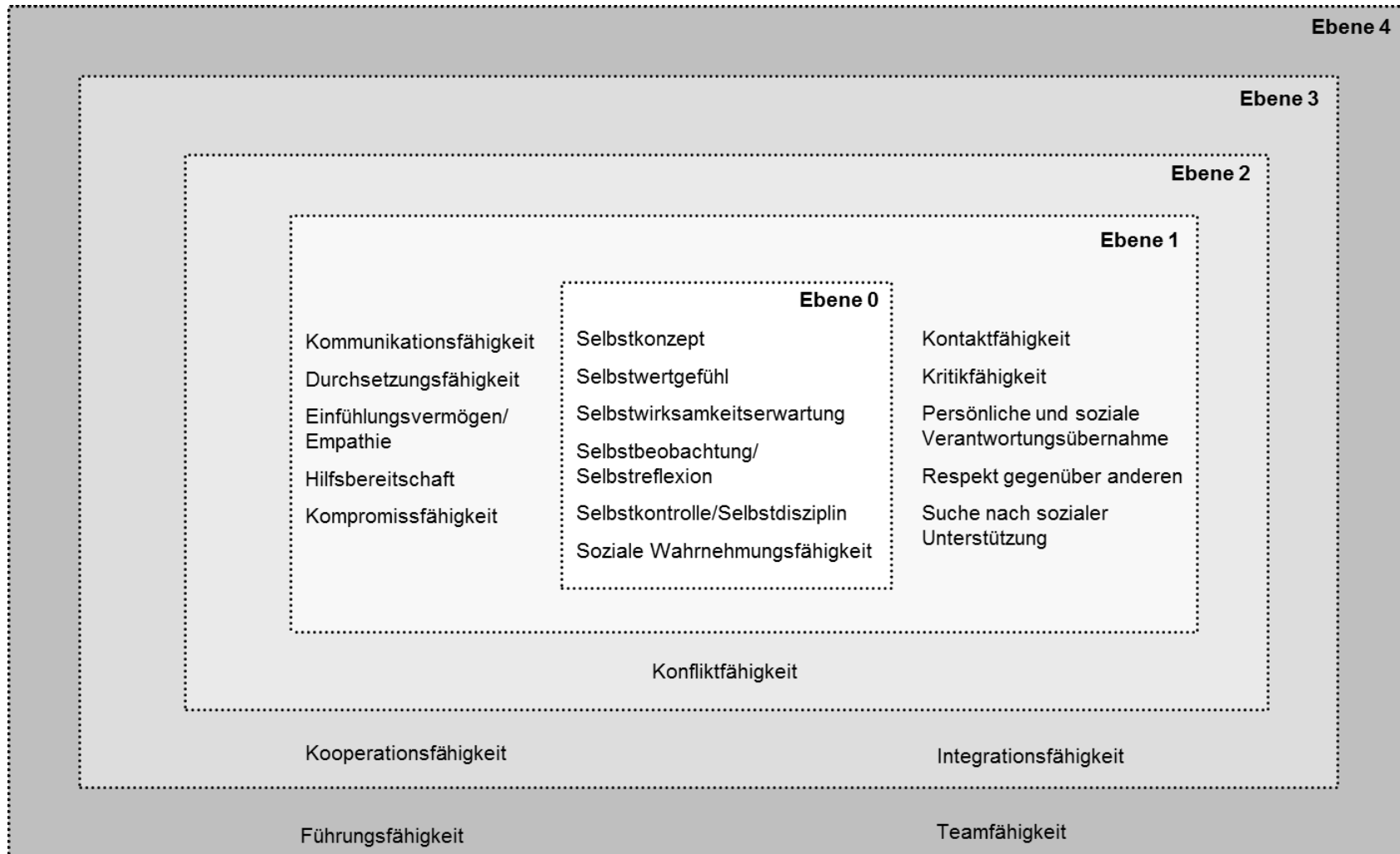


Abbildung 9. Eigenes Arbeitsmodell sozialer Kompetenzen.

Für die Zusammenstellung der interpersonellen Kompetenzen (Ebene 1-4) flossen insbesondere die Modelle von Caldarella und Merrell (1997), Frank et al. (Entwurf 2015) und Payton et al. (2000) ein. Auch die interpersonellen Kompetenzen wurden nach ihrer Komplexität angeordnet (Vgl. Holtz, 1994).

In Anlehnung Malti und Perren (2008) sind die Ebenen durchlässig und interaktiv (gepunktete Linien zwischen den Ebenen in Abbildung 9). Selbst innerhalb der Abstraktionsebenen sind weitere Unterordnungen denkbar. So könnte z.B. Empathie auch als Voraussetzung für Kommunikationsfähigkeit oder für Hilfsbereitschaft gewertet werden. Das Arbeitsmodell erhebt kein Anspruch auf Vollständigkeit und ist heuristisch einzuordnen. Die Definitionen der im Modell enthaltenen sozialen Kompetenzen enthält Anhang A.

3.5. Messung sozialer Kompetenzen

Seit den 1920er Jahren orientierte sich die Messung sozialer Kompetenz an dem Ansatz von Thorndike (1920), der zwischen der sozialen Sensitivität und der sozialen Handlungskompetenz unterscheidet. Die soziale Sensitivität gibt an wie gut eine Person andere Menschen versteht. Sie ist eine notwendige, aber keine hinreichende Bedingung für die soziale Handlungskompetenz. Die soziale Handlungskompetenz beinhaltet wie gut eine Person soziale Problemsituationen löst. Die Erfassung der Sensitivität durch Zeichnungen, Hörproben oder Filme scheiterte am Nachweis der konvergen-ten Validität (Amelang, 1987; Rosenthal, Hall, Archer, DiMatteo & Rogers, 1979). Zur Erhebung der sozialen Handlungskompetenz werden drei Ansätze herangezogen (A-sendorf & Neyer, 2012; Bastians & Runde, 2002).

Ein Ansatz liegt in der Selbst- und Fremdbeurteilung sozialer Fertigkeiten via Fragebogen. Dieser Ansatz stellt keine direkte Kompetenzmessung dar, sondern erfasst die selbstwahrgenommene soziale Kompetenz, die je nach Situation und Person anders gefärbt sein kann. Von dieser wird auf die soziale Kompetenz einer Person geschlossen (Erpenbeck, 1997; Riemann & Allgöwer, 1993). Vorteile der Methode liegen in der Ökonomie, der Standardisierung und darin, dass Personen durch die Selbstbeschreibung eigenes Weiterentwicklungspotenzial aufdecken können (Bastians & Runde, 2002). Während Sonntag und Schäfer-Rauser (1993) Beschönigungstendenzen als Nachteil von Selbsteinschätzungsverfahren aufführen, sehen andere Autoren die Selbstdarstellung als eigene soziale Kompetenz (Hogan, Barrett & Hogan, 2007; Kanning, 2009) und heben den Zusammenhang mit der Arbeitsleistungen hervor (Morgeson et al., 2007). Seyfried (1995) kritisiert, dass bei Selbst- und Fremdeinschätzungen unklar ist, ob eine Person ihre Einstellung, ihre Handlungsweise oder ihre Fähigkeit

beurteilt. Bergmann (2007) argumentiert demgegenüber, dass Kompetenzmessungen durch Selbsteinschätzung durchaus valide sind, wenn sie anonym, dimensions- und verhaltensorientiert erhoben werden und die Person Erfahrung in der Selbsteinschätzung mitbringt.

Der zweite Ansatz ist die Simulation von sozialen Problemsituationen durch verbale, schriftliche oder videobasierte Situationsbeschreibungen. Aus der Reaktion einer Person werden Rückschlüsse auf das Handeln in zukünftigen Situationen gezogen (*Behavioral Consistency Theory*, Campbell, 1963). Ein bekanntes Verfahren aus der Personalauswahl ist das *Situational Interview* (Latham, Saari, Pursell & Campion, 1980), bei dem den Bewerbern eine konkrete Situation beschrieben wird. Diese müssen daraufhin schildern, wie sie sich in dieser Situation verhalten würden. Eine weitere dazugehörige Verfahrensgruppe sind *SJTs*, auf die in Kapitel 4 ausführlich eingegangen wird.

Der dritte Ansatz ist die Verhaltensbeobachtung, mit Hilfe derer das Sozialverhalten einer Person im künstlichen oder realen Rollenspiel beurteilt wird. Aufgrund der Beobachtung von konkreten Verhaltensweisen wird auf zugrunde liegende Konstrukte geschlossen. Um die Einflüsse einer konkreten Situation ausschließen zu können, sollten Verhaltensbeobachtungen immer über mehrere Situationen vorgenommen werden (Kanning, 2003). Die Beobachtung von sozialen Verhalten wurde in den 1970er und 1980er Jahren vor allem im klinischen Bereich in standardisierten Settings durchgeführt (z.B. de Mynck & Ullrich, 1987). In der Personalauswahl etablierte sich das *Assessment Center* als standardisiertes Beobachtungsverfahren für soziale Kompetenzen (Asendorpf & Neyer, 2012). Die Verhaltensbeobachtung gilt als am besten geeignete Methode zur Erfassung sozialer Kompetenzen (Goldfried & D'Zurilla 1969; Kauffeld et al., 2002). Dies geht darauf zurück, dass Kompetenzen „(...) nicht direkt prüfbar sind, sondern nur aus der Realisierung der Dispositionen erschließbar und evaluierbar“ sind (Erpenbeck, 1997, S.311). Der Nachteil der Beobachtung liegt allerdings im Ressourcenaufwand.

Darüber hinaus führt Kanning (2003) Kompetenzindikatoren als weiteren Ansatz zur Messung sozialer Kompetenzen an. Kompetenzindikatoren repräsentieren Konsequenzen des Sozialverhaltens, von denen wiederum auf entsprechende soziale Kompetenzen rückgeschlossen wird. Wenn eine Person z.B. über ein gut funktionierendes soziales Netzwerk verfügt, so wird ihr eine höhere soziale Kompetenz zugeschrieben als einer Person, die über kein funktionierendes soziales Netzwerk verfügt.

4. Der Situational Judgement Test

Zur Ergänzung der medizinischen Studierendenauswahl in Heidelberg soll ein *SJT* entwickelt werden, der in der Lage ist, soziale Kompetenzen von Studienbewerbern zu erheben und gleichzeitig über die Anforderungen des Medizinstudiums an die eigenen sozialen Kompetenzen zu informieren. Die folgenden Abschnitte stellen die Grundlagen für die Planung und Durchführung der eigenen *SJT*-Entwicklung in Kapitel 5 und 6 dar, auf die auch bei der Hypothesengenerierung und den Diskussionen der Validierungsstudien (Kapitel 7 und 8) Bezug genommen wird.

Kapitel 4.1 schildert zunächst Ursprung, Aufbau und Funktionsweise von *SJT*s, bevor in Kapitel 4.2 auf deren Messtheorie eingegangen wird. Die Schritte bei der *SJT*-Entwicklung beschreibt Kapitel 4.3, welche unmittelbar die Gütekriterien von *SJT*s beeinflussen (Kapitel 4.4). Kapitel 4.5 gibt anschließend einen kurzen Abriss über die Vor- und Nachteile des Verfahrens. Abschließend stellt Kapitel 4.6 den derzeitigen Einsatz von *SJT*s in der medizinischen Auswahl vor.

4.1. Ursprung, Aufbau und Funktionsweise

Der *SJT* ist ein in der Personalpsychologie eingesetztes Auswahlverfahren, welches Bewerbern arbeitsbezogene Situationen in mündlicher, schriftlicher oder videobasierter Form präsentiert, um eine möglichst realistische Vorschau auf die Arbeitstätigkeit zu geben. Die Aufgabe des Bewerbers besteht darin, aus mehreren Antwortalternativen zu einer präsentierten Problemsituation die Handlung auszuwählen, die er für am effektivsten hält (wissensbezogene Instruktion) bzw. am ehesten ausführen würde (verhaltensbezogene Instruktion). Ein Beispiel-Item aus dem Tätigkeitsfeld eines Managers (Motowidlo et al., 1990, S. 642) zeigt Abbildung 10.

*SJT*s besitzen einen starken intuitiven Aufforderungscharakter, bei dem die Testteilnehmer gebeten werden, sich in eine bestimmte Situation hineinzusetzen (Stemler & Sternberg, 2006). Schriftliche *SJT*s werden den *Low Fidelity* Verfahren zugeordnet (Gessner & Klimoski, 2006; Hauenstein, Findlay & McDonald, 2010). Sie bilden realistische Szenarien ab. Der Testteilnehmer muss eine Auswahl bezüglich der Antwortoptionen treffen, aber nicht direkt mit einem Verhalten reagieren. Die Funktionsweise des *SJT*s basiert auf der Behavioral Consistency Theory (Campbell, 1963; Motowidlo et al., 1990). Sie besagt, dass früheres Verhalten der beste Prädiktor für zukünftiges Verhalten ist. Videobasierte *SJT*s weisen durch die verfilmten Situationsbeschreibungen eine mittlere *Fidelity* auf (Funke & Schuler, 1998).

You and someone from another department are jointly responsible for coordinating a project involving both departments. This other person is not carrying out his share of the responsibilities. You would. . . Most Likely/Least Likely

- a) Discuss the situation with your manager and ask him to take it up with the other person's manager.
- b) Remind him that you need his help and that the project won't be completed effectively without a full team effort from both of you.
- c) Tell him that he is not doing his share of the work, that you will not do it all yourself, and that if he doesn't start doing more; you'll be forced to take the matter to his manager.
- d) Try to find out why he is not doing his share and explain to him that this creates more work for you and makes it harder to finish the project.
- e) Get someone else from his department to help with the project.

Abbildung 10. *SJT*-Item aus dem Management Bereich.

SJTs können für ein breites Spektrum an Konstrukten eingesetzt werden (McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001). Überwiegend werden Führungsfähigkeiten (38%), gemischte Konstrukte (33%) des Services, Verkaufs und Managements, soziale Fähigkeiten (13%), Persönlichkeitseigenschaften (10%), Teamwork (4%) und arbeitsplatzbezogenem Wissen (3%) in *SJTs* abgebildet (Christian, Edwards & Bradley, 2010; Polyhart & McKenzie, 2011). Der Ursprung des modernen *SJTs* mit geschlossenem Antwortformat (*Multiple Choice*) liegt in den 1920er Jahren. Der erste bekannte und breit eingesetzte *SJT* ist der *George Washington Social Intelligence Test* zur Erfassung von Urteilsfähigkeit in sozialen Situationen (McDaniel et al., 2001; Moss, 1926). Die ersten *SJTs* wurden häufig aufgrund der hohen Korrelation zu kognitiven Fähigkeiten und ihrer Multidimensionalität kritisiert (File & Remmers, 1971; Northrop, 1989; Thorndike & Stein, 1937). Einen neuen Aufschwung erlebte der *SJT* in den 1990er Jahren, als die Arbeitsgruppe um Motowidlo (1990) den *SJT* als *Low Fidelity* Verfahren neu einführte und dessen prädiktive Validität von Arbeitsverhalten belegte (Polyhart & McKenzie, 2011).

4.2. Was messen SJTs?

Dieser Frage gehen Autoren seit den 1990er Jahren nach. Ein Ansatz besteht darin, dass *SJTs* kontextspezifisches Wissen einer Person erfassen (Weekley & Jones, 1997; McDaniel & Nguyen, 2001). Schmidt und Hunter (1993) vermuten, dass *SJTs* die generelle kognitive Fähigkeit, arbeitsplatzbezogenes Wissen oder die Fähigkeit, situative Entscheidungen zu treffen, messen. Brooks und Highhouse (2006) gehen auf die Entscheidungsfähigkeit genauer ein und behaupten, dass *SJTs* quasi-rationales Entscheiden in domänenspezifischen Dilemmata messen. Unter quasi-rationalen Entscheidungen verstehen sie die intuitiven und analytischen Elemente, die bei der Entscheidungsfindung zum Tragen kommen (Hammond, 1955). Motowidlo et al. (1990) behaupten, dass *SJTs* Verhaltensintentionen einer Person messen. Wiederum andere gehen davon aus, dass *SJTs* multidimensionale Messmethoden sind, die ähnlich wie *Assessment-Center* und Interviews mehrere latente Merkmale gleichzeitig erfassen (z.B. McDaniel et al., 2001; Polyhart, 2006). Nachfolgend werden zwei der einflussreichsten Ansätze vorgestellt – die Theorie des Tacit Knowledge (Sternberg, 1997, 1999) und die Theorie des prozeduralen Wissens (Motowidlo, Hooper & Jackson, 2006).

Tacit Knowledge Theory. Die Gruppe um Sternberg geht davon aus, dass *SJTs* *Tacit Knowledge* erfassen (Sternberg, 1997, 1999; Stemler & Sternberg, 2006). Zum besseren Verständnis des Begriffs muss zunächst kurz auf Sternbergs Intelligenztheorie (Abbildung 11, 1988, 1997, 1999) eingegangen werden. Diese beschreibt die kreative, die analytische und die praktische Intelligenz als drei übergeordnete Intelligenzformen. Unter der kreativen Intelligenz versteht Sternberg (1988), die Fähigkeit Ideen für Problemlöseprozesse zu generieren, während die analytische Intelligenz notwendig ist, um die Umsetzbarkeit der Ideen zu prüfen. Die praktische Intelligenz kommt schließlich zum Tragen, wenn die geprüften Ideen in die Realität umgesetzt werden.

Tacit Knowledge ist eine Komponente der praktischen Intelligenz. Die praktische Intelligenz beinhaltet eine kognitive und eine behaviorale Komponente. Die kognitive Komponente umfasst explizites Wissen und implizites Wissen (*Tacit Knowledge*) über effektives Handeln in Alltagssituationen. Explizites Wissen wird formal z.B. in der Schule gelernt und kann verbal wiedergegeben werden. Implizites Wissen wird dagegen durch Erfahrungen angeeignet und ist nur schwer oder gar nicht artikulierbar (Anderson, 1983).

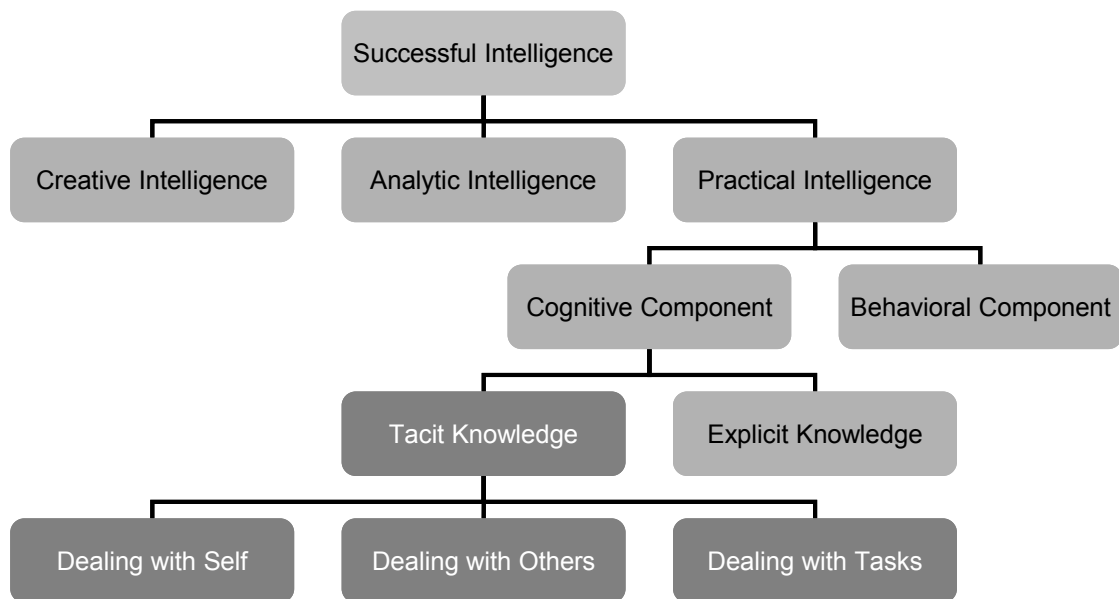


Abbildung 11. Tacit Knowledge Theory (Sternberg, 1997, 1999; Stemler & Sternberg, 2006).

Tacit Knowledge ist das Vermögen einer Person, die Problemlösungsstrategie auszuwählen, die auf eine kontextspezifische Situation passt. Dies betrifft Situationen im Umgang mit der eigenen Person, mit anderen Personen und im Umgang mit Aufgaben (Stemler et al., 2006). *Tacit Knowledge* stellt somit kontextspezifisches Wissen über adäquates Verhalten dar und bezieht sich lediglich auf alltägliche Problemsituationen.

Theory of SJT procedural knowledge. Ein verwandter Ansatz zum zugrunde liegenden Mechanismus von *SJTs* ist die Theorie des prozeduralen Wissens (Motowidlo et al., 2006). Wie der Name vermuten lässt, gehen die Autoren davon aus, dass *SJTs* prozedurales Wissen erfassen. Darunter verstehen sie das Wissen einer Person über effektives Verhalten in Situationen. Dabei messen *SJTs* unabhängig von der Instruktionsform prozedurales Wissen. Wissensbasierte Instruktionen fragen nach dem effektivsten Verhalten in einer Situation und erheben somit das Wissen einer Person über effektive Verhaltensweisen (vgl. auch: Weekley & Jones, 1997; McDaniel & Nguyen, 2001). Verhaltensbasierte Instruktionen fragen nach der präferierten Handlungsintention einer Person. Insbesondere in Bewerbungssituationen wird der Testteilnehmer abwägen, welche Verhaltensweise die effektivste für die dargebotene Situation ist. Folglich misst die verhaltensbasierte Instruktion ebenfalls prozedurales Wissen. Dabei erfassen verschiedene *SJTs* je nach inhaltlichem Schwerpunkt des jeweiligen *SJTs* unterschiedliches prozedurales Wissen. Motowidlo et al. (2006) untersuchten nicht nur die Frage, welche Wissensart *SJTs* messen, sondern analysierten auch die vermitteln-

den Prozesse zwischen der Persönlichkeit eines Individuums (*Traits*) und seinem prozeduralen Wissen (Punktzahl im *SJT*). Dafür führten sie den Begriff der *Implicit Trait Policies* ein (*ITPs*; Abbildung 12).

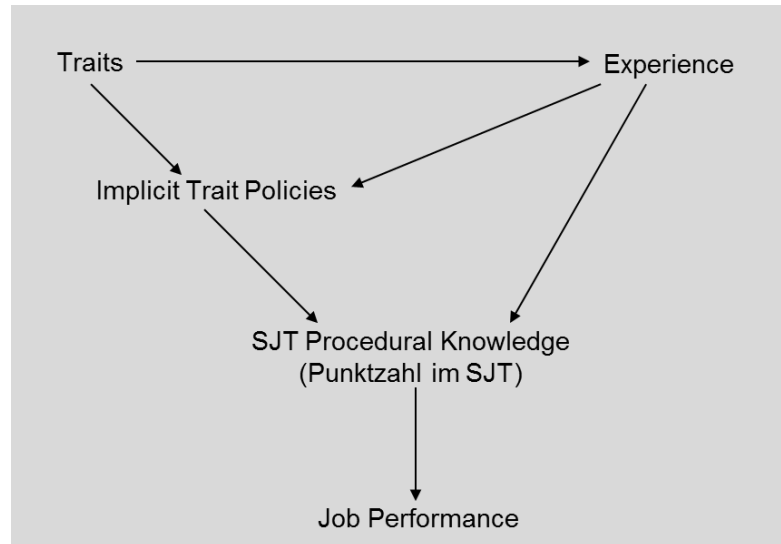


Abbildung 12. Theory of *SJT* procedural knowledge (Motowidlo et al., 2006, S. 61).

Diese definieren sie als „(...) implicite beliefs about the effectiveness of different level of trait expression“ (S.57). Personen sind dabei nur schwer oder gar nicht in der Lage, diese Annahmen in Worte zu fassen (Anderson, 1983). Im *SJT* bewerten Personen mit hoher Ausprägung in einer Persönlichkeitseigenschaft die Verhaltensweisen als effektiver, die ein höheres Ausmaß dieser Persönlichkeitseigenschaft repräsentieren. So würde eine gewissenhafte Person das Einhalten eines Projektplans als effektiver bewerten als eine weniger gewissenhafte Person und in einem *SJT* für Gewissenhaftigkeit die Handlungen bevorzugen, die ein hohes Maß an Gewissenhaftigkeit widerspiegeln. Die Autoren behaupten zudem, dass Personen mit hoher Ausprägung in einer Persönlichkeitseigenschaft besser zwischen sehr effektiven und wenig effektiven Verhaltensoptionen bezüglich dieser Persönlichkeitseigenschaft im *SJT* differenzieren. Dies führen sie darauf zurück, dass die Personen über ein großes prozedurales Wissen zu dieser Persönlichkeitseigenschaft verfügen. Die *ITPs* werden durch Erfahrungen beeinflusst, die wiederum von den Persönlichkeitseigenschaften einer Person gelenkt werden. So suchen Personen bevorzugt Situationen auf, in denen sie ihre Persönlichkeitsausprägung gut ausleben können. Beispielsweise wird die gewissenhafte Person eher eine Arbeitsstelle bevorzugen, die das Einhalten eines Projektplans als wichtig erachtet. Hält sie den Projektplan beim Arbeiten ein, wird sie durch den Vorgesetzten verstärkt. Diese Erfahrung wirkt sich positiv auf ihre Annahme über effektives Verhalten aus und verstärkt ihr Arbeitsverhalten.

Ein Konsens zu einem gemeinsamen theoretischen Hintergrund für alle *SJT*s existiert bisher nicht (Stemler & Sternberg, 2006, McDaniel & Nguyen, 2001; Weekley & Jones, 1999). Als Gemeinsamkeiten der aufgeführten Theorien können das prozedurale kontextspezifische Wissen und die situative Entscheidungsfähigkeit festgehalten werden. *SJT*s weisen inkrementelle Validität über Persönlichkeitseigenschaften und kognitive Fähigkeiten auf, sodass sie vermutlich Aspekte messen, die nicht mit Persönlichkeitsfragebögen oder mit allgemeinen kognitiven Fähigkeitstests erfasst werden können. Faktorenanalysen liefern meist einfaktorielle Lösungen, welche nur einen geringen Prozentsatz der Gesamtvarianz wiedergeben (Schmitt & Chan, 2006). Zur Veranschaulichung der Größenordnung soll die Studie von Gillespie, Oswald, Schmitt, Mannheim und Kim (2002) herangezogen werden. In dieser zeigte die Faktorenanalyse für einen zwölf Konstrukte umfassenden *SJT* einen Faktor mit 12% Varianzaufklärung, einen zweiten mit 5% Varianzaufklärung, gefolgt von vielen Faktoren mit geringerem Aufklärungsgehalt. Folglich kann zwar von einer Art Generalfaktor für den *SJT* ausgegangen werden, aber der Großteil der Varianz wird durch viele andere spezifische Faktoren erklärt (Schmitt & Chan, 2006).

4.3. Entwicklung eines *SJT*s

Die Entwicklung eines *SJT*s erfolgt in fünf Schritten (Weekley, Polyhart & Holtz, 2006; Abbildung 13). Zuerst werden die Situationen (Schritt 1) und die Antwortoptionen (Schritt 2) generiert. Die folgenden Schritte sind stark miteinander verknüpft. In diesen wird die Instruktionsform (Schritt 3) in Abhängigkeit von der Bewertungsmethode (Schritt 4) und dem Bilden des Auswertungsschlüssels (Schritt 5) festgelegt. Bei video-basierten *SJT*s schließt sich als letzter Schritt die Verfilmung an. Die Schritte werden im Folgenden näher erläutert. Die Beschreibung der herkömmlichen *SJT*-Entwicklung folgt Polyhart und MacKenzie (2011). Die Variationen der unterschiedlichen Testentwicklungsaspekte sind dem Artikel von Weekley et al. (2006) entnommen. Auf weitere Autoren wird im Text ausdrücklich verwiesen.

SJT Entwicklung				
Situation	Antwortoption	Instruktion	Bewertungs- methode	Antwort- effektivität
- Quelle - Komplexität - Realitätsnähe - Inhalt	- Quelle - Konstruktbezug - Komplexität - Realitätsnähe - Verfälschbarkeit	Verhaltens- basiert vs. Wissensbasiert	Forced Choice vs. kontinuierlich	rationaler, empirischer, theoretischer oder hybrider Auswertungsschlüssel

Abbildung 13. Schritte und Aspekte der SJT-Entwicklung (in Anlehnung an Weekley et al., 2006, S.158).

Generieren der Situationen. Zwei Ansätze werden gewöhnlich zum Generieren von Situationen verwendet. Der induktive Weg erfolgt über das Sammeln von erfolgskritischen Arbeitssituationen nach der *Critical Incident Technique* (Flanagan, 1954). Stelleninhaber sammeln typische herausfordernde Arbeitssituationen, um zwischen geeigneten und weniger geeigneten Personen differenzieren zu können. Die erfolgskritischen Situationen sind nach dem Drei-Satz: Antezedenz, Verhalten und Konsequenz aufgebaut. Die Antezedenz stellt den Auslöser für die kritische Situation dar und führt zum aktuellen Verhalten einer Person, welches wiederum Konsequenzen hervorruft. Die Antezedenz wird im SJT als Itemstamm ausformuliert, die Verhaltensalternativen werden in die im nächsten Schritt beschriebenen Antwortoptionen überführt. Die Konsequenz bildet sich in der Bewertung der Handlung als Punktwert ab.

Beim deduktiven Weg werden theoriebasiert Konstrukte hergeleitet, die für die Tätigkeit maßgeblich sind. Auf deren Grundlage werden für die Itementwicklung erfolgskritische Situationen abgeleitet. Bei der Situationsentwicklung können die Quelle und die Komplexität variieren. Die Quelle kann im induktiven und deduktiven Ansatz liegen und je nach Zusammensetzung der Expertengruppe variieren. In der Expertengruppe sollten die Personen vertreten sein, deren Perspektive auf die Tätigkeit relevant ist und in die SJT-Entwicklung eingehen soll. Die Komplexität der Situation betrifft die Sprache, Detailliertheit und Länge der Situationsbeschreibung. Derzeit ist nicht geklärt, welche Komplexität angemessen ist. Detailliert formulierte Itemstämme scheinen die Validität von SJTs zum Teil zu erhöhen (z.B. Reynolds, Sydell, Scott & Winter, 2000), zum Teil aber auch zu senken (McDaniel et al., 2001). Komplex formulierte Itemstämme können die Anforderung an die Lesekompetenz erhöhen und Einflüsse auf die Testfairness bezüglich ethnischer Subgruppen hervorrufen (z.B. Sacco, Scheu, Ryan & Schmitt, 2000a). Ein weiterer Aspekt, der bei der Itemstamentwicklung berücksichtigt werden sollte, ist die physische und psychische Realitätsnähe (Fidelity). Die physische Realitätsnähe bezieht sich darauf, ob der SJT die gleichen Verhaltensoperationen er-

fordert, die auch die reale Tätigkeit abverlangen würde. Die psychische Realitätsnähe betrachtet, ob der *SJT* die gleichen kognitiven Prozesse hervorruft wie die reale Tätigkeit. Videobasierte *SJTs* besitzen durch die gesteigerte Kontextinformation von non-verbale Hinweisreizen eine höhere physische und psychische Realitätsnähe als schriftliche *SJTs* (Bass & Barrett, 1972). Große Variationsbreite kommt darüber hinaus dem Inhalt der Situationen zu. Die meisten *SJTs* beinhalten zwischenmenschliche Szenarien. Sie können jedoch auch prozedurales oder deklaratives Wissen unabhängig von sozialen Situationen erfassen und trainieren.

Generieren der Antwortoptionen. Zu den Situationen entwickeln Experten oder Testentwickler effektive und ineffektive Antwortoptionen. Es wird vermutet, dass Experten qualitativ und quantitativ höhere Antwortoptionen entwickeln als Testentwickler, da sie aufgrund ihres Erfahrungshintergrundes realistischere und vielfältigere Antworten generieren. Die Antwortoptionen in den finalen *SJTs* werden gewöhnlich auf vier bis sechs beschränkt. Ein wichtiger Aspekt bei der Antwortentwicklung ist der Konstruktbezug. *SJTs*, deren Antwortoptionen niedrige und hohe Ausprägungen des zu messenden Konstrukts abbilden, zeigen eine höhere inkrementelle Validität als *SJTs* deren Antworten keinen/kaum Konstruktbezug aufweisen (Polyhart & Ryan, 2000; Trippe & Foti, 2003). Daher wird empfohlen einige Antworten mit hoher und einige mit niedriger Ausprägung des zu messenden Konstrukts zu entwerfen (Motowidlo et al., 2006). Abbildung 14 gibt ein Beispiel für dieses Vorgehen am Konstrukt der Verträglichkeit (Polyhart & Ryan, 2000, S. 242):

You are working at the cash register on a very busy shift. There is a long line of customers waiting to check out, and you can see several individuals are getting impatient. What do you do?

- a) Tell the waiting customers that you appreciate their patience.
- b) Regularly look at those waiting so that they know you are aware of them.
- c) Once in a while look at those waiting so that they know you are aware of them.
- d) Focus on your job as a cashier and don't look at those waiting.

Abbildung 14. Konstruktbezogene Antworten durch hohe und niedrige Ausprägung von Verträglichkeit.

Basierend auf der *Tacit Knowledge* Theorie (Sternberg, 1997, 1999) empfiehlt die Arbeitsgruppe um Stemler (2006) ein alternatives Vorgehen zur Entwicklung von Antwortoptionen. Nach ihnen können Personen sieben Problemlösungsstrategien anwenden,

um soziale Situationen zu bewältigen. Diese sind nachzugeben (*comply*), sich zu beraten (*consult*), Rücksprache halten (*confer*), vermeiden (*avoid*), delegieren (*delegate*), Regeln aufstellen (*legislate*) und sich rächen (*retaliate*). Die Methode verdeutlicht Abbildung 15 an einem Beispiel aus dem Schulbereich (Stemler & Sternberg, 2006, S. 125).

Mr. Thompson usually gets along well with his colleagues. One day, in a departmental meeting about the curriculum, a colleague personally attacks him because Mr. Thompson expresses a different option about a new program than most of his colleagues.

Given the situation, please rate the quality of the following statements.

(1= strongly disagree to 7= strongly agree)

- a) Mr. Thompson should reiterate his option about the curriculum but state that he is willing to go along with the group. (*comply*)
- b) After the meeting, Mr. Thompson should ask one of the other teachers how he or she thinks he should deal with the colleague's comments. (*consult*)
- c) Mr. Thompson should talk privately with his colleague and say that he felt the personal attack was inappropriate and out of line. (*confer*)
- d) Mr. Thompson should ignore the attack and continue his discussion with another teacher. (*avoid*)
- e) Mr. Thompson should ask the principal speak to the colleague about his behaviour. (*delegate*)
- f) Mr. Thompson should propose the establishment of formal rules of order for faculty meetings. (*legislate*)
- g) Mr. Thompson should state that he is not interested in responding to petty personal attacks, but will be happy to answer questions about his opinion of the program. (*retaliate*)

Abbildung 15. Konstrukt bezogene Antworten durch Problemlösestrategien.

Wie die Situationsbeschreibung können auch die Antwortoptionen in der Komplexität variieren. Vermutet wird, dass detaillierte spezifische Antworten eine höhere Validität, aber eine geringere Generalisierbarkeit aufweisen als allgemein formulierte Antworten. Systematische Forschungsergebnisse liegen jedoch diesbezüglich nicht vor. Die Realitätsnähe der Antworten kann erhöht werden, indem die Testteilnehmer selbst ihre Antworten formulieren (Walker & Goldenberg, 2004). Die Transparenz stellt einen weiteren

Faktor bei der Antwortentwicklung dar. Unter der Transparenz ist zu verstehen wie leicht die Testteilnehmer die richtige(n) Antwort(en) im *SJT* erkennen können (Hooper, Cullen & Sackett, 2006). Um die Transparenz zu reduzieren, wird empfohlen, möglichst nur sozial akzeptable Antworten zu formulieren (Weekley et al., 2006).

Wahl der Instruktion und der Bewertungsmethode. Nach der Entwicklung der Antworten entscheidet sich der Testentwickler für eine Instruktionsform. Diese lassen sich in ein Vier-Felder-Schema einordnen, wobei zwischen wissensbezogenen (*should do*) vs. verhaltensbezogenen (*would do*) und *Forced Choice* vs. kontinuierlichem *Rating* unterschieden wird (Polyhart & Ehrhart, 2003; Ployhart & MacKenzie, 2011). Tabelle 8 veranschaulicht die Instruktionsformen mit Beispielen:

Tabelle 8: *Instruktionsformen von SJTs*

Instruktionsform	wissensbezogen	verhaltensbezogen
Forced Choice:	How should you react?	How would you react?
Wahl einer Antwort	Which reaction is the best/the most effective?	How have you reacted in the past?
Wahl zwischen zwei Antworten	Which reaction is the best/the most effective and which is the worst/least effective?	How would you most and least likely react?
Wahl 3 aus 8 Antworten	Choose the three most appropriate actions to take in this situation.	-
Kontinuierliches Rating:		
Beurteilung aller Antworten	Rate the effectiveness of each response.	Rate how likely you would do each response.

McDaniel, Hartman, Whetzel und Grubb (2007) weisen die wissensbezogenen Instruktionen der *maximal performance* und die verhaltensbezogenen Instruktionen der *typical performance* nach dem Konzept von Cronbach (1960) zu. Die *maximal performance* wird über Fähigkeitsmessungen, häufig über Wissenstests und Intelligenztests erhoben und bildet die maximale Leistung eines Probanden ab. Die *typical performance* wird mit nicht-kognitiven Verfahren wie Persönlichkeits- oder Motivationstests erhoben und bildet typische Leistungen eines Probanden ab. Bei den Instruktionen mit *Forced Choice*-Format werden die Testteilnehmer gebeten, entweder eine Antwort (z.B. Motowidlo et al., 1990; Weekley & Jones, 1997) oder zwei Antworten (z.B. Weekley & Jones, 1999)

pro Item auszuwählen. Als neueres Format wird auch die Wahl von drei aus acht Antwortoptionen bei wissensbasierten Instruktionen verwendet (Patterson et al., 2012). Bei Instruktionen mit kontinuierlichem *Rating* beurteilen die Probanden jede Antwortoption nach der Effektivität bzw. die Wahrscheinlichkeit mit der sie diese ausführen würden auf einer Likert-Skala (Polyhart & Ehrhart, 2003; Sacco et al., 2000). Bezüglich der sozialen Erwünschtheit wird von einigen Forschern vermutet, dass wissensbezogene Instruktionen verfälschungsresistenter sind als verhaltensbezogene Instruktionen (Nguyen, Biderman & McDaniel, 2005). Dies konnte jedoch nicht hinreichend bestätigt werden. Eher scheinen wissens- vs. verhaltensbezogene Instruktionen die Kriteriumsvalidität zu beeinflussen bzw. die Beziehung zwischen der *SJT*-Punktzahl, kognitiven Fähigkeiten und Persönlichkeitseigenschaften zu beeinflussen (Polyhart & MacKenzie, 2011). Die Wahl der Instruktionsform wirkt sich auf das Datenniveau des Auswertungsschlüssels aus (bezeichnet als Bewertungsmethode in Abbildung 13). Instruktionen mit kontinuierlichem *Rating* bieten gegenüber *Forced Choice*-Formaten den Vorteil, dass die Daten eine höhere Varianz aufweisen, was sich positiv auf die Reliabilität und Validität des *SJT*s auswirkt (McDaniel & Nguyen, 2001; Polyhart & Ehrhart, 2003).

Generieren des Auswertungsschlüssels. Die Entwicklung des Auswertungsschlüssels ist notwendig, um das Testergebnis zu ermitteln. Dafür erfolgt ein Abgleich zwischen den Antworten der Teilnehmer und dem optimalen Antwortmuster. Die Entwicklung des Auswertungsschlüssels erfolgt am häufigsten auf rationale, empirische, theoretische oder hybride Art (Bergman, Drasgow, Donovan, Henning & Juraska, 2006; Weekley & Polyhart, 2011).

Bei der rationalen Methode beurteilen Experten des Gebietes die Antworten nach ihrer Effektivität. Die Größe der Expertengruppe variiert meist zwischen 10 und 50 Personen (vgl. Chan & Schmitt, 1997; De Meijer, Born, Van Zielst & Van der Molen, 2010; Weekley & Jones, 1997). Die Beurteilung der Effektivität findet im *Forced Choice* Format oder als kontinuierliches *Rating* statt (Polyhart & McKenzie, 2011). Bei ersterem bewerten die Experten jede Antwort entweder als effektiv oder als ineffektiv, während bei letzterem die Experten die Effektivität auf einer Likert-Skala beurteilen. Anschließend wird geprüft, wie übereinstimmend die Experten die Effektivität der Antworten beurteilt haben. Die Beurteilerübereinstimmung wird gewöhnlich über den *Intra-class Coefficienten (ICC)* berechnet (z.B. De Meijer et al., 2010; Lievens & Sackett, 2006; Motowidlo et al., 1990). Antworten, die eine zufriedenstellende Beurteilerübereinstimmung ($> .70$) erzielen, verbleiben im *SJT*. Kritisch an der Beurteilerübereinstimmung ist, dass Items mit hoher Augenscheinvalidität oder mit hoher sozialer Erwünschtheit leichter übereinstimmend von Experten bewertet werden als weniger of-

fensichtliche Items (Krokos, Meade, Cantwell, Pond & Wilson, 2004). Bei der rationalen Methode sollte darüber hinaus beachtet werden, dass die *SJTs* durch die verwendeten Expertengruppen Einflüsse einer bestimmten Einrichtung ausgesetzt sind. Wenn dies nicht ausdrücklich erwünscht ist, sollten Experten unterschiedlicher Einrichtungen einbezogen werden.

Die Effektivität der Antwortoptionen wird bei der empirischen Methode über Korrelationen der Antwortoptionen zu tätigkeitsrelevanten Kriterien gebildet (Hogan, 1994). Notwendige Schritte sind die Kriterienauswahl, das Treffen von Entscheidungsregeln, die Gewichtung der Items und die Kreuzvalidierung der Ergebnisse (Bergman et al., 2006). Bei den Entscheidungsregeln muss festgelegt werden, wie viele Personen mindestens eine Antwort auswählen müssen und wie hoch die minimale Korrelation der Antworten mit dem Außenkriterium sein soll. Die Gewichtung der Antworten erfolgt über die Zuweisung der Zahlen „1“ (positive Korrelation mit Kriterium), „-1“ (negative Korrelation mit Kriterium) oder „0“ (Nullkorrelation mit Kriterium). Bei der Kreuzvalidierung werden die Ergebnisse von Subgruppen mit den Ergebnissen der Gesamtgruppe vorgenommen. Empirische Methoden zeigen oft hohe Kriteriumsvaliditäten verglichen mit anderen Methoden (Hogan, 1994). Die Nachteile der empirischen Methode liegen jedoch in der notwendigen Stichprobengröße (Weekley et al., 2006), der Abhängigkeit von der Qualität der Kriterien (Campbell, 1990), der Stabilität und Generalisierbarkeit der Ergebnisse (Mumford & Owens, 1987), sowie dem damit verbundenen Problem des Bias-Varianz-Dilemmas (Cureton, 1950; Rey & Wender, 2008). Bei diesem werden zufällige Variationen im Datensatz mit im Regressionsmodell erfasst, sodass eine gute Modellübereinstimmung zur Untersuchungsstichprobe besteht, das Modell jedoch nicht vorbehaltlos zur Vorhersage von anderen Stichproben verwendet werden kann. Abhilfe schaffen Replikationen und Kreuzvalidierungen.

Bei einem theoretischen Auswertungsschlüssel wird die Effektivität der Antwortoptionen theoriebasiert bestimmt. Antworten, die der zugrunde liegenden Theorie entsprechen, werden als korrekt gewertet. Antworten, die konträr zur Theorie sind, werden als inkorrekt bewertet. Antwortoptionen, die keinen Theoriebezug aufweisen, gehen nicht in die Punktwertung ein. Vorteile dieser Methode liegen in der höheren Generalisierbarkeit der Ergebnisse und in der theoretischen Herleitung der *SJTs*. Nachteile können die höhere Transparenz der Items und die Fehlerhaftigkeit der Theorie sein (Hough & Paullin, 1994). Die theoretische Methode wird am seltensten bei der *SJT*-Entwicklung genutzt (Weekley et al., 2006).

Die hybride Methode integriert zwei der beschriebenen Vorgehensweisen, um die Stärken einzelner Methoden zu bündeln und nicht verfügbare Daten zu kompensieren (Bergman et al., 2006). Ein standardisiertes Vorgehen zur Bildung eines hybriden

Auswertungsschlüssel existiert aufgrund der unterschiedlichen praktischen Rahmenbedingungen nicht (Hauenstein et al., 2010). So kann sowohl die Reihenfolge als auch die Gewichtung der Methoden variieren. Die hybride Methode verspricht eine höhere prädiktive Validität des SJTs (Olson-Buchanan et al., 1998), erfordert jedoch komplexe theoretische und praktische Entscheidungen zum Vorgehen.

Zum Vergleich der Methoden empfehlen Bergman et al. (2006) folgende Kriterien heranzuziehen: Ein Auswertungsschlüssel sollte eine hohe Kriteriumsvalidität und inkrementelle Validität aufweisen. Er sollte mit verwandten Konstrukten höher korrelieren als mit inhaltlich fremden Konstrukten. Er sollte keine Subgruppeneffekte hervorrufen. Bisherige Forschungsergebnisse liefern inkonsistente Befunde. So zeigte beispielweise die Studie von Weekley und Jones (1997), dass der rationale Auswertungsschlüssel dem empirischen Auswertungsschlüssel leicht in der Vorhersage des Jobverhaltens über kognitive Fähigkeiten und Berufserfahrung hinaus überlegen war ($\Delta R^2 = 9.6\%$ vs. 5.7%). Andere Studien fanden hingegen keine bedeutsamen Validitätsunterschiede zwischen dem rationalen und empirischen Auswertungsschlüssel (MacLaine, Barton, Holloway-Lundy & Nickles, 2001; Paullin & Hanson, 2001). Such und Schmidt (2004) stellten eine leichte Überlegenheit des empirischen gegenüber dem rationalen Auswertungsschlüssel fest. Vergleiche zwischen allen Methoden anhand eines SJTs für Führungsfähigkeit zeigte, dass der rationale und der empirische Auswertungsschlüssel vergleichbare Korrelationen zum Jobverhalten aufwiesen und Jobverhalten über kognitive Fähigkeiten und Persönlichkeitstests hinaus aufklärten. Obwohl der theoretische Auswertungsschlüssel nicht mit dem Jobverhalten korrelierte, wies der hybride Schlüssel, gebildet aus empirischen und theoretischen Antworten, Zusammenhänge zum Jobverhalten auf (Bergman, Drasgow, Donovan & Juraska, 2003). In einer Folgestudie zum gleichen *SJT* untersuchten die Autoren zwei rationale (Experten vs. Novizen), einen empirischen, sechs theoretische und drei hybride theoretisch-empirische Auswertungsschlüssel auf deren Effektivität. Die Ergebnisse legten dar, dass der empirische, der rationale expertenbasierte und ein hybrider Auswertungsschlüssel Führungsfähigkeit vorhersagten. Diese und ein weiterer hybrider Auswertungsschlüssel zeigten zudem inkrementelle Validität über kognitive Fähigkeiten und Persönlichkeitstests und wiesen keine Subgruppenunterschiede bezüglich des Geschlechts auf. Alle elf Auswertungsschlüssel wiesen eine diskriminante Validität auf. Anhand der dargestellten Studien ist es verfrüht, eindeutige Schlussfolgerungen über die Effektivität der unterschiedlichen Auswertungsschlüssel zu treffen. Als erstes Fazit scheinen jedoch rationale und empirische Auswertungsschlüssel in ihrer Validität vergleichbar (Weekley et al., 2006). Die meist geringe Korrelation zwischen den rationalen

und empirischen Auswertungsschlüsseln deutet darauf hin, dass sie unterschiedliche Aspekte erfassen (vgl. Weekley & Jones, 1997).

Verfilmen der Situationen. Für videobasierte *SJTs* folgt der zusätzliche Schritt der Verfilmung. Nachfolgende Beschreibungen und Empfehlungen richten sich nach Olson- Buchanan und Drasgow (2006). Bevor die schriftlichen Szenarien in Filmsequenzen überführt werden, müssen Skripte für die einzelnen Situationen geschrieben werden. Die Skripte werden mit Laienschauspielern oder professionellen Schauspielern verfilmt. Mimik, Gestik, Stimmlage und Betonung der Schauspieler sollten dabei der ursprünglichen Intention des schriftlichen *SJTs* entsprechen. Zudem sollten eine möglichst neutrale Umgebung und unauffällige Requisiten gewählt werden, die zeitlich nicht schnell überholt wirken. Jede Situation eines videobasierten *SJTs* wird durch einen Erzähler eingeleitet und anschließend abgespielt. An der kritischen Stelle stoppt die Szene und der Testteilnehmer wird aufgefordert, sich innerhalb von 25 Sekunden zwischen den Antwortalternativen zu entscheiden (Lievens & Sackett, 2006). Die Antwortalternativen werden visuell dargeboten und gleichzeitig durch einen Erzähler vorgelesen oder videobasiert präsentiert.

4.4. Gütekriterien des SJT

4.4.1. Objektivität

Die Durchführungs- und die Auswertungsobjektivität sind aufgrund der standardisierten PC-gestützten Durchführung und Auswertung gegeben. Der Auswertungsschlüssel wird wie im vorherigen Kapitel dargelegt gewöhnlich über Expertenratings entwickelt, welche die Antwortoptionen nach deren Angemessenheit beurteilen. Die Interpretationsobjektivität wird durch den Abgleich zwischen den Antworten der Teilnehmer und denen der Experten (kriterielle Bezugsnorm) oder zwischen den Antworten eines Teilnehmers mit denen der anderen Teilnehmer (soziale Bezugsnorm) hergestellt.

4.4.2. Reliabilität

Die interne Konsistenz von *SJTs* ist aufgrund ihrer Multidimensionalität und komplexen Natur der zu messenden Eigenschaften gewöhnlich relativ gering, es sei denn, die *SJTs* haben eine sehr hohe Itemanzahl (Legree, 1995; Polyhart & Ehrhart, 2003; Schmitt & Chan, 2006). Kanning (2005) berichtet für einen 11 Items umfassenden *SJT* zur Erhebung von spezifischen sozialen Kompetenzen Polizeibeamter eine interne Konsis-

tenz von .45. Weekley und Jones (1997) geben interne Konsistenzen zwischen .29 und .32 für einen SJT mit 20 Items, .73 für einen SJT mit 34 Items und .87 für einen SJT mit 46 Items an. Ähnliche Größenordnungen zeigte die Analyse von Clevenger et al. (2001). SJTs mit 33 Items erreichten Cronbachs Alphas von .63, SJTs mit 39 Items von .82 und SJTs mit 43 Items von .90. Die interne Konsistenz variiert in Abhängigkeit von der Instruktion (Polyhart & Ehrhart, 2003; Tabelle 9).

Tabelle 9: *Reliabilitäten in Abhängigkeit von der Instruktionsform des SJTs*

Instruktion	Interne Konsistenz	Retest-Reliabilität
Wissensbasiert vs. verhaltensbasiert		
Most/least effective vs. Would most/least likely do	.60 vs. .65	.63 vs. .66
Rate how effective vs. Rate how likely would do	.73 vs. .69	.20 vs. .92
Should do	.32	.24
Have done in past	.44	.65

Werden die Teilnehmer aufgefordert jede Antwortoption zu bewerten (kontinuierliches *Rating*), fallen die internen Konsistenzen größer aus, als wenn sie eine Zweifach- oder Einfachauswahl (*Forced Choice*-Formate) vornehmen. Dieser Effekt trat unabhängig von den verhaltensbasierten bzw. wissensbasierten Instruktionen auf (Polyhart & Ehrhart, 2003). Auch eine zeitlich frühere Studie stützt dieses Ergebnis. In dieser fiel die interne Konsistenz eines wissensbasierten SJT mit einer kontinuierlichen Instruktion höher aus als bei einem SJT mit fünf verschiedenen *Forced Choice*-Formaten ($\alpha = .80$ vs. .44-.76; Legree, 1995). Interne Konsistenzen von SJTs sollten gerade wegen ihrer Multidimensionalität jedoch nur im Zusammenhang mit anderen Homogenitätsprüfungen z.B. Faktorenanalysen interpretiert werden (Hattie, 1958). Whetzel und McDaniel (2009) betonen, dass die Reliabilität von SJTs bei der Berechnung der internen Konsistenz aufgrund der Konstrukteterogenität unterschätzt wird. Sie empfehlen daher, die Reliabilität über die Retest-Reliabilität oder Paralleltest-Reliabilität zu erfassen.

Videobasierte SJTs erzielen verglichen mit schriftlichen aufgrund der stärkeren Realitätsnähe eine höhere Akzeptanz der Teilnehmer, was mit einem reliableren Antwortverhalten einhergeht (Chan & Schmitt, 1997; Richman-Hirsch, Olson-Buchanan & Drasgow, 2000). Die höhere Realitätsnähe ist dabei auf die gesteigerte physische und psychische Abbildung einer Situation von videobasierten verglichen zu papierbasierten Verfahren zurückzuführen (Bass & Barrett, 1972). Somit scheinen videobasierte SJTs besser geeignet, soziale Fähigkeiten mit ihren subtilen sozialen Hinweisreizen und Kontextinformationen zu messen (Christian et al., 2010).

Die Retest-Reliabilität scheint, in Abhängigkeit der Rating-Instruktion zu schwanken. Polyhart und Ehrhart (2003) fanden Retest-Reliabilitäten zwischen .20 und .92 bei unterschiedlichen Antwortformaten, wobei die Retest-Reliabilität höher bei verhaltensbezogenen als bei wissensbezogenen Instruktionen ausfiel. Die höchste Retest-Reliabilität wurde erzielt, wenn die Teilnehmer aufgefordert wurden, jede der Antwortoptionen nach ihrer präferierten Handlungsintention zu bewerten (Tabelle 9). Nach Angabe der Autoren betrug der Abstand der Messung mehrere Wochen. Theoretisch ist noch ungeklärt, welche Höhe der Retest-Reliabilität bei *SJTs* als angemessen gilt. Falls *SJTs* Persönlichkeitsfaktoren erfassen, die relativ stabil sind, sollte die Retest-Reliabilität eher hoch ausfallen. Falls sie überwiegend prozedurales Wissen und Erfahrungen erfassen, würden die individuellen Messwerte mehr schwanken und somit in einer geringeren Retest-Reliabilität resultieren (Schmitt & Chan, 2006). Andere Autoren argumentieren, dass die Stärke von *SJTs* nicht in der Reliabilität, sondern in der Kriteriumsvalidität liegt (Kanning, Grewe, Hollenberg & Hadouch, 2006).

4.4.3. Validität

Aufgrund ihrer tätigkeitszentrierten Konstruktionsweise unter Einbezug von Experten des jeweiligen Tätigkeitsbereichs besitzen *SJTs* eine hohe Inhaltsvalidität (Ployhart & MacKenzie 2011). *SJTs*, die auf Grundlage von Anforderungsanalysen entwickelt wurden, haben eine höhere Konstruktvalidität als die, die nicht mit Hilfe von Anforderungsanalysen konzipiert wurden ($r = .38$ vs. $.29$; McDaniel & Nguyen, 2001; McDaniel et al., 2001).

SJTs zeigen in Metaanalysen niedrige bis mittlere Zusammenhänge mit allgemeinen kognitiven Fähigkeiten ($r = .29$), dem arbeitsplatzbezogenen Wissen ($r = .29$) und der beruflichen Leistungsfähigkeit ($r = .20$; McDaniel & Nguyen, 2001). *SJTs* korrelieren niedrig mit der Persönlichkeitseigenschaft Verträglichkeit ($r = .22$), Gewissenhaftigkeit ($r = .23$), Extraversion ($r = .13$), und Offenheit für Erfahrungen ($r = .11$; McDaniel et al., 2001; 2007). Die Kriteriumsvaliditäten sind vergleichbar mit anderen Methoden der Eignungsbeurteilung (McDaniel et al., 2001; 2007). Die niedrigen Zusammenhänge mit den kognitiven Kriterien und den Persönlichkeitsfaktoren deuten darauf hin, dass *SJTs* etwas anderes als allgemeine kognitive Fähigkeiten und der Persönlichkeit messen (Ployhart & MacKenzie, 2011).

Differenzierte Analysen der Kriteriums- und Konstruktvalidität erfolgten durch die Gegenüberstellung von schriftlichen vs. videobasierten *SJTs*, sowie *SJTs* mit wissens- vs. verhaltensbasierter Instruktion. Videobasierte *SJTs* erzielten höhere Zusammenhänge zu sozialen Fähigkeiten ($r = .36$ vs. $.20$) und Führungsfähigkeiten ($r = .25$

vs. .21) als schriftliche *SJTs* (Christian et al., 2010). *SJTs* mit wissensbezogener Instruktion korrelieren höher mit allgemeinen kognitiven Fähigkeiten als *SJTs* mit verhaltensbezogenen Instruktionen ($r = .35$ vs. $.19$; McDaniel & Nguyen, 2001). Letztere wiesen hingegen höhere Zusammenhänge mit Verträglichkeit ($r = .37$ vs. $.19$), Gewissenhaftigkeit ($r = .34$ vs. $.24$) und emotionale Stabilität ($r = .35$ vs. $.12$) auf (McDaniel & Nguyen, 2001). Konträr dazu wiesen Polyhart und Ehrhart (2003) höhere Korrelationen von *SJTs* mit verhaltensbasierter als mit wissensbasierter Instruktion mit dem GPA nach ($r = .30$ -.40 vs. $.04$ -.16). Auch wurden keine signifikanten Unterschiede zwischen *SJTs* mit wissens- vs. verhaltensbezogener Instruktion und allgemeinen kognitiven Fähigkeiten gefunden (Lievens, Sackett & Buyse, 2009b). Aufgrund der derzeitigen inkonsistenten Befundlage sind keine eindeutigen Schlussfolgerungen über die Kriteriumsvalidität von *SJTs* mit wissens- vs. verhaltensbasierter Instruktion zu ziehen (Polyhart & MacKenzie, 2011). Kritik ernten die Metaanalysen dahingehend, dass sie unabhängig von den Konstruktbereichen durchgeführt werden, welche die *SJTs* erfassen. Eine getrennte Analyse nach Konstrukt zeigte, dass *SJTs*, die zwischenmenschliche Fähigkeiten, Teamwork oder Führungsfähigkeiten messen, höhere Kriteriumsvaliditäten erreichten als heterogene *SJTs* ($r = .17$ /.27/.19 vs. $.14$).

Subgruppenunterschiede treten zwischen Rassen und in Abhängigkeit vom Geschlecht auf. Mehrheitlich berichten Studien, dass Afro-Amerikaner ($d = .40$), Lateinamerikaner ($d = .37$) und Asiaten ($d = .47$) schlechter im *SJT* abschneiden als weiße Amerikaner. Diese Subgruppenunterschiede sind ca. halb so groß wie bei kognitiven Fähigkeitstests, aber um das 2-5-fache so groß wie bei Persönlichkeitstests (Polyhart & Holtz, 2008; Polyhart & McKenzie, 2011). Ein entscheidender Moderator der Rassenunterschiede ist die verbale Fähigkeit. So erhöhen sich die Rassenunterschiede zwischen afroamerikanischen und weißen Personen, wenn die Leseanforderung gesteigert wird oder schriftliche statt videobasierte *SJTs* verwendet werden (Sacco et al., 2000a; Sacco, Schmidt & Rogg, 2000b). Aufgrund dieser Befunde wird häufig geschlossen, dass *SJTs* fairer als kognitive Fähigkeitstests sind. Dies ist nur eingeschränkt richtig, da *SJTs* weniger reliabel als kognitive Tests sind und somit der wahre Einfluss auf unterschiedliche Probandengruppen nicht explizit belegt werden kann (Polyhart & McKenzie, 2011). Frauen erzielen geringfügig höhere Punktwerte im *SJT* als Männer mit Effektstärken zwischen $.11$ und $.32$ (Lievens, 2013; Lievens & Coetsier, 2002; Weekley & Jones, 1997, 1999). Auch aktuellere Studien belegen diesen Effekt (Hartman, McDaniel, Grubb, Lawrence, 2007; Nguyen et al., 2005; O'Connell, Whetzel, McDaniel & Nguyen, 2008). Dieser Effekt kann teilweise durch die Nähe des *SJTs* zu Persönlichkeitsfaktoren erklärt werden. *SJTs*, die höher mit Gewissenhaftigkeit und Verträglichkeit korrelieren, zeigen höhere Punktwerte bei Frauen als bei Männern

(Whetzel et al., 2008). Der Geschlechterunterschied ist verglichen mit kognitiven Tests minimal höher und verglichen mit Persönlichkeitstests um das 2-3-fache geringer ausgeprägt.

Bewerber erzielen höhere Punktwerte im *SJT* als Stelleninhaber. Letzter Befund ist jedoch mit Vorsicht zu betrachten, da die meisten Validierungsstudien mit Stelleninhabern statt mit Bewerbern stattfinden (McDaniel & Nguyen, 2001).

Zur prädiktiven und inkrementellen Validität liegen derzeit zwei Metaanalysen von McDaniel et al. (2001, 2007) vor. *SJTs* haben eine prädiktive Validität von .14 für die Vorhersage von Arbeitserfolg. Sie weisen eine geringfügig höhere prädiktive Validität von .02-.07 über kognitive Fähigkeiten, Persönlichkeitseigenschaften und deren Kombination auf. Wissens- vs. verhaltensbezogene Instruktionen zeigen keinen Einfluss auf die inkrementelle Validität von *SJTs*. Eine getrennte Untersuchung zwischen schriftlichen und videobasierten *SJTs* wurde in den Metaanalysen nicht vorgenommen. Lievens und Sackett (2006) stellten in einer Studie fest, dass ein videobasierter *SJT* zur Erfassung von zwischenmenschlichem Verhalten von Medizinstudierenden eine höhere prädiktive und inkrementelle Validität aufwies als sein schriftliches Pendant. Der schriftliche *SJT* korrelierte gering mit der Note im Kurs *zwischenmenschliches Verhalten* (.08), während der videobasierte *SJT* deutlich höher korrelierte (.34). Zudem zeigte der videobasierte im Gegensatz zum schriftlichen *SJT* eine inkrementelle Validität von .11 über einen allgemeinen kognitiven Test und der Bearbeitung eines medizinischen Texts. Die starke Differenz zwischen einem videobasierten und einem schriftlichen *SJT* könnte jedoch durch das Konstrukt des zwischenmenschlichen Verhaltens hervorgerufen worden sein, welches sich durch einen hohen Anteil an nonverbalen Reizen auszeichnet. Nonverbale Aspekte lassen sich realistischer durch Verfilmung als durch schriftliche Ausformulierung darstellen. In einer späteren Längsschnittstudie untersuchte Lievens (2013) die prädiktive Validität desselben videobasierten *SJTs* über die Gesamtdauer des Medizinstudiums. Er stellte fest, dass der *SJT* Kursleistungen mit Patientenkontakt vorhersagen konnte und einen inkrementellen Beitrag leistete. Im Einzelnen betraf dies die Leistung in Kommunikationskursen ($r = .21$, $R^2 = .04$), im OSCE ($r = .12$, $R^2 = .01$), in fallbasierten Interviews der Abschlussprüfung ($r = .19$, $R^2 = .03$) und dem Arbeitsverhalten während der Zusatzausbildung zum Allgemeinmediziner neun Jahre nach der Zulassung zum Studium ($r = .15$, $R^2 = .02$). Eine vergleichende Studie zwischen nicht-kognitiven Verfahren in der allgemeinen Facharztauswahl zeigte, dass der *SJT* einem Problemlösungstest und klinischen Fallvignetten in der Vorhersageleistung überlegen war ($R^2 = .21$ vs. $.09$ vs. $.07$; Patterson et al., 2009).

Die dargestellten Befunde beruhen überwiegend auf Metaanalysen, bei deren zugrunde liegenden Studien es weitgehend an vergleichenden experimentellen Designs fehlt, um die Validität des *SJT*s differenziert zu erforschen (Arthur & Villado 2008; Ployhart & MacKenzie 2011). Darüber hinaus kritisieren die Autoren, dass vergleichende Studien an Bewerbern und Stelleninhabern ausstehen.

4.5. Vor- und Nachteile des *SJT*

In den letzten zwei Dekaden nahm die Beliebtheit von *SJT*s aufgrund der moderaten Kriteriumsvalidität, hohen Augenscheinvalidität, der Testfairness bezüglich Nationalität und Geschlecht, sowie der fortschreitenden Videotechnologie zu (De Meijer et al., 2010; Ployhart & MacKenzie, 2011). *SJT*s zeigen eine mittlere prädiktive Validität vergleichbar zu anderen Methoden der Eignungsbeurteilung und weisen eine geringfügig höhere prädiktive Validität über kognitive Fähigkeiten, Persönlichkeitseigenschaften und deren Kombination auf (McDaniel et al., 2001, 2007). Bezüglich der Testfairness sind sie weniger sensitiv gegenüber ethnischen Minderheiten als traditionelle kognitive Fähigkeitstests (Clevenger et al., 2001; Motowidlo et al., 1990). Die Bearbeitungszeit von *SJT*s ist vergleichbar mit kognitiven Fähigkeitstests (Polyhart & MacKenzie, 2011). Im Gegensatz zu anderen Simulationsverfahren wie *Assessment Center*, Arbeitsproben oder Situational Interviews sind *SJT*s weniger kostenintensiv in der Administration und Auswertung (Christian et al., 2010). Dank PC-gestützter Auswertung ist ein flächendeckender Einsatz in frühen Recruiting-Phasen möglich (Weekley & Ployhart, 2006), der es ermöglicht, den Bewerbern ihre Ergebnisse sofort zurückzumelden (Schmitt, Gilliland, Landis & Devine, 1993). Die Stärke von videobasierten *SJT*s liegt darin, komplexe soziale Situationen realitätsnah abzubilden (Bass & Barrett, 1972). Allerdings ist das Verfilmen von *SJT*s zeit- und kostspielig. Vor- und Nachteil zugleich ist, dass *SJT*s mehrere Konstrukte gleichzeitig erfassen und somit ein multidimensionales Messinstrument darstellen (Ployhart & MacKenzie, 2011). Eine Folge der Konstruktheterogenität ist die schwächere Reliabilität von *SJT*s gegenüber klassischen Testverfahren (Kanning et al., 2006). Auch besteht kein Konsens über das zugrunde liegende Messmodell von *SJT*s (Stemler & Sternberg, 2006, McDaniel & Nguyen, 2001; Weekley & Jones, 1999).

4.6. SJTs in der medizinischen Studierenden- und Facharztauswahl

Patterson et al. (2012) halten den Einsatz von *SJTs* in der medizinischen Studierenden- und Facharztauswahl insbesondere deshalb für geeignet, weil sie nicht-kognitive Eigenschaften erheben und vorhersagen können, die für die medizinische Ausbildung und Praxis relevant sind. Die *SJTs* können flexibel in den medizinischen Kontext eingebettet werden. Die hohen Bewerberzahlen auf einen medizinischen Studienplatz erfordern kostengünstige PC-gestützte Verfahren wie den *SJT*. Dadurch, dass *SJTs* den Erwerb von prozeduralem Wissen fördern, sind sie auch für den Einsatz in der medizinischen Ausbildung geeignet (Patterson et al., 2012). Derzeit werden *SJTs* in der Medizin überwiegend in Großbritannien, Amerika und Kanada für die Facharztauswahl in schriftlicher Form eingesetzt. Diese setzen fachliches Wissen bei den Bewerbern voraus. Ein Item der Facharztauswahl in Großbritannien gibt Abbildung 16 wieder, welches in einem *SJT* zur Erfassung von Empathie, Integrität und Belastbarkeit angewandt wird (Patterson et al., 2009, S. 52).

A 55-year-old woman with ischemic heart disease has smoked 20 cigarettes per day for 40 years. She requests nicotine replacement patches. She has had these previously but has been inconsistent in their use and has often continued to smoke while using the patches.

Rank in order the following immediate actions in response to this situation (1 = most appropriate; 5 = least appropriate)

- a) Emphasize the dangers of smoking but do not prescribe.
- b) Enquire about the difficulties she has with stopping smoking and any previous problems with patches.
- c) Insist on a period of abstinence before prescribing any further patches.
- d) Prescribe another supply of patches and explain how they should be used.
- e) Suggest that nicotine replacement therapy is not suitable for her but explore alternative therapies.

Abbildung 16. *SJT*-Item für die Facharztauswahl.

In der medizinischen Studierendenauswahl ist die Universität von Ghent ein Vorreiter, der seit 1997 einen videobasierten *SJT* für interpersonelles Verhalten und Kommunikation verwendet (Lievens & Sackett, 2006). In den letzten Jahren nimmt der Trend zu, *SJTs* in die Studierendenauswahl zu integrieren. So wurde z.B. in Graz 2010 ein schriftlicher *SJT* auf Grundlage des biopsychosozialen Modells für die medizinische Studierendenauswahl eingesetzt (Manhal, Neges, Reibnegger, Ithaler & Smolle, 2012).

In Großbritannien wird ein *SJT* als Bestandteil des UK Clinical Aptitude Test (UKCAT) eingesetzt (<http://www.ukcat.ac.uk/about-the-test/situational-judgement/>). Auch die National University of Singapore lässt die Studienplatzbewerber einen *SJT* absolvieren (http://www.geraldton.com/premed/Applying_to_NUS_YLL_School_of_Medicine.html, Zugriff am 15.03.2014) und die Association of American Medical Colleges entwickelt derzeit einen *SJT* für die Studierendenauswahl (<https://www.aamc.org/initiatives/admissionsinitiative/sjt/>, Zugriff am 15.03.2014). An deutschen medizinischen Fakultäten wird bisher kein *SJT* als Auswahlverfahren verwendet.

5. Schlussfolgerung für die SJT-Entwicklung an der Medizinischen Fakultät Heidelberg

Aus den theoretischen Darlegungen wurden Schlussfolgerungen für die SJT-Entwicklung an der Medizinischen Fakultät Heidelberg gezogen. Dem *SJT* sollte eine Anforderungsanalyse vorausgehen, da auf diese Weise höhere Validitäten erreicht werden als ohne Anforderungsanalyse (McDaniel et al., 2001). Um die Multidimensionalität einzuschränken, wurde der *SJT* konstruktbezogen geplant. Bei der Entwicklung wurden Personengruppen der Medizinischen Fakultät einbezogen, die herausfordernde Lehrsituationen aus dem Human- und Zahnmedizinstudium kennen. Hintergrund ist, dass *SJTs* eine hohe Inhaltsvalidität besitzen, wenn Experten des jeweiligen Tätigkeitsbereichs an der Entwicklung beteiligt sind (Ployhart & MacKenzie 2011). Um soziale Interaktionen möglichst umfassend und realitätsnah abzubilden, fiel die Entscheidung für einen videobasierten *SJT*, der verglichen mit schriftlichen *SJTs* aufgrund subtiler Hinweise und Kontextinformationen höhere Reliabilitäten aufweist (Chan & Schmitt, 1997; Christian et al., 2010; Richman-Hirsch et al., 2000).

Da das Auswahlverfahren der Medizinischen Fakultät Heidelberg stark auf kognitiven Kriterien beruht, soll der *SJT* weniger die kognitiven als die persönlichen Aspekte einer Person erfassen. Daher wurde eine verhaltensbezogene Instruktionsform gewählt, die in der Regel höhere Zusammenhänge zu Persönlichkeitseigenschaften und niedrigere zu kognitiven Faktoren aufweist (McDaniel & Nguyen, 2001). Für die Instruktionsform und den Auswertungsschlüssel wurde ein kontinuierliches *Rating* verwendet, da diese Variante gegenüber *Forced Choice*-Formaten eine höhere interne Konsistenz zeigt (Polyhart & Ehrhart, 2003). Auf der Grundlage, dass die rationale und die empirische Methode vergleichbare Validitäten für die Auswertungsschlüssel zeigen (Bergman et al., 2003; Weekley et al., 2006), wurde die expertenbasierte rationale Methode gewählt, die praktikabler umsetzbar erscheint (vgl. Campbell, 1990; Mumford & Owens, 1987; Rey & Wender, 2008; Weekley et al., 2006).

Um der Vielfalt sozialer Reaktionsmöglichkeiten gerecht zu werden und der sozialen Erwünschtheit entgegen zu wirken, wurden für den *SJT* nicht nur Antwortoptionen unterschiedlicher, sondern auch gleicher sozialer Angemessenheit generiert. Benbasat und Baumal (2007) empfehlen zudem, nicht-kognitive Kriterien lediglich für die Selbstselektion von Bewerbern einzusetzen. Daher wurde der *SJT* als freiwilliges *Self-Assessment* und nicht als Kriterium im Auswahlverfahren integriert. Dies entspricht auch der Stellungnahme der Deutschen Gesellschaft für Psychologie e. V. (2004), beratende *Self-Assessments* zur Auseinandersetzung der Studieninteressenten mit der eigenen Person und dem Studienangebot und -anforderungen anzubieten.

6. Die Entwicklung des SJTs der Medizinischen Fakultät Heidelberg

Dieses Kapitel beschreibt die Entwicklung des SJTs zur Erfassung von sozialen Kompetenzen von medizinischen Studienplatzbewerbern in Heidelberg. Das methodische Vorgehen baut dabei auf den vorausgegangenen Kapiteln 2-5 auf.

Kapitel 6.1 erläutert die Anforderungsanalyse zur Erhebung der relevanten sozialen Kompetenzen im Human- und Zahnmedizinstudium. Anschließend wird der Workshop zur Sammlung von erfolgskritischen Situationen und Verhaltensweisen für den SJT beschrieben (Kapitel 6.2). Kapitel 6.3 stellt die Vortestung an Gymnasiasten dar und die daraus resultierenden Überarbeitungen der schriftlichen SJT-Items. Kapitel 6.4 erläutert die Entwicklung des Auswertungsschlüssels. Kapitel 6.5 geht auf die Verfilmung der schriftlichen Items ein. Das Vorgehen veranschaulicht Abbildung 17 grafisch:

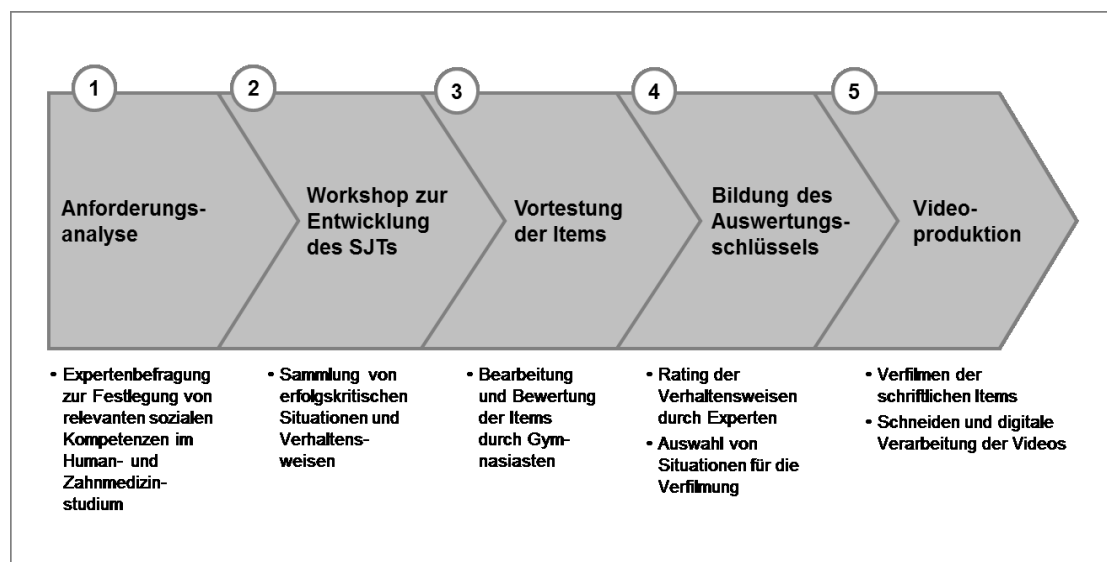


Abbildung 17. Entwicklungsablauf des SJTs für soziale Kompetenzen.

6.1. Anforderungsanalyse

Aufgrund der Erkenntnis, dass *SJTs*, die auf Anforderungsanalysen basieren, eine höhere Konstruktvalidität aufweisen als die, die ohne Anforderungsanalyse konzipiert wurden (McDaniel et al., 2001), erfolgte für den geplanten *SJT* eine ausführliche Anforderungsanalyse. Ziel war es, die im Arbeitsmodell (Kapitel 3.4) enthaltenen sozialen Kompetenzen für die *SJT*-Entwicklung zu reduzieren, indem ihre Wichtigkeit für das Medizinstudium bewertet wurde. Es wurde darauf geachtet, dass die Wichtigkeit für das Studium und nicht für den Arztberuf beurteilt wird, da der zu entwickelnde *SJT* pri-

mär eine Eignung zum Studium erfassen soll. Um die Umfrageteilnehmer dafür zu sensibilisieren und möglichst differenzierte Bewertungen zu erhalten, wurden sie gebeten, die Wichtigkeit sozialer Kompetenzen getrennt für das Studium und den Beruf zu beurteilen. Da das Verständnis und das Beurteilen von sozialen Kompetenzen stark kontextabhängig sind (Albrecht, 1997; Holtz, 1994; Reschke, 1995), wurden sowohl allgemeine Definitionen sozialer Kompetenzen als auch kontextspezifische Beispiele aus dem Medizinstudium und -beruf in das Erhebungsinstrument eingebaut.

6.1.1. Forschungsfragen

Die zentrale Frage der Anforderungsanalyse richtete sich darauf, welche sozialen Kompetenzen des Arbeitsmodells für das Human- bzw. Zahnmedizinstudium relevant sind. Zudem wurde untersucht, ob sich Unterschiede zwischen den Antworten der Human- und Zahnmediziner ergaben. Aufgrund der hohen Überschneidung der Tätigkeitsanforderungen, die an Studierende bzw. Berufstätige der Human- und Zahnmedizin gestellt werden, wurde vermutet, dass keine wesentlichen Unterschiede in der Einschätzung der Wichtigkeit sozialer Kompetenzen für die beiden Studiengänge bestehen. Methodisch wurde darüber hinaus untersucht, ob die gewählten kontextspezifischen Beispiele die Beurteilungen sozialer Kompetenzen verzerren.

6.1.2. Eingesetzte Verfahren und Untersuchungsvariablen

Für die Expertenbefragung wurde ein Onlinefragebogen eingesetzt (Anhang B-1). Ausgehend von dem Arbeitsmodell sozialer Kompetenzen konzentrierte sich dieser auf die zehn interpersonellen sozialen Kompetenzen der ersten Ebene, da sie im Modell die basalen interaktionsbezogenen Kompetenzen bilden. Zusätzlich wurden die zwei intrapersonellen Kompetenzen Selbstbeobachtung/Selbstreflexion und Selbstkontrolle/Selbstdisziplin aufgenommen, da sie zentral für ein erfolgreiches Studium sind (z.B. De Fruyte & Mervielde, 1996; Konegen-Grenier, 2001; O'Connor & Pannonen, 2007). Es wurden drei Fragebogenversionen eingesetzt (Tabelle 10), um methodische Artefakte allein aufgrund des Kontextbezugs der sozialen Kompetenzen einzuschränken. Alle drei Versionen wurden durch eine allgemeine Definition der sozialen Kompetenzen eingeleitet. Version 2 veranschaulichte die sozialen Kompetenzen durch ein zusätzliches Beispiel aus dem Medizinstudium, während Version 3 ein zusätzliches Beispiel aus dem Arztberuf enthielt. Die Probanden beurteilten die Wichtigkeit der Kompetenzen auf einer 7-stufigen Likert-Skala (1 = sehr unwichtig bis 7 = sehr wichtig). Die randomisierte Zuweisung der Teilnehmer zu den Fragebogenversionen erfolgte anhand

des Geburtsmonats, welches als unabhängig von Persönlichkeitseigenschaften gilt (Ihme, Jia, Penzold & Zimmermann, 2008). Beim Ausfüllen des Fragebogens bestand kein Antwortzwang, sodass einzelne Items übersprungen werden konnten. Die Teilnehmer hatten darüber hinaus die Möglichkeit freie Anmerkungen zu machen. Abschließend wurden die soziodemographischen Variablen Geschlecht, Tätigkeitsbereich, Anzahl der Arbeits- und Lehrjahre als Kontrollvariablen erhoben. Drei Mediziner und zwei Diplom-Psychologen prüften den Fragebogen auf Verständlichkeit, Mehrdeutigkeit und Adäquatheit der medizinischen Beispiele.

Tabelle 10: *Fragebögen der Anforderungsanalyse am Beispiel der Verantwortungsübernahme*

Version	Erklärung der sozialen Kompetenz	Zuweisung nach Geburtsmonat
1	<i>Definition:</i> Bereitschaft und Fähigkeit, verantwortungsbewusst gegenüber der eigenen Person und der Gesellschaft zu handeln.	Januar - April
2	<i>Definition:</i> Bereitschaft und Fähigkeit, verantwortungsbewusst gegenüber der eigenen Person und der Gesellschaft zu handeln. <i>Beispiel aus dem Medizinstudium:</i> für die eigenen Prüfungsleistungen Verantwortung übernehmen	Mai - August
3	<i>Definition:</i> Bereitschaft und Fähigkeit, verantwortungsbewusst gegenüber der eigenen Person und der Gesellschaft zu handeln. <i>Beispiel aus dem Arztberuf:</i> Verantwortung für das eigene Handeln im Umgang mit Patienten und anderen übernehmen	September - Dezember

6.1.3. Stichprobenrekrutierung und -beschreibung

Die Befragung richtete sich an approbierte Human- und Zahnmediziner aller Fachrichtungen. Die approbierten Mediziner wurden als Experten angesehen, da sie das Medizinstudium bereits erfolgreich absolviert haben und rückblickend beurteilen können, welche sozialen Kompetenzen im Studium relevant sind. Um eine Verzerrung durch die Institution der medizinischen Fakultät zu vermeiden, wurden neben fakultätsinternen auch externe Mediziner befragt (Ployhart & MacKenzie 2011). Die Stichprobenrekrutie-

rung erfolgte auf drei Wegen: Fakultätsintern wurde der Link zum Fragebogen per Email über die Lehrbeauftragten aller Fächer an die Universitätsmitarbeiter versendet. Fakultätsextern wurden 315 Humanmediziner und 315 Zahnmediziner mit ausgeglichenerem Geschlechterverhältnis über das Berufsnetzwerk Xing angeschrieben. Die Auswahl der Personen erfolgte nach der Reihenfolge der Trefferliste. Zudem wurde der Link in vier deutsche Medizinerforen eingestellt. Allen Teilnehmern wurde eine Rückmeldung zu den Umfrageergebnissen angeboten. An der von April bis Juli 2011 stattgefundenen Befragung nahmen 128 Mediziner teil, darunter 55 Humanmediziner ($N_{\text{männlich}} = 24$, $N_{\text{weiblich}} = 29$) und 59 Zahnmediziner ($N_{\text{männlich}} = 27$, $N_{\text{weiblich}} = 26$). Keine Angabe zum Arztberuf machten 14 Personen, keine Angaben zum Geschlecht 22 Personen. Die durchschnittliche Arbeitserfahrung lag bei den Humanmedizinern über alle Fragebogenversionen hinweg bei 9.2 Jahren ($SD = 8.64$), bei den Zahnmedizinern bei 12.0 Jahren ($SD = 7,56$). Die durchschnittliche Lehrerfahrung im Studentenunterricht betrug bei den Humanmedizinern über alle Fragebogenversionen hinweg 3.9 Jahre ($SD = 6.70$), bei den Zahnmedizinern 2.0 Jahre ($SD = 4.53$). Die erste Fragebogenversion bearbeiteten 35, die zweite 47 und die dritte 46 Personen. Die soziodemografischen Angaben getrennt nach Human- und Zahnmedizinern sowie nach Fragebogenversion zeigt Tabelle 11.

Tabelle 11: *Demografie der Teilnehmer der Anforderungsanalyse*

Gruppe	Version	Geschlecht		Arbeitsjahre		Lehrjahre	
		w	m	M	SD	M	SD
Humanmediziner	1	7	6	9.46	8.09	2.18	3.16
	2	10	12	8.93	8.65	3.94	7.24
	3	12	6	9.35	9.51	5.06	7.93
Zahnmediziner	1	7	8	9.69	6.70	2.50	5.34
	2	9	10	13.11	5.96	1.29	2.53
	3	10	12	12.83	9.46	2.38	5.68

Anmerkung. w = weiblich, m = männlich.

Die erste Fragebogenversion wurde von weniger Human- und Zahnmedizinern bearbeitet als die anderen zwei Versionen. In der zweiten und dritten Version der zahnmedizinischen Stichprobe war die Arbeitserfahrung höher als in der ersten Fragebogenversion. Humanmediziner, die Fragebogenversion 3 bearbeitet hatten, zeigten eine längere Lehrerfahrung als die anderen human- und zahnmedizinischen Teilnehmer.

Die Teilnehmer waren zum Zeitpunkt der Untersuchung in Tätigkeitsfeldern mit und ohne Patientenkontakt tätig (Tabelle 12).

Tabelle 12: *Tätigkeitsfelder der Teilnehmer der Anforderungsanalyse*

Facharztrichtung Humanmedizin	N	%
Allgemeinmedizin	6	10.9
Anästhesie	3	5.5
Gynäkologie und Urologie	5	9.1
Innere Medizin	11	20.0
Psychiatrie und Psychotherapie	8	15.5
Neurologie	3	5.5
Sonstige mit Patientenkontakt	10	18.2
Sonstige ohne Patientenkontakt	5	9.1
Facharztrichtung Zahnmedizin		
Allgemeine Zahnmedizin	30	50.8
Sonstige mit Patientenkontakt	5	8.5

Vertreter der Inneren Medizin waren bei den Humanmedizinern am häufigsten vertreten. Bei den Zahnmedizinern überwog die allgemeine Zahnheilkunde. Die Verteilung der Tätigkeitsfelder getrennt nach den Fragebogenversionen zeigte deskriptiv keine auffälligen Unterschiede, daher wurde auf die Darstellung hier verzichtet (Anhang B-2).

6.1.4. Statistische Auswertung

Deskriptive Statistik und Voraussetzungsprüfungen. Die Datenanalyse erfolgte mit SPSS.19 (SPSS Inc., Chicago, USA). Für kategoriale Untersuchungsvariablen wurde die Häufigkeit und für metrische Variablen der arithmetische Mittelwert, die Standardabweichung, sowie Minimum und Maximum berechnet. Die Normalverteilungsvoraussetzung wurde mit dem Kolmogorov-Smirnov Test unter Berücksichtigung von Häufigkeits-Diagrammen überprüft. Die abhängigen Variablen waren weder in der Gesamtgruppe, noch getrennt nach Human- und Zahnmedizinern normalverteilt ($p < .05$). Zusammenhänge zwischen der Einschätzung der sozialen Kompetenzen untereinander wurden mittels bivariater Korrelationsanalysen untersucht. Als Korrelationskoeffizient diente Spearmans Rho, da für die Mehrheit der abhängigen Variablen die Normalverteilungsvoraussetzung nicht erfüllt war. Mit gleicher Prüfgröße wurde untersucht, ob die Kontrollvariablen Geschlecht, Tätigkeitsbereich und Anzahl der Arbeits- und Lehrjahre einen Einfluss auf die Einschätzung der Wichtigkeit sozialer Kompetenzen

hatten. Die Anpassung des Signifikanzniveaus erfolgte durch die Bonferroni-Holm-Korrektur.

Inferenzstatistik. Unterschiede zwischen den Fragebogenversionen wurden mit dem Kruskal-Wallis-Test analysiert, der für den Vergleich von mehr als zwei unabhängigen Gruppen mit nicht-parametrischen Daten geeignet ist. Die Unterschiedstestung zwischen Human- und Zahnmedizinern wurde anhand des Kolmogorov-Smirnov-Tests zum Vergleich zweier unabhängiger Gruppen mit nicht-parametrischer Verteilung vorgenommen. Des Weiteren wurde mit der einfaktoriellen univariaten Varianzanalyse mit Messwiederholung geprüft, ob sich die Einschätzung der Wichtigkeit der sozialen Kompetenzen untereinander innerhalb des Studiums bzw. innerhalb des Berufs signifikant voneinander unterschied. Als Messwiederholungsfaktor galt die Zeit, wobei die Faktorstufen durch die zwölf sozialen Kompetenzen definiert wurden. Um zu kontrollieren, ob die Varianzen der Differenzen zwischen jeweils zwei Faktorstufen homogen sind, wurde der Mauchly-Test auf Sphärizität als Voraussetzungsprüfung für die Varianzanalyse durchgeführt. Im Falle von inhomogenen Varianzen der Differenzen wird empfohlen, die korrigierten Werte zur Analyse der Haupteffekte nach Greenhouse-Geisser zu verwenden (Rasch, Friese, Hofmann & Naumann, 2006). Die praktische Bedeutsamkeit der aufgetretenen Unterschiede wurde über das partielle Eta-Quadrat (η_p^2) bestimmt. Für dessen Interpretation wurde die Konvention von Cohen (1988) herangezogen. Dieser empfiehlt Werte ab $\eta_p^2 = .01$ als kleinen Effekt, Werte ab $\eta_p^2 = .06$ als mittleren Effekt und Werte ab $\eta_p^2 = .14$ als großen Effekt zu gewichten. Der Post-Hoc-Vergleich zwischen den einzelnen Faktorstufen erfolgte anhand paarweiser Vergleiche der sozialen Kompetenzen. Die Unterschiede zwischen den sozialen Kompetenzen im Studium vs. Beruf wurde anhand des t-Tests für Differenzen ermittelt. Dieser kann für den Vergleich zweier abhängiger nicht-parametrischer Stichproben eingesetzt werden, wenn die Stichprobe mehr als 30 Personen umfasst (Bortz & Döring, 2006). Die Einschätzung der praktischen Bedeutsamkeit erfolgte anhand der Effektstärke d . Als Interpretationsgrundlage diente die Empfehlung nach Cohen (1988), der eine Effektstärke von $d = .20$ als gering, von $d = .50$ als mittel und $d = .80$ als groß deutet. Aufgrund der Vielzahl von Einzelvergleichen wurde bei allen Unterschiedstestungen das Signifikanzniveau mittels Bonferroni-Holm-Korrektur (Holm, 1979) adjustiert, um eine α -Fehler-Inflation auszuschließen. Dieses Verfahren hat im Vergleich zur einfachen Bonferroni-Korrektur den Vorteil, dass es weniger konservativ testet (Bühner & Ziegler, 2009).

6.1.5. Ergebnisse

Zunächst wurde untersucht, ob Unterschiede in der Beurteilung der Wichtigkeit sozialer Kompetenzen zwischen den drei Fragebogenversionen in den human- und zahnmedizinischen Stichproben auftraten.

Die Humanmediziner bewerteten die Wichtigkeit sozialer Kompetenzen im Studium ähnlich pro Fragebogenversion (oberer Abschnitt Tabelle 13). Der größte deskriptive Unterschied bestand zwischen der Bewertung der Kommunikationsfähigkeit in der ersten und dritten Fragebogenversion, die um einen Skalenpunkt differierte. Die inferenzstatistische Analyse bestätigte, dass sich die Bewertung der einzelnen sozialen Kompetenz im Studium nicht zwischen den Fragebogenversionen unterschied. Die Humanmediziner, die die zweite und dritte Version bearbeitet hatten, beurteilten die sozialen Kompetenzen im Beruf als wichtiger als die, die die erste Version bearbeitet hatten (unterer Abschnitt Tabelle 13). Die Unterschiedstestung ergab, mit Ausnahme der sozialen Unterstützung, keine signifikanten Effekte in der eingeschätzten Wichtigkeit der einzelnen sozialen Kompetenzen zwischen den Versionen. Nach der Bonferroni-Holm-Korrektur trat die Signifikanz für die Wichtigkeit der Suche nach sozialer Unterstützung nicht mehr auf.

Tabelle 13: *Wichtigkeit sozialer Kompetenzen in Studium und Beruf (Humanmediziner)*

	Version 1		Version 2		Version 3		p ^a
	M	SD	M	SD	M	SD	
Im Studium							
Kommunikationsfähigkeit	5.00	1.16	4.68	1.25	3.95	1.43	.11
Durchsetzungsfähigkeit	5.15	1.35	4.77	1.31	4.25	1.71	.28
Empathie	5.00	1.73	4.81	1.75	4.30	1.53	.35
Hilfsbereitschaft	4.62	1.76	4.77	1.74	4.95	1.82	.82
Kompromissfähigkeit	4.38	1.56	4.50	1.66	4.45	1.43	.88
Kontaktfähigkeit	4.92	1.75	4.86	1.46	5.10	1.68	.78
Kritikfähigkeit	5.31	1.60	5.14	1.46	5.25	1.29	.92
Verantwortungsübernahme	5.23	1.54	5.09	1.60	4.68	1.86	.69
Respekt	5.85	1.41	5.23	1.69	5.45	1.76	.52
Soziale Unterstützung	4.62	1.50	5.55	1.37	5.25	1.92	.20
Selbstbeobachtung/Selbstreflexion	5.46	1.13	5.67	1.56	5.70	1.53	.61
Selbstkontrolle/Selbstdisziplin	5.38	1.50	5.36	1.43	5.35	1.66	1.00

Fortsetzung Tabelle 13:

Im Beruf							
Kommunikationsfähigkeit	6.00	2.08	6.20	1.74	6.10	1.80	.80
Durchsetzungsfähigkeit	5.23	2.05	5.64	1.50	5.70	1.46	.93
Empathie	5.85	2.23	6.18	1.76	6.45	1.15	.99
Hilfsbereitschaft	5.08	1.98	5.68	1.43	5.95	1.32	.41
Kompromissfähigkeit	5.00	1.83	5.32	1.25	5.95	1.32	.12
Kontaktfähigkeit	5.62	1.76	5.64	1.62	4.95	1.67	.22
Kritikfähigkeit	5.62	2.14	6.05	1.79	5.95	1.36	.42
Verantwortungsübernahme	5.69	1.97	6.23	1.63	6.25	1.59	.32
Respekt	5.69	1.97	6.09	1.82	6.25	1.48	.49
Soziale Unterstützung	5.69	1.73	5.64	1.43	6.15	1.63	.02*
Selbstbeobachtung/Selbstreflexion	5.77	1.70	6.10	1.68	6.30	1.42	.35
Selbstkontrolle/Selbstdisziplin	5.62	1.81	5.91	1.27	5.80	1.61	.94

Anmerkung. Bewertung von 1 = sehr unwichtig bis 7 = sehr wichtig. $n_{\text{Version1}} = 13$, $n_{\text{Version2}} = 22$, $n_{\text{Version3}} = 20$. ^a Irrtumswahrscheinlichkeit des Kruskal-Wallis-Tests; $p \leq .05$.

Die Zahnmediziner beurteilten die Wichtigkeit sozialer Kompetenzen im Studium in den Fragebogenversionen ähnlich hoch wie die Humanmediziner (oberer Abschnitt Tabelle 14). Die stärkste Differenz trat in der Bewertung der Hilfsbereitschaft auf, die sich in der ersten und dritten Fragebogenversion um 1.1 Skalenpunkte unterschied. Die Signifikanztestung zeigte keine Unterschiede zwischen den Fragebogenversionen ($p > .05$). Hinsichtlich des Berufs schätzten die Zahnmediziner, welche die erste Version bearbeitet hatten, die sozialen Kompetenzen etwas weniger wichtig ein als die, die die zweite oder dritte Version bearbeiteten (unterer Abschnitt Tabelle 14). Die Unterschiede zeigten keine Signifikanz.

Tabelle 14: *Wichtigkeit sozialer Kompetenzen in Studium und Beruf (Zahnmediziner)*

Im Studium	Version 1		Version 2		Version 3		p^a
	M	SD	M	SD	M	SD	
Kommunikationsfähigkeit	5.19	1.17	5.05	1.84	5.00	1.48	.92
Durchsetzungsfähigkeit	4.81	1.64	4.84	1.64	4.73	1.52	.96
Empathie	4.50	1.90	4.84	1.89	5.05	1.79	.66
Hilfsbereitschaft	5.81	1.56	5.21	1.78	4.70	1.87	.19
Kompromissfähigkeit	5.88	0.96	4.95	1.55	4.86	2.08	.25
Kontaktfähigkeit	5.94	0.85	5.74	1.37	5.33	1.71	.71

Fortsetzung Tabelle 14:

	Version 1		Version 2		Version 3		p ^a
	M	SD	M	SD	M	SD	
Im Studium							
Kritikfähigkeit	5.75	1.00	5.68	1.49	5.76	1.79	.60
Verantwortungsübernahme	5.06	1.34	5.26	1.79	5.45	1.87	.58
Respekt	5.80	1.90	5.32	1.64	5.68	1.62	.34
Soziale Unterstützung	5.64	1.50	5.37	1.42	5.68	1.76	.47
Selbstbeobachtung/Selbstreflexion	5.25	1.88	5.32	1.73	5.68	1.73	.58
Selbstkontrolle/Selbstdisziplin	5.75	1.73	6.00	1.53	5.73	1.67	.88
Im Beruf							
Kommunikationsfähigkeit	6.00	1.50	6.10	1.43	6.10	1.64	.90
Durchsetzungsfähigkeit	5.23	1.49	5.60	1.46	5.70	1.36	.72
Empathie	5.85	1.54	6.10	1.43	6.45	1.76	.91
Hilfsbereitschaft	5.08	1.41	5.60	1.77	5.95	1.76	.82
Kompromissfähigkeit	5.00	1.15	5.30	1.34	5.95	1.56	.28
Kontaktfähigkeit	5.62	1.46	5.60	1.63	4.95	1.45	.43
Kritikfähigkeit	5.62	1.12	5.95	1.74	5.95	1.60	.43
Verantwortungsübernahme	5.69	1.41	6.15	1.53	6.25	1.81	.63
Respekt	5.69	1.66	6.00	1.86	6.25	1.87	.69
Soziale Unterstützung	5.00	1.33	5.55	1.51	6.15	1.76	.06
Selbstbeobachtung/Selbstreflexion	5.77	1.53	6.10	1.40	6.30	1.59	.80
Selbstkontrolle/Selbstdisziplin	5.62	1.54	5.85	1.43	5.80	1.46	.73

Anmerkung. Bewertung von 1 = sehr unwichtig bis 7 = sehr wichtig. n_{Version1} = 16, n_{Version2} = 19, n_{Version3} = 22. ^a Irrtumswahrscheinlichkeit des Kruskal-Wallis-Tests.

Nachdem festgestellt wurde, dass sich die Ergebnisse nicht systematisch zwischen den eingesetzten Fragebogenversionen unterschieden, wurden die Beurteilungen über die Versionen zusammengefasst. Über die Versionen hinweg beurteilten die Zahnmediziner die Wichtigkeit sozialer Kompetenzen im Studium bis auf die Fähigkeit zur Selbstbeobachtung/Selbstdisziplin zwar etwas höher als die Humanmediziner (oberer Abschnitt Tabelle 15), die Signifikanztestung zeigte jedoch keine bedeutsamen Unterschiede. Ausnahme bildete die Kompromissfähigkeit ($p = .02$), welche jedoch nach der Bonferroni-Holm-Korrektur keine Signifikanz mehr aufwies. Die Bewertung der Wichtigkeit sozialer Kompetenzen im Beruf fiel zwischen den Human- und Zahnmedizinern ähnlich aus (unterer Abschnitt Tabelle 15). Eine signifikante Differenz trat bezüglich der Selbstkontrolle auf, die die Zahnmediziner mit 0.5 Skalenpunkten wichtiger als die Humanmediziner bewerteten. Der signifikante Unterschied persistierte aber nach der Alpha-Fehler-Korrektur nicht.

Tabelle 15: *Wichtigkeit sozialer Kompetenzen in Studium vs. Beruf (Human- vs. Zahnmediziner)*

Studium	Humanmedizin		Zahnmedizin		p ^a
	M	SD	M	SD	
Kommunikationsfähigkeit	4.49	1.35	5.07	1.51	.37
Durchsetzungsfähigkeit	4.67	1.49	4.79	1.57	.99
Empathie	4.67	1.66	4.82	1.83	.96
Hilfsbereitschaft	4.80	1.75	5.20	1.78	.45
Kompromissfähigkeit	4.45	1.53	5.18	1.68	.02*
Kontaktfähigkeit	4.96	1.59	5.64	1.39	.28
Kritikfähigkeit	5.22	1.41	5.73	1.47	.19
Verantwortungsübernahme	4.98	1.67	5.28	1.70	.75
Respekt	5.45	1.64	5.59	1.68	1.00
Soziale Unterstützung	5.22	1.63	5.56	1.56	.32
Selbstbeobachtung/Selbstreflexion	5.63	1.43	5.44	1.75	.99
Selbstkontrolle/Selbstdisziplin	5.36	1.51	5.82	1.62	.14
Beruf					
Kommunikationsfähigkeit	6.11	1.81	6.39	1.51	.94
Durchsetzungsfähigkeit	5.56	1.61	5.72	1.41	1.00
Empathie	6.20	1.68	6.30	1.57	1.00
Hilfsbereitschaft	5.64	1.54	5.47	1.64	.90
Kompromissfähigkeit	5.47	1.45	5.42	1.38	.99
Kontaktfähigkeit	5.38	1.67	5.80	1.49	.67
Kritikfähigkeit	5.91	1.71	5.65	1.52	.19
Verantwortungsübernahme	6.11	1.69	6.29	1.57	1.00
Respekt	6.05	1.73	5.74	1.79	.53
Soziale Unterstützung	5.67	1.61	5.42	1.58	.77
Selbstbeobachtung/Selbstreflexion	6.09	1.57	6.16	1.49	1.00
Selbstkontrolle/Selbstdisziplin	5.80	1.51	6.28	1.45	.02*

Anmerkung. Bewertung von 1 = sehr unwichtig bis 7 = sehr wichtig. n_{Humanmediziner} = 55, n_{Zahnmediziner} = 57. ^a Irrtumswahrscheinlichkeit des Kolmogorov-Smirnov-Tests; * p ≤ 0.05.

Aufgrund der nicht bedeutsamen Unterschiede in den Urteilen der Human- und Zahnmediziner wurden die Beurteilungen der sozialen Kompetenzen für die Human- und Zahnmediziner zusammengeführt (Tabelle 16).

Tabelle 16: *Wichtigkeit sozialer Kompetenzen in Studium vs. Beruf (Gesamtgruppe)*

	Studium		Beruf		t	p	d
	M	SD	M	SD			
Kommunikationsfähigkeit	4.72	1.49	6.19	1.71	-9.72	.00**	.92
Durchsetzungsfähigkeit	4.72	1.53	5.63	1.54	-6.67	.00**	.59
Empathie	4.69	1.81	6.15	1.69	-8.52	.00**	.83
Hilfsbereitschaft	4.93	1.80	5.46	1.63	-3.53	.00**	.31
Kompromissfähigkeit	4.83	1.63	5.42	1.43	-4.03	.00**	.39
Kontaktfähigkeit	5.22	1.62	5.52	1.62	-2.58	.00**	.19
Kritikfähigkeit	5.44	1.48	5.78	1.61	-3.17	.00**	.22
Verantwortungsübernahme	5.09	1.68	6.19	1.60	-7.19	.00**	.67
Respekt	5.46	1.71	5.87	1.73	-3.24	.00**	.24
Soziale Unterstützung	5.36	1.64	5.52	1.64	-1.69	.09*	.10
Selbstbeobachtung/Selbstreflexion	5.47	1.67	6.07	1.56	-5.58	.00**	.37
Selbstkontrolle/Selbstdisziplin	5.64	1.57	6.06	1.48	-4.64	.00**	.28

Anmerkung. Bewertung von 1 = sehr unwichtig bis 7 = sehr wichtig. $n_{\text{Humanmediziner}} = 55$, $n_{\text{Zahnmediziner}} = 57$; * $p \leq .05$; ** $p \leq .01$.

Die Teilnehmer bewerteten die Wichtigkeit sozialer Kompetenzen für das Studium im Mittel zwischen 4.7 und 5.6. Die niedrigste Bewertung erhielt die Empathie, gefolgt von Kommunikationsfähigkeit und Durchsetzungsfähigkeit. Die Selbstkontrolle/Selbstdisziplin wurde als am wichtigsten bewertet. Hinsichtlich der Wichtigkeit sozialer Kompetenzen im Beruf lagen die Werte zwischen 5.4 und 6.2. Von geringster Wichtigkeit wurde die Kompromissfähigkeit bewertet, von höchster die Kommunikationsfähigkeit und Verantwortungsübernahme. Alle sozialen Kompetenzen wurden für den Beruf als signifikant wichtiger als für das Studium der Human- und Zahnmedizin eingeschätzt. Nach der Bonferroni-Holm-Korrektur erreichte die soziale Unterstützung nicht mehr die Grenze der Signifikanz. Die statistischen Unterschiede zeigten für die Kommunikationsfähigkeit und Empathie eine hohe praktische Bedeutsamkeit ($d > .80$). Für die Durchsetzungsfähigkeit und Verantwortungsübernahme resultierten mittlere Effektstärken ($d > .50$). Die restlichen Unterschiede zwischen den sozialen Kompetenzen im Studium vs. Beruf zeigten eine geringe praktische Bedeutsamkeit ($d < .50$).

Aufbauend auf die deskriptiven Daten der Tabelle 16 wurde untersucht, ob sich die Mittelwerte der beurteilten sozialen Kompetenzen untereinander im Studium und im Beruf signifikant unterschieden. Der als Voraussetzungsprüfung für die Varianzanalyse eingesetzte Mauchly-Test auf Sphärizität zeigte, dass die Differenzen zwischen den Faktorstufen bei der Bewertung der sozialen Kompetenzen im Studium und im Beruf

heterogen ausfielen (Studium: Mauchly-W(65) = .15, $\chi^2(65) = 205.88$, $p = .00$; Beruf: Mauchly-W(65) = .12, $\chi^2(65) = 222.75$, $p = .00$), sodass die korrigierten Werte zur Analyse der Haupteffekte nach Greenhouse-Geisser verwendet wurden.

Die Varianzanalyse mit Messwiederholung zeigte bei den sozialen Kompetenzen im Studium ein hochsignifikantes Ergebnis ($F(1,8) = 10.4$, $p = .00$). Somit wurden die sozialen Kompetenzen für das Studium global unterschiedlich wichtig eingeschätzt. Der signifikante Effekt hatte eine mittlere praktische Bedeutsamkeit ($\eta^2 = .09$). Um zu analysieren, welche der sozialen Kompetenzen signifikant unterschiedlich bewertet wurden, wurden paarweise Vergleiche (Tabelle 17) durchgeführt. Diese ergaben, dass die Selbstkontrolle/Selbstdisziplin, Selbstbeobachtung/Selbstreflexion, Respektfähigkeit, Kontaktfähigkeit, Kritikfähigkeit und Suche nach sozialer Unterstützung als signifikant wichtiger für das Studium eingeschätzt wurden als die Kommunikationsfähigkeit, Durchsetzungsfähigkeit, Empathie, Hilfsbereitschaft und Kompromissfähigkeit. Die sechs als am wichtigsten eingeschätzten sozialen Kompetenzen unterschieden sich, mit Ausnahme der Selbstkontrolle/Selbstdisziplin, nicht signifikant untereinander. Die Selbstkontrolle/Selbstdisziplin wurde als wichtiger eingeschätzt als die Kontaktfähigkeit. Die Verantwortungsübernahme wurde nicht eindeutig bewertet. Sie wurde als signifikant wichtiger als die Kommunikationsfähigkeit und Empathie eingeschätzt, aber signifikant weniger wichtig als die Selbstkontrolle/Selbstdisziplin, Selbstbeobachtung/Selbstreflexion und Respektfähigkeit.

Die Wichtigkeit sozialer Kompetenzen für den Beruf wurden ebenfalls signifikant unterschiedlich mit einer mittleren praktischen Bedeutsamkeit bewertet ($F(1,8) = 13.8$; $p = .00$; $\eta^2 = .11$). Die paarweisen Vergleiche (Tabelle 17) zeigten, dass Kommunikationsfähigkeit, Empathie, Verantwortungsübernahme, Selbstbeobachtung/Selbstreflexion, Selbstkontrolle/Selbstdisziplin als am wichtigsten für den Beruf als Human- und Zahnmediziner eingeschätzt wurden. Sie wurden signifikant wichtiger als Durchsetzungsfähigkeit, Hilfsbereitschaft, Kompromissfähigkeit und Kontaktfähigkeit beurteilt und überwiegend wichtiger als Kritikfähigkeit, Respektfähigkeit und Suche nach sozialer Unterstützung. Die am wichtigsten klassifizierten fünf sozialen Kompetenzen unterschieden sich nicht signifikant untereinander.

Tabelle 17: *Paarweiser Vergleich der Wichtigkeit sozialer Kompetenzen in Studium und Beruf*

	1	2	3	4	5	6	7	8	9	10	11	12
Ausdrucksfähigkeit (1)	-	0.55	0.03	0.73	0.79	0.66	0.46	0.03	0.33	0.76	0.15	0.13
Durchsetzungsfähigkeit (2)	-0.04	-	-0.52	0.08	0.25	0.11	-0.08	-0.52	-0.22	0.21	-0.40	-0.36
Empathie (3)	-0.05	-0.01	-	0.70	0.76	0.63	0.44	0,00	0,30	0.73	0.12	0.16
Hilfsbereitschaft (4)	-0.31	-0.28	-0.27	-	0.06	-0.07	-0.26	-0.70	-0.40	0.03	-0.58	-0.55
Kompromissfähigkeit (5)	-0.07	-0.04	-0.03	-0.24	-	-0.14	-0.33	-0.76	-0.46	-0.04	-0.65	-0.61
Kontaktfähigkeit (6)	-0,58	-0,55	-0,53	-0,27	-0,51	-	-0.19	-0.63	-0.33	0.10	-0.51	-0.47
Kritikfähigkeit (7)	-0,72	-0,69	-0,68	-0,41	-0,65	-0.14	-	-0.44	-0.14	0.29	-0.32	-0.28
Verantwortungsübernahme (8)	-0,38	-0,34	-0,33	-0,06	-0,30	0.21	0.35	-	0.30	0.73	0.12	0.16
Respekt (9)	-0,83	-0,80	-0,79	-0,52	-0,76	-0.25	-0.11	-0,45	-	0.43	-0.13	-0.15
Soziale Unterstützung (10)	-0,63	-0,60	-0,59	-0,32	-0,56	-0.05	0.09	-0.26	0.20	-	-0.61	-0.57
Selbstbeobachtung/Selbstreflexion (11)	-0,85	-0,81	-0,80	-0,54	-0,77	-0.27	-0.13	-0.47	-0.02	-0.21	-	0.04
Selbstkontrolle/Selbstdisziplin (12)	-0,96	-0,92	-0,91	-0,64	-0,88	-0,38	-0.23	-0,58	-0.13	-0.32	-0.11	-

Anmerkung. N = 112. Unterhalb der Diagonale werden mittlere Differenzen zwischen den sozialen Kompetenzen im Studium dargestellt. Oberhalb der Diagonale werden mittlere Differenzen zwischen den sozialen Kompetenzen im Beruf dargestellt. Fett markierte Werte zeigen signifikante Unterschiede auf dem 1%-Niveau. Alle Signifikanzwerte wurden mittels Bonferroni-Korrektur adjustiert.

Zur Absicherung der Ergebnisse wurde der Einfluss der Kontrollvariablen Geschlecht, Anzahl der Arbeitsjahre und Lehrjahre analysiert (Tabelle 18). Ein Einfluss der Kontrollvariablen konnte mit der Ausnahme des Geschlechts ausgeschlossen werden. Das Geschlecht korrelierte mit der Wichtigkeit der Verantwortungsübernahme im Studium. Frauen schätzten die Verantwortungsfähigkeit dabei signifikant ($p = .00$) wichtiger für das Studium ein als Männer ($M_w = 5.6$ vs. $M_m = 4.6$, $t(104) = -3.22$). Die praktische Bedeutsamkeit des Effekts war gering ($\eta^2 = .30$).

Tabelle 18: *Zusammenhang Wichtigkeit sozialer Kompetenzen mit Kontrollvariablen*

	Geschlecht		Arbeitsjahre		Lehrjahre	
	Studium	Beruf	Studium	Beruf	Studium	Beruf
Kommunikationsfähigkeit	.15	.12	-.09	-.16	.07	-.07
Durchsetzungsfähigkeit	.02	-.06	.03	-.14	-.02	-.03
Empathie	.21	.11	-.13	-.13	.19	.14
Hilfsbereitschaft	.08	.11	.00	-.08	.09	.03
Kompromissfähigkeit	.21	.11	-.11	-.05	-.01	-.06
Kontaktfähigkeit	.25	.05	-.08	-.07	-.03	-.09
Kritikfähigkeit	.07	.12	-.04	.01	-.02	.01
Verantwortungsübernahme	.28**	.16	-.07	-.08	.03	.03
Respekt	.25	.23	-.05	-.21	.06	.20
Soziale Unterstützung	.05	.16	-.08	-.18	.05	-.04
Selbstbeobachtung/Selbstreflexion	.07	.01	-.14	.01	.18	.09
Selbstkontrolle/Selbstdisziplin	-.01	-.04	-.01	.01	-.04	-.01

Anmerkung. N = 112; ** $p \leq .01$.

6.1.6. Diskussion

Die Diskussion widmet sich zuerst der gewählten Stichprobe und dem Fragebogen. Anschließend geht sie auf die als am wichtigsten beurteilten sozialen Kompetenzen für das Medizinstudium ein und betrachtet abschließend den aufgetretenen Unterschied in der Wichtigkeit sozialer Kompetenzen in Studium vs. Beruf.

Stichprobe und Fragebogen. Durch die unterschiedlichen Wege der Stichprobenrekrutierung wurde versucht, Personen aus unterschiedlichen Fachrichtungen und Institutionen zu berücksichtigen, um einer institutionellen Verzerrung der Daten vorzubeugen (Ployhart & MacKenzie 2011). Zum Schutz der Anonymität wurde bei der Datenerhebung jedoch nicht erfasst, ob die Teilnehmer fakultätsintern oder -extern waren,

da anhand des Geschlechts, des medizinischen Tätigkeitsbereiches, sowie der Anzahl der Arbeits- und Lehrjahre in kleinen Fächern der Medizinischen Fakultät Rückschlüsse auf Personen möglich wären. Ob die approbierten Mediziner am besten geeignet waren die Wichtigkeit sozialer Kompetenzen zu bewerten, kann nicht eindeutig belegt werden. Zwar können sie aufgrund ihres erfolgreich absolvierten Studiums mit ihren Studienerfahrungen als Experten für das Studium gelten, aber inwieweit sie dies in Bezug auf soziale Kompetenzen waren, ist ungewiss. Vermutlich wurden jedoch mit der freiwilligen Teilnahme eher die Mediziner angesprochen, die an der Thematik sozialer Kompetenzen interessiert sind und ihnen sensibler gegenüberstehen. Was für die Validität der Ergebnisse spricht, ist, dass sowohl die Gruppe der Human- als auch Zahnmediziner, sowie die Gesamtgruppe die Wichtigkeit sozialer Kompetenzen gleich einschätzten. Als alternative Personengruppe hätten Lehrende unterschiedlicher Fakultäten, die auf das Training von Arzt-Patienten-Kommunikation in der medizinischen Ausbildung spezialisiert sind, befragt werden können. Dies hätte jedoch nur ein kleines Spektrum der zwischenmenschlichen Beziehungen während des Studiums abgebildet. Für die Konstruktvalidität des Fragebogens spricht, dass keine Unterschiede zwischen den Fragebogenversionen auftraten, sodass die gewählten Beispiele aus Studium und Beruf die Ergebnisse nicht verzerrt zu haben scheinen.

Wichtigkeit sozialer Kompetenzen im Studium. Als am relevantesten für das Medizinstudium wurden die Selbstkontrolle/Selbstdisziplin, Selbstbeobachtung/Selbstreflexion, Respektfähigkeit, Kontaktfähigkeit, Kritikfähigkeit und Suche nach sozialer Unterstützung beurteilt. Die Selbstkontrolle/Selbstdisziplin wurde unter diesen nochmals wichtiger als die Kontaktfähigkeit bewertet. Nachfolgend werden die Ergebnisse argumentativ untermauert und mit den für die medizinische Ausbildung relevanten CanMed-Rollen (Frank et al., Entwurf 2015; Kapitel 3.3) in Verbindung gebracht. Die Selbstkontrolle/Selbstdisziplin als Selbststeuerungsprozess ist eine Voraussetzung zum selbstregulierten Lernen (Mace, Belfiore & Shea, 1998; Schiefele & Pekrun, 1996). Das Medizinstudium erfordert ein hohes Maß an selbstreguliertem Lernen, um die hohe Stoffmenge und die Prüfungsanforderungen zu bewerkstelligen. Diese selbstregulatorischen Kompetenzen sind in der CanMed-Rolle des *Professionals* abgebildet, der seine eigenen Gefühle, Gedanken und Impulse kontrollieren muss, um professionell zu handeln. Die Selbstreflexion ist eine wichtige Voraussetzung für selbstreguliertes und auf Erfahrung beruhendes Lernen. Der eigene Lernstand wird durch Selbstreflexion bewusst und kann dadurch weiterentwickelt werden (Ellis & Davidi, 2005; Ellis, Carette, Anseel & Lievens, 2014; Tashiro, Shimpuku, Naruse, Matsutani & Matsutani, 2013). Die Selbstreflexion findet sich in den CanMed-Rollen des *Collaborators* und des

Scholars wieder, die ihre eigenen Handlungen, Kenntnisse und Einstellungen reflektieren müssen, um effektiv zu arbeiten. Wenn eine Person beim Lernen allein nicht mehr weiterkommt, ist die Suche nach sozialer Unterstützung entscheidend, um angemessene Hilfe von Kommilitonen oder Lehrenden aufzusuchen. Soziale Unterstützung hilft den Studierenden, Unsicherheit zu reduzieren und ihre Umwelt zu kontrollieren (Albrecht & Adelman, 1987; Cohen & Wills, 1985; Lakey & Cohen, 2000; Thoits, 1995). Die soziale Unterstützung hat einen direkten und einen indirekten Effekt - moderiert über das Stressempfinden und die akademische Selbstwirksamkeit - auf das Lernen und die Zufriedenheit im Studium (Tsay, 2013). Die Suche nach sozialer Unterstützung bildet sich in der *CanMed*-Rolle des *Scholars* ab, da dieser für seinen Lernprozess registrieren muss, wann er Hilfe von anderen benötigt. Um andere Personen um Unterstützung zu bitten, ist es notwendig auf andere zuzugehen und die Kontakte aufrecht zu erhalten. Die Kontaktfähigkeit ist darüber hinaus für die soziale Integration des Studierenden an der Hochschule relevant (Bean, 1980, 1985; Heublein et al., 2003; Tinto, 1975, 1993). Diese bildet sich insbesondere in den Rollen des *Communicators*, *Collaborators* und *Health Advocate* ab, da diese Rollen ein hohes Maß verlangen, auf andere Personen zugehen zu können. Lernen setzt auch ein gewisses Maß an Kritikfähigkeit voraus, um Feedback anzunehmen und zu geben. Die Medizinstudierenden erhalten nicht nur qualitatives Feedback in Übungssituationen und quantitatives Feedback über Prüfungsleistungen, sondern werden auch aufgefordert, ihren Kommilitonen in Übungssituationen wie dem medizinischen Kommunikationstraining konstruktives Feedback zu geben. Konstruktives Feedback steigert Motivation und Leistungen (Fishbach & Finkelstein, 2012; Hattie & Timperley, 2007; Zhi-Feng & Lee, 2013). Die Kritikfähigkeit wird durch die *CanMed*-Rolle des *Scholars* und des *Collaborators* repräsentiert, da sie Feedback geben und empfangen und die Arbeit im Team Konfliktpotenzial bereithält. Dass Respektfähigkeit als wichtig für das Studium eingeschätzt wurde, geht vermutlich darauf zurück, dass sie ein integraler Bestandteil des ethischen und professionellen Verhaltens in der medizinischen Ausbildung und Praxis ist (Beach, Duggan, Cassel & Geller, 2007). Dies umfasst nicht nur respektvolles Verhalten gegenüber Patienten, sondern betrifft alle Personengruppen, mit denen zusammengearbeitet wird (Association of American Medical Colleges, 1998; Frank et al., Entwurf 2015; Robins, Braddock, Fryer-Edwards, 2002). Die Respektfähigkeit kommt insbesondere in den *CanMed*-Rollen des *Communicators* und *Professionals* zum Tragen, da sie einen engen Patientenkontakt beinhalten. Insgesamt erscheinen die Ergebnisse hinsichtlich der Literatur schlüssig.

Wichtigkeit sozialer Kompetenzen in Studium vs. Beruf. Alle sozialen Kompetenzen, mit Ausnahme der Suche nach sozialer Unterstützung, wurden für den Arztberuf als wichtiger als für das Medizinstudium eingeschätzt, wobei die Kommunikationsfähigkeit, Empathie, Durchsetzungsfähigkeit und Verantwortungsübernahme eine große bis mittlere praktische Bedeutsamkeit zeigten. Insbesondere Kommunikationsfähigkeit, Empathie und Verantwortungsübernahme stellen soziale Kompetenzen dar, die für die Patientenversorgung im ärztlichen Beruf zentral sind. Die Durchsetzungsfähigkeit spielt vermutlich im Beruf ebenfalls eine größere Rolle als im Studium, da z.B. Befunde und Therapievorhaben gegenüber dem Team und den Vorgesetzten vertreten werden müssen. Insgesamt kann es als schlüssig gewertet werden, dass die Kompetenzen wichtiger für den Beruf als für das Studium beurteilt wurden, da der Arztberuf stärker die Arzt-Patienten-Beziehung in den Fokus stellt und eine höhere Verantwortung beinhaltet als unter den Ausbildungsbedingungen des Studiums. Auch entwickeln sich soziale Kompetenzen durch Rückkopplungsprozesse über die Jahre (Malti & Perren, 2008). Vor diesem Hintergrund könnte die Wichtigkeit sozialer Kompetenzen im Beruf als höher gewichtet worden sein. Für die Suche nach sozialer Unterstützung lässt sich vermuten, dass sie für das Studium und den Beruf gleich relevant ist, da die Unterstützung von anderen, sei es zur Prüfungsvorbereitung im Studium oder zum Treffen von wichtigen Entscheidungen im Arztberuf, hilfreich ist. Die sechs als am wichtigsten für das Medizinstudium eingeschätzten sozialen Kompetenzen sollen als Grundlage für die *SJT*-Entwicklung dienen.

6.2. Workshops zur Itementwicklung

Das Generieren der Situationen und der Antwortoptionen fand innerhalb zweier Workshops an der Medizinischen Fakultät Heidelberg im März und April 2012 statt. Das Vorgehen orientierte sich an den in Kapitel 4.3 beschriebenen Testentwicklungsschritten von Ployhart und MacKenzie (2011) unter Berücksichtigung der Testentwicklungsaspekte von Weekley et al. (2006). Zur Gewinnung von Teilnehmern wurden Mitarbeiter der Medizinischen Fakultät und des Universitätsklinikums Heidelberg angesprochen, die direkt oder indirekt in die medizinische Lehre eingebunden waren. Dabei wurden Mitarbeiter verschiedener Berufsgruppen, Fachbereiche und Positionen berücksichtigt, um vielfältige Perspektiven bei der Testentwicklung einzubeziehen. Zur Rekrutierung von fortgeschrittenen Studierenden wurde über die studentischen Fachschaften, Kurssprecher und Aushänge geworben.

6.2.1. Generieren der Situationen

Am ersten Workshop im März 2012 nahmen elf Personen, darunter zwei Humanmediziner, drei Zahnmediziner, zwei Naturwissenschaftler, zwei Pflegekräfte, ein Diplom-Psychologe und eine Erziehungswissenschaftlerin teil. Eine genaue Übersicht der Teilnehmer nach Geschlecht, Berufsgruppe, Fachgebiet und Position befindet sich im Anhang C-1. Der Workshop hatte zum Ziel, möglichst viele typische und erfolgskritische soziale Situationen des Human- und Zahnmedizinstudiums für die sechs in der Anforderungsanalyse als am wichtigsten für ein Medizinstudium beurteilten sozialen Kompetenzen zu entwickeln. Dies betraf die sozialen Kompetenzen Selbstkontrolle/Selbstdisziplin, Selbstbeobachtung/Selbstreflexion, Respektfähigkeit, Kontaktfähigkeit, Kritikfähigkeit und Suche nach sozialer Unterstützung.

Nach der Begrüßung und Vorstellung der Teilnehmer erfolgte eine theoretische Einführung zu sozialen Kompetenzen in der medizinischen Studierendenauswahl, dem *SJT* als Messverfahren, der *SJT*-Entwicklung und der vorausgegangenen Anforderungsanalyse. Anschließend führten die Teilnehmer ein individuelles Brainstorming zu kritischen Situationen für zwei der sechs sozialen Kompetenzen durch. Sie wurden aufgefordert, möglichst typische, herausfordernde Situationen zu entwerfen, die Studierende im Umgang mit Kommilitonen, Lehrenden, Ärzten, Pflegepersonal, Patienten oder deren Angehörigen wiedergeben. In den zwei darauf folgenden Arbeitsphasen fanden sich jeweils zwei Teilnehmer zusammen, die für dieselbe soziale Kompetenz zuständig waren. Die Tandems wurden gebeten, voneinander unabhängig Situationen auf einem für Abiturienten angemessenen Sprachniveau zu formulieren. Zwischen den Arbeitsphasen fanden kurze Diskussionsrunden statt, um einen Vergleich der Arbeitsergebnisse zwischen den Tandems vorzunehmen. Der Workshop wurde durch ein Blitzlicht zum Arbeitsergebnis und einem Ausblick auf den nächsten Workshop abgeschlossen.

Die Situationen wurden anschließend aufbereitet. Neben der Korrektur von Rechtschreib- und Kommafehlern wurden Fremdwörter und medizinische Fachbegriffe durch einfache Wörter ersetzt. Komplexe Satzstrukturen wurden in mehrere Hauptsätze aufgelöst und sprachlich nicht ausgereifte Formulierungen abgeändert. Im Sinne der Genderfairness wurden Wörter ausgetauscht, die nicht beide Geschlechter ansprachen. Des Weiteren wurden alle Situationen als direkte Anrede umformuliert (z.B. „*Sie* sind Zahnmedizinstudierende(r) im 4. Semester...“). Die Situationen, welche sowohl Szenen des Human- als auch des Zahnmedizinstudiums wiedergeben, wurden durch die Namen beider Studiengänge ergänzt (z.B. „Als Studierende(r) der *Human-/Zahnmedizin*...“). Stichpunktartige Situationsbeschreibungen wurden ausformuliert

oder mit einer Bitte zur weiteren Konkretisierung für den folgenden Workshop versehen. Situationen, die für die Erhebung mehrerer sozialer Kompetenzen geeignet waren, wurden zusätzlich einer anderen sozialen Kompetenz zugeordnet, um auf einen großen Itempool zurückgreifen zu können. Nach der Aufbereitung lagen 71 Situationsbeschreibungen vor.

6.2.2. Generieren der Verhaltensweisen

Der zweite Workshop fand im April 2012 statt. Die 15 Teilnehmer setzten sich aus fünf Humanmedizinern, vier Zahnmedizinern, zwei Naturwissenschaftlern, zwei Pflegekräften, einem Diplom-Psychologen und einer Erziehungswissenschaftlerin zusammen. Zehn dieser Personen hatten bereits am ersten Workshop teilgenommen. Eine genaue Übersicht der Teilnehmer nach Geschlecht, Berufsgruppe, Fachgebiet und Position befindet sich im Anhang C-2. Das Ziel dieses Workshops bestand darin, für die im ersten Workshop entwickelten Situationsbeschreibungen möglichst viele qualitativ hochwertige Antwortoptionen zu entwerfen, die die sechs sozialen Kompetenzen wieder spiegeln.

Nach der Begrüßung und Vorstellung der Teilnehmer folgte eine theoretische Einführung zur Entwicklung von Antwortoptionen eines *SJT*s. Danach erhielten die Teilnehmer den Auftrag, so viele Handlungsalternativen wie möglich für die im ersten Workshop entwickelten Situationen zu formulieren. Personen, die bereits am ersten Workshop teilgenommen hatten, wurden zusätzlich gebeten, ihre ggf. unkonkret formulierten Situationen zu ergänzen. Für die Entwicklung der Handlungsmöglichkeiten erhielten die Teilnehmer eine Checkliste (Abbildung 18), um die in Kapitel 4.3 erläuterten Aspekte der Antwortentwicklung einzubeziehen. Um vielfältige Verhaltensweisen zu generieren, erhielten die Teilnehmer als Anregung zusätzlich die sieben Strategien im Umgang mit der eigenen oder anderen Person (Stemler & Sternberg 2006; Stemler et al., 2006).

- ✓ Antworten sollten niedrige, mittlere und hohe Ausprägungen der zu messenden sozialen Kompetenz abbilden (Konstruktbezug).
- ✓ Die Antworten sollten möglichst sozial akzeptabel formuliert sein (Reduktion der Transparenz).
- ✓ Die beschriebenen Handlungen sollten realistisch sein (Realitätsnähe).
- ✓ Die Antworten sollten so konkret wie möglich und nötig formuliert werden (Reduktion der Komplexität).
- ✓ Die Länge der Antworten sollte möglichst gleich lang sein (Vermeidung von Hinweisreizen).
- ✓ Die Sprache sollte für die Zielgruppe der Abiturienten angemessen sein. Bitte keine medizinischen Fachbegriffe verwenden (Testfairness).

Abbildung 18. Checkliste zur Entwicklung der Antwortoptionen des SJTs.

Für die Antwortentwicklung stand den Teilnehmern ein Arbeitsblatt zur Verfügung (Anhang C-3). Dieses wurde auch für die anschließende Arbeitsphase im Tandem verwendet. In dieser sollten die Teilnehmer einschätzen, welche Ausprägung der zu messenden sozialen Kompetenz die einzelnen Antwortoptionen widerspiegeln. Darüber hinaus sollten sie anhand der Checkliste (Abbildung 18) Vorschläge zur Optimierung der Antwortoptionen geben. Bei Bedarf tauschten sich die Teilnehmer über Herausforderungen bei der Antwortentwicklung untereinander und mit der Testentwicklerin aus. Abschließend wurden die Arbeitsergebnisse im Plenum diskutiert, weitere Ideen zur Antwortentwicklung aufgeworfen und ein Ausblick auf die folgenden Testentwicklungsschritte gegeben. Zum Ende des Workshops lagen 65 Situationen mit 1-9 Antwortoptionen vor. Diese wurden nachbereitet, indem Rechtschreib- und Kommafehler korrigiert und Fremdwörter ersetzt wurden. Die Antwortoptionen wurden so umformuliert, dass sie mit der ersten Person Singular beginnen (z.B. „*Ich* sage, dass...“) und beide Geschlechter abbildeten (z.B. „Sie arbeiten als *Praktikant(in)*“). Antworten, die reine Gedanken wiedergaben (z.B. „Ist mir egal“), wurden zu Handlungen umformuliert und die Gedanken als Begründung angehängt (z.B. „Ich mache nichts, da es für mich nicht wichtig ist“). Um möglichst sozial akzeptable Antworten zu erhalten, wurden negativ durch positiv geprägte Wörter ausgetauscht oder nachvollziehbare Begründungen für das Handeln hinzugefügt. Unpräzise ausformulierte Situationen wurden ergänzt. Situationen mit nur ein bis zwei Antwortoptionen wurden aussortiert. Die Antworten der verbliebenen Items wurden in eine zufällige Reihenfolge angeordnet. Nach der Bearbeitung lagen 61 Situationen mit vier bis acht Antwortoptionen vor.

6.2.3. Limitationen

Im folgenden Abschnitt soll der Workshop kritisch reflektiert werden, um zukünftigen Entwicklern Vor- und Nachteile der gewählten Vorgehensweise aufzuzeigen. Positiv zu werten ist, dass Personen nach Empfehlungen von Polyhart und MacKenzie (2011) und Weekley et al. (2006) rekrutiert wurden, die ihre Erfahrungen von der studentischen Ausbildung in der Medizin einbringen konnten. Um möglichst vielfältige Perspektiven bei der Testentwicklung einzubeziehen, wurden Personen aus verschiedenen Berufsgruppen, medizinischen Fachbereichen und Positionen berücksichtigt. Die Studierendenperspektive konnte nicht berücksichtigt werden, da trotz intensiver Werbung keine Studierenden zur Teilnahme gewonnen werden konnten. Bezüglich des Vorgehens innerhalb der Workshops erwies sich das freie individuelle Brainstorming von Situationen als vorteilhaft, da zahlreiche und vielfältige Situationen gesammelt wurden. Um die Intention des Ideengebers zu erhalten und den Nachbereitungsaufwand zu reduzieren, sollten die Situationen von den Teilnehmern differenzierter formuliert werden. Dafür bietet es sich an, dieselbe Person parallel an den Situationen und Antwortoptionen arbeiten zu lassen, um sie wechselseitig aneinander anzupassen. Im Workshop wurden die Situationen und Antworten in getrennten Arbeitsschritten von verschiedenen Personen entwickelt. Dies führte teilweise zu Verunsicherungen bei unklar beschriebenen Situationen. Für eine einheitlichere Situationsbeschreibung bietet sich eine Vorlage an, um den Nachbereitungsaufwand zu reduzieren. In dieser könnte festgelegt werden wie die Anrede und Genderfairness gestaltet werden soll (z.B. „*Sie sind Studierende(r)* im vorklinischen Studienabschnitt...“), und ob die Situation für das Human- und/oder Zahnmedizinstudium zutreffend ist. Nicht weiterzuempfehlen ist es, dieselbe Situation mehreren Konstrukten zuzuordnen, da die Situationsbeschreibung letztendlich nur einmal in den *SJT* eingeht. Für die Entwicklung der Antwortoptionen bietet sich ebenfalls eine Vorlage an, die vorgibt, in welcher Person die Antworten formuliert werden sollen (z.B. erste vs. dritte Person Singular). Auch sollte auf bessere und weniger geeignete Antworten hingewiesen werden. So kam es häufiger vor, dass Zwischenlösungen (z.B. „Ich spreche erst einmal mit einem guten Freund darüber.“) generiert wurden, die wenig geeignet sind, um zwischen Personen mit hoher und niedriger Merkmalsausprägung zu differenzieren. Um den Konstruktbezug der Antworten zu stärken, wäre es hilfreich, die Teilnehmer bei der Entwicklung begründen zu lassen, warum die Antworten ein niedriges bis hohes Ausmaß eines Konstrukts widerspiegeln. Im nächsten Schritt könnten Itemstamm und Antwortoptionen durch andere Personen ergänzt und überarbeitet werden. Zusammenfassend empfiehlt es sich von Anfang an

stärker an der Qualität der Items zu arbeiten, um anschließende Überarbeitungsschleifen und Itemausschlüsse reduzieren zu können.

6.3. Vortestung der Items

Die 61 schriftlichen Items (Anhang D) wurden an Gymnasiasten getestet, um deren Verständlichkeit und Unabhängigkeit vom medizinischen Vorwissen zu untersuchen. Die Instruktionsform wurde auf Praktikabilität geprüft und es wurden alternative Handlungsmöglichkeiten erfragt, um typische Reaktionen der Zielgruppe aufzunehmen.

6.3.1. Stichprobenbeschreibung

Die Testung fand in zwei elften Klassen eines beruflichen Gymnasiums in Heidelberg im Mai 2012 statt. An der Testung nahmen insgesamt 32 Gymnasiasten (25 weibliche, 7 männliche) im Alter von 16 bis 24 Jahren ($M_{\text{Alter}} = 17.7$) teil. Unter diesen waren sechs Personen, die an einem Human- oder Zahnmedizinstudium interessiert waren, sechzehn Personen, die nicht interessiert waren und zehn Personen, die ihr Interesse nicht einzuschätzen wussten.

6.3.2. Methodik

Die Testung wurde jeweils in zwei Schulstunden eingebunden, die das Messen von sozialen Kompetenzen thematisierte. Den Testteilnehmern wurden das Ziel des Tests und die Instruktion zum Antworten mündlich erläutert. Hierbei wurden die Gymnasiasten aufgefordert, pro Verhaltensalternative anzugeben mit welcher Wahrscheinlichkeit sie wie folgt reagieren würden. Pro Item mussten insgesamt 100% vergeben werden. Abbildung 19 gibt ein Instruktionsbeispiel. Die Instruktionsform wurde entwickelt, da bisherige kontinuierliche, verhaltensorientierte Instruktionen („*Rate how likely you would do*“) Personen im vorliegenden *SJT* benachteiligen würden, da er mehrere Verhaltensmöglichkeiten mit gleicher Adäquatheit zulässt. Auf das Problem wird in Kapitel 6.4.6 genauer eingegangen. Bisher enthielten *SJT*-Items nur Antworten, bei denen jede Adäquatheit nur einmal vergeben werden durfte. Zudem werden die Items mit mittlerer Wertigkeit nicht gewertet (vgl. Pereira & Harvey, 1999). Die Instruktion wurde darüber hinaus gewählt, weil Instruktionen, die den Teilnehmer auffordern, jede Antwortoption in den Antwortprozess einzubeziehen, eine höhere interne Konsistenz und eine höhere Retest-Reliabilität aufweisen als Instruktionen im Forced choice Format (siehe Kapitel 4.4.2).

Sie möchten Humanmedizin studieren und müssen dafür ein Pflegepraktikum in einem Krankenhaus absolvieren. Sie sind auf einer unfallchirurgischen Station mit einem anderen Studierenden eingeteilt, der/die schon seit einem Monat auf der Station im Praktikum ist. Sie helfen ihm/ihr beim Aufsagen eines Gipsverbandes, wobei diese(r) den Patienten mit der Säge leicht an der Haut verletzt. Nachdem der Patient gegangen ist, beschuldigt er Sie, das Bein falsch gehalten zu haben, obwohl Sie es genau nach seiner Anweisung ausgeführt haben. Wie wahrscheinlich würden Sie wie folgt auf die Kritik reagieren?

Die Prozentangaben müssen insgesamt 100% ergeben.

- | | |
|---|------|
| a) Ich frage ihn, was ich das nächste Mal besser machen kann. | ___% |
| b) Ich denke mir meinen Teil und schweige. | ___% |
| c) Ich beschwere mich beim Stationsarzt über den Vorfall. | ___% |
| d) Ich frage einen Arzt, ob der andere Studierende das Gipsabnehmen beherrscht. | ___% |
| e) Ich sage ihm, dass ich genau nach seinen Anweisungen gehandelt habe. | ___% |

Abbildung 19. Instruktion des SJTs in der Vortestung.

Jeder Schüler bearbeitete fünf Items einer sozialen Kompetenz (Beispielbogen im Anhang E-1). Dabei wurden sie gebeten, sich ehrlich einzuschätzen. Die Schüler erhielten vorher die Definition der entsprechenden sozialen Kompetenz. Nach der Itembeantwortung wurden sie befragt wie sie in dieser Situation vielleicht noch handeln würden, um weitere Antwortmöglichkeiten zu gewinnen. Anschließend bewerteten die Schüler die Verständlichkeit der Situationsbeschreibung auf einer fünf-stufigen Likert-Skala (1= sehr leicht verständlich bis 5= sehr schwer verständlich) und wurden gebeten anzugeben, was ggf. nicht verständlich war. Auch beurteilten sie mit den Kategorien „ja“, „nein“ und „weiß nicht“, ob sie zum Verständnis der Situationen medizinisches Vorwissen benötigten. Zusätzlich konnten freie Kommentare gegeben werden. Schnell arbeitenden Personen wurde ein weiterer Testbogen zur Bearbeitung angeboten. Die Testung wurde mit einer mündlichen Diskussion abgerundet. Diese thematisierte, was den Gymnasiasten bei der Itembearbeitung generell aufgefallen war, wie sie mit der Instruktionsform zu Recht kamen und wie sie die genderfaire Formulierung fanden.

Die statistische Auswertung erfolgte auf Itemebene, indem für jedes Item die Wahrscheinlichkeit, der arithmetische Mittelwert und die Standardabweichung pro Antwortoption ermittelt wurden. Items, die nicht mit einer Gesamtwahrscheinlichkeit von 100% beurteilt wurden, wurden ausgeschlossen. Alternative Handlungsmöglichkeiten wurden als qualitative Daten für die Optimierung und Neugewinnung von Antwortoptionen verwendet. Für die Unabhängigkeit vom medizinischen Vorwissen wurden die Häufigkei-

ten pro Kategorie bestimmt. Für die Verständlichkeit der Items wurde der Mittelwert und der Range errechnet und freie Anmerkungen zur Verständlichkeit aufgenommen. Wurde von einer oder mehreren Person(en) beim medizinischen Vorwissen die Kategorie „ja“ angekreuzt oder die Verständlichkeit mit ≥ 3 gewertet, wurde das Item noch einmal genauer hinsichtlich inhaltlicher und sprachlicher Vereinfachung begutachtet. Mündliche Diskussionsbeiträge wurden notiert.

6.3.3. Ergebnisse

Fünf Items wurden von einem Schüler, sieben Items von zwei Schülern, 36 von drei Schülern und 13 Items von vier Schülern bearbeitet. Bei allen Items wurde die Möglichkeit genutzt, die Wahrscheinlichkeiten für das eigene Verhalten auf die dargebotenen Antwortalternativen zu verteilen. Bei sieben Bearbeitungen wurden die gesamten 100% an nur eine Antwortoption vergeben. Dies entspricht 4% der gesamt bearbeiteten Items. Dies betraf Items mit vier bis fünf Antwortoptionen. Bei den insgesamt 192 bearbeiteten Items musste acht Mal ein Item für die Mittelwerts- und Standardabweichungsberechnung ausgeschlossen werden (4% fehlerhafte Antworten), da nicht die Gesamtwahrscheinlichkeit von 100% vergeben wurde. Dies betraf Items mit fünf bis sieben Antwortoptionen. Die mittleren Wahrscheinlichkeitsangaben pro Antwortoption aller Items können im Anhang E-2 eingesehen werden. Bei 21 Items wurde eine Antwortoption, bei neun Items wurden zwei Antwortoptionen und bei zwei Items wurden drei Antwortoptionen von keinem Schüler ausgewählt. Dabei wurden nur die Items berücksichtigt, die von mindestens zwei Schülern bearbeitet wurden. In der mündlichen Diskussion begrüßten die Gymnasiasten, mit Ausnahme von einer Person, die Möglichkeit in Wahrscheinlichkeiten zu antworten. Das Bearbeiten der Items sei zügig gegangen und das Rechnen bis 100 sei kein Problem gewesen. Die Schüler machten insgesamt 48 alternative Handlungsvorschläge und drei Anmerkungen zur Verständlichkeit. Die mittlere Verständlichkeit der Items lag zwischen 1.0 und 4.0 (Tabelle 19). Bei 47 Items lagen die Minima und Maxima im sehr verständlichen bis verständlichen Bereich (Werte von 1-2). Vierzehn Items wurden von einer oder mehreren Personen mit einem Wert ≥ 3 bewertet. Die Gymnasiasten beurteilten, dass 8 von 61 Items nur mit medizinischem Vorwissen zu verstehen waren.

Tabelle 19: *Verständlichkeit der SJT-Items*

Item	N	Verständlichkeit			Vorwissen		
		M	Min	Max	ja	nein	weiß nicht
SBSR1	3	1.67	1	2	0	3	0
SBSR2	3	1.67	1	2	0	2	1
SBSR3	3	2.33	1	4	1	2	0
SBSR4	3	1.67	1	2	0	3	0
SBSR5	3	2.67	1	4	0	3	0
SBSR6	2	1.00	1	1	0	2	0
SBSR7	2	1.00	1	1	1	1	0
SKSD1	3	1.00	1	1	0	3	0
SKSD2	3	1.33	1	2	0	3	0
SKSD3	3	1.33	1	2	0	3	0
SKSD4	3	1.33	1	2	0	2	1
SKSD5	3	2.00	1	3	0	2	1
SKSD6	4	1.25	1	2	0	3	0
SKSD7	4	1.25	1	2	0	4	0
SKSD8	3	1.67	1	2	0	3	0
SKSD9	3	1.00	1	1	0	3	0
SKSD10	3	1.67	1	3	1	2	0
SKSD11	3	1.00	1	1	2	1	0
Kri1	4	1.25	1	2	1	3	0
Kri2	4	1.00	1	1	0	4	0
Kri3	4	1.00	1	1	0	4	0
Kri4	4	1.00	1	1	0	4	0
Kri5	4	1.00	1	1	0	3	0
Kri6	4	1.25	1	2	0	4	0
Kri7	4	1.25	1	2	0	4	0
Kri8	4	1.00	1	1	1	3	0
Kri9	4	1.25	1	2	0	4	0
Kri10	4	1.00	1	1	0	4	0
Kri11	3	1.33	1	2	0	3	0
Kri12	3	1.33	1	2	1	2	0
R1	4	1.25	1	2	0	4	0
R2	4	1.75	1	3	0	4	0
R3	4	1.00	1	1	0	4	0
R4	4	1.25	1	2	0	4	0
R5	4	1.75	1	3	0	4	0
R6	4	1.25	1	2	0	4	0
R7	4	2.00	2	2	0	3	1

Fortsetzung Tabelle 19:

Item	N	Verständlichkeit			Vorwissen		
		M	Min	Max	ja	nein	weiß nicht
R8	4	2.00	1	3	0	4	0
SU1	3	2.00	1	3	0	3	0
SU2	3	1.67	1	2	0	3	0
SU3	3	2.00	1	3	0	3	0
SU4	3	1.00	1	1	1	2	0
SU5	3	1.33	1	2	0	3	0
SU6	2	2.00	2	2	0	2	0
SU7	2	1.00	1	1	0	2	0
SU8	1	3.00	3	3	0	0	1
SU9	1	4.00	4	4	0	1	0
SU10	1	2.00	2	2	0	1	0
SU11	2	1.00	1	1	0	2	0
KF1	3	1.00	1	1	0	3	0
KF2	3	1.33	1	2	0	3	0
KF3	3	2.00	1	3	0	3	0
KF4	3	1.33	1	2	0	3	0
KF5	3	1.67	1	3	0	3	0
KF6	2	2.50	1	4	0	2	0
KF7	2	1.00	1	1	0	2	0
KF8	2	1.50	1	2	0	1	0
KF9	2	1.50	1	2	0	2	0
KF10	2	1.50	1	2	0	2	0
KF11	3	1.67	1	2	0	3	0
KF12	3	1.67	1	2	0	3	0

Anmerkung. 1 = sehr leicht verständlich bis 5 = sehr schwer verständlich; SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Fett markierte Werte heben die Items hervor, bei denen mindestens eine Person „ja“ beim Vorwissen angekreuzt hat oder/und mindestens eine Person die Verständlichkeit ≥ 3 bewertet hat.

In der mündlichen Diskussion machten die Gymnasiasten auf Schwächen des *SJTs* aufmerksam. So wurde neben der sozialen Erwünschtheit diskutiert, ob ein freies Antwortformat Vorteile zum MC-Format bieten würde und ob Interviews besser zur Auswahl geeignet wären. Bezüglich der Genderfairness würden sich die Schüler unabhängig von der männlichen oder weiblichen Personenform angesprochen fühlen. Teilweise

würde die gleichzeitige Nennung der weiblichen und männlichen Person den Lesefluss stören. Sie unterbreiteten den Vorschlag abwechselnd männliche und weibliche Personen in den Situationsbeschreibungen agieren zu lassen.

6.3.4. Ergebnisdiskussion und Itemüberarbeitung

Die Ergebnisse haben eine eingeschränkte Aussagekraft, da nicht jeder Schüler alle Items bearbeitete. Dies lag zum einen daran, dass die Unterrichtseinheit auf 45 Minuten begrenzt war, zum anderen war es schwer möglich die Motivation aufrechtzuerhalten, 61 Items zu begutachten. Das Ziel der Untersuchung richtete sich jedoch von vornherein mehr auf qualitative Aspekte zur Itemoptimierung. Die Untersuchung wurde an Gymnasiasten in der Oberstufe durchgeführt, welche die Zielpersonen des *SJT*s darstellen. Es handelte sich jedoch um eine Gelegenheitsstichprobe an einem beruflichen Gymnasium. Der Lehrer schätzte die Klasse als eher schwach in der schulischen Leistung ein. Somit könnten sich im Vergleich zur Grundgesamtheit der Gymnasiasten und insbesondere zur Gruppe potentieller Bewerber um einen Medizinstudienplatz eher leistungsschwächere Personen in der Stichprobe befunden haben. Zur Untersuchung der Verständlichkeit und der Unabhängigkeit von medizinischem Vorwissen kann dies jedoch eher positiv bewertet werden. Um hinsichtlich der Lesekomplexität testfaire Items zu erhalten (Sacco et al., 2000a, 2000b), sollten möglichst verständliche *SJT*-Items generiert werden. Aufgrund dessen wurde auch ein strenger Maßstab für die Begutachtung der Verständlichkeit angesetzt. Items, die von einer oder mehreren Personen mit einem Wert ≥ 3 bewertet wurden, wurden sprachlich und inhaltlich vereinfacht. Die Mehrheit der Items wurde jedoch als sehr leicht bis leicht verständlich beurteilt. Die wenigen qualitativen Anmerkungen zur Verständlichkeit wurden eingearbeitet. Aus der mündlichen Diskussion mit den Schülern wurde der Vorschlag zur Genderfairness aufgenommen, abwechselnd männliche und weibliche Personen agieren zu lassen, um den Lesefluss zu vereinfachen. Die Mehrzahl der Items konnte ohne medizinisches Vorwissen verstanden werden. Die Items, die medizinisches Wissen voraussetzten, bildeten meist Student-Patienten-Interaktionen ab. Die Frage war anscheinend missverständlich formuliert, da wiederholt gefragt wurde, wie sie zu verstehen sei. So beurteilten einige Schüler, ob der Studierende in der präsentierten Situation medizinisches Vorwissen benötigt, um die Situation zu meistern. Dadurch wurden vermutlich mehr Situationen als abhängig vom medizinischen Vorwissen eingeschätzt als tatsächlich vorlagen. Dies hatte jedoch keine negative Konsequenz, da die Items nicht ausgeschlossen, sondern inhaltlich und sprachlich begutachtet und adaptiert wurden.

Die Schüler schienen insgesamt gut mit der Instruktion zu Recht zu kommen. Die Rechenfehler fielen gering aus, sollen aber in zukünftigen Untersuchungen durch automatische Berechnungen ganz vermieden werden. Die Rechenfehler betrafen dabei überwiegend Items mit mehr als vier Antworten. Das Angebot in Wahrscheinlichkeiten pro Antwortoption zu antworten, scheint attraktiv gewesen zu sein. Bei nur 4% der bearbeiteten Items wurde die Gesamtwahrscheinlichkeit an nur eine Antwort vergeben. Dies betraf Items mit vier bis fünf Antworten. Hier könnte der Ankereffekt eine Rolle gespielt haben, durch den die Probanden durch für die Entscheidung irrelevante Kontextbedingungen beeinflusst worden sein könnten (Tversky & Kahneman, 1974). In diesem Fall stellte die Kontextbedingung die Anzahl an dargebotenen Antwortoptionen pro Item dar. So ist es gut möglich, dass Probanden eher weniger Antworten auswählen, wenn weniger Antwortoptionen dargeboten werden, als wenn mehr Antwortoptionen zur Verfügung stehen. Für die angestrebte Anzahl an Antwortoptionen wird sich an der typischen Antwortanzahl von vier bis sechs orientiert (Polyhart & MacKenzie, 2011, Weekley et al., 2006), damit die Antwortvielfalt erhalten, aber gleichzeitig die Beantwortung praktikabel bleibt.

Die Antwortoptionen, die von den Schülern nie ausgewählt wurden, spiegelten Verhaltensweisen wieder, die weniger sozial kompetent waren. Diese wurden versucht, sozial akzeptabler zu formulieren. Falls das Item mehrere Antworten enthielt, die nicht ausgewählt wurden, wurden diese Antworten mit einer anderen zusammengefasst oder gestrichen. Um die Antwortqualität und -vielfalt zu steigern, wurden die Gymnasiasten nach alternativen Handlungsmöglichkeiten gefragt. Die 48 vorgeschlagenen Handlungsalternativen wurden als zusätzliche Antworten aufgenommen oder in bestehende Antworten integriert, wenn sie neue Verhaltensweisen wiedergaben und realistisch waren. Während der Itemüberarbeitung wurden zusätzlich die Eignung der Items zur Erhebung der sozialen Kompetenz und die Konstruktausprägung der Antwortoptionen mit einem Psychologen aus dem Team der Testentwicklerin diskutiert. Dies führte dazu, dass Items einer anderen sozialen Kompetenz zugeordnet (N=1) oder ausgeschlossen wurden (N=8). Aus der Erkenntnis, dass ein Item trotz unterschiedlicher Auslegung nicht zweimal in den *SJT* eingehen kann, wurden Items, die doppelt sozialen Kompetenzen zugeordnet waren, der geeigneteren Kompetenz zugeordnet (N=4). Des Weiteren wurde die Formulierung einiger Items abstrahiert, um für möglichst alle Zielpersonen zutreffend zu sein. Zum Beispiel wurde der anstehende „attraktive Mountainbike-Urlaub“ zu einem „attraktiven Urlaub“ abstrahiert. Sinngemäß gleiche Antworten wurden zusammengefasst (z.B. eine Person kontaktieren statt mailen oder anrufen). Zudem wurden Antworten ausgeschlossen, die nur einen Zwischenschritt (z.B. „Ich rede mit einem guten Freund darüber.“), aber keine Lösung der Situation beinhal-

teten. Bei der Diskussion wurde durch den Einbezug eines weiteren Psychologen versucht, die Subjektivität einzuschränken, jedoch könnten durch eine andere Personenkonstellation unterschiedliche Entscheidungen getroffen worden sein.

Da während der Nachbereitung auffiel, dass die Antwortoptionen teilweise keine idealen Lösungen der Situation beinhalteten, wurde ein weiterer Bearbeitungsschritt eingeschoben. Einer bis dahin unbeteiligten Psychologin und einer Humanmedizinerin wurden dafür die Situationsbeschreibungen ohne die Antwortmöglichkeiten vorgelegt und gebeten, eine Ideal-Lösung pro Item niederzuschreiben. Die gleiche Humanmedizinerin begutachtete anschließend noch die vorhandenen Antwortoptionen. Offene Diskussionspunkte bei zahnmedizinischen Items wurden zusätzlich mit zwei Zahnmedizinern besprochen, offene Punkte bei Items bezüglich der Pflege von Patienten mit einem Krankenpfleger. Die Reihenfolge der Antwortoptionen wurde pro Item zufällig festgelegt. Nach der Bearbeitung lagen 48 Situationen vor (Item Pool Anhang D).

6.4. Bildung des Auswertungsschlüssels

Die Funktion eines Auswertungsschlüssels liegt darin, den Antwortoptionen und Items Punkte zuweisen zu können, um letztendlich eine Gesamtpunktzahl für den *SJT* zu erhalten. Aufgrund der Forschungsbefunde des Kapitels 4.3 wird für die Entwicklung des Auswertungsschlüssels die rationale, expertenbasierte Methode unter Verwendung eines kontinuierlichen *Ratings* verwendet. Um eventuelle Verzerrungen durch die ausgewählte Expertengruppe einzugrenzen, wurden die Expertenurteile anschließend von der Testkonstrukteurin und zwei Mitarbeitern diskutiert. Die Untersuchung erfolgte am schriftlichen *SJT*. Im Zuge der Untersuchung wurden die Experten zudem aufgefordert, die Realitätsnähe der Situationen zu beurteilen, um unrealistische Items aus dem *SJT* zu entfernen.

6.4.1. Stichprobenrekrutierung und -beschreibung

Als Experten nahmen approbierte Human- und Zahnmediziner teil. Diese Gruppe wurde ausgewählt, da sie das Medizinstudium bereits erfolgreich abgeschlossen haben und vermutlich ähnliche soziale Situationen wie im vorliegenden *SJT* aus ihrem Studium kennen und beurteilen können. Die Werbung der Experten erfolgte während der hochschuldidaktischen Schulung der Medizinischen Fakultät Heidelberg, an der Mediziner aller Fachrichtungen teilnehmen. Um einen reinen Heidelberger Einfluss auszuschließen, wurden außerdem Personen unterschiedlicher Einrichtungen berücksichtigt. Dafür schrieb die Testentwicklerin 64 Human- und Zahnmediziner aus Deutschland an,

die in der Anforderungsanalyse ihre Emailadresse für Rückmeldungen und Folgestudien hinterlassen hatten. Zusätzlich bat sie Mediziner während der Tagung der Gesellschaft für Medizinische Ausbildung (GMA) im September 2012 um ihre Teilnahme. Des Weiteren kontaktierte sie Psychologen in fortgeschrittener therapeutischer Ausbildung, da ihr Beruf aufgrund des Patientenkontaktes Ähnlichkeit mit dem Arztberuf aufweist und sie darin geschult sind, menschliches Verhalten zu beobachten und zu beurteilen. Dafür versendete sie einen Umfrage-Link über den Emailverteiler dreier Jahrgänge von drei psychotherapeutischen Ausbildungsinstituten (Heidelberg, Mannheim, Dresden) mit unterschiedlicher Therapieausrichtung. An der Umfrage beteiligten sich 65 Personen, wobei ein Teilnehmer ausgeschlossen wurde, der lediglich zwei Antwortoptionen eines Items beurteilt hatte. Die Stichprobe setzte sich aus 22 Humanmedizinern, 15 Zahnmedizinern und 29 Psychologen zusammen (Tabelle 20). In der Gruppe der Human- und Zahnmediziner bestand ein ausgeglichenes Geschlechterverhältnis, während in der Gruppe der Psychologen die weiblichen Teilnehmer überwogen. In allen Gruppen zeigte sich eine starke Streuung hinsichtlich der mittleren Berufs- und Lehrerfahrung. Die Berufserfahrung der Mediziner war höher als die der Psychologen. Die Berufserfahrung von weiblichen Human- oder Zahnmedizinern war geringer als die ihrer männlichen Kollegen. Die Human- und Zahnmediziner verfügten über mehr Berufs- als Lehrerfahrung im Studentenunterricht. Die Lehrerfahrung war bei den Humanmedizinern höher als bei den Zahnmedizinern.

Tabelle 20: *Teilnehmerdemografie Auswertungsschlüssel*

Berufsgruppe		Geschlecht		Berufserfahrung			Lehrerfahrung		
		N	%	M	SD	Spanne	M	SD	Spanne
Human- mediziner	w	11	50.0	4.50	3.11	0.5-12	3.00	3.29	0-11
	m	11	50.0	11.09	9.34	1-32	5.55	5.29	0-18
	ges	22	100.0	7.80	7.59	0.5-32	4.21	4.44	0-18
Zahnmediziner	w	7	46.7	6.00	4.24	1-11	1.86	3.63	0-10
	m	8	53.3	14.25	7.56	2-24	2.43	2.51	0-5
	ges	15	100.0	10.40	7.38	1-24	2.14	3.01	0-10
Psychologen	w	23	82.1	4.08	7.03	0-32	-	-	-
	m	4	14.3	3.75	2.63	0-6	-	-	-
	ges	28	96.4	4.42	6.63	0-32	-	-	-

Anmerkung. w = weiblich, m = männlich, ges = gesamt.

6.4.2. Untersuchungsablauf

Die Befragung der Expertengruppe erfolgte online. Items, die sowohl im Human- als auch im Zahnmedizinstudium vorkommen (N = 33), wurden beiden medizinischen Gruppen zur Beurteilung vorgelegt. Speziell humanmedizinische Items (N = 11) wurden nur den Humanmedizinerinnen und speziell zahnmedizinische (N = 4) nur den Zahnmedizinerinnen zugewiesen. Alle eingegangenen Items enthält Anhang D. Die Medizinerinnen wurden zunächst gebeten, die Realitätsnähe der Situationen auf einer fünfstufigen Likert-Skala (1 = vollkommen realistisch bis 5 = überhaupt nicht realistisch) zu bewerten. Anschließend beurteilten sie die Angemessenheit jeder Handlungsoption hinsichtlich Ausprägung der jeweiligen sozialen Kompetenz. Dafür stand ihnen dieselbe fünfstufige Ratingskala (1 = sehr angemessen bis 5 = sehr unangemessen) wie den Testkonstruktoren und die Zusatzkategorie „kein Zusammenhang mit der sozialen Kompetenz“ zur Verfügung. Zudem konnten die Situationen und Antwortoptionen frei kommentiert werden. Einen Beispielbogen enthält Anhang F-1. Die Psychologinnen beurteilten nur die Angemessenheit der Reaktionen. Aufgrund der hohen Item-Anzahl (N = 48) wurde den Teilnehmerinnen freigestellt wie viele Items sie bearbeiten wollten. Die Items erschienen in randomisierter Reihenfolge, sodass für jedes Item die gleiche Chance bestand beurteilt zu werden. Auf der GMA-Tagung wurden die Items jeweils auf einem DIN A4 Blatt zum Bearbeiten dargeboten. Die Medizinerinnen beantworteten zudem demografische Fragen nach Geschlecht, sowie Berufs- und Lehrerfahrung im Studentenunterricht. Psychologinnen wurden nach Geschlecht und der Berufserfahrung in der Arbeit mit Patientinnen inklusive Praktika und Psychotherapiestunden gefragt.

6.4.3. Statistische Auswertung

Die mittlere Realitätsnähe der Items wurde über den arithmetischen Mittelwert bestimmt. Items mit einer beurteilten Realitätsnähe ≥ 3.5 (Skala von 1= vollkommen realistisch bis 5= überhaupt nicht realistisch) wurden ausgeschlossen. Für die Antwortoptionen wurden die Häufigkeitsverteilung, sowie die gemittelten Expertenurteile (Mediane) für die soziale Angemessenheit erhoben. Die Beurteilerübereinstimmung der Experten wurde über den *Intraclass Coeffizienten (ICC)* berechnet, da für Ratingskalen, die nicht grob gegen die Äquidistanz der Skalenpunkte verstoßen, ein metrisches Übereinstimmungsmaß verwendet werden kann (Wirtz & Caspar, 2002). Für den ICC wurde der absolute (unjustierte) Wert zur Beurteilung herangezogen. Dieser gibt an, inwieweit ein Wert eines Beurteilers exakt mit dem Mittelwert der anderen Beurteiler übereinstimmt und gilt als präzise Methode der Reliabilitätsprüfung (Wirtz & Caspar,

2002). Als Voraussetzungsprüfung wurden Trennschärfenanalysen und der Tukey Test auf Additivität durchgeführt. Letzterer prüft, ob eine Interaktion zwischen den beurteilten Objekten und den Beurteilern ausgeschlossen werden kann. Bei Verletzung der Nicht-Additivität darf dennoch der *ICC* berechnet werden, wenn eine homogene korrigierte Trennschärfe vorliegt, da in diesem Falle die Reliabilität nicht über-, sondern unterschätzt wird (Asendorpf & Wallbott, 1979; Wirtz & Caspar, 2002). Im Falle von inhomogenen korrigierten Trennschärfen empfehlen selbige Autoren, die Beurteiler auszuschließen, deren Trennschärfe negativ mit den anderen Urteilern korreliert oder gering ausfällt. Aufgrund dieser Empfehlung wurden in der vorliegenden Untersuchung Urteiler mit negativen Trennschärfen oder Trennschärfen $< .30$ ausgeschlossen. *ICCs* $\geq .70$ gelten im Allgemeinen als zufriedenstellend (vgl. z.B. Lievens & Sackett, 2006; Weekley & Jones, 1997).

6.4.4. Ergebnisse

Die Bewertung der Realitätsnähe und der Beurteilerübereinstimmung stellt Tabelle 21 dar. Von 48 Situationen wurden 46 als realistisch für das Medizinstudium eingeschätzt, wobei die Einzelurteile bei neun Items über die gesamte Skala streuten (SBSR6, SKSD7, Kri 4, Kri7, R2, SU3, SU12, KF6, KF9). Acht Items wurden nur von ein bis zwei Experten bezüglich der Realität eingeschätzt (SBSR5, SBSR7, SKSD4, SKSD5, SKSD11, Kri1, SU1, SU12). Item SKSD7 überschritt den Cut-Off-Wert von 3.5 und wurde somit ausgeschlossen. Ein Grenzfall stellte das Item SBSR5 dar, welches genau den Cutt-off-Wert erreichte. Da es jedoch lediglich von zwei Personen bewertet wurde, wurde es im Item Pool belassen. Die Beurteilerübereinstimmung zwischen den Experten erreichte Werte zwischen .28 und .77. Eine zufriedenstellende Beurteilerübereinstimmung $\geq .70$ zeigten neun Items. Die Voraussetzung der Nicht-Additivität war für fünf Items nicht erfüllt (SBSR5, SKSD6, SKSD10, Kri2, R6). Für diese konnte nach Asendorpf und Wallbott (1979), sowie Wirtz und Caspar (2002) dennoch der *ICC* berechnet werden, indem die Urteiler mit inhomogenen Trennschärfen ausgeschlossen wurden.

Tabelle 21: *Realitätsnähe und Beurteilerübereinstimmung*

Item	Realitätsnähe				Beurteilerübereinstimmung		
	N	M	SD	Spanne	N	N ^a	ICC
SBSR1	11	1.09	0.30	1-2	13	5	.40*
SBSR3	9	1.89	0.93	1-3	13	3	.60**
SBSR4	7	2.86	0.69	2-4	17	4	.41**
SBSR5	2	3.50	0.71	3-4	10	0	.72**
SBSR6	8	2.13	1.46	1-5	11	3	.04
SBSR7	1	1.00	0.00	1-1	9	0	.72**
SKSD1	11	1.55	0.82	1-3	16	2	.52**
SKSD4	2	1.00	0.00	1-1	9	0	.51**
SKSD5	2	2.50	2.12	1-4	8	0	.65**
SKSD7	14	3.57	1.22	1-5	21	0	.76**
SKSD8	3	1.67	0.58	1-2	13	2	.33**
SKSD9	7	1.57	0.79	1-3	11	1	.64**
SKSD10	4	1.75	0.96	1-3	12	0	.71**
SKSD11	2	1.50	0.71	1-2	12	1	.65**
Kri1	2	2.50	0.71	2-3	10	1	.62**
Kri2	8	1.75	0.89	1-3	10	1	.62**
Kri3	7	2.29	0.49	2-3	11	3	.28**
Kri4	9	1.78	1.30	1-5	13	3	.63**
Kri6	10	1.80	1.14	1-4	13	3	.29**
Kri7	11	2.09	1.30	1-5	17	2	.58**
Kri9	5	1.80	0.84	1-3	11	4	.34**
Kri10	9	1.33	0.71	1-3	13	2	.69**
Kri11	3	1.33	0.58	1-2	9	0	.77**
Kri12	6	1.33	0.52	1-2	12	2	.55**
R2	12	2.08	1.44	1-5	16	1	.52**
R3	9	2.33	1.00	1-4	16	4	.56**
R6	7	3.00	0.58	2-4	13	0	.54**
R8	10	2.20	0.92	1-4	16	4	.58**
SU1	2	2.50	2.12	1-4	10	0	.62**
SU2	11	1.18	0.40	1-2	15	0	.77**
SU3	8	2.88	1.13	1-5	11	1	.64**
SU4	4	1.50	0.58	1-2	10	0	.76**
SU5	11	1.73	1.10	1-4	15	0	.56**
SU6	12	1.83	0.94	1-4	16	6	.46**
SU8	8	1.00	0.00	1-1	10	0	.63**
SU9	11	1.27	0.47	1-2	15	4	.53**

Fortsetzung Tabelle 21:

Item	Realitätsnähe				Beurteilerübereinstimmung		
	N	M	SD	Spanne	N	N ^a	ICC
SU10	6	2.00	0.63	1-3	12	0	.61**
SU12	2	3.00	2.83	1-5	11	1	.67**
KF1	13	1.38	0.65	1-3	17	0	.42**
KF2	13	1.54	0.66	1-3	18	2	.66**
KF3	9	1.78	0.97	1-4	14	2	.28**
KF4	10	1.30	0.48	1-2	13	0	.68**
KF5	6	1.83	0.75	1-3	10	1	.57**
KF6	7	2.29	1.45	1-5	12	2	.68**
KF7	9	1.78	0.83	1-3	9	0	.77**
KF8	10	2.00	0.94	1-3	8	2	.43**
KF9	11	2.55	1.13	1-5	11	2	.60**
KF11	10	1.70	0.95	1-3	16	0	.76**

Anmerkung. N^a = Anzahl ausgeschlossener Expertenurteile aufgrund inhomogener Trennschärfen. ICC = Intra-Class-Coefficient; *p ≤ .05, **p ≤ .01. SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/ Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

6.4.5. Diskussion der sozialen Angemessenheit der Antwortoptionen

Die Expertenurteile wurden als Basis für die Diskussion der Angemessenheit der Handlungsoptionen verwendet. Aufgrund der Kritik an den ICCs, dass Items mit hoher Augenscheinvalidität oder mit hoher sozialer Erwünschtheit übereinstimmender von Beurteilern bewertet werden und Items mit mehr Verhaltensalternativen schwieriger übereinstimmend zu bewerten sind (Krokos et al., 2004; Möller, 2010), wurde der *Cutoff*-Wert des ICCs auf .50 reduziert. Alle Items mit kleinerem Wert wurden ausgeschlossen (SBSR4, SBSR6, SKSD8, Kri6, Kri9, SU6, KF3). Alle Items mit größerem Wert wurden von der Testkonstrukteurin gemeinsam mit einem psychologischen Kollegen und einer wissenschaftlichen Hilfskraft mit sozialwissenschaftlichem Hintergrund hinsichtlich des Grades der sozialen Angemessenheit diskutiert. Die Gruppe wird im Folgenden Diskussionsteam genannt. Stimmtten die gemittelten Expertenurteile mit der Einschätzung der Angemessenheit des Diskussionsteams genau überein, wurde das Urteil für die Antworten beibehalten. Werteten mindestens zwei Personen des Diskussionsteams das Verhalten als angemessener bzw. unangemessener als die Experten, wurde das Urteil als Kompromisslösung um maximal einen Skalenpunkt verschoben.

Antworten, die keinen Konstruktbezug aufwiesen, wurden überarbeitet oder ausgeschlossen. Hier wurden auch die freien Anmerkungen der Experten berücksichtigt und bei Bedarf die Situationsbeschreibungen angepasst. Die zwei Kontaktfähigkeitsitems KF2 und KF4 wurden in diesem Schritt der Suche nach sozialer Unterstützung zugewiesen, da sie besser zu deren Definition passten (KF2 heißt nun SU13, KF4 heißt nun SU14). Das Kritikfähigkeitsitem Kri11 wurde aufgrund der polarisierenden Thematik der Ausländerfeindlichkeit ausgeschlossen. Aufgrund der mangelnden Passung zum Konstrukt wurden weiterhin R8, KF5 und KF9 ausgeschlossen. Item Kri3 wurde aus Imagegründen ausgeschlossen, da kein negatives Bild von ungepflegten Studierenden bei den Studienplatzbewerbern hervorgerufen werden soll.

Pro Item wurde anschließend geprüft, ob deren Handlungsoptionen in der sozialen Angemessenheit variierten, um ein Spektrum der sozialen Kompetenz abbilden und letztendlich zwischen Personen mit hoher und niedriger Merkmalsausprägung differenzieren zu können. Dafür wurde festgelegt, dass die Antworten eines Items mindestens drei Skalenpunkte der fünfstufigen Ratingskala abdecken sollten. Dementsprechend sollte z.B. eine Antwort ein sehr angemessenes Verhalten repräsentieren, eine andere eine mittlere Angemessenheit und eine weitere ein unangemessenes Sozialverhalten. Alle Items, deren Antwortoptionen weniger als drei Angemessenheitsgrade aufwiesen, wurden ausgeschlossen (SBSR1, SBSR3, SBSR4, Kri3, R6, SU3, KF1, KF7, KF8).

Abschließend wurden Items mit inhaltlicher Dopplung ausgeschlossen, um einen möglichst hohen Informationsgrad des *SJT*s in Abhängigkeit von der Itemanzahl zu erhalten (SKSD5, SKSD9, SU4). Im *SJT* verblieben insgesamt 24 Items.

6.4.6. Punktbildung

Der Angemessenheit der Verhaltensoptionen wurde jeweils einem Faktor zugewiesen. Wurde ein Verhalten als sehr sozial angemessen beurteilt, wurde es mit dem Faktor 5 multipliziert. Wurde ein Verhalten als sozial angemessen bewertet, erhielt es den Multiplikator 4. Wurde das Verhalten als teils/teils angemessen beurteilt, wurde es mit dem Faktor 3 multipliziert. Unangemessenes Verhalten wurde mit dem Faktor 2 und sehr unangemessenes mit dem Faktor 1 multipliziert. Dadurch wurde die Auswahl sozial angemessenen Verhaltens belohnt, während sozial unangemessenes Verhalten durch niedrige Multiplikatoren bestraft wurde. Jeder Faktor durfte mehrfach vergeben werden. Somit wurde gewährleistet, dass mehrere Verhaltensweisen mit der gleichen sozialen Angemessenheit pro Item zulässig waren. Um die gleiche Punktzahl pro Item trotz unterschiedlicher Antwortanzahl zu erhalten, musste mindestens einmal der Faktor 1 und

5 vergeben worden sein. Dafür mussten ggf. sehr unangemessene (Faktor 1) bzw. sehr angemessene (Faktor 5) Verhaltensoptionen ergänzt werden. Die Teilnehmerantwort pro Verhaltensoption wurde mit dem zugewiesenen Faktor multipliziert und für eine leichtere Handhabung durch 100 dividiert. Der Punktwert pro Item errechnet sich durch Addition der Punktwerte pro Antwort. Pro Item konnten ein bis fünf Punkte erreicht werden. Die Berechnung folgte der Formel (1) und wird beispielhaft am Item SBSR5 in Tabelle 22 dargestellt.

$$\text{Punkte pro Antwortoption} = \frac{\text{Faktor} \times \text{Teilnehmerantwort}}{100} \quad (1)$$

Wählt eine Person die sozial angemessenste Antwort erhält sie die maximale Punktzahl von 5 Punkten (Beispiel 1). Wählt sie die angemessenste Antwort z.B. mit 70% und die zweitangemessenste mit 30% so erhält sie 4.7 von 5 Punkten (Beispiel 2). Die Punktzahl sinkt, wenn die Person unangemessenere Verhaltensweisen auswählt bis zu einem Minimum von einem Punkt bei der Auswahl der unangemessensten Antwort zu 100% (Beispiel 3 und 5). Die Verrechnung benachteiligt keine Personen, die statt einer Antwort, zwei Antworten der gleichen Angemessenheit wählen (Beispiel 5 vs. 6). Verteilt die Person die Wahrscheinlichkeit ihres Handelns auf alle Antworten gleichmäßig, erhält sie eine mittlere Punktzahl (Beispiel 7). Sie erhält dadurch weniger Punkte als die Personen, die nur angemessenere Antworten wählen (Beispiel 1 vs. 2), aber mehr als die, die nur unangemessenere Antworten aussuchen (Beispiel 3, 5 vs. 6) oder die, die die angemessenste und unangemessenste Antwort wählen (Beispiel 4).

Zur Bildung eines alternativen Auswertungsschlüssels wurde simuliert, dass die Teilnehmer auf einer fünfstufigen Ratingskala (1 = sehr unwahrscheinlich bis 5 = sehr wahrscheinlich) angeben, wie wahrscheinlich sie jede Verhaltensweise ausführen würden (Werte in Klammern in Tabelle 22). Die Simulation des alternativen Auswertungsschlüssels zeigte, dass Personen eine höhere Punktzahl erzielen würden, wenn sie alle Antworten z.B. mit 5 raten würden, als wenn sie nur die angemessenste Antwort mit 5 raten würden (Beispiel 7 vs. 1). Der Auswertungsschlüssel würde somit Personen bevorzugen, die keine Wahl zwischen den angemessenen und unangemessenen Verhaltensweisen treffen. Zudem wären Personen benachteiligt, die sich für die angemessenste Antwort statt für mehrere angemessene Antworten entscheiden würden (Beispiel 1 vs. 2). Darüber hinaus läge das Problem vor, dass mit den Items mit höherer Antwortanzahl eine höhere Punktzahl erreicht werden könnte als mit antwortärmeren Items. Die Items wären somit nicht gleichwertig. Der alternative Auswertungsschlüssel wurde daraufhin verworfen.

Tabelle 22: *Punktbildung*

Antwortoption	Faktor	Beispiel						
		1	2	3	4	5	6	7
a Ich mache mir nichts weiter aus dem Verhalten der Arzthelferinnen. Wichtiger ist, dass die Hausärztin mit mir zufrieden ist.	2	0% (1)	0% (1)	0% (1)	0% (1)	100% (5)	50% (5)	20% (5)
b Ich verhalte mich den Arzthelferinnen gegenüber genauso wie sie es mir gegenüber tun.	1	0% (1)	0% (1)	100% (5)	50% (5)	0% (1)	0% (1)	20% (5)
c Ich suche ein klärendes Gespräch mit den Arzthelferinnen.	5	100% (5)	70% (5)	0% (1)	50% (5)	0% (1)	0% (1)	20% (5)
d Ich schildere der Hausärztin meine Wahrnehmung und bitte sie um ihre Einschätzung.	4	0% (1)	30% (5)	0% (1)	0% (1)	0% (1)	0% (1)	20% (5)
e Ich suche nicht das Gespräch. Vermutlich geht die wenig herzliche Aufnahme der Arzthelferinnen auf andere Ursachen zurück, die nichts mit meiner Arbeitsweise zu tun haben.	2	0% (1)	0% (1)	0% (1)	0% (1)	0% (1)	50% (5)	20% (5)
Punktbildung:		5 (34)	4.7 (50)	1 (20)	3.5 (38)	2 (22)	2 (30)	2.8 (70)

Anmerkung. Beispiel 1: Die angemessenste Antwort wird gewählt. Beispiel 2: Die angemessenste und die zweitangemessenste Antwort werden gewählt. Beispiel 3: Die unangemessenste Antwort wird gewählt. Beispiel 4: Die angemessenste und unangemessenste Antwort werden gewählt. Beispiel 5: Die zweitunangemessenste Antwort wird gewählt. Beispiel 6: Zwei gleichwertige zweitunangemessenste Antworten werden gewählt. Beispiel 7: Alle Antworten werden mit gleicher Wahrscheinlichkeit gewählt. Die Werte in Klammern bilden das Vorgehen des alternativen Auswertungsschlüssels ab.

6.4.7. Limitationen

Die Auswahl der Experten basierte auf Gelegenheitsstichproben, was zu Ergebnisverzerrungen geführt haben könnte. So wurden etwa die Psychologischen Ausbildungsinstitute angeschrieben, zu denen persönliche Kontakte vorlagen. Andererseits wurde durch die verschiedenen Rekrutierungswege versucht, eine möglichst heterogene Gruppe aus fakultätsinternen und externen Experten zu gewinnen. Insgesamt beteiligten sich 64 Experten, wobei jedes Item von 8-21 Experten beurteilt wurde. Zur Absicherung der Ergebnisse wäre ein Vergleich zwischen den zwei Expertengruppen der Mediziner vs. Psychologen von Vorteil. Eine aussagekräftige Analyse war jedoch aufgrund der geringen Beurteileranzahl pro Item nicht möglich. Durch die zusätzliche Diskussion der Angemessenheit der Verhaltensoptionen durch das Diskussionsteam wurde ein Abgleich und Kompromiss zwischen der Gruppe der Experten und der der Testkonstrukteure vorgenommen. Somit floss sowohl die Alltagsexpertise von Ärzten und Psychologen als auch die Expertise zur Testentwicklung ein. Abweichungen zwischen dem Diskussionsteam und den Experten könnten zum einen darauf zurückgeführt werden, dass die Experten die Angemessenheit der Verhaltensweisen vermutlich eher allgemein statt kompetenzspezifisch beurteilt haben. So merkten zwei Teilnehmer an, dass sie es als anspruchsvoll erlebt hatten, das Verhalten jeweils auf eine spezifische soziale Kompetenz hin zu bewerten. Mit der Fokussierung auf eine spezifische soziale Kompetenz wurde versucht, der Empfehlung nachzukommen, *SJTs* konstruktbasierend zu entwickeln (Motowidlo et al., 2006; Polyhart & Ryan, 2000; Trippe & Foti, 2003). Neben approbierten Medizinerinnen und Psychologen hätten Patienten die Verhaltensoptionen ebenfalls bewerten können. Jedoch wäre dies nur für acht Items mit Patienten-Studenten-Interaktion denkbar gewesen. Die übrigen 40 Items enthielten Interaktionen zwischen Kommilitonen und Lehrenden, für die Patienten nicht die passenden Experten gewesen wären. Die Anzahl der Beurteiler pro Item stand mit der Expertenanzahl anderer Studien in Einklang (vgl. Chan & Schmitt, 1997; De Meijer et al., 2010; Weekley & Jones, 1997). Die Realitätsnähe der Items wurde von weniger Experten bewertet als die Antwortoptionen. Zum einen lag dies daran, dass nur die Mediziner diese beurteilten, zum anderen hob sich die Frage nicht gut vom restlichen Fragebogentext ab, sodass die Instruktion vermutlich häufiger übersehen wurde. Um der Kritik entgegen zu wirken, dass Items mit hoher Augenscheinvalidität oder mit hoher sozialer Erwünschtheit von Beurteilern übereinstimmender und Items mit mehr Verhaltensalternativen weniger übereinstimmend bewertet werden (Krokos et al., 2004; Möller, 2010), wurde der *Cut-off-Wert* des *ICCs* von den üblichen .70 (vgl. z.B. Lievens & Sackett, 2006; Weekley & Jones, 1997) auf .50 herabgesetzt. Dies ist die Schwelle, bei der sich

50% der Beurteiler einig sind und 50% nicht. Auf diese Weise sollten Items mit geringerer Augenscheinvalidität die Chance erhalten, im *SJT* zu verbleiben, aber gleichzeitig Items mit sehr geringer Übereinstimmung ausgeschlossen werden. Neben dem *ICC* wurden die Realitätsnähe der Situationen und die Variation in der Angemessenheit der Verhaltensweisen berücksichtigt, um zwischen Personen mit hoher und niedriger Merkmalsausprägung differenzieren zu können. Dass die Antworten eines Items daher mindestens drei Skalenpunkte der fünfstufigen Ratingskala abdecken sollten, war jedoch beliebig gesetzt und gestaltete sich im Nachhinein als wenig hilfreich. Um gleichgewichtete Items zu erhalten, wurden bei einigen Items nachträglich neue Antwortoptionen ergänzt. Somit konnte eine stärkere Streuung der Angemessenheit der Verhaltensweisen auch im Nachhinein erreicht werden. Die Auswertungsschlüssel derzeitiger *SJTs* enthalten nur Verhaltensoptionen unterschiedlicher Angemessenheit oder Effektivität. Der geplante Auswertungsschlüssel ist in der Lage, mehrere, sozial gleichrangige Handlungsoptionen zu integrieren. Dies entspricht eher der menschlichen Reaktionsvielfalt und der Komplexität sozialen Verhaltens als die Darbietung von nur einer möglichen Verhaltensweise pro Angemessenheitsgrad. Der Auswertungsschlüssel basiert darüber hinaus auf empfohlenen kontinuierlichen *Ratings*, welche gewöhnlich mit höheren internen Konsistenzen einhergehen (Polyhart & Ehrhart, 2003). Durch den Auswertungsschlüssel wurde zudem erreicht, dass alle Items unabhängig von ihrer Antwortanzahl die gleiche Punktspanne (1-5 Punkte) aufwiesen und somit untereinander gleichwertig waren.

6.5. Verfilmen der Items

Nach der Bildung des Auswertungsschlüssels wurden die verbliebenen 24 Situationen mit Hilfe eines professionellen Filmteams verfilmt. Die Verfilmung orientierte sich nach den Empfehlungen von Olson-Buchanan und Drasgow (2006). Für die Situationen wurde ein Drehbuch entworfen, in dem Dialoge, Regieanweisungen und notwendige Requisite festgehalten wurden. Abbildung 20 gibt einen Drehbuchauszug.

Originalitem SU1: Sie beginnen den ersten Tag Ihres Pflegepraktikums auf einer Krankenhausstation. Nach einer kurzen Begrüßung vom Arzt- und Pflegepersonal gehen alle hektisch Ihrer Arbeit nach. Nur Sie wissen gar nicht, was sie eigentlich tun sollen. Ein Ansprechpartner wurde Ihnen nicht zugewiesen.

Szene SU1-1

Personen:

STUDENT

2 Ärzte und ein Pfleger

Requisite:

Jacke und Rucksack des Studenten

Einleitung: Für das Humanmedizinstudium müssen Sie ein 3-monatiges Pflegepraktikum vor Beginn des Studiums oder in den unterrichtsfreien Zeiten absolvieren. In diesem lernen Sie die Arbeit von Pflegekräften und den Stationsablauf kennen.

IM FLUR EINER KRANKENHAUSSTATION

INNEN/MORGENS

STUDENT mit Jacke und Rucksack bekleidet kommt schnell durch den Flur und sucht das Stationszimmer. Als er es gefunden hat, klopft er an die Tür.

ÄRZTE VON INNEN
Herein!

Szene SU1-2

Personen:

STUDENT

2 ÄRZTE und ein PFLEGER

Requisite:

- Ärztekittel und Pflegebekleidung für alle Ärzte und Pfleger
- Ein paar Patientenakten
- Jacke und Rucksack des Studenten

IM STATIONSZIMMER

INNEN/MORGENS

STUDENT öffnet die Tür und schließt sie wieder hinter sich. ARZT- und PFLEGEPERSONAL steht versammelt bei der morgendlichen Übergabe mit Patientenakten in der Hand.

ÄRZTIN
Einen Moment bitte.
(Besprechen die Patientin zu Ende.)

ARZT
Ah, Sie müssen der neue Medizinstudent für das Pflegepraktikum sein.

STUDENT
Ja, das bin ich.
(nickt)

Fortsetzung Abbildung 20:

ARZT
Herzlich willkommen!

Ich bin Herr Rupprecht.
(gibt Hand)

ÄRZTIN
Hallo, Dörr mein Name.
(gibt Hand)

PFLEGER
Hallo, Schmidt.

ARZT
Gut, wir hätten alles soweit besprochen. Gehen wir an die Arbeit. Bis
später.

PFLEGER
Ich habe Ihnen noch Kleidung vorbereitet. Dort ist die Umkleide. Bis
später.
(drückt ihm Kleidung in die Hand und geht)

(Alle ÄRZTE und PFLEGER verlassen den Raum, um ihrer Arbeit nachzugehen. STUDENT steht verlassen immer noch mit Jacke bekleidet, Rucksack auf dem Rücken und Pflegebekleidung in den Händen im Stationszimmer.)

Abbildung 20. Drehbuchauszug des SJTs.

Das Drehbuch wurde von einer Medizinstudentin in einem höheren klinischen Semester auf fachliche Richtigkeit und Realitätsnähe geprüft. Der Dreh fand an 10 Drehtagen in Räumlichkeiten der Medizinischen Fakultät und des Universitätsklinikums Heidelberg statt. Die 78 Laienschauspieler wurden über den privaten und beruflichen Freundes- und Bekanntenkreis der Autorin gewonnen. Auch Medizinstudenten und Ärzte der Fakultät beteiligten sich am Dreh. Die Organisation und Regie übernahm die Autorin. Video- und Tonaufnahmen, sowie Schnitt und Nachbearbeitung übernahm das gecharterte Filmteam. Die Filmsequenzen wurden durch eine kurze schriftliche Erklärung eingeleitet, wenn der Dialog nicht alle zum Verständnis notwendigen Informationen enthielt. An der kritischen Stelle stoppten die Filmszenen, um den Testteilnehmer aufzufordern, die 4-6 schriftlichen Handlungsoptionen nach der Wahrscheinlichkeit zu bewerten, mit der er sie ausführen würde. Auf eine zeitliche Limitation bei der Beantwortung der Items wurde verzichtet, da der *SJT* für den geplanten Einsatzbereich keinen Leistungstest darstellt. Bei der Verfilmung wurde darauf geachtet, dass Mimik, Gestik und Sprache zu den ursprünglichen Items passten. Nach der Verfilmung wurden die schriftlichen Items in der Anzahl und dem Geschlecht der Akteure adaptiert, damit sie mit den videobasierten Versionen übereinstimmten.

7. Validierung an medizinischen Studienplatzbewerbern und -interessenten

Zielsetzung der Studie war es, die Faktorenstruktur des videobasierten *SJT*s an Studienplatzbewerbern und Studieninteressenten der Human- und Zahnmedizin zu prüfen. Des Weiteren wurden Itemstatistiken der klassischen Testtheorie und die Kriteriumsvalidität durch den Zusammenhang mit Selbsteinschätzungsskalen sozialer Kompetenzen und der allgemeinen kognitiven Leistungsfähigkeit (Abiturnotendurchschnitt, TMS-Ergebnis) analysiert. Auch die Nützlichkeit und Akzeptanz des *SJT*s wurde via Fragebogen erhoben. Die Durchführung der Studie wurde durch ein Ethikvotum der Medizinischen Fakultät Heidelberg genehmigt (Ethikantrag S-162/2011).

7.1. Hypothesen

Aufgrund des der Arbeit zugrunde liegenden Arbeitsmodells sozialer Kompetenzen wird vermutet, dass mit einer konfirmatorischen Faktorenanalyse ein zweifaktorielles Modell nachgewiesen werden kann (Abbildung 21).

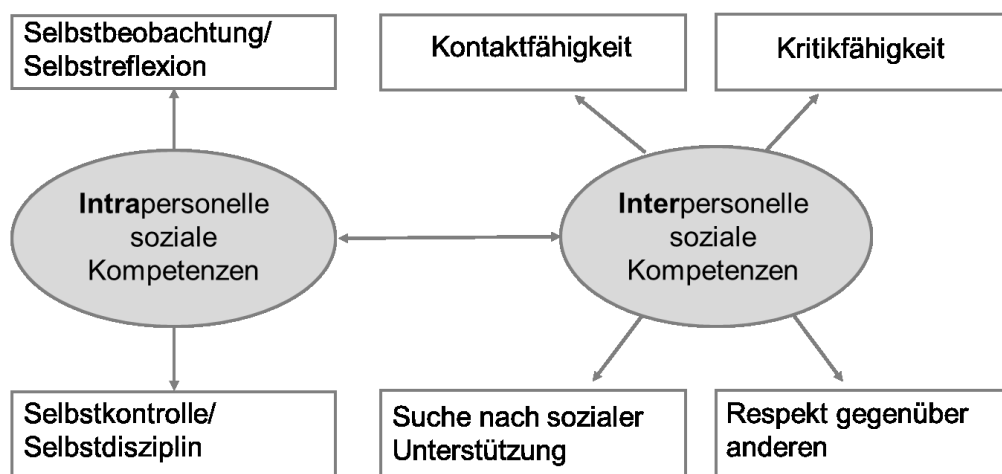


Abbildung 21. Angenommene zweifaktorielle Struktur des *SJT*s.

Das Modell spiegelt die sozialen Kompetenzen der nullten und ersten Abstraktionsebenen des Arbeitsmodells wieder (Kapitel 3.4). Ein Faktor bildet die intrapersonellen Kompetenzen (0. Ebene des Arbeitsmodells) ab, die im *SJT* durch die Items der Selbstkontrolle/Selbstdisziplin und Selbstbeobachtung/Selbstreflexion wiedergegeben werden. Der andere Faktor repräsentiert die interpersonellen Kompetenzen der ersten Ebene des Arbeitsmodells. Diese werden im *SJT* durch die Items der Respektfähigkeit, Kontaktfähigkeit, Kritikfähigkeit und Suche nach sozialer Unterstützung abgebildet. Die Faktoren sind abhängig voneinander, da die intrapersonellen Kompetenzen eine Vo-

oraussetzung für die interpersonellen Kompetenzen sind. Die Hypothese (H) lautet daher:

H1: Dem videobasierten *SJT* liegen zwei Faktoren zugrunde. Ein Faktor setzt sich aus den zwei sozialen Kompetenzen Selbstkontrolle/Selbstdisziplin und Selbstbeobachtung/Selbstreflexion zusammen. Der zweite Faktor bildet die vier sozialen Kompetenzen Respektfähigkeit, Kontaktfähigkeit, Kritikfähigkeit und Suche nach sozialer Unterstützung ab.

Basierend auf der *Multitrait-Multimethod*-Methode (Campell & Fiske, 1959) sollten Verfahren, die gleiche Konstrukte erheben, miteinander stärker korrelieren als Verfahren die unterschiedliche Konstrukte erfassen. Die konvergenten Validitätskoeffizienten sollten signifikant größer null, die diskriminanten Validitätskoeffizienten signifikant kleiner als die konvergenten Validitätskoeffizienten sein. Da der *SJT* und die Selbsteinschätzungsskalen dieselben sozialen Kompetenzen zu messen beanspruchen, sollten diese miteinander signifikant größer null korrelieren. Neben den Annahmen der *Multitrait-Multimethod*-Methode beeinflusst auch die Instruktionsform des *SJT*s die Hypothesenbildung. *SJT*s mit verhaltensbezogener Instruktion korrelieren höher mit Persönlichkeitsfaktoren und geringer mit allgemeinen kognitiven Fähigkeiten als *SJT*s mit wissensbezogener Instruktion (McDaniel & Nguyen, 2001). Dies sollte auch für den vorliegenden *SJT* mit verhaltensbezogener Instruktion gelten. Folgende Hypothesen werden daher getestet:

H2: Der *SJT* korreliert signifikant mit den Selbsteinschätzungsskalen der sozialen Kompetenzen.

H3: Der *SJT* zeigt Nullkorrelationen bis niedrige Korrelationen mit den Kriterien der allgemeinen kognitiven Leistungsfähigkeit (Abiturnotendurchschnitt, TMS-Ergebnis).

Frauen erzielen gewöhnlich geringfügig höhere Punktwerte als Männer im *SJT* (Hartman et al., 2007; Lievens, 2012; Nguyen et al., 2005). Daher wird auch für den vorliegenden *SJT* der Geschlechtsunterschied erwartet:

H4: Frauen erreichen höhere Punktzahlen im *SJT* als Männer.

Videobasierte *SJTs* erzielen verglichen mit schriftlichen aufgrund der stärkeren Realitätsnähe eine höhere Akzeptanz der Teilnehmer (Chan & Schmitt, 1997; Richman-Hirsch, Olson-Buchanan & Drasgow, 2000). In Anlehnung daran, sollte der vorliegende videobasierte *SJT* von den Studienplatzbewerbern als ansprechend erlebt und akzeptiert werden. Demzufolge wird angenommen:

H5: Der *SJT* wird von den Studienplatzbewerbern akzeptiert.

Ein Zweck von *SJTs* besteht darin, eine möglichst realistische Vorschau auf die Arbeitstätigkeit zu geben (Motowidlo et al., 1990). Der vorliegende *SJT* sollte daher die Studieninteressenten und -bewerber über die sozialen Anforderungen informieren, die das Medizinstudium an sie stellt. Folgende Hypothese wird getestet:

H6: Der *SJT* ist in der Lage, die Teilnehmer über die Anforderungen zu informieren, die das Medizinstudium an die sozialen Kompetenzen der Studierenden stellt.

Aufgrund des Einsatzes als *Self-Assessment*-Instrument sollte der *SJT* die Auseinandersetzung der Teilnehmer mit der eigenen Person, dem Studienangebot und den Studienanforderungen fördern, um eine systematische Studienwahlentscheidung zu fördern (Organisation for Economic Cooperation and Development, 2004). In diesem Fall sollte er zum Reflektieren der eigenen sozialen Kompetenzen anregen und den Wunsch, Medizin zu studieren, verstärken oder schwächen. Im Optimalfall werden dabei Personen mit hohen *SJT*-Werten bestärkt, während Personen mit niedrigen Punktzahlen im *SJT* abgeschreckt werden. Die Hypothesen lauten daher:

H7: Der *SJT* regt die Teilnehmer zum Reflektieren der eigenen sozialen Kompetenzen an.

H8: Personen, mit hohen *SJT*-Werten, fühlen sich bestärkt Medizin zu studieren, während Personen mit niedrigen *SJT*-Werten von einem Medizinstudium abgeschreckt werden.

7.2. Untersuchungsablauf und eingesetzte Verfahren

Die Untersuchung erfolgte an zwei Stichproben. Die erste Stichprobe setzte sich aus Studieninteressenten und -bewerbern zusammen, die während der Bewerbungsphase (01.06.-15.07.2013) für das Human- und Zahnmedizinstudium im Wintersemester 2013/2014 am *SJT* teilnahmen. Diese wird folgend als erste Stichprobe bezeichnet.

Die zweite Stichprobe bestand aus Personen, die nach Ablauf der Bewerbungsphase den *SJT* bearbeiteten (16.07.2013-31.03.2014). Diese wird folgend als zweite Stichprobe bezeichnet. Die Erhebung fand online über das Programm q-set statt.

Zu Beginn des Tests beantworteten die Teilnehmer soziodemografische Fragen und gaben ihre freiwillige Einwilligung zur wissenschaftlichen Datenauswertung. Die Teilnehmer wurden zu Beginn und am Ende der gesamten Erhebung gefragt, ob sie beabsichtigen, den Test wahrheitsgemäß zu bearbeiten. Damit sollten Personen ausgeschlossen werden, die den Test nur ausprobieren wollten. Zudem wurde erhoben, ob die Personen zum ersten oder wiederholtem Mal am Test teilnahmen, um letztere ausschließen zu können. Anschließend bearbeiteten die Studienplatzbewerber und -interessenten den *SJT* und schätzten sich anhand von Selbsteinschätzungsskalen bezüglich der sechs sozialen Kompetenzen des *SJTs* ein. Abschließend evaluierten sie den *SJT* selbst. Die 90-minütige Testbearbeitung konnte durch Zwischenspeichern unterbrochen und zu einem späteren Zeitpunkt wieder aufgenommen werden. Unmittelbar nach Beendigung des Fragebogens erhielten die Bewerber eine Rückmeldung zum Testergebnis. Die Rückmeldung verglich die gewählten Verhaltensoptionen des Teilnehmers mit der sozialen Angemessenheit der Verhaltensoptionen auf Basis der Expertenurteile. Die eingesetzten Verfahren werden nun in der Reihenfolge des Untersuchungsablaufs genauer beschrieben.

Soziodemografische Angaben. Die soziodemografischen Angaben umfassten das Alter und das Geschlecht der Teilnehmer. Außerdem wurde gefragt, ob die Personen zum ersten oder wiederholten Male am Test teilnehmen und ob sie eine ernsthafte Bearbeitung des Tests beabsichtigten. Diese Angaben wurden als Kontrollvariablen in die Untersuchung einbezogen, um potenzielle Einflüsse auf die Bearbeitung des *SJTs* zu prüfen. Darüber hinaus wurde eine Einwilligung zur wissenschaftlichen Verwertung der Daten eingeholt. Als kognitive Leistungsindikatoren wurden der Abiturnotendurchschnitt und der standardisierte Testwert im TMS erfragt, der bereits in Kapitel 2.1 beschrieben wurde.

Videobasierte SJT. Der eingesetzte *SJT* enthielt 24 verfilmte Situationen aus dem Human- und Zahnmedizinstudium, welche in der zeitlichen Reihenfolge des Studiums dargestellt wurden. Der *SJT* setzte sich aus zwei Filmen zur Selbstbeobachtung/Selbstreflexion, vier zur Selbstkontrolle/Selbstdisziplin, sechs zur Kritikfähigkeit, einen zur Respektfähigkeit, neun zur Suche nach sozialer Unterstützung und zwei zur Kontaktfähigkeit zusammen. Die Filmsequenzen wurden durch eine kurze schriftliche Erklärung eingeleitet, wenn der Dialog nicht alle, zum Verständnis notwendigen, Infor-

mationen enthielt. An der jeweils erfolgskritischen Stelle stoppten die Filme und die Bewerber wurden aufgefordert, anzugeben wie sie an Stelle des dargestellten Studierenden reagieren würden. Dafür standen ihnen pro Item vier bis sechs Verhaltensalternativen zur Verfügung, die in zufälliger Reihenfolge angeordnet waren. Pro Verhaltensalternative sollten sie angeben, mit welcher Wahrscheinlichkeit sie diese ausführen würden. Insgesamt mussten 100 Prozent pro Item vergeben werden. Die bereits vergebenen Prozente wurden automatisch angezeigt. Die Filme konnten bei Bedarf erneut angeschaut werden. Eine zeitliche Limitation zur Beantwortung der Items bestand nicht. Eine kritische Filmstelle und die Instruktion des SJTs veranschaulicht Abbildung 22.



Wie wahrscheinlich würden Sie wie folgt an Stelle des Studenten auf die Äußerung der Patientin reagieren? Bitte vergeben Sie insgesamt 100 Prozent.

- | | |
|---|-------------|
| a) Ich frage sie, ob sie schon schlechte Erfahrungen gemacht hat, um die Situation zu verstehen. | ___% |
| b) Ich sage höflich, dass ich ihr Verhalten übertrieben finde. | ___% |
| c) Ich sage, dass sie sich keine Sorgen machen muss, da ich das Blutabnehmen beherrsche. | ___% |
| d) Ich akzeptiere ihre Meinung und gebe der Ärztin Bescheid, dass die Patientin von einem Arzt behandelt werden möchte. | ___% |
| Summe | ___% |

Abbildung 22. Kritische Filmstelle und Instruktion des SJTs.

Selbsteinschätzungsskalen sozialer Kompetenzen. Die Selbsteinschätzungsskalen bestanden aus siebenstufigen Likert-Skalen (1 = geringer als bei Gleichaltrigen bis 7 = höher als bei Gleichaltrigen), die die Teilnehmer aufforderten, sich in den sechs sozialen Kompetenzen des *SJT* - verglichen mit Gleichaltrigen - einzuschätzen. Falls sich die Teilnehmer in einer Kompetenz nicht einzuschätzen wussten, stand ihnen die Option „kann ich nicht beurteilen“ zur Verfügung. Im Vorfeld wurde jeweils die Definition der sozialen Kompetenz gegeben, damit die Teilnehmer möglichst das selbe unter der Begrifflichkeit verstanden.

Evaluation des SJTs. Die Evaluationsfragen wurden verwendet, um die Nützlichkeit und Akzeptanz des *SJT* zu erheben. Dazu beurteilten die Teilnehmer Aussagen zum *SJT* auf einer fünfstufigen Likert-Skala (1 = trifft zu bis 5 = trifft nicht zu). Die Aussagen bezogen sich darauf, ob die Teilnehmer eine Vorstellung von den Anforderungen an ihre sozialen Kompetenzen im Studium bekommen haben und ob der *SJT* zur Selbstreflexion des eigenen Verhaltens in sozialen Situationen anregt hat. Des Weiteren wurde erfasst, ob die Studienplatzbewerber in ihrer Entscheidung, Medizin zu studieren bestärkt oder abgeschreckt wurden und ob sie den *SJT* weiterempfehlen würden. Zusätzlich konnten die Teilnehmer Kritik zum *SJT* äußern.

7.3. Stichprobenrekrutierung und -beschreibung

Die medizinischen Studieninteressenten und -bewerber der ersten Stichprobe wurden während der Bewerbungsphase auf zwei Wegen rekrutiert. Zum einen wurde der Link zum Selbsteinschätzungstest auf der Homepage für Studieninteressenten der Medizinischen Fakultät Heidelberg platziert (<http://www.medizinische-fakultaet-hd.uni-heidelberg.de/index.php?id=110069>, Zugriff am 01.04.2013). Zum anderen wurde den Bewerbern vor der Eingabe der Zulassungsdaten für das AdH-Verfahren angeboten, am freiwilligen Selbsteinschätzungstest teilzunehmen. Sie wurden explizit darauf hingewiesen, dass der *SJT* keinen Einfluss auf ihre Bewerbung hat. Die zweite Stichprobe bestand aus Studieninteressenten, die seit August 2013 bis März 2014 am *SJT* teilnahmen. Die Teilnehmer gelangten nur über die Homepage für Studieninteressenten zum Test.

Von den 913 vollständigen Datensätzen der ersten Stichprobe wurden 17 Personen ausgeschlossen, die angaben, den Test nicht ehrlich beantwortet zu haben. Des Weiteren wurden 23 Personen ausgeschlossen, die den *SJT* mehr als einmal bearbeitet hatten und 104 Personen, die sich gegen eine wissenschaftliche Verarbeitung ihrer Daten aussprachen. Somit gingen 769 Personen in die Auswertung ein. Die erste

Stichprobe setzte sich aus 530 weiblichen (69%) und 239 männlichen (31%) Studienplatzbewerbern und -interessenten zusammen. Das Durchschnittsalter betrug 20.1 Jahre ($SD = 3.3$) mit einer Altersspanne von 15 bis 54 Jahren. Getrennt nach Altersgruppen stellten die 18-20 Jährigen die am stärksten vertretene Gruppe dar (Tabelle 23). Getrennt nach Geschlecht zeigten sich keine Unterschiede im mittleren Alter (Frauen vs. Männer: 20.1 vs. 19.9, $p = .39$). Das maximale Alter lag bei den Frauen jedoch höher als bei den Männern (54 vs. 36 Jahre).

Tabelle 23: *Altersverteilung der Stichproben*

Alter	Stichprobe 1		Stichprobe 2	
	N	%	N	%
15-17	55	7.2	36	22.1
18-20	511	66.4	74	45.4
21-23	120	15.6	26	16.0
24-26	49	6.4	8	4.9
27-29	15	2.0	8	4.9
>30	15	2.0	11	6.7

Von den 211 vollständigen Datensätzen der zweiten Stichprobe wurden vier Personen ausgeschlossen, die angaben, den Test nicht ehrlich beantwortet zu haben und vier Personen, die den *SJT* mehr als einmal bearbeitet hatten. Zudem wurden 40 Personen ausgeschlossen, die sich gegen die wissenschaftliche Auswertung ihrer Daten aussprachen. Die resultierende Stichprobe setzte sich aus 163 Teilnehmern zusammen, darunter 120 Frauen (74%) und 43 Männer (26%). Das Durchschnittsalter betrug 21.0 Jahre ($SD = 6.5$) mit einer Altersspanne von 15 bis 53 Jahren. Die 18-20 Jährigen stellten gefolgt von den 15-17 Jährigen die am stärksten vertretene Gruppe dar (Tabelle 23). Das Durchschnittsalter zeigte in der weiblichen und männlichen Stichprobe keinen Unterschied (20.4 vs. 20.1, $p = .67$).

Beim Vergleich der ersten und zweiten Stichprobe konnte festgestellt werden, dass der Geschlechteranteil in beiden Gruppen zu Gunsten der Frauen ausfiel. In beiden Stichproben waren die 18-20 Jährigen, die am häufigsten vertretene Gruppe, wobei der Anteil in der ersten Stichprobe höher lag als in der zweiten (66% vs. 45%). In der zweiten Stichprobe waren die 15-17 Jährigen häufiger vertreten als in der ersten Stichprobe (22% vs. 7%). Die Teilnehmer waren in der ersten Stichprobe ca. ein Jahr jünger als in der zweiten, die Streuung in der zweiten Stichprobe war höher ($M_{SP1} = 20.1$; $SD_{SP1} = 3.3$ vs. $M_{SP2} = 21.0$; $SD_{SP2} = 6.5$). In beiden Stichproben bestand kein Unterschied im

mittleren Alter von Frauen und Männern ($M_{SP1} = 20.1$ vs. 19.9 , $p = .39$; $M_{SP2} = 20.4$ vs. 20.1 , $p = .67$).

7.4. Statistische Auswertung

Das Hintergrundwissen und die Empfehlungen zur statistischen Auswertung stammten aus Bühner (2011). Kamen darüber hinaus andere Quellen zum Einsatz, werden diese gesondert aufgeführt.

Stichprobenumfang. Zur Berechnung des Stichprobenumfangs wurde das Programm *G* Power 3* (Faul, Erdfelder, Lang & Buchner, 2007) verwendet. Die statistische Power wurde nach Empfehlungen von Rost (2013) für alle Berechnungen auf .80 gesetzt. Für die bivariaten zweiseitigen Korrelationsanalysen werden bei Erwartung von mittleren Effektstärken 82 Testpersonen benötigt. Für die Unterschiedstestung zwischen den Geschlechtern werden bei einseitiger Testung und einem erwarteten mittleren Effekt, 102 Personen benötigt. Zur Schätzung des Stichprobenumfangs bei Faktorenanalysen wurden die Richtlinien von MacCallum, Widaman, Zhang und Hong (1999) herangezogen. Diese ordnen Stichprobengrößen von 100 als ausreichend, von 200 als befriedigend, von 300 als gut, von 500 als sehr gut und von 1000 als exzellent ein. Um mindestens eine zufriedenstellende Eignung zu erzielen, sollten die Stichproben mindestens 200 Personen umfassen.

Deskriptive Statistik. Der Auswertungsschlüssel des *SJT*s wurde anhand des beschriebenen Vorgehens in Kapitel 6.4.6 gebildet. Zur Exploration der Daten wurden Minima, Maxima, Häufigkeiten, Mittelwerte und Standardabweichungen ermittelt. Bei den *SJT*-Items wurden zusätzlich die Schiefe und Kurtosis analysiert. Die Prüfung auf Normalverteilung der Daten fand mittels Kolmogorov-Smirnov-Test und der Betrachtung der Histogramme statt. Extremwerte und Ausreißer wurden anhand der Boxplots analysiert. Bei der Analyse der *SJT*-Items wurde zusätzlich zum arithmetischen Mittelwert der M-Schätzer nach Huber ermittelt. Fallen der arithmetische Mittelwert und der M-Schätzer ähnlich aus, kann dies als Hinweis gelten, dass sich keine gravierenden Ausreißer in der Stichprobe befinden.

Konfirmatorische Faktorenanalyse (KFA). Zur Prüfung der Modellannahme des *SJT*s wurde eine konfirmatorische Faktorenanalyse mit der *Maximum-Likelihood*-Methode (*ML*-Methode) mit dem Programm SPSS AMOS (Version 21, IBM) durchgeführt. Als Voraussetzung für die Durchführung einer KFA sollte die Mindeststichpro-

bengröße ca. 200 Personen umfassen und die Daten intervallskaliert sein. Mit dem Mardia-Test wurde geprüft, ob die Daten auf Grundlage der Schiefe und Kurtosis multivariat normalverteilt waren. Fällt der *critical ratio*, der als z-Wert interpretiert werden kann, $> 1,96$ aus, ist die multivariate Normalverteilung nicht gegeben. In diesem Falle wurde das *Bollen-Stine-Bootstrap*-Verfahren bei der ML-Schätzung angewandt. Items mit einer Schiefe > 2 und einer Kurtosis > 7 sollten für die KFA ausgeschlossen werden (West, Finch & Curran, 1995). Eine Überprüfung der Ausreißer erfolgte bereits während der deskriptiven Analyse. Des Weiteren wurden die Daten auf multivariate Ausreißer getestet. Der in AMOS bereitgestellte Test hat jedoch den Nachteil, dass er mit steigender Stichprobengröße mehr Personen als Ausreißer deklariert. Um eine angemessenere Faktorenlösung und eine exakte und stabile Parameterschätzung zu erzielen, wird darüber hinaus empfohlen, mindestens vier Items pro latente Variable in eine KFA aufzunehmen.

Die Bewertung des Gesamtmodells erfolgte auf Grundlage des Chi-Quadrat-Tests (χ^2 -Test). Bei signifikantem Testergebnis unterscheidet sich die Modellannahme von der empirischen Verteilung der Stichprobe. Mit steigender Stichprobengröße führen kleine Modellabweichungen schnell zur Ablehnung des Modells durch den χ^2 -Test. Daher wurden Fit-Indizes zur globalen Beurteilung des Modells einbezogen. Nach Empfehlungen von Beauducel und Wittmann (2005) wurden der *Comparative-Fit-Index (CFI)*, der *Root-Mean-Square-Error of Approximation (RMSEA)* und der *Standardized-Root-Mean-Residual (SRMR)* verwendet. Bei dem *CFI* gelten Werte $\geq .95$ als gut. Der *RMSEA* sollte Werte $< .08$ für Stichproben ≤ 250 Personen und Werte $< .06$ für Stichproben von mehr als 250 Personen annehmen. Der *SRMR*-Wert sollte $< .11$ ausfallen. Werden diese *Cut-off*-Werte nicht erreicht, muss das Modell verworfen werden.

Zur Untersuchung des lokalen Modell-Fits wurden die unstandardisierten und standardisierten Ladungen herangezogen. Fallen die unstandardisierten Ladungen signifikant aus, ist die Mindestvoraussetzung für den lokalen Fit gegeben. Wenn die standardisierten Ladungen deutlich über eins liegen, weist dies auf Schätzprobleme hin.

Explorative Faktorenanalyse. Im Falle der Modellverwerfung durch die KFA wurden exploratorische Faktorenanalysen (EFA) herangezogen, um die Datenstruktur zu analysieren und eine Skalenbildung zu ermöglichen. Der EFA gingen Eignungsprüfungen der Daten voraus. Mit dem Bartlett-Test, dem Kaiser-Meyer-Olkin-Koeffizient (KMO) und dem *MSA*-Koeffizienten (*Measure of Sample Adequacy*) wurde getestet, ob die Daten substantielle Korrelationen untereinander aufwiesen. Der Bartlett-Test sollte ein signifikantes Testergebnis aufweisen. Der KMO- und der *MSA*-Koeffizient können Werte zwischen null und eins annehmen. Werte $\geq .90$ gelten als sehr gut geeignet zur

Durchführung einer EFA, Werte zwischen .80 und .89 als gut, zwischen .70 und .79 als mittel, zwischen .60 und .69 als mäßig, zwischen .50 und .59 als schlecht und kleiner .50 als ungeeignet. Als weitere Voraussetzung für die EFA gilt die Analyse von Ausreißern. Als Methode zur Faktorenextraktion wurde die verwendet, die am besten die empirische Stichprobenkorrelation wiedergibt. Zur leichteren Interpretation der Faktorenlösung wurden Ladungen unter .30 im Output unterdrückt. Als Abbruchkriterium für die Faktorenextraktion wurde ein Vergleich zwischen dem Kaiser-Guttman-Kriterium, der Parallelanalyse (Horn, 1965) und dem *Minimum Average Partial-Test* (MAP; Velicer, 1976) vorgenommen. Die beiden letzten Verfahren gelten als objektive Methoden und sind den anderen in der Interpretation zu bevorzugen. Als Rotationsmethode wird Promax verwendet, da die Faktoren als abhängig voneinander angenommen werden.

Reliabilitätsanalysen. Zur Analyse der Reliabilität wurden die interne Konsistenz, die part-whole korrigierte Trennschärfe und die Itemschwierigkeit berechnet. Die interne Konsistenz wurde anhand des Alpha-Koeffizienten (α) erhoben (Cronbach, 1951). Beträgt *Cronbachs* α zwischen .50 und .70 sollte ein Test nur für Gruppenvergleiche verwendet werden. Beträgt die interne Konsistenz $\geq .70$ kann ein Test hingegen auch auf Individualebene eingesetzt werden (Hossiep et al., 2000; Lienert & Raatz, 1998). Die interne Konsistenz gibt jedoch keinen eindeutigen Hinweis auf die Eindimensionalität einer Skala. Hohe Werte könnten auch dadurch verursacht worden sein, dass hoch korrelierende Messfehler vorliegen. Die Trennschärfe bildet besser die Homogenität der Testitems ab. Umso höher die Trennschärfe, desto besser kann zwischen leistungsstarken und -schwachen Probanden differenziert werden. Negative Trennschärfen deuten darauf hin, dass ein Item entgegen der Erwartungen mit der Skala korreliert. Als Faustregel gilt, dass die Höhe der Trennschärfe mindestens .30 betragen soll (Backhaus, Erichson, Plinke & Weiber, 2008). Da ein Item für einen Test nicht trennscharf ist, wenn es immer oder nie korrekt gelöst wird, wurde darüber hinaus geprüft wie das Antwortverhalten der einzelnen Items ausfiel. Dazu wurde analysiert wie viel Prozent der Teilnehmer eine Antwortoption zu null oder zu 100 Prozent auswählten. Betreffende Items sollten ausgeschlossen werden. Die Trennschärfe eines Items sollte nicht losgelöst von seiner Schwierigkeit betrachtet werden (Lienert & Raatz, 1998; S. 105 ff.). Items mittlerer Aufgabenschwierigkeit bringen bessere Voraussetzungen für hohe Trennschärfen mit als leichte oder schwierige Items. Die Schwierigkeit der intervallskalierten *SJT*-Items wurde berechnet, indem der Mittelwert eines Items durch den maximal möglichen Punktwert des Items dividiert wurde. Itemschwierigkeiten sollten nach Fisseni (2004) im Bereich von .20 bis .80 liegen. Die mittlere Itemschwierigkeit sollte um die .50 betragen. Werte um .70 werden jedoch auch als akzeptabel angesehen.

hen (Kelava & Moosbrugger, 2008). Generell sind Item-Kennwerte deskriptive Größen, die Empfehlungen, aber keine strengen Regeln zum Ein- und Ausschluss der Items darstellen. Daher sollten stets inhaltliche Überlegungen in die Entscheidung eingehen.

Validitätsanalysen. Die Zusammenhänge zwischen der *SJT*-Skala und den Validitätskriterien wurden mittels bivariater Korrelationen nach Pearson und Spearman vergleichend analysiert. Das parametrische Verfahren wurde aufgrund seiner höheren statistischen Power eingesetzt (Bortz, Lienert & Boehnke, 2000). Das nicht-parametrische Verfahren wurde verwendet, weil die Daten nicht normalverteilt waren. Zur Analyse ob sich die Punktzahl des *SJT*s zwischen Männern und Frauen unterschied, wurde der t-Test verwendet, der als robust gegenüber Verletzungen der Normalverteilung gilt (Bortz, 2005). Als Voraussetzungsprüfung wurde der F-Test nach Fisher auf Varianzgleichheit in der Grundgesamtheit durchgeführt. Zur besseren Absicherung wurde zusätzlich der nicht parametrische Kolmogorov-Smirnov-Test eingesetzt. Der Einfluss des Alters auf die Punktzahl im *SJT* wurde mit einer univariaten Varianzanalyse mit vorausgehendem Levene-Test auf Varianzhomogenität durchgeführt. Die praktische Bedeutsamkeit des Effekts wurde mit dem partiellen Eta-Quadrat (η^2) erhoben. Die Kommentare der Teilnehmer zur Akzeptanz und Nützlichkeit des *SJT*s wurden mit der qualitativen Datenanalyse nach Mayring (1990) ausgewertet.

7.5. Ergebnisse

Die Ergebnisdarstellung erfolgt zunächst auf Itemebene, in der die deskriptiven Itemstatistiken und die Ergebnisse der faktorenanalytischen Untersuchungen beider Stichproben berichtet werden. Anschließend werden die Ergebnisse auf Skalenebene beschrieben.

7.5.1. Ergebnisse auf Itemebene

Deskriptive Statistik. Tabelle 24 zeigt die Verteilung der *SJT*-Items beider Stichproben. Die Item-Mittelwerte der Stichproben lagen zwischen 3.1 und 4.6 Punkten mit Standardabweichungen zwischen 0.2 und 1.0 Punkten. Verglichen mit der theoretisch möglichen Punktzahl von 1-5 Punkten pro Item erreichten die Probanden überwiegend hohe Werte. Der M-Schätzer nach Huber zeigte in der ersten Stichprobe eine maximale Differenz von 0.2 Punkten zum Mittelwert und in der zweiten von 0.3. Beide Maximaldifferenzen betrafen das Item KF6 mit der höchsten Standardabweichung. Die geringen Abweichungen deuteten darauf hin, dass keine Ausreißer in der Stichprobe vor-

handen waren bzw. die Verzerrungen durch Ausreißer zu vernachlässigen sind (Bühner, 2011). Die *Boxplots* kennzeichneten in der ersten Stichprobe 0-17 Werte pro Item als Ausreißer und in den zweiten 0-12 Werte. Ausreißer waren vor allem die gegenüber dem Mittelwert niedriger ausgeprägten Werte. Da keine inhaltlichen Gründe zum Ausschluss der Ausreißer vorlagen und diese nur einen geringen Verzerrungseffekt zu haben schienen, wurden sie nicht ausgeschlossen.

Tabelle 24: *Deskriptive Statistik auf Itemebene*

	M		M-Schätzer ^a		SD		Min		Max	
	SP1	SP2	SP1	SP2	SP1	SP2	SP1	SP2	SP1	SP2
SBSR5	3.94	3.98	4.03	4.11	0.80	0.82	1.3	2.0	5.0	5.0
SBSR7	4.18	4.12	4.24	4.29	0.68	0.67	1.0	1.3	5.0	5.0
SKSD1	3.50	3.52	3.49	3.51	0.81	0.78	1.0	2.0	5.0	5.0
SKSD4	3.10	3.17	3.10	3.18	0.60	0.66	1.2	1.7	5.0	5.0
SKSD10	4.42	4.47	4.49	4.64	0.55	0.61	2.0	2.0	5.0	5.0
SKSD11	4.17	4.18	4.29	4.32	0.69	0.73	1.0	1.8	5.0	5.0
Kri1	3.62	3.73	3.68	3.82	0.72	0.67	1.0	1.5	5.0	5.0
Kri2	4.01	4.13	4.02	4.20	0.69	0.72	2.0	2.0	5.0	5.0
Kri4	4.06	4.09	4.09	4.18	0.56	0.66	2.0	2.0	5.0	5.0
Kri7	4.32	4.36	4.40	4.42	0.54	0.54	2.0	1.0	5.0	5.0
Kri10	4.27	4.36	4.38	4.66	0.78	0.82	1.0	1.0	5.0	5.0
Kri12	4.26	4.31	4.29	4.39	0.54	0.60	2.6	1.8	5.0	5.0
R2	4.15	4.15	4.24	4.26	0.80	0.82	1.0	1.4	5.0	5.0
SU1	3.92	3.94	3.95	3.97	0.48	0.48	1.0	1.8	5.0	5.0
SU2	4.61	4.62	4.62	4.64	0.24	0.23	3.4	3.8	5.0	5.0
SU5	3.63	3.62	3.63	3.63	0.29	0.29	2.4	3.0	5.0	4.2
SU8	4.43	4.50	4.50	4.58	0.54	0.53	1.0	2.2	5.0	5.0
SU9	3.66	3.69	3.69	3.73	0.45	0.50	2.0	2.2	5.0	5.0
SU10	4.38	4.36	4.48	4.51	0.52	0.60	1.5	2.0	5.0	5.0
SU12	4.39	4.47	4.47	4.60	0.60	0.62	1.0	2.1	5.0	5.0
SU13	3.60	3.55	3.64	3.61	0.65	0.70	1.5	1.5	5.0	4.8
SU14	4.27	4.34	4.29	4.35	0.29	0.35	2.6	2.6	5.0	5.0
KF6	4.14	4.18	4.33	4.50	0.93	1.01	1.0	1.0	5.0	5.0
KF11	4.20	4.27	4.33	4.47	0.72	0.77	1.0	1.4	5.0	5.0

Anmerkung. SP = Stichprobe. $N_{SP1} = 769$; $N_{SP2} = 163$. ^a M-Schätzer nach Huber.

Itembezeichnung: SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/ Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Bei allen Items wurde die Möglichkeit genutzt, die Gesamtwahrscheinlichkeit von 100 Prozent auf die Antwortoptionen zu verteilen. Bei insgesamt 18 456 Antworten (jeder Teilnehmer beantwortet 24 Items) in der ersten bzw. 3 912 in der zweiten Stichprobe wurden die Gesamtwahrscheinlichkeit 1 850 Mal in der ersten und 482 Mal in der zweiten Stichprobe an nur eine Antwort vergeben. Dies entspricht einem Anteil von 10% bzw. 12% der gesamt bearbeiteten Antworten.

Da ein Item für einen Test nicht trennscharf ist, wenn es immer oder nie gelöst wird, wurde das Antwortverhalten pro Item geprüft. Dazu wurde analysiert wie viel Prozent der Teilnehmer eine Antwortoption mit 0% bzw. mit 100% auswählten. Sollte eine Antwort nie oder immer mit 100% gewählt worden sein, sollte sie ausgeschlossen werden. Dies traf für keine der Antwortoptionen beider Stichproben zu (Anhang G-1). Es gab jedoch Antworten, die von mehr als 90% der Teilnehmer in beiden Stichproben nie ausgewählt wurden (SU2c, SU5c, SU5d, SU9c, SU14a, SU14e). In der zweiten Stichprobe wurde zusätzlich SKSD1c von mehr als 90% der Teilnehmer nie gewählt. Diese Antworten entsprachen denjenigen, welche von den Experten als wenig sozial kompetent eingestuft wurden. Am beliebtesten waren die Antworten Kri10c und KF6c, die von 33% bis 47% der Teilnehmer zu 100% ausgewählt wurden. Diese Antworten entsprachen den Antworten, die von den Experten als am sozial kompetentesten eingestuft wurden. Zusammenfassend gab es Antworten, die selten, aber nicht nie oder immer mit 100% gewählt wurden, was für den Verbleib aller Antwortoptionen im *SJT* sprach.

Bei dem Antwortverhalten fiel weiterhin auf, dass 78% der Teilnehmer in der ersten und 74% der Teilnehmer in der zweiten Stichprobe Prozentzahlen in 5er- oder 10er-Schritten vergaben. Prozentangaben in 1er-Schritten verwendeten 22% der Teilnehmer in der ersten und 26% in der zweiten Stichprobe.

Konfirmatorische Faktorenanalyse. Die Voraussetzung von intervallskalierten Daten und der Mindeststichprobengröße von 200 Personen war in der ersten Stichprobe gegeben. Die zweite Stichprobe lag mit 163 Teilnehmern unter dieser Empfehlung. Die Schiefe war bei allen Items außer SKSD1 und SKSD4 in beiden Stichproben rechtssteil (Tabelle 25). SKSD1 wies keine Abweichung in der Normalverteilung hinsichtlich der Schiefe in beiden Stichproben auf. SKSD4 war in der zweiten Stichprobe leicht linkssteil. Die Kurtosis fiel in beiden Stichproben für 15 von 24 Items steiler als die Normalverteilung aus. In der ersten Stichprobe zeigte SU8 die höchste Kurtosis mit 4.9, während in der zweiten Stichprobe Kri7 die größte Kurtosis mit 8.8 aufwies. Sieben Items zeigten eine leicht negative bis neutrale Kurtosis in beiden Stichproben. Die Schiefen und Kurtosen beider Stichproben lagen bis auf die Kurtosis des Items Kri7 in der zweiten Stichprobe innerhalb der Richtwerte von West et al. (1995). Die multivaria-

ten Kurtoses und die *critical ratios* überschritten in beiden Stichprobenden den z-Wert von 1.96, was gegen die multivariate Normalverteilung sprach. Bei der Berechnung des χ^2 -Wertes wurde daher die *Bollen-Stine-Bootstrap*-Korrektur des p-Wertes durchgeführt. Der Kolmogorov-Smirnov-Test mit Lilliefors-Korrektur wies alle Items, außer SKSD4 der zweiten Stichprobe, als nicht normalverteilt aus ($p \leq .001$).

Tabelle 25: *Multivariate Normalverteilung*

	Stichprobe 1				Stichprobe 2			
	Schiefe	c.r.	Kurtosis	c.r.	Schiefe	c.r.	Kurtosis	c.r.
SBSR5	-0.63	-7.11	-0.31	-1.74	-0.71	-3.68	-0.38	-0.98
SBSR7	-1.30	-14.68	3.18	17.98	-1.01	-5.26	1.72	4.47
SKSD1	0.04	0.47	-0.60	-3.40	-0.00	-0.02	-0.71	-1.84
SKSD4	0.03	0.37	-0.10	-0.58	0.14	0.71	-0.06	-0.16
SKSD10	-1.00	-11.34	1.03	5.82	-1.23	-6.43	1.03	2.68
SKSD11	-1.18	-13.34	1.78	10.08	-1.09	-5.65	0.66	1.71
Kri1	-0.52	-5.90	-0.05	-0.28	-0.70	-3.64	0.17	0.44
Kri2	-0.23	-2.65	-0.61	-3.48	-0.66	-3.42	-0.08	-0.21
Kri4	-0.45	-5.13	-0.01	-0.03	-0.64	-3.36	-0.06	-0.16
Kri7	-1.22	-13.75	1.99	11.29	-2.03	-10.55	8.54	22.25
Kri10	-1.18	-13.37	1.41	7.97	-1.37	-7.12	1.68	4.39
Kri12	-0.36	-4.11	-0.55	-3.12	-0.97	-5.06	1.16	3.02
R2	-0.83	-9.34	0.01	0.06	-0.83	-4.35	0.14	0.35
SU1	-0.75	-8.54	2.89	16.38	-0.73	-3.79	1.86	4.86
SU2	-0.75	-8.49	1.34	7.59	-0.75	-3.91	0.97	2.54
SU5	-0.02	-0.20	0.92	5.19	-0.14	-0.75	-0.76	-1.99
SU8	-1.62	-18.29	4.85	27.47	-1.45	-7.56	2.32	6.03
SU9	-0.29	-3.30	0.21	1.17	-0.38	-1.99	0.36	0.94
SU10	-1.51	-17.11	3.07	17.39	-1.80	-9.38	3.53	9.21
SU12	-1.19	-13.48	1.97	11.18	-1.36	-7.06	1.62	4.21
SU13	-0.29	-3.28	-0.23	-1.32	-0.52	-2.72	-0.33	-0.85
SU14	-0.90	-10.18	2.92	16.50	-0.82	-4.29	3.22	8.38
KF6	-1.15	-13.00	0.82	4.66	-1.37	-7.15	1.25	3.25
KF11	-1.34	-15.20	2.45	13.85	-1.41	-7.35	1.72	4.48
Multivariate			176.18	69.15			101.15	18.28

Anmerkung. $N_{SP1} = 769$; $N_{SP2} = 163$; c.r. = critical ratio.

Itembezeichnung: SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Die Testung auf multivariate Ausreißer wies 99 Personen in der ersten und 100 in der zweiten Stichprobe als Ausreißer aus. Da keine theoretischen Gründe für den Ausschluss der Ausreißer vorlagen, verblieben alle Werte in den Stichproben. Zumal steigt mit zunehmender Stichprobengröße die Wahrscheinlichkeit an, multivariate Ausreißer zu entdecken (Bühner, 2011). Die KFA wurde auf Basis des theoretisch angenommenen zweifaktoriellen Modells durchgeführt (Abbildung 21, S. 107). Das Modell konnte die Daten beider Stichproben nicht angemessen repräsentieren (Tabelle 26). Die χ^2 -Werte fielen signifikant aus und die *Modell-Fit* überschritten die empfohlenen Grenzwerte. Der *RMSEA* der zweiten Stichprobe lag innerhalb der empfohlenen Grenze von .08 für Stichproben mit weniger als 250 Personen, überschritt aber im 90%-igen Konfidenzintervall den oberen Grenzwert.

Tabelle 26: *Modelfit-Indizes*

	χ^2	df	p^a	CFI	RMSEA	CI90	SRMR
Stichprobe 1	1134.68	252	.001	.69	.07	.06-.07	.12
Stichprobe 2	449.32	230	.003	.57	.08	.07-.09	.12

Anmerkung. $N_{SP1} = 769$, $N_{SP2} = 163$; χ^2 = Chi-Quadrat-Wert, df = Freiheitsgrade; ^a p = Irrtumswahrscheinlichkeit mit Bollen-Stine Bootstrap-Korrektur; CI90 = 90%-iges Konfidenzintervall.

Die unstandardisierten Ladungen waren in der ersten Stichprobe signifikant, sodass die Voraussetzung für den lokalen *Modell-Fit* gegeben war (Tabelle 27). In der zweiten Stichprobe zeigten sechs Items keine signifikanten Ladungen. Die betroffenen Items repräsentierten überwiegend die intrapersonellen Kompetenzen im angenommenen Modell.

Die standardisierten Regressionskoeffizienten (β) beider Stichproben zeigten niedrige bis hohe Korrelationen zwischen den Items und dem intra- bzw. interpersonellen Faktor. Es traten keine Ladungen > 1 auf, sodass Schätzprobleme unwahrscheinlich waren (Bühner, 2011). Teilweise unterschieden sich die Regressionskoeffizienten eines Items in den zwei Stichproben (z.B. SBR5, SKSD11). Die Varianzaufklärung durch die zwei Faktoren war in beiden Stichproben gering ($R^2_{SP1} = .03-.34$; $R^2_{SP2} = .02-.46$) und ging mit hohen Fehlervarianzen einher ($1 - R^2_{SP1} = .66-.97$; $1 - R^2_{SP2} = .54-.98$). Insgesamt musste H2 über das zweifaktorielle Modell des *SJTs* verworfen werden. Die Faktorenstruktur des *SJTs* wurde daraufhin mit einer explorativen Faktorenanalyse weiter untersucht.

Tabelle 27: Regressionskoeffizienten und Kommunalitäten der KFA

Item	Faktor	b		β		R^2	
		SP1	SP2	SP1	SP2	SP1	SP2
SBSR5	Intrapersonelle	1.20**	4.45	.49	.15	.24	.27
SBSR7	Kompetenzen	1.00**	1.00	.48	.52	.23	.02
SKSD1		0.43**	1.24	.17	.15	.03	.02
SKSD4		0.46**	3.06	.25	.45	.06	.20
SKSD10		0.93**	2.67	.54	.42	.30	.18
SKSD11		0.84**	1.37	.40	.18	.16	.03
Kri1	Interpersonelle	1.78**	1.40**	.50	.35	.25	.12
Kri2	Kompetenzen	1.66**	1.65**	.49	.39	.24	.15
Kri4		1.00**	1.40**	.36	.36	.13	.13
Kri7		1.36**	-	.52	-	.27	-
Kri10		2.16**	1.99*	.57	.41	.32	.17
Kri12		1.00**	1.00	.38	.28	.14	.08
R2		0.79**	0.32*	.20	.07	.04	.00
SU1		0.64**	1.24**	.28	.44	.08	.19
SU2		0.33**	0.33*	.28	.24	.08	.06
SU5		0.28**	0.38*	.20	.22	.04	.05
SU8		1.21**	1.17**	.46	.37	.21	.14
SU9		0.89**	0.73*	.40	.25	.16	.06
SU10		1.46**	1.66**	.57	.46	.33	.22
SU12		1.26**	1.40**	.43	.38	.18	.14
SU13		1.71**	2.70**	.54	.65	.29	.42
SU14		0.44**	0.57*	.30	.27	.09	.08
KF6		2.66**	4.05**	.58	.68	.34	.46
KF11		1.82**	2.99**	.52	.66	.27	.43

Anmerkung. SP = Stichprobe; $N_{SP1} = 769$; $N_{SP2} = 163$; b = unstandardisierter Regressionskoeffizient; β = standardisierter Regressionskoeffizient; R^2 = Varianzaufklärung der Items durch den Faktor (Kommunalität).

Itembezeichnung: SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Explorative Faktorenanalyse. Als Voraussetzungsprüfung für die Durchführung der Faktorenanalyse wurden die Interkorrelationen der SJT-Items berechnet (Tabelle 28). Die Korrelationen zwischen den Items betragen in der ersten Stichprobe zwischen -.07 bis .45 und in der zweiten -.13 bis .61. Überwiegend lagen niedrige bis mittlere Interkorrelationen in beiden Stichproben vor. Die Items SKSD1, R2 und SU5 wiesen in

beiden Stichproben die niedrigsten Korrelationen zu den anderen Items auf. Sie zeigten überwiegend Nullkorrelationen bis zu vernachlässigende negative Zusammenhänge. Während dies in der ersten Stichprobe auch auf SKSD4 und SU1 zutraf, zeigten diese Items in der zweiten Stichprobe Korrelationen im niedrigen Bereich zu den anderen Items. In der zweiten Stichprobe wies SKSD11 mehr Nullkorrelationen auf als in der ersten. In beiden Stichproben korrelierten die zwei Kontaktfähigkeits-Items KF6 und KF11 am höchsten miteinander ($r_{SP1} = .45$ vs. $r_{SP2} = .61$), wobei der Zusammenhang in der zweiten Stichprobe höher ausfiel. Korrelationen um .40 lagen in beiden Stichproben für Item SU9 und SU10 vor, die von der Prüfungsvorbereitung handeln. In der zweiten Stichprobe zeigte zudem SU13 Korrelationen um .50 zu den zwei Kontaktfähigkeits-Items. Der in beiden Stichproben signifikant ausfallende Bartlett-Test bestätigte die Minimalanforderung an die Daten (SP1: $\chi^2(276) = 3060.26$; $p = .00$; SP2: $\chi^2(276) = 770.25$; $p = .00$). Der Kaiser-Meyer-Olkin-Koeffizient gab für die erste Stichprobe eine gute (.89) und für die zweite eine mittlere Eignung (.74) zur Faktorenanalyse an. Auf Itemebene (Diagonale der Tabelle 28) zeigten in der ersten Stichprobe dreizehn Items eine sehr gute ($MSA \geq .90$), sieben eine gute ($.80 \leq MSA \leq .89$) und vier eine mittlere ($.70 \leq MSA \leq .79$) Eignung für die Faktorenanalyse. Letztere Items entsprachen bis auf SU9 denjenigen Items, die geringe Korrelationen zu den anderen Items aufwiesen (SKSD1, R2, SU5). In der zweiten Stichprobe fiel die Eignung der Daten für eine Faktorenanalyse geringer aus. Zwei Items zeigten eine gute ($.80 \leq MSA \leq .89$), 15 eine mittlere ($.70 \leq MSA \leq .79$), drei eine mäßige ($.60 \leq MSA \leq .69$) und vier eine schlechte Eignung für die Faktorenanalyse ($.50 \leq MSA \leq .59$). Letztere entsprachen denjenigen Items, die viele Nullkorrelationen mit den anderen Items aufwiesen (SKSD1, R2, SU5, SKSD11).

Tabelle 28: Interkorrelationen und MSA-Koeffizienten der SJT-Items

Stichprobe 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
SBSR5 (1)	.91 ^a																									
SBSR7 (2)	.22	.90 ^a																								
SKSD1 (3)	.04	.16	.70 ^a																							
SKSD4 (4)	.22	.06	.00	.85 ^a																						
SKSD10(5)	.26	.27	.08	.13	.93 ^a																					
SKSD11 (6)	.19	.19	.09	.06	.22	.92 ^a																				
Kri1 (7)	.37	.14	.07	.16	.23	.14	.91 ^a																			
Kri2(8)	.25	.27	.03	.06	.34	.15	.24	.93 ^a																		
Kri4 (9)	.22	.23	.04	.12	.28	.14	.21	.21	.90 ^a																	
Kri7(10)	.30	.25	.12	.10	.30	.18	.25	.28	.15	.92 ^a																
Kri10 (11)	.39	.24	.08	.12	.33	.23	.28	.29	.27	.29	.93 ^a															
Kri12 (12)	.19	.23	.10	.04	.27	.19	.17	.20	.20	.17	.26	.89 ^a														
R2 (13)	.09	.18	.12	-.04	.11	.10	.07	.11	.02	.07	.16	.27	.75 ^a													
SU1 (14)	.07	.14	-.03	.02	.17	.08	.18	.17	.10	.08	.12	.14	.13	.84 ^a												
SU2 (15)	.14	.16	.12	.04	.16	.05	.14	.12	.16	.12	.19	.14	.03	.04	.90 ^a											
SU5 (16)	.11	.00	-.07	.07	.00	.05	.16	.00	.05	.12	.08	-.07	-.01	.00	.06	.72 ^a										
SU8 (17)	.24	.30	.12	.09	.27	.18	.21	.25	.18	.24	.26	.20	.06	.19	.10	.05	.89 ^a									
SU9 (18)	.20	.11	.14	.09	.17	.07	.22	.14	.14	.15	.24	.09	.02	.10	.17	.12	.29	.79 ^a								
SU10 (19)	.31	.22	.09	.09	.29	.16	.26	.28	.14	.35	.34	.19	.09	.13	.18	.18	.24	.42	.87 ^a							
SU12 (20)	.22	.19	.01	.09	.25	.19	.25	.24	.13	.26	.20	.19	.12	.17	.11	.05	.23	.13	.23	.93 ^a						
SU13 (21)	.31	.22	.06	.13	.27	.16	.29	.27	.18	.28	.25	.16	.09	.12	.15	.20	.26	.14	.33	.21	.92 ^a					
SU14 (22)	.13	.17	.09	.04	.22	.11	.15	.22	.19	.22	.21	.22	.13	.11	.09	-.07	.09	.04	.07	.16	.15	.87 ^a				
KF6(23)	.30	.19	.09	.10	.30	.14	.30	.25	.18	.29	.31	.22	.13	.20	.12	.14	.25	.20	.31	.24	.31	.16	.90 ^a			
KF11 (24)	.29	.12	.04	.16	.21	.08	.23	.22	.14	.26	.28	.14	.05	.08	.15	.19	.19	.17	.27	.19	.37	.10	.45	.86 ^a		

Fortsetzung Tabelle 28 Stichprobe 2:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
SBSR5 (1)	.82 ^a																								
SBSR7 (2)	.08	.79 ^a																							
SKSD1 (3)	.09	.04	.54 ^a																						
SKSD4 (4)	.26	-.08	.09	.68 ^a																					
SKSD10(5)	.22	.18	.00	.16	.78 ^a																				
SKSD11 (6)	.01	.10	.09	.14	.09	.55 ^a																			
Kri1 (7)	.32	.08	-.04	.29	.14	.10	.75 ^a																		
Kri2(8)	.27	.11	.09	.19	.31	-.02	.18	.79 ^a																	
Kri4 (9)	.21	.10	.14	.18	.17	.22	.17	.10	.78 ^a																
Kri7(10)	.16	.11	.23	.03	.15	.08	.17	.09	.24	.73 ^a															
Kri10 (11)	.30	.16	-.02	.13	.29	-.03	.26	.26	.11	.17	.83 ^a														
Kri12 (12)	.13	.14	.04	.21	.16	.14	.17	.27	.23	.04	.23	.74 ^a													
R2 (13)	.11	.10	.04	-.02	-.04	.01	.03	.09	-.07	.01	.20	.17	.55 ^a												
SU1 (14)	.23	.19	.01	.16	.10	.09	.22	.08	.24	.18	.23	.08	-.03	.78 ^a											
SU2 (15)	.04	.05	.08	.05	.01	-.07	.06	.05	.18	.08	-.03	-.08	-.03	.06	.62 ^a										
SU5 (16)	.15	.08	-.13	.07	-.06	.07	.06	-.03	-.03	-.07	.04	-.08	.07	.14	.15	.54 ^a									
SU8 (17)	.20	.15	.17	.04	.25	.20	.10	.20	.21	.24	.25	.19	-.02	.25	-.04	.02	.78 ^a								
SU9 (18)	-.01	.22	.24	-.01	.01	.16	.04	.01	.20	.04	.07	.02	.04	.21	.10	.12	.14	.67 ^a							
SU10 (19)	.18	.25	.16	.01	.21	.08	.19	.14	.17	.11	.21	.07	.07	.28	.17	.23	.18	.40	.78 ^a						
SU12 (20)	.22	.26	.12	.18	.24	.12	.21	.16	.15	.10	.18	.27	.05	.33	.13	.06	.12	.07	.15	.77 ^a					
SU13 (21)	.30	.11	.18	.14	.21	.09	.07	.29	.24	.34	.20	.12	.02	.20	.20	.27	.24	.14	.35	.17	.79 ^a				
SU14 (22)	.19	.21	.04	.07	.22	.05	.21	.35	.05	-.02	.21	.25	.16	.20	-.08	-.01	.08	.11	.15	.33	.07	.73 ^a			
KF6(23)	.27	.07	-.02	.12	.28	.15	.23	.18	.21	.25	.25	.10	-.02	.28	.20	.18	.16	.09	.26	.23	.48	.06	.78 ^a		
KF11 (24)	.38	.03	.06	.15	.31	-.01	.22	.25	.17	.22	.20	.14	-.01	.17	.17	.07	.25	.08	.19	.16	.51	.12	.61	.74 ^a	

Anmerkung. ^a Maß der Stichprobeneignung = MSA-Koeffizient. Fett markierte Werte zeigen die höchsten Interkorrelationen zwischen den Items an.

Für die Faktorenextraktion wurden drei Verfahren herangezogen. Nach dem Kaiser-Guttman-Kriterium sollten in der ersten Stichprobe fünf Faktoren mit Eigenwerten > 1 extrahiert werden, die zusammen 42% der Varianz in den Daten aufklärten. In der zweiten Stichprobe sollten acht Faktoren mit einer Varianzaufklärung von 59% extrahiert werden (Tabelle 29). In beiden Datensätzen klärte der erste Faktor den Großteil der Varianz auf, während die folgenden Faktoren kleine Varianzbeiträge lieferten.

Tabelle 29: Erklärte Gesamtvarianz nach dem Kaiser-Guttman-Kriterium

	Stichprobe 1						Stichprobe 2					
	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
	Gesamt	% der Varianz	Kum. %	Gesamt	% der Varianz	Kum. %	Gesamt	% der Varianz	Kum. %	Gesamt	% der Varianz	Kum. %
1	5.19	21.63	21.63	5.19	21.63	21.63	4.54	18.91	18.91	4.54	18.91	18.91
2	1.59	6.62	28.26	1.59	6.62	28.26	1.84	7.66	26.56	1.84	7.66	26.56
3	1.20	5.02	33.27	1.20	5.02	33.27	1.66	6.93	33.49	1.66	6.93	33.49
4	1.12	4.67	37.94	1.12	4.67	37.94	1.47	6.12	39.61	1.47	6.12	39.61
5	1.06	4.41	42.35	1.06	4.41	42.35	1.31	5.46	45.07	1.31	5.46	45.07
6	0.99	4.15	46.50				1.14	4.76	49.83	1.14	4.76	49.83
7	0.97	4.04	50.54				1.09	4.56	54.39	1.09	4.56	54.39
8	0.94	3.92	54.46				1.05	4.36	58.75	1.05	4.36	58.75
9	0.89	3.71	58.17				0.96	4.00	62.76			
10	0.86	3.58	61.75				0.88	3.67	66.43			
11	0.84	3.48	65.23				0.83	3.45	69.87			
12	0.82	3.43	68.67				0.81	3.38	73.23			
13	0.77	3.21	71.88				0.74	3.09	76.34			
14	0.75	3.11	74.99				0.70	2.93	79.27			
15	0.71	2.96	77.95				0.68	2.85	82.12			
16	0.70	2.93	80.88				0.67	2.81	84.92			
17	0.68	2.83	83.71				0.61	2.53	87.46			
18	0.64	2.68	86.39				0.55	2.31	89.76			
19	0.62	2.59	88.98				0.54	2.26	92.02			
20	0.60	2.49	91.47				0.47	1.97	93.99			
21	0.57	2.38	93.85				0.42	1.76	95.75			
22	0.53	2.22	96.07				0.41	1.69	97.44			
23	0.50	2.07	98.14				0.34	1.43	98.87			
24	0.45	1.86	100.00				0.27	1.13	100.00			

Anmerkung. Extraktionsmethode: Hauptkomponentenanalyse. Kum. = Kumuliert.

Nach dem Ergebnis der Parallelanalyse (Tabelle 30) sollten in der ersten Stichprobe zwei und in der zweiten vier Faktoren extrahiert werden, deren empirischer Eigenwerteverlauf über dem zufälligen Eigenwerteverlauf lag.

Tabelle 30: *Ergebnisse Parallelanalyse*

Faktor	Stichprobe 1		Stichprobe 2	
	empirische Eigenwerte	zufällige Eigenwerte	empirische Eigenwerte	zufällige Eigenwerte
1	5.19	>	1.33	4.54 > 1.77
2	1.59	>	1.28	1.84 > 1.64
3	1.20	<	1.24	1.66 > 1.54
4	1.12	<	1.21	1.47 > 1.46
5	1.06	<	1.18	1.31 < 1.39
6	0.99	<	1.15	1.14 < 1.32
7	0.97	<	1.12	1.09 < 1.25
8	0.94	<	1.10	1.05 < 1.19
9	0.89	<	1.07	0.96 < 1.14
10	0.86	<	1.05	0.88 < 1.08
11	0.84	<	1.02	0.83 < 1.03
12	0.82	<	1.02	0.81 < 0.98
13	0.77	<	1.00	0.74 < 0.93
14	0.74	<	0.98	0.70 < 0.89
15	0.71	<	0.95	0.68 < 0.84
16	0.70	<	0.93	0.67 < 0.8
17	0.68	<	0.91	0.61 < 0.75
18	0.64	<	0.89	0.55 < 0.71
19	0.62	<	0.87	0.54 < 0.66
20	0.60	<	0.85	0.47 < 0.62
21	0.57	<	0.82	0.42 < 0.58
22	0.53	<	0.80	0.40 < 0.53
23	0.50	<	0.77	0.34 < 0.48
24	0.45	<	0.74	0.27 < 0.43

Anmerkung. Die Unterschiede sind auf dem 5%-Niveau signifikant. Fett markierte Werte zeigen die Faktorenanzahl, bei der die empirischen Eigenwerte größer als die zufälligen Eigenwerte ausfielen.

Nach dem *MAP*-Test (Tabelle 31) sollte in beiden Stichproben ein Faktor extrahiert werden, da die mittlere quadrierte Partialkorrelation nach der Auspartialisierung von einem Faktor am geringsten ausfiel und ab der Auspartialisierung des zweiten Faktors wieder anstieg.

Tabelle 31: *Ergebnisse MAP-Test*

Faktor	Mittlere quadrierte Partialkorrelation	
	Stichprobe 1	Stichprobe 2
0	.0360	.0307
1	.0057	.0125
2	.0066	.0129
3	.0089	.0133
4	.0115	.0155
5	.0145	.0178
6	.0180	.0208
7	.0215	.0246
8	.0268	.0283
9	.0311	.0327
10	.0363	.0373
11	.0444	.0427
12	.0521	.0502
13	.0604	.0582
14	.0713	.0693
15	.0845	.0821
16	.0965	.1007
17	.1201	.1195
18	.1495	.1545
19	.1956	.1915
20	.2596	.2572
21	.3136	.3320
22	.4788	.5126
23	1.000	1.000

Anmerkung. Fett markierte Wert gibt die Faktorenanzahl mit der geringsten mittleren quadrierten Partialkorrelation an.

Zusammengefasst sollten nach dem Kaiser-Guttman-Kriterium fünf Faktoren in der ersten und acht Faktoren in der zweiten Stichprobe extrahiert werden. Die Parallelanalyse empfahl, zwei in der ersten und vier Faktoren in der zweiten Stichprobe zu extra-

hieren. Nach den *MAP*-Tests sollte für beide Stichproben ein Faktor extrahiert werden. Da die beiden letzteren Verfahren als objektiver gelten (Bühner, 2011), wurden für die weiteren Analysen zwei Faktoren für die erste Stichprobe und vier Faktoren für die zweite extrahiert. Anschließend wurden die einfaktoriellen Lösungen beider Stichproben analysiert. Für die Wahl der Methode wurde ein Vergleich der nicht reproduzierten Residuen der Hauptkomponenten-, Hauptachsen- und *ML*-Methode herangezogen (Tabelle 32). Die beobachtbaren empirischen Ladungen wurden in beiden Stichproben jeweils mit der Hauptachsen- und *ML*-Methode am besten abgebildet, da sie die geringsten Anzahlen an nicht reproduzierbaren Residuen aufwiesen. Deshalb wurden diese zur Faktorenextraktion verwendet.

Tabelle 32: *Prozent der nicht reproduzierten Residuen*

extrahierte Faktoren	Stichprobe 1			Stichprobe 2		
	HKA	HAA	ML	HKA	HAA	ML
4	-	-	-	51%	31%	31%
2	39%	14%	14%	-	-	-
1	44%	26%	26%	61%	58%	58%

Anmerkung. HKA = Hauptkomponentenanalyse; HAA = Hauptachsenanalyse; ML = Maximum-Likelihoodanalyse.

Für die erste Stichprobe wurden zunächst zwei Faktoren extrahiert. Die unrotierte zweifaktorielle Lösung zeigte für beide Methoden, dass der erste Faktor 18 von 24 Items repräsentierte (Tabelle 33). Für fünf Items fielen die Ladungen $< .30$ aus (SKSD1, SKSD4, SU1, SU2, R2). Item SU5 lud als einziges Item negativ auf den zweiten Faktor. Kri12 lud auf beiden Faktoren. Der Eigenwert des ersten Faktors fiel stärker aus als der des zweiten.

Nach der obliquen Rotation repräsentierte der erste Faktor elf Items bei der Hauptachsenanalyse und zehn bei der *ML*-Methode. Der zweite Faktor repräsentierte bei beiden Methoden zehn Items. Die Faktorladungen lagen zwischen $.30$ und $.63$, wobei SKSD4 die niedrigste und KF11 die höchste Ladung aufwiesen. Neben SKSD1, SU1 und SU2, die bereits bei der unrotierten Lösung Ladungen $< .30$ aufwiesen, zeigten nun auch SU8 und SU12 Ladungen unter $.30$. SKSD4 und R2 wiesen hingegen nach der Rotation Ladungen $\geq .30$ auf. Zwei Items luden auf beiden Faktoren (Kri10, SU5), wobei SU5 auf den ersten Faktor positiv und auf dem zweiten negativ lud. Der Eigenwert des ersten Faktors fiel höher aus als der des zweiten Faktors. Beide Faktoren korrelierten zu $.66$. Die Ergebnisse der Hauptachsen- und *ML*-Methode waren bis auf kleine Abweichungen deckungsgleich.

Tabelle 33: zweifaktorielle Lösungen Stichprobe 1

Faktor	unrotierte Lösung				Promax-rotierte Lösung			
	HAA		ML		HAA		ML	
	1	2	1	2	1	2	1	2
SBSR5	.55		.55		.50		.49	
SBSR7	.45		.44			.52		.52
SKSD1								
SKSD4					.30			
SKSD10	.56		.55			.47		.48
SKSD11	.33		.33			.32		.33
Kri1	.50		.50		.48		.47	
Kri2	.50		.50			.39		.40
Kri4	.39		.39			.30		.32
Kri7	.52		.52		.36		.36	
Kri10	.58		.58		.33	.31	.32	.32
Kri12	.41	.32	.41	.32		.60		.58
R2						.42		.41
SU1								
SU2								
SU5		-.40		-.40	.57	-.41	.58	-.42
SU8	.48		.47					
SU9	.38		.38		.43		.42	
SU10	.56		.56		.56		.55	
SU12	.43		.43					
SU13	.53		.53		.53		.54	
SU14	.32		.32			.47		.47
KF6	.56		.56		.49		.51	
KF11	.49		.49		.63		.65	
Eigenwert	9.91	0.72	9.90	0.72	5.20	3.39	4.87	3.42

Anmerkung. HAA = Hauptachsenanalyse; ML = Maximum-Likelihoodanalyse.

Itembezeichnung; SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Die zweifaktorielle Lösung konnte nicht nach dem theoretischen Modell interpretiert werden. Weder spiegelte der eine Faktor die intrapersonellen und der andere die interpersonellen sozialen Kompetenzen wieder, noch ließ sich die Lösung anders interpretieren. Daher wurde im nächsten Schritt ein Faktor gemäß dem MAP-Test für die erste

Stichprobe extrahiert. Zuvor wird jedoch die vierfaktorielle Lösung der zweiten Stichprobe beschrieben (Tabelle 34).

Tabelle 34: *Unrotierte vierfaktorielle Lösung Stichprobe 2*

Faktor	Hauptachsenanalyse				Maximum-Likelihood-Analyse			
	1	2	3	4	1	2	3	4
SBSR5	.51				.50			
SBSR7	.30		.32			.34		
SKSD1				.36				.46
SKSD4								
SKSD10	.46				.45			
SKSD11								
Kri1	.40				.37			
Kri2	.43				.41			
Kri4	.41				.38			
Kri7	.37			.32	.37			.32
Kri10	.46				.43			
Kri12	.34	.39				.32		
R2								
SU1	.45				.43			
SU2								
SU5				-.42				-.40
SU8	.42				.40			
SU9			.56				.51	
SU10	.49		.38		.47		.41	
SU12	.44				.40			
SU13	.61	-.35			.65			
SU14	.35	.45				.43		
KF6	.61	-.32			.68	-.33		
KF11	.61		-.34		.68	-.35		
Eigenwert	7.66	1.51	1.60	1.10	6.62	1.77	0.92	1.18

Anmerkung. Itembezeichnung: SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/ Selbstdisziplin; Kri = Kritikfähigkeit; R = Respekffähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontakffähigkeit.

In der unrotierten vierfaktoriellen Lösung der zweiten Stichprobe repräsentierte der erste Faktor 17 Items bei der Hauptachsenanalyse und 14 bei der ML-Methode. Die anderen drei Faktoren repräsentierte jeweils drei bis fünf Items. Für vier Items fielen die Ladungen $< .30$ aus (SKSD4, SKSD11, R2, SU2). Beide Methoden zeigten ein

ähnliches, aber nicht exakt übereinstimmendes Ergebnis. Der Eigenwert des ersten Faktors war höher als der der anderen Faktoren bei beiden Methoden.

Nach der obliquen Rotation (Tabelle 35) repräsentierte der erste Faktor acht und die drei anderen Faktoren zwischen vier bis sechs Items bei beiden Methoden. Die zwei Kontaktfähigkeits-Items zeigten jeweils die höchsten Ladungen. Drei Items wiesen Ladungen $< .30$ auf (SKSD4, SKSD11, R2).

Tabelle 35: *Promaxrotierte vierfaktorielle Lösung Stichprobe 2*

Faktor	Hauptachsenanalyse				Maximum-Likelihood-Analyse			
	1	2	3	4	1	2	3	4
SBSR5	.35	.33			.33	.31		
SBSR7			.38				.35	
SKSD1				.48				.58
SKSD4								
SKSD10	.39				.39			
SKSD11								
Kri1	.39				.42			
Kri2	.51				.48			
Kri4				.41				.33
Kri7				.39				.38
Kri10	.48				.48			
Kri12	.53				.53			
R2								
SU1							.31	
SU2		.33						
SU5		.34	.36	-.33			.41	-.32
SU8				.34				.31
SU9			.58				.57	
SU10			.56				.58	
SU12	.43				.45			
SU13		.63				.62		
SU14	.63				.62			
KF6		.71				.76		
KF11		.70				.77		
Eigenwert	3.71	3.04	1.88	1.95	3.70	2.46	2.22	1.92

Anmerkung. Itembezeichnung: SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/ Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Unterschiede in den Extraktionsmethoden traten für SU2 und SU5 auf. SU2 lud bei der Hauptachsenanalyse schwach auf dem zweiten Faktor, unterschritt aber in der ML-Methode die Ladungshöhe von .30. SU5 wurde in der Hauptachsenanalyse von drei und in der ML-Methode von zwei Faktoren repräsentiert. Zudem zeigte SU5 bei beiden Methoden positive und negative Ladungen zugleich. Dies deutet darauf hin, dass das Item teils gegensätzliche Aspekte misst. Der Eigenwert des ersten Faktors fiel etwas höher aus als bei den drei anderen Faktoren.

Tabelle 36 zeigt die Interkorrelation der vier extrahierten Faktoren. Die Korrelation des ersten, zweiten und vierten Faktors untereinander war höher als die mit dem dritten Faktor.

Tabelle 36: *Interkorrelation vierfaktorielle Lösung Stichprobe 2*

Faktor	1	2	3	4
1	-			
2	.41	-		
3	.22	.22	-	
4	.36	.32	.14	-

Die vierfaktorielle Lösung entsprach nicht dem angenommenen Modell. Es zeigte sich keine Unterteilung in intra- und interpersonelle soziale Kompetenzen, auch keine hinsichtlich der Item-Thematik, noch ließ sich die Lösung anders interpretieren. Daher wurde im nächsten Schritt ein Faktor gemäß dem MAP-Test extrahiert.

Die einfaktorielle Lösung repräsentierte 18 Items in der ersten und 16-17 Items in der zweiten Stichprobe, deren Ladungen zwischen .30 und .63 betragen (Tabelle 37). Die Ladungen der vier Items SKSD1, SU2, SU5 und R2 fielen in beiden Stichproben $< .30$ aus. In der ersten Stichprobe betraf dies zusätzlich SKSD4 und SU1, in der zweiten zusätzlich SKSD11, SBSR7 und SU9. Dabei handelte es sich um die Items, die überwiegend Nullkorrelationen bis zu vernachlässigende negative Korrelationen zu den anderen Testitems und eine mäßige bis schlechte Stichprobeneignung besaßen. Die mittlere Faktorladung fiel in der ersten Stichprobe etwas höher aus als in der zweiten (Eigenwert SP1 = 8.5 vs. Eigenwert SP2 = 7.2-7.5). Dies spiegelte sich auch in der Varianzaufklärung wieder, die in der ersten Stichprobe bei 18.3% und in der zweiten bei 15.7% lag.

Tabelle 37: *Einfaktorielle Lösungen Stichprobe 1 und 2*

	Stichprobe 1		Stichprobe 2	
	HAA	ML	HAA	ML
SBSR5	.55	.55	.51	.51
SBSR7	.44	.44		
SKSD1				
SKSD4			.30	
SKSD10	.55	.55	.46	.46
SKSD11	.33	.33		
Kri1	.50	.50	.41	.39
Kri2	.50	.50	.43	.42
Kri4	.39	.39	.40	.39
Kri7	.52	.52	.36	.37
Kri10	.59	.59	.46	.45
Kri12	.40	.40	.34	.31
R2				
SU1			.46	.44
SU2				
SU5				
SU8	.48	.47	.42	.41
SU9	.38	.38		
SU10	.56	.56	.46	.45
SU12	.43	.43	.44	.42
SU13	.53	.53	.59	.62
SU14	.32	.31	.33	.30
KF6	.56	.56	.59	.63
KF11	.48	.49	.58	.62
Eigenwert	8.51	8.50	7.54	7.19

Anmerkung. HAA = Hauptachsenanalyse; ML = Maximum Likelihood Methode.

Itembezeichnung: SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Für die Auswertung auf Skalenebene wurden alle Items bis auf die vier zusammengefasst, deren Ladungen in beiden Stichproben unter .30 lagen (SKSD1, SU2, SU5, R2). Aus inhaltlicher Sicht sprach für den Ausschluss von SKSD1, dass es als einziges Item eine quantitative statt qualitative Abstufung der Antwortoptionen besaß. So mussten die Teilnehmer entscheiden, ob sie trotz der notwendigen Prüfungsvorbereitung eine

Woche, zwei Wochen oder gar nicht in den Urlaub fahren. Dies könnte ein anderes Antwortverhalten als bei den anderen Items hervorgerufen haben. R2 war vermutlich unterrepräsentiert, da es als einziges Item für respektvolles Verhalten in den *SJT* einging. SU2 handelte von Finanzierungsproblemen im Studium. Dieses Thema war nicht eng mit den Thematiken der anderen Items verbunden, die eher den Umgang mit anderen und die Prüfungsvorbereitung fokussierten. Da es kein essenzielles Thema für das Medizinstudium darstellte, wurde es ausgeschlossen. SU5 sprach als einziges Item eine stark emotionale Komponente - die Angst vor dem Umgang mit Leichen im Fach Anatomie - an, die wichtig für das Medizinstudium ist, aber stark auf den emotionalen statt sozialen Umgang mit Menschen fokussierte. Die Items, die nur in einer der beiden Stichproben Ladungen unter .30 aufwiesen, verblieben im *SJT* (SKSD4, SKSD11, SBSR7, SU1, SU9). SKSD4, SKSD11 und SBSR7 stellten typische Interaktionen zwischen Studierenden und Patienten und die Reflexion klinischer Settings dar. Sie wurden daher auch aufgrund der wichtigen Thematik beibehalten. SU1 wurde als erstes Item im *SJT* möglicherweise als Aufwärm-Item genutzt.

7.5.2. Ergebnisse auf Skalenebene

Die folgenden Analysen beziehen sich auf die durch die Faktorenanalysen entstandene 20 Item umfassende *SJT*-Skala. Zunächst werden die deskriptiven Daten und Reliabilitäten der Skala betrachtet, bevor auf die Validität eingegangen wird.

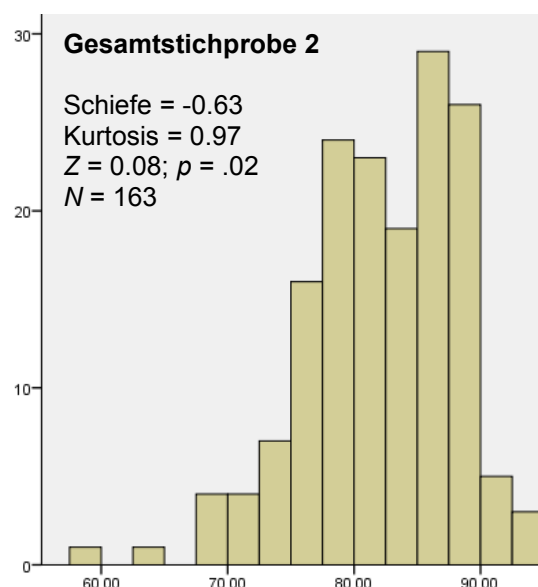
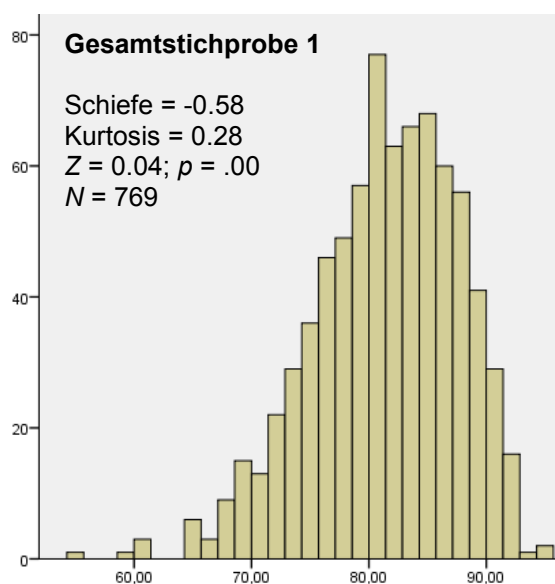
Deskriptive Statistik. Die Teilnehmer beider Stichproben erreichten im Mittel 81-82 Punkte auf der *SJT*-Skala (Tabelle 38). Von der theoretisch möglichen Ausnutzung von 20-100 Skaleneinheiten wurde in beiden Stichproben eher der obere Skalenbereich ausgenutzt. Den maximalen Punktwert von 100 erreichte kein Proband. Getrennt nach weiblichen und männlichen Studienteilnehmern erzielten in beiden Stichproben Frauen im Durchschnitt einen Punkt mehr als Männer. In der zweiten Stichprobe erzielten die Männer einen höheren Minimalwert verglichen mit den Frauen in beiden Stichproben und den Männern der ersten Stichprobe. Der M-Schätzer wich in beiden Stichproben vom Mittelwert zwischen 0.2 und 0.5 Punkten ab. Der Boxplot zeigte für die erste Stichprobe fünf Ausreißer in der Gesamtgruppe, vier in der weiblichen und drei in der männlichen Gruppe. In der zweiten Stichprobe wurden zwei Ausreißer für die Gesamtgruppe und zwei für die weiblichen Teilnehmer gekennzeichnet. Es bestand kein theoretischer Grund die Ausreißer auszuschließen.

Tabelle 38: Deskriptive Daten der SJT-Skala

	M		M-Schätzer ^a		SD		Min		Max	
	SP1	SP2	SP1	SP2	SP1	SP2	SP1	SP2	Sp1	SP2
ges	81.36	82.29	81.81	82.66	6.20	6.15	54.45	58.15	95.10	96.60
w	81.65	82.57	82.08	83.08	5.98	6.52	58.85	58.15	95.10	96.60
m	80.71	81.50	81.23	81.33	6.64	4.96	54.45	71.17	94.65	89.97

Anmerkung. SP = Stichprobe; ges. = gesamt, w = weiblich; m = männlich. $N_{SP1} = 769$; $n_{SP1 w} = 530$; $n_{SP1 m} = 239$; $N_{SP2} = 163$; $n_{SP2 w} = 120$; $n_{SP2 m} = 43$. ^a M-Schätzer nach Huber.

In beiden Stichproben war die Verteilung der SJT-Skala in der Gesamtgruppe sowie in den weiblichen und männlichen Teilgruppen eingipflig und leicht rechtssteil (Abbildung 23). Eine Ausnahme bildeten die männlichen Teilnehmer der zweiten Stichprobe, welche normalverteilt hinsichtlich der Schiefe waren. Die Gesamtstichprobe beider Datensätze, die männliche Stichprobe 1, sowie die weibliche Stichprobe 2 waren steiler als die Normalverteilung. Die Kurtosis der weiblichen Stichprobe 1 entsprach der Normalverteilung. Die Kurtosis der männlichen Stichprobe 2 verlief flacher als die Normalverteilung. Nach dem Kolmogorov-Smirnov-Test mit Lilliefors-Korrektur waren alle Gruppen bis auf die männliche Stichprobe 2 nicht normalverteilt.



Fortsetzung Abbildung 23:

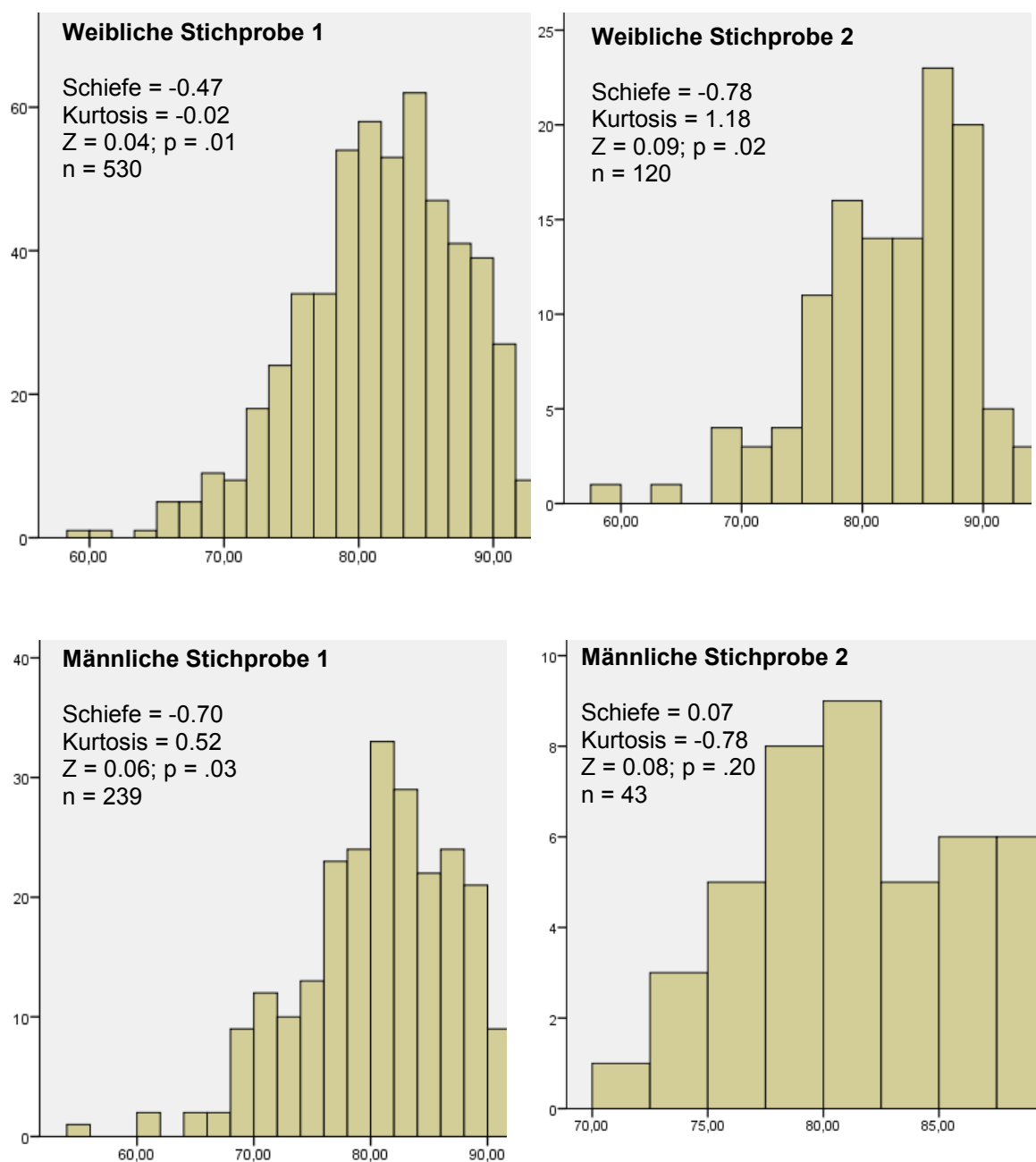


Abbildung 23. Verteilung der SJT-Skala.

Abszisse = Häufigkeit; Ordinate = Punktzahl im SJT.

Reliabilität. Die interne Konsistenz der SJT-Skala betrug .83 in der ersten und .81 in der zweiten Stichprobe. Interne Konsistenzen $\geq .70$ gelten als zufriedenstellend (Hossiep et al., 2000). Die Trennschärfen in der ersten Stichprobe lagen zwischen .21 und .53, die der zweiten zwischen .19 und .51 (Tabelle 39). Drei Items der ersten Stichprobe (SKSD4, SU1, SU14) und vier der zweiten (SBSR7, SKSD4, SKSD11, SU9) unterschritten das geforderte Mindestmaß von .30 (Backhaus et al., 2008). Die durchschnittliche Trennschärfe war in der ersten Stichprobe mit .41 höher als in der

zweiten mit .37. Bei Entfernung eines Items würde die interne Konsistenz in beiden Stichproben konstant bleiben oder um ein bis zwei Hundertstel absinken. Ausnahme bildete die Entnahme von SKSD4, welche zu einem Anstieg der internen Konsistenz um ein Hundertstel in der ersten Stichprobe führen würde.

Tabelle 39: *Trennschärfe und Itemschwierigkeit der SJT-Skala*

	r_{it}		α , falls Item entfernt		p_i	
	SP1	SP2	SP1	SP2	SP1	SP2
SBSR5	.51	.45	.82	.79	.79	.80
SBSR7	.39	.25	.83	.80	.84	.82
SKSD4	.21	.28	.84	.80	.62	.63
SKSD10	.51	.43	.82	.79	.88	.89
SKSD11	.30	.19	.83	.81	.83	.84
Kri1	.46	.39	.83	.80	.72	.75
Kri2	.46	.38	.83	.80	.80	.83
Kri4	.36	.38	.83	.80	.81	.82
Kri7	.47	.32	.83	.80	.86	.87
Kri10	.53	.41	.82	.79	.85	.87
Kri12	.36	.33	.83	.80	.85	.86
SU1	.25	.42	.83	.80	.78	.79
SU8	.44	.39	.83	.80	.89	.90
SU9	.33	.21	.83	.80	.73	.74
SU10	.50	.40	.82	.80	.88	.87
SU12	.40	.40	.83	.80	.88	.89
SU13	.48	.51	.82	.79	.72	.71
SU14	.29	.31	.83	.80	.85	.87
KF6	.51	.51	.82	.79	.83	.84
KF11	.44	.51	.83	.79	.84	.85

Anmerkung. SP = Stichprobe; r_{it} = korrigierte Trennschärfe; p_i = Itemschwierigkeit.

Itembezeichnung: SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Die Itemschwierigkeiten beider Stichproben lagen im Bereich von .60-.90. Dies bedeutet, dass die Items eher leicht zu lösen waren. Das leichteste Item war in beiden Stichproben SU8 und das schwierigste SKSD4. Die Itemschwierigkeit unterschied sich nicht in Abhängigkeit der Antwortanzahl pro Item (SP1: $F = 0.06$; $p = .94$; SP2: $F = 0.17$; $p = .84$). So zeigten Items mit vier Antworten ($N = 4$) eine mittlere Schwierigkeit von .81 in

beiden Stichproben, Items mit fünf Antworten ($N = 9$) eine von .82 in beiden Stichproben und Items mit sechs Antworten ($N = 7$) eine von .80 in der ersten und eine von .81 in der zweiten Stichprobe. Für sieben Items der ersten und fünf Items der zweiten Stichprobe traf der empfohlene Schwierigkeitsbereich für Leistungstests zwischen .20 und .80 (Fisseni, 2004) zu. Die mittlere Schwierigkeit von .81 in der ersten und .82 in der zweiten Stichprobe überschritt den empfohlenen Bereich zwischen .50 und .70 (Kelava & Moosbrugger, 2008). Da die Trennschärfen und Schwierigkeiten Empfehlungen, aber keine strengen Regeln zum Ausschluss von Items darstellen (Bühner, 2009), verblieben alle 20 Items im *SJT*, zumal keine inhaltlichen Gründe für einen Ausschluss vorlagen.

Validität. Zur Analyse der Zusammenhänge wird zunächst die deskriptive Statistik der Validierungskriterien beschrieben (Tabelle 40). Die Teilnehmer beider Stichproben beurteilten die Ausprägung ihrer sozialen Kompetenzen höher als bei Gleichaltrigen. Dabei wurde die Ausprägung soziale Unterstützung aufzusuchen am geringsten und die Respektfähigkeit am höchsten bewertet.

Tabelle 40: *Deskriptive Statistik der Validitätskriterien*

	N		M		SD		Min		Max	
	SP1	SP2	SP1	SP2	SP1	SP2	SP1	SP2	SP1	SP2
Selbstbeobachtung/ Selbstreflexion	735	151	5.32	5.42	0.98	1.04	2	3	7	7
Selbstkontrolle/ Selbstdisziplin	726	149	5.29	5.22	1.20	1.27	2	2	7	7
Kritikfähigkeit	747	157	4.97	5.02	1.01	1.12	2	2	7	7
Respekt	734	155	5.70	5.75	0.99	1.04	2	3	7	7
Suche nach sozialer Unterstützung	713	149	4.48	4.54	1.24	1.27	1	2	7	7
Kontaktfähigkeit	731	149	4.79	4.73	1.36	1.43	1	2	7	7
Allgemeine soziale Kompetenzen	712	147	5.17	5.10	1.02	1.19	1	1	7	7
Abiturnoten- durchschnitt	714	99	1.74	1.85	0.58	0.62	0.8	1	3.5	3.7
TMS	352	19	108.09	111.47	11.43	11.17	75	82	136	124

Anmerkung. Skala der sozialen Kompetenzen von 1 = geringer als bei Gleichaltrigen bis 7 = höher als bei Gleichaltrigen. SP = Stichprobe. TMS-Werte entsprachen T-Werten.

Der Abiturnotendurchschnitt lag in beiden Stichproben im sehr guten bis guten Bereich und streute von sehr guten bis befriedigenden Werten. Die Teilnehmer beider Stichproben erzielten durchschnittliche bis überdurchschnittliche TMS-Werte mit einer Streuung von weit unterdurchschnittlichen bis weit überdurchschnittlichen Werten.

Die *SJT*-Skala korrelierte gering bis mittel mit allen Selbsteinschätzungsskalen der sozialen Kompetenzen mit Ausnahme der Respektfähigkeit in der ersten Stichprobe (Tabelle 41). Die selbsteingeschätzte Kontaktfähigkeit wies gefolgt von der Einschätzung der allgemeinen sozialen Kompetenzen die höchste Korrelation auf. Die Korrelationen blieben nach der Bonferroni-Holm-Korrektur signifikant. In der zweiten Stichprobe zeigten sich niedrige Korrelationen mit der Suche nach sozialer Unterstützung, mit der Kontaktfähigkeit und den allgemeinen sozialen Kompetenzen, welche nach der Alpha-Fehler-Adjustierung ihre Signifikanz verloren. In der ersten Stichprobe traten mehr hypothesenkonforme Zusammenhänge auf als in der zweiten. H2 kann somit nur bedingt bestätigt werden. Zwischen der *SJT*-Skala, dem Abiturnotendurchschnitt und dem TMS bestanden Nullzusammenhänge bis niedrige Korrelationen in beiden Stichproben. Somit konnte H3 bestätigt werden.

Tabelle 41: *Korrelationen der SJT-Skala mit Validierungskriterien und Kontrollvariablen*

	N		r		r _s	
	SP1	SP2	SP1	SP2	SP1	SP2
Selbstbeobachtung/Selbstreflexion	735	151	.15**	.03	.14**	.03
Selbstkontrolle/Selbstdisziplin	726	149	.12**	.09	.10**	.05
Kritikfähigkeit	747	157	.13**	.10	.13**	.07
Respekt	734	155	.06	.13	.05	.06
Suche nach sozialer Unterstützung	713	149	.15**	.19*	.15**	.18*
Kontaktfähigkeit	731	149	.33**	.20*	.28**	.16*
Allgemeine soziale Kompetenzen	712	147	.27**	.21*	.21**	.13
Abiturnotendurchschnitt	714	99	.11**	.01	.11**	.04
TMS	364	19	-.13*	-.05	-.11*	-.13
Alter	765	163	.17**	.04	.18**	.07

Anmerkung. SP = Stichprobe; r_s = Korrelationskoeffizient nach Spearman; * p ≤ .05; ** p ≤ .001. Fett markierte Korrelationen blieben nach Bonferroni-Holm-Korrektur signifikant.

In der ersten Stichprobe korrelierte die *SJT*-Skala gering mit dem Alter, nicht jedoch in der zweiten. Daraufhin wurde der Zusammenhang zwischen dem *SJT* und dem Alter in der ersten Stichprobe genauer untersucht. Deskriptiv stieg die Punktzahl über die Al-

tersgruppen bis 29 Jahre an und sank bei den 30-54 Jährigen wieder ab (Tabelle 42). Eine Varianzanalyse zeigte signifikante Unterschiede zwischen den Altersgruppen ($F(5, 759) = 9.02; p = .00$) und der Punktzahl im *SJT*, wobei lediglich 6% der Variation im *SJT* auf das Alter zurückzuführen war. Aufgrund der geringen praktischen Bedeutsamkeit wurden keine weiteren Post-Hoc-Tests zwischen den Altersgruppen durchgeführt.

Tabelle 42: *SJT-Punktzahl abhängig vom Alter in Stichprobe 1*

Alter	N	M	SD
15-17	55	79.11	6.84
18-20	511	80.74	6.15
21-23	120	82.76	5.46
24-26	49	85.06	5.29
27-29	15	85.15	5.12
>30	15	83.35	7.45

Frauen erzielten im *SJT* durchschnittlich einen Punkt mehr als Männer (SP1: 81.7 vs. 80.7; SP2: 82.6 vs. 81.5). Für den Unterschiedstest wurde der t-Wert für gleiche Varianzen verwendet, da der F-Test nach Fisher signifikant ausfiel (SP1: $F(767) = 2.8; p = .10$; SP2: $F(161) = 2.4, p = .12$). In der ersten Stichprobe konnte ein signifikanter Unterschied zwischen den Geschlechtern ($t(767) = 1.95, p = .052$) festgestellt werden, nicht aber in der zweiten ($t(161) = 0.98, p = .33$). Die alternative Prüfung mit dem nicht parametrischen Kolmogorov-Smirnov-Test zeigte keine signifikanten Unterschiede in der *SJT*-Punktzahl von Frauen und Männern ($p_{SP1} = .19, p_{SP2} = .16$). Aufgrund der überwiegend nicht signifikanten Ergebnisse und dem nur knapp signifikanten Ergebnis des parametrischen Verfahrens in der ersten Stichprobe wurde H_4 abgelehnt.

Hinsichtlich der Nützlichkeit und Akzeptanz konnte festgestellt werden (Abbildung 24), dass 77% der Teilnehmer der ersten und 72% der zweiten Stichprobe die sozialen Anforderungen bewusst geworden sind, die das Medizinstudium an sie stellt; (N SP1 = 597; N SP2 = 117). Demzufolge konnte H_6 bestätigt werden. Zum Nachdenken über die eigenen sozialen Kompetenzen angeregt wurden 57% der Teilnehmer beider Stichproben (N SP1 = 448; N SP2 = 94), sodass H_7 nur für knapp zwei Drittel der Teilnehmer belegt werden konnte. Den *SJT* weiterempfehlen würden 68% der Teilnehmer in der ersten und 71% in der zweiten Stichprobe (N SP1 = 534; N SP2 = 115). Somit konnte H_5 bestätigt werden. In ihrer Entscheidung verstärkt Medizin zu studieren, wurden 61% der Personen beider Stichproben (N SP1 = 473; N SP2 = 98). Neun

Prozent der Teilnehmer der ersten und 14% der zweiten Stichprobe wurden abgeschreckt, ein Medizinstudium aufzunehmen (N SP1 = 69; N SP2 = 22). Die absoluten Häufigkeitsangaben für die Bewertung der Nützlichkeit und Akzeptanz können bei Bedarf dem Anhang G-2 entnommen werden.

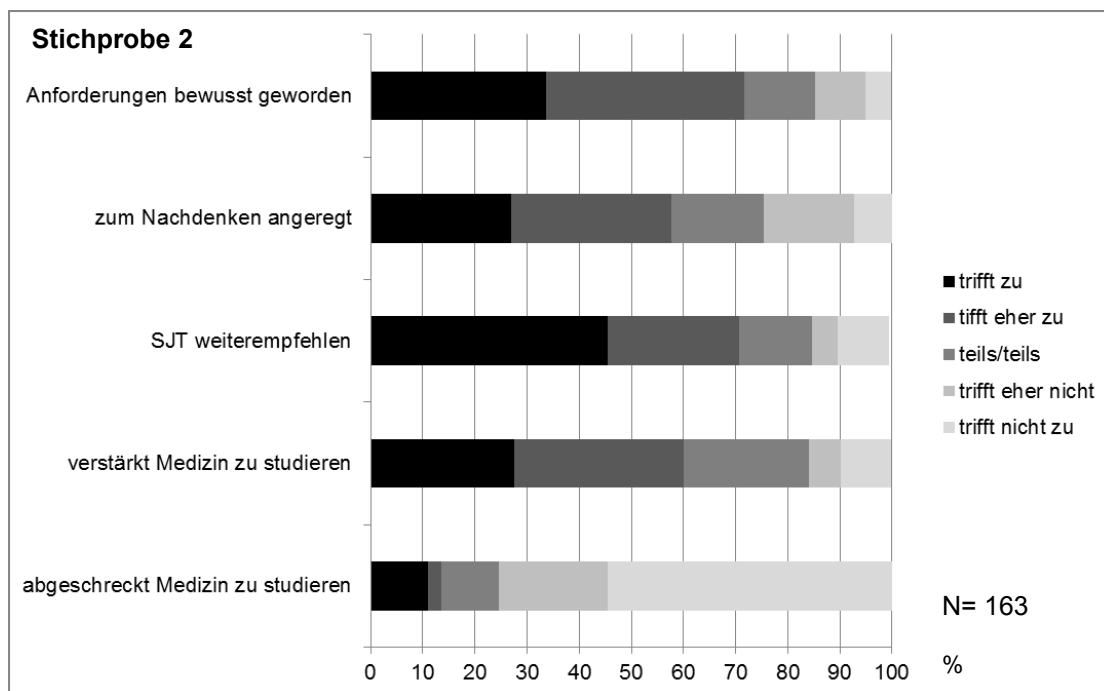
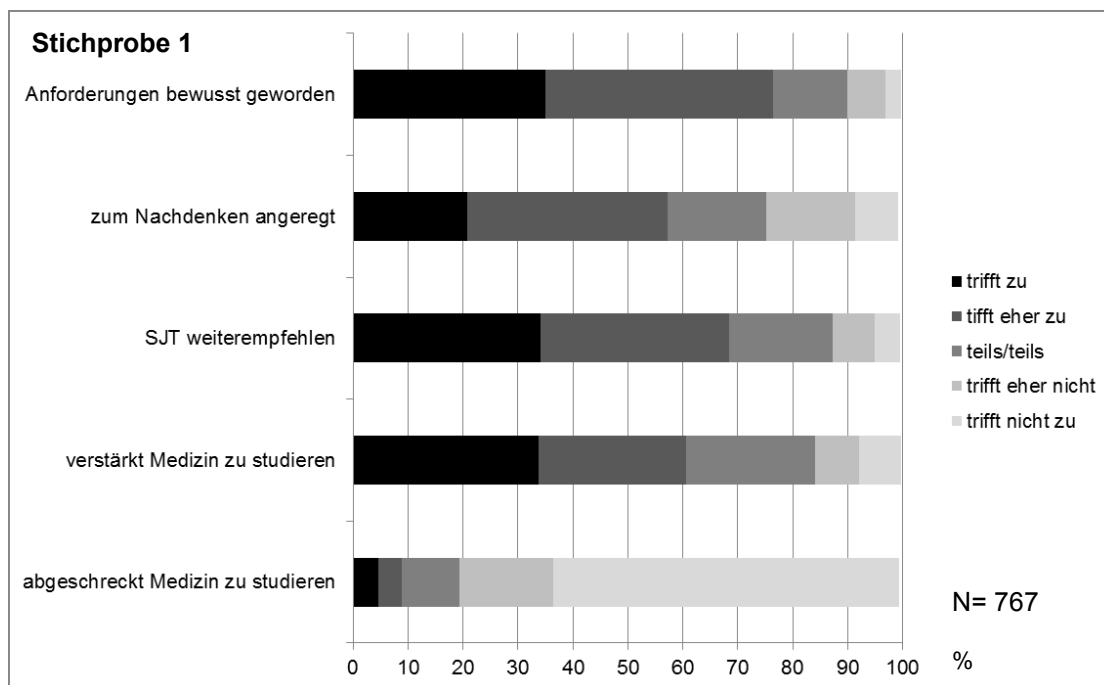


Abbildung 24. Nützlichkeit und Akzeptanz des SJTs.

Darüber hinaus wurde analysiert, ob die Personen, die sich bestärkt fühlten Medizin zu studieren, höhere Punktwerte im *SJT* aufwiesen. Dafür wurden die Kategorien *trifft zu* und *trifft eher zu* zu der Gruppe der Bestärkten ($N_{SP1} = 469$; $N_{SP2} = 98$) und die Kategorien *trifft eher nicht zu* und *trifft nicht zu* zu der Gruppe der Nicht-Bestärkten ($N_{SP1} = 119$; $N_{SP2} = 26$) zusammengefasst. Die Gruppe der Bestärkten zeigte in beiden Stichproben einen höheren *SJT*-Wert als die Nicht-Bestärkten ($M_{SP1} = 82.0$ vs. 80.4 ; $M_{SP2} = 82.8$ vs. 81.5). Der Unterschied war in der ersten Stichprobe signifikant, aber zeigte nur eine geringe praktische Bedeutsamkeit ($F(1,762) = 6.3$; $p = .01$; $\eta^2_{SP1} = .01$). In der zweiten Stichprobe trat kein signifikanter Unterschied zwischen den bestärkten und nicht-bestärkten Teilnehmern auf ($F(1,161) = 0.8$; $p = .36$). Gegensätzlich dazu wurde getestet, ob Personen, die abgeschreckt wurden Medizin zu studieren, niedrigere Punktwerte im *SJT* erzielten. Dafür wurden die Kategorien *trifft zu* und *trifft eher zu* zu der Gruppe der Abgeschreckten ($N_{SP1} = 67$; $N_{SP2} = 22$) und die Kategorien *trifft eher nicht zu* und *trifft nicht zu* zu der Gruppe der Nicht-Abgeschreckten ($N_{SP1} = 616$; $N_{SP2} = 123$) zusammengefasst. Die Gruppe der Abgeschreckten zeigte in beiden Stichproben einen um etwa einen Punkt niedrigeren *SJT*-Wert als die Gruppe der Nicht-Abgeschreckten ($M_{SP1} = 80.5$ vs. 81.8 ; $M_{SP2} = 81.5$ vs. 82.5). Der Unterschied in den Gruppen war weder in der ersten ($F(1,762) = 2.5$; $p = .12$), noch in der zweiten Stichprobe signifikant ($F(1,161) = 0.5$; $p = .46$). H_8 muss somit verworfen werden.

In der ersten Stichprobe wurde zusätzlich das schriftliche Feedback der Teilnehmer zum *SJT* ausgewertet. Insgesamt gaben 20% der Teilnehmer ($N = 153$) eine Rückmeldung. Tabelle 43 fasst die Aussagen geordnet nach positiver und negativer Kritik zusammen.

Tabelle 43: *Qualitative Rückmeldungen zum SJT*

Aspekt	Positive Rückmeldung	Negative Rückmeldung
Videos:	<ul style="list-style-type: none"> - qualitativ hochwertig, realitätsnah, anschaulich, authentisch ($N = 63$) - geben Einblicke in das Studium und Alltagsleben von Medizinstudierenden ($N = 34$) - bilden wichtige Inhalte aus Medizinstudium ab ($N = 13$) - Ermöglichen es, sich in die Hauptperson hineinzusetzen ($N = 11$) 	<ul style="list-style-type: none"> - Akustik verbessern ($N = 5$)

Fortsetzung Tabelle 43:

Aspekt	Positive Rückmeldung	Negative Rückmeldung
Instruktion:	- Prozentvergabe ermöglichen differenziertes und realistisches Antworten ($N = 13$)	- Prozentvergabe umständlich ($N = 7$) - Bevorzugung von anderen geschlossenen ($N = 8$) oder freien ($N = 7$) Antwortformaten - Zweifelhafte Aussagekraft ($N = 2$)
Antwortoptionen:	- präferierte Reaktion gut abgebildet ($N = 4$) - "guten" Reaktionen nicht leicht zu erkennen ($N = 4$)	- präferierte Reaktion nicht abgebildet ($N = 8$) - "guten" Reaktionen leicht zu erkennen ($N = 6$) - undifferenzierte Antworten ($N = 6$) - widersprüchliche Antworten ($N = 1$) - Antwort sollte mehrere Handlungsschritte enthalten ($N = 1$)
Allgemein:	- gute Idee ($N = 16$) - klar strukturiert und verständlich aufgebaut ($N = 8$)	- Testkürzung ($N = 20$) - Testverlängerung ($N = 14$) - Trennung nach Personen mit vs. ohne medizinische Ausbildung ($N = 5$) bzw. Human- vs. Zahnmediziner ($N = 2$) - technische Fehler bei der automatischen Rückmeldefunktion ($N = 4$) - Rechtschreibfehler ($N = 2$)
Effekt:	- regt zur Selbstreflexion an ($N = 21$)	

Die qualitativen Aussagen stützen die quantitativen Ergebnisse zur Akzeptanz des *SJT*s (H5), seiner informierenden Funktion (H6) und dass er in der Lage ist, die Teilnehmer zum Reflektieren der eigenen sozialen Kompetenzen anzuregen (H7).

7.6. Diskussion

In den folgenden Abschnitten wird zuerst auf Einschränkungen hinsichtlich der Stichproben eingegangen, bevor die Deskriptiva und Reliabilität des *SJT*s diskutiert werden. Darauf folgt die Auseinandersetzung mit den Ergebnissen zur Validität. Diese werden in der Reihenfolge der Forschungshypothesen kommentiert.

Stichproben. Für die Studie wurden zwei Stichproben von Studieninteressenten und -bewerbern herangezogen. Die zweite Stichprobe wurde als Kreuzvalidierung untersucht. Die Stichproben waren im Geschlechteranteil vergleichbar. Die Geschlechterverteilung war zudem für die Grundgesamtheit der Studienplatzbewerber an der Medizinischen Fakultät repräsentativ. In der zweiten Stichprobe lag ein höherer Anteil an 15-17-Jährigen vor. Dies könnte darin begründet liegen, dass nach Abschluss der Bewerbungsphase für das Wintersemester 2013/14 vor allem Personen teilgenommen haben könnten, die sich noch in der Schulausbildung der Oberstufe befanden. Hier könnte ein systematischer Unterschied zwischen den Stichproben bestehen. Bei den Teilnehmern der ersten Stichprobe war vermutlich das Vorhaben, sich zu bewerben, konkreter ausgeprägt als bei den Teilnehmern der zweiten Stichprobe. Ein Unterschied lag in der Teilnehmeranzahl, die in der ersten Stichprobe um das 4.7 fache größer war als in der zweiten Stichprobe. Der Stichprobenumfang der ersten Stichprobe kann nach MacCallum et al. (1999) für die Faktorenanalysen als sehr gut und der Umfang der zweiten als ausreichend bis zufriedenstellend gewertet werden. Die Mindestteilnehmeranzahl für die Korrelations- und Unterschiedsanalysen war in beiden Stichproben nach den poweranalytischen Berechnungen gegeben. Positiv zu werten ist, dass die Untersuchung an der Zielgruppe der Studienplatzbewerber und nicht an Stelleninhabern durchgeführt wurde, wie es meist der Fall ist (McDaniel & Nguyen, 2001).

Antwortverhalten. Die Verteilung der Wahrscheinlichkeit auf mehrere Antwortoptionen wurde häufig genutzt. So wurde in nur 10-12% der Fälle nur eine Antwortoption ausgewählt. In den Freitexten äußerten sich die Teilnehmer widersprüchlich zur Instruktionsform. Teilweise wurde die Vergabe von Prozenten aufgrund des differenzierten und realistischeren Antwortens bevorzugt, teilweise wurde es als umständlich erlebt. Die Instruktionsform förderte die Prozentvergabe in 5er- und 10er-Schritten, wobei ein Drittel der Teilnehmer auch 1er-Schritte verwendete. Trotz automatischer Anzeige bereits vergebener Prozente, lässt dies vermuten, dass die Probanden im Kopf mitgerechnet haben. Potenzial besteht in der Erforschung von weiteren Instruktionsversionen wie einer freien Antworteingabe oder wissensbasierter Formen. In der nächsten Studie soll das Augenmerk zunächst jedoch stärker auf die Konstruktvalidierung des *SJTs* gelegt werden.

Faktorenanalysen. Die erste Stichprobe war aufgrund ihrer Stichprobengröße besser für eine Faktorenanalyse geeignet als die zweite. Die Ergebnisse der zweiten Stichprobe sollten daher mit höherer Vorsicht interpretiert werden. Beide Stichproben konnten das angenommene zweifaktorielle Modell nicht bestätigen. Die aufgeklärte

Varianz fiel in beiden Stichproben gering und die Fehlervarianz groß aus. Schon während der Testentwicklung fiel es schwer, Situationen und insbesondere Antwortoptionen zu konstruieren, die nur eine soziale Kompetenz wiedergaben. Oftmals überlappen sich soziale Kompetenzen. Wenn z.B. eine Person die Hilfe einer anderen Person aufsucht, muss sie zwangsläufig die Person ansprechen, was einen Teil der Kontaktfähigkeit wiedergibt. Hier spiegelt sich vermutlich die Schwierigkeit einer klaren Abgrenzung sozialer Kompetenzen wieder, die auch bei der Definition und Modellbildung sozialer Kompetenzen auftreten (Darden & Gintner, 1996; Döpfner et al., 1981; Kauffeld et al., 2002; Mathews et al., 2002).

Die Ergebnisse der Verfahren zur Faktorenextraktion in der explorativen Faktorenanalyse variierten stark. Lediglich der als objektiv geltende *MAP*-Test (Bühner, 2011) empfahl für beide Stichproben einen Faktor zu extrahieren. Die Faktoren klärten einen geringen Varianzanteil auf. Dies ist nicht verwunderlich, da überwiegend niedrige bis mittlere Korrelationen zwischen den Items vorlagen. Die einfaktoriellen Lösungen waren in beiden Stichproben sinnhafter zu interpretieren als die vorgeschlagenen mehrfaktoriellen Lösungen der anderen Verfahren. Einfaktorielle Lösungen mit geringer Varianzaufklärung scheinen typisch für *SJTs* zu sein (Gillespie et al., 2002; Schmitt & Chan, 2006). Somit kann nach denselben Autoren auch für den vorliegenden *SJT* von einer Art Generalfaktor ausgegangen werden, aber der Großteil der Varianz wird durch viele andere spezifische Faktoren erklärt.

Reliabilität. Die interne Konsistenz der *SJT*-Skala lag in beiden Stichproben mit .83-.84 über dem geforderten Mindestmaß von .70 (Hossiep et al., 2000). Somit kann die Skala zur Individualdiagnostik eingesetzt werden. Weekley und Jones (1997) berichteten für schriftliche *SJTs* mit 20 Items interne Konsistenzen zwischen .29 und .32. Clewenger et al. (2001) gaben für einen *SJT* mit 33 Items ein *Cronbachs alpha* von .63 an. Die interne Konsistenz des vorliegenden 20 Item umfassenden *SJTs* kann somit als hoch interpretiert werden. Auch verglichen zu internen Konsistenzen von .69 für verhaltensbasierte schriftliche *SJTs* mit kontinuierlichem *Rating* schnitt der *SJT* gut ab (Polyhart & Ehrhart, 2003) und lässt den Schluss nahe, dass der *SJT* ein relativ einheitliches Konstrukt erfasst. Jedoch könnte die hohe interne Konsistenz auch auf hochkorrelierende Messfehler zurückgeführt werden, da hohe Fehlervarianzen bei der konfirmatorischen Faktorenanalyse vorlagen.

Bis auf drei Items der ersten und vier der zweiten Stichprobe erreichten die Items akzeptable Trennschärfen ($\pi = .19-.51$) über dem Mindestmaß von .30 (Backhaus et al., 2008). Bei den Items, die unterhalb des Mindestmaßes lagen, handelte es sich überwiegend um Items, die das Konstrukt Selbstkontrolle/Selbstdisziplin und Suche

nach sozialer Unterstützung wiedergeben sollten. Da die Entnahme der Items jedoch nicht zu einer wesentlichen Steigerung der internen Konsistenz geführt hätte, verblieben sie im *SJT*. Die teils niedrigen Trennschärfen spiegelten sich in den hohen Itemschwierigkeiten wieder. Die Items waren in beiden Stichproben leicht zu lösen. Die rechtssteilen und steiler als die Normalverteilung ausfallenden Daten haben wahrscheinlich die Trennschärfen und Schwierigkeiten beeinflusst (Lienert & Raatz, 1998). Die Itemschwierigkeit war unabhängig von der Antwortanzahl pro Item. Somit scheint die Antwortanzahl kein Ablenkreuz zu sein, der die Item-Beantwortung erschwert. Forschungsbefunde zu Itemschwierigkeiten und Trennschärfen von *SJTs* werden nur vereinzelt berichtet. So zeigte ein schriftlicher *SJT* für die medizinische Studierendenauswahl Itemschwierigkeiten von .55-.99 und Trennschärfen von .13-.39 (Manhal et al., 2012). Für den videobasierten *SJT* für Kundenorientierung (Möller, 2010) wurden in einer Studie an Auszubildenden im Hotelfach Trennschärfen von .13-.47 und Schwierigkeiten von .05-.79 mit einer mittleren Schwierigkeit von .54 berichtet. Somit fielen die Trennschärfen des vorliegenden *SJTs* besser als in den zwei berichteten Studien aus. Die Schwierigkeiten des vorliegenden *SJTs* können zwischen beiden Studien eingeordnet werden. Hohe Itemschwierigkeiten und niedrige Trennschärfen scheinen nicht untypisch für Messinstrumente sozialer Kompetenzen zu sein (Frey & Balzer, 2003; Lang, 2008). Zudem sollte bedacht werden, dass *SJTs* aufgrund der Konstruktheterogenität bessere Kriteriumsvaliditäten als Reliabilitäten zeigen (Kanning et al., 2006; Whetzel & McDaniel, 2009). Um die Itemschwierigkeit zu reduzieren, könnte die Attraktivität der Antwortoptionen gesteigert werden, die von mehr als 90% der Teilnehmer nie ausgewählt wurden.

Validität. Eine weitere Schwäche liegt in der Messung der konvergenten Validität, die über die Korrelation zu globalen nicht-standardisierten Selbsteinschätzungsskalen erhoben wurde, obwohl dieses Vorgehen methodisch als eher schwach einzuordnen ist. Ursprünglich sollte neben diesen das standardisierte Inventar Sozialer Kompetenzen (ISK, Kanning, 2009) eingesetzt werden. Aufgrund des Lizenzschutzes durfte dieses Verfahren jedoch online nur über das Hogrefe-Test-System durchgeführt werden, was eine finanzielle Grenze für das Projekt darstellte. Der ISK soll daher in der als Vorort-Testung geplanten Folgestudie nach Einwerbung entsprechender Drittmittel eingesetzt werden.

Die *SJT*-Skala korrelierte mit Ausnahme der Respektfähigkeit gering bis mittel mit den Selbsteinschätzungsskalen der sozialen Kompetenzen in der ersten Stichprobe. Zwischen der selbsteingeschätzten Kontaktfähigkeit und der allgemeinen sozialen Kompetenz bestanden die höchsten Korrelationen mit der *SJT*-Skala. In der zweiten

Stichprobe zeigte sich ein ähnliches Korrelationsmuster, nur fielen die Korrelationen geringer aus und waren nach der Alpha-Fehler-Korrektur nicht mehr signifikant. Dies könnte auf die geringere Stichprobengröße zurückzuführen sein. Verglichen mit Kriteriumsvaliditäten um .30 zu Persönlichkeitsfaktoren und sozialen Fähigkeiten fielen die Korrelationen zu den Selbsteinschätzungsskalen kleiner aus (z.B. Christian et al., 2010; McDaniel & Nguyen, 2001). Der Grund für die Nullkorrelation der *SJT*-Skala zur Respektfähigkeit könnte darin liegen, dass die Items zu Respektfähigkeit aufgrund der faktorenanalytischen Untersuchungen verworfen worden waren. Die eher niedrigen Zusammenhänge zwischen der *SJT*-Skala und den Selbsteinschätzungsskalen könnten auch darauf zurückgeführt werden, dass sie unterschiedliche Messmethoden darstellen. Der *SJT* erfasst anhand von Videos situationsspezifische Reaktionen eines Teilnehmers, während die Selbsteinschätzungsskalen situationsunabhängige Beurteilungen der eigenen sozialen Kompetenzen erfragen. Zudem stellen die Selbsteinschätzungsskalen globale Messverfahren dar, welche nach Bergmann (2007) zu einer geringeren Validität geführt haben könnten als dimensionsorientierte Selbsteinschätzungen. Aus den Ergebnissen kann nicht sicher abgeleitet werden, was der *SJT* misst. Am ehesten scheint er Kontaktfähigkeit oder ein Mischkonstrukt aus allgemeinen sozialen Kompetenzen zu messen, da der *SJT*-Wert am höchsten mit der selbsteingeschätzten Kontaktfähigkeit gefolgt von der selbsteingeschätzten allgemeinen sozialen Kompetenz korrelierte.

Die Prüfung der diskriminanten Validität erfolgte über zwei Leistungsparameter, zum einen über den Abiturnotendurchschnitt als Leistungsaggregat der letzten zwei Schuljahre, zum anderen über den TMS als punktuelle Leistungsmessung. Die *SJT*-Skala zeigte in beiden Stichproben Nullkorrelationen bis niedrige Zusammenhänge zu den beiden kognitiven Kriterien. Diese Zusammenhänge ließen sich nicht auf eingeschränkte Varianzen der kognitiven Kriterien zurückführen. Sowohl der Abiturnotendurchschnitt als auch das TMS-Ergebnis wiesen breite Streuungen auf. Die Korrelationshöhe zwischen dem *SJT* und den kognitiven Kriterien sind vergleichbar zu anderen *SJT*-Studien. Möller (2010) berichtete ein Nullzusammenhang zwischen dem Notendurchschnitt des Abschlusszeugnisses und einem videobasierten *SJT* für Kundenorientierung. McDaniel und Nguyen (2001) fanden in ihrer Metaanalyse Zusammenhänge von .19 zwischen *SJT*s mit verhaltensbezogener Instruktion und allgemeinen kognitiven Fähigkeiten. Polyhart und Ehrhart (2003) berichteten Nullkorrelationen zwischen dem amerikanischen Studieneingangstest (*SAT*) und mehreren *SJT*s. Somit erfasst der *SJT* vermutlich keine kognitiven Eigenschaften. Eine umfassendere Prüfung der konvergenten und diskriminanten Validität soll in einer Folgestudie vorgenommen werden.

Der *SJT* scheint weder Frauen noch Männer zu bevorzugen. Zwar zeigten Frauen einen geringfügig höheren Punktwert als Männer, aber der Unterschied war in drei von vier Testungen nicht signifikant. Nur der parametrische Test in der ersten Stichprobe war knapp signifikant. Das Ergebnis zeigt die Tendenz, aber entspricht nicht den in der Forschungsliteratur berichteten Effekt, dass Frauen signifikant höhere Punktwerte im *SJT* erzielen als Männer (Hartman et al., 2007; Lievens, 2012; Nguyen et al., 2005). Es liegen jedoch auch Studien vor, die keine Geschlechtsunterschiede im *SJT* beschreiben (Motowidlo et al., 1990; Motowidlo & Tippins, 1993; Möller, 2010), sodass der nicht vorhandene Geschlechtsunterschied im vorliegenden *SJT* nicht als Argument gegen die Konstruktvalidität interpretiert wird. Im positiven Sinn scheint der entwickelte *SJT* genderfair zu messen. Zur Absicherung soll der *SJT* noch einmal in einer Folgestudie auf Geschlechtsunterschiede untersucht werden.

Als Kontrollvariable wurde das Alter untersucht. Dieses korrelierte gering mit der *SJT*-Skala in der ersten, aber nicht in der zweiten Stichprobe. Denkbar ist, dass ältere Personen bereits mehr Erfahrungen im Umgang mit anderen Menschen gesammelt haben als jüngere. Die genauere Analyse zwischen verschiedenen Altersklassen in der ersten Stichprobe ergab einen signifikanten, aber praktisch nicht bedeutsamen Effekt. In der Stichprobe überwogen die 18-20-Jährigen mit über 500 Teilnehmern sehr stark. Die Gruppe der über 30-Jährigen war hingegen mit 15 Personen sehr gering besetzt, sodass der Einfluss des Alters vermutlich nicht adäquat untersucht werden konnte. Insgesamt liegt jedoch der Schwerpunkt der Zielgruppe auf jüngere Personen, die sich im Anschluss an ihre Schulausbildung um einen Studienplatz bewerben möchten. Somit scheint das Alter der Zielgruppe einen geringen Einfluss auf den Punktwert im *SJT* zu haben. Als weitere Kontrollvariablen sollen in den Folgestudien Deutsch als Erstsprache und die medizinische Ausbildungen vor dem Studium aufgenommen werden, um zu prüfen, ob sie einen Einfluss auf die Punktzahl im *SJT* haben. Der *SJT* könnte aufgrund der schriftlichen Antwortoptionen, die eine gewisse Lesekompetenz voraussetzen, Personen benachteiligen, deren Erstsprache nicht Deutsch ist (Sacco et al., 2000a; 2000b). Auch könnten Personen durch eine medizinische Ausbildung vor dem Studium in den klinischen Situationen des *SJT*s bevorzugt werden. Einige Teilnehmer wünschten sich sogar separate Items für Personen mit bzw. ohne medizinische Ausbildung.

Nützlichkeit und Akzeptanz. Mehr als zwei Drittel der Teilnehmer sind die sozialen Anforderungen bewusst geworden, die das Medizinstudium an sie stellt. Auch haben sie einen Einblick in das Studium und das Alltagsleben von Medizinstudierenden bekommen. Daher kann die informierende Funktion des *SJT*s als gegeben gewertet

werden. Zudem ist der *SJT* in der Lage, über die Hälfte der Teilnehmer zum Nachdenken über die eigenen sozialen Kompetenzen anzuregen. Auch scheint er den Entscheidungsprozess für oder gegen ein Medizinstudium zu beeinflussen. Dabei bestärkt der *SJT* Personen mehr, ein Medizinstudium aufzunehmen als sie davon abzuschrecken. Der *SJT* war jedoch nicht in der Lage, Personen mit niedrigen Punktwerten signifikant mehr abzuschrecken als Personen mit hohen Werten im *SJT*. Die Teilnehmer erzielten mehrheitlich Werte im oberen Skalenbereich, sodass die Varianz der Skala eingeschränkt war. Eine Prüfung des *SJT*s anhand einer Zufallsstichprobe an Stelle von einer Gelegenheitsstichprobe könnte darüber mehr Aufklärung bringen.

Die Akzeptanz des *SJT*s durch die Studienplatzbewerber und -interessenten scheint gegeben. Zwei Drittel der Teilnehmer würden den *SJT* weiterempfehlen und erlebten ihn als ansprechend und realitätsnah. Dieser Befund ist konsistent zu der Erkenntnis, dass videobasierte *SJT*s aufgrund der stärkeren Realitätsnähe eine hohe Akzeptanz der Teilnehmer erzielen (Chan & Schmitt, 1997; Richman-Hirsch et al., 2000). Auch kann die tätigkeitszentrierte Entwicklung mit Experten, die Inhaltsvalidität begünstigt haben (Ployhart & MacKenzie, 2011).

Kritisiert wurde hingegen die Länge des *SJT*s. Diese verkürzte sich bei der Skalenbildung durch die Entnahme von vier Items um ca. 10-12 Minuten. Die neue Testbearbeitungszeit liegt nun bei ca. 60 Minuten. Weitere Kürzungen sind nicht vorgesehen, um die informierende Funktion des *SJT*s nicht einzuschränken. Einige Teilnehmer regten sogar zu zusätzlichen Situationen an. Bevor neue Items generiert werden, sollen vorerst die vorhandenen Items weiter geprüft werden. Die teils bemängelte Akustik der Videos wurde verbessert und die technischen Fehler bei der automatischen Rückmeldefunktion behoben.

Insgesamt zeigte sich eine gute Übereinstimmungen in den Ergebnissen beider Stichproben. Dies betraf die Ergebnisse zur Reliabilität, zum Geschlechterunterschied, sowie zur Nützlichkeit und Akzeptanz des *SJT*s. Bei der Faktorenanalyse bestanden kleine Abweichungen bei der empfohlenen Faktorenextraktion und in den einfaktoriellen Lösungen. Der Zusammenhang mit den Selbsteinschätzungsskalen fiel in der kleinen Stichprobe geringer aus als in der großen. Diese Studie richtete den Fokus auf die Faktorenstruktur, die Reliabilität und die Akzeptanz der Teilnehmer. In der folgenden soll nun die Konstruktvalidität intensiver analysiert werden.

8. Validierung an Medizinstudierenden

Die Validierungsstudie hatte zum Ziel, das Konstrukt des *SJT*s zu untersuchen. Hierzu wurde der Zusammenhang mit sozialen Kompetenzen, der Intelligenz, Persönlichkeitseigenschaften und dem Studienerfolg erforscht. Die Durchführung der Studie wurde durch ein Ethikvotum der Medizinischen Fakultät Heidelberg genehmigt (Ethikantrag S-162/2011).

8.1. Hypothesen

Die Hypothesennummerierung setzt die Nummerierung der Vorstudie fort. Stimmen die Hypothesen zwischen den Studien überein, erhalten sie dieselben Nummern. Das Ergebnis der letzten Studie, dass sich weibliche und männliche Studienplatzbewerber und -interessierte nicht in der Punktzahl im *SJT* unterschieden, wurde erneut überprüft. In der Literatur erzielten Frauen im *SJT* geringfügig höhere Werte als Männer (Hartman et al., 2007; Lievens, 2012; Nguyen et al., 2005) insbesondere dann, wenn *SJT*s mit Gewissenhaftigkeit und Verträglichkeit korrelieren (Whetzel et al., 2008). Daher wurde folgende Hypothese (H) erneut getestet:

H1: Frauen erreichen eine höhere Punktzahl im *SJT* als Männer.

Zur Prüfung der konvergenten Validität des videobasierten *SJT*s wurden konstruktnahe Verfahren herangezogen. So wurde angenommen, dass der *SJT*, der beabsichtigt soziale Kompetenzen zu erfassen, mit anderen sozialen Kompetenzverfahren korreliert. Ein standardisiertes Verfahren zur Erfassung sozialer Kompetenzen stellt das Inventar sozialer Kompetenzen (ISK; Kanning, 2009) dar. Einige Skalen des ISKs weisen eine definitorische Nähe zu den bei der Entwicklung zugrunde liegenden Konstrukten des vorliegenden *SJT*s auf. So ähnelt die Definition der Primärskalen direkte und indirekte Selbstaufmerksamkeit des ISKs dem im *SJT* enthaltenem Konstrukt Selbstbeobachtung/Selbstreflexion. Die Primärskalen Selbstkontrolle und emotionale Stabilität der Sekundärskala Selbststeuerung des ISKs kommen dem Konstrukt Selbstkontrolle/Selbstdisziplin des *SJT*s nahe. Die Primärskala Extraversion des ISKs ist vergleichbar mit den Konstrukten Kontaktfähigkeit und Suche nach sozialer Unterstützung des *SJT*s. Aufgrund der dargestellten inhaltlichen Nähe des *SJT*s mit dem ISK wurde angenommen:

H2: Der *SJT* korreliert mit den Skalen des ISKs.

Als unstandardisiertes Verfahren wurden Selbsteinschätzungsskalen eingesetzt, die den bei der *SJT*-Entwicklung zugrunde liegenden Konstrukten entsprechen. Für den Zusammenhang zwischen *SJT* und den Selbsteinschätzungsskalen wurde vermutet:

H3: Es zeigen sich positive Zusammenhänge zwischen dem *SJT* und den Selbsteinschätzungsskalen sozialer Kompetenzen.

Für die Validierung des ISKs nutzte Kanning (2009) Außenkriterien, die Anhaltspunkte für die soziale Integration einer Person als Folge von kompetentem bzw. inkompetentem Sozialverhalten geben. Die zwei ISK-Skalen Selbststeuerung und Offensivität korrelierten mit dem *Bemühen Netzwerke zu bilden*, der *Größe des Netzwerkes* und der *sozialen Integration* der Teilnehmer. Des Weiteren ergaben sich negative niedrige Zusammenhänge zwischen den drei Sekundärskalen soziale Orientierung, Offensivität und Selbststeuerung und der Höhe des Fernsehkonsums in der Freizeit. Das Ausmaß an friedfertig gelösten Konflikten korrelierte positiv mit den Sekundärskalen soziale Orientierung und Selbststeuerung. Konform dazu ergaben sich negative Zusammenhänge zwischen denselben Skalen mit dem Ausmaß aufbrausend gelöster Konflikte. Der Anteil an sozialen Interaktionen, in denen sich eine Person unsicher fühlt, korrelierte negativ mit den Faktoren Selbststeuerung, Offensivität und soziale Orientierung. Infolgedessen wurde angenommen, dass signifikante Korrelationen zwischen dem *SJT* und den Außenkriterien vorliegen. Die Vermutungen waren im Einzelnen:

H4: Der *SJT* korreliert positiv mit dem Bemühen Netzwerke zu bilden/erhalten, der sozialen Unterstützung bei Problemen, der sozialen Eingebundenheit und dem Ausmaß an friedfertig gelösten Konflikten.

H5: Der *SJT* korreliert negativ mit dem alleinigen Medienkonsum in der Freizeit, dem Ausmaß an aufbrausend gelösten Konflikten und der Unsicherheit in sozialen Interaktionen.

Der vermutete Zusammenhang zwischen dem *SJT* und den Persönlichkeitseigenschaften veranschaulicht folgendes Zitat: „Erwartungsgemäß gehen hohe Werte der sozialen Kompetenz (...) nahezu immer mit hohen Werten auf solchen Persönlichkeitsskalen einher, die sich positiv auf die Gestaltung sozialer Interaktionen auswirken dürften (Extraversion, Offenheit, Verträglichkeit, Gewissenhaftigkeit). Die Umkehrung gilt für die Dimension Neurotizismus, was nicht verwundert, da emotionale Instabilität dem Konzept des sozial kompetenten Verhalten zuwider läuft“ (Kanning, 2005, S. 144). ASENDORPF und NEYER (2012) gehen insbesondere auf die Ursache des Zusammenhangs

zwischen den sozialen Kompetenzen und der Dimension Offenheit ein. Nach ihnen korreliert sie mit sozialen Kompetenzen, da sie die intellektuelle Fähigkeit repräsentiert, sich in andere hineinzusetzen und soziale Konflikte zu lösen. Metaanalysen zum Zusammenhang zwischen *SJTs* und den Persönlichkeitseigenschaften zeigten, dass *SJTs* zu .22 mit Verträglichkeit, zu .23 mit Gewissenhaftigkeit, zu .13 mit Extraversion und zu .11 mit Offenheit für Erfahrungen korrelierten (McDaniel et al., 2001; 2007). *SJTs* mit verhaltensbasierter Instruktion wiesen dabei höhere Zusammenhänge mit den Persönlichkeitseigenschaften Verträglichkeit (.37 vs. .19), Gewissenhaftigkeit (.34 vs. .24) und emotionale Stabilität (.35 vs. .12) auf als *SJTs* mit wissensbasierter Instruktion (McDaniel & Nguyen, 2001). Daher wurde angenommen:

- H6: Der *SJT* zeigt positive Zusammenhänge zu den Persönlichkeitseigenschaften Verträglichkeit, Gewissenhaftigkeit, emotionale Stabilität und Offenheit für Erfahrungen.
- H7: Es besteht ein negativer Zusammenhang zwischen dem *SJT* und der Persönlichkeitseigenschaft Neurotizismus.

Der Zusammenhang zwischen dem vorliegenden *SJT* und der Intelligenz wird wie folgt angenommen: Zum einen stehen *SJTs* nach der *Tacit Knowledge* Theorie (Sternberg, 1997, 1999; Stemler & Sternberg, 2006) der praktischen Intelligenz nahe, die in kontextspezifischen Problemlösesituationen zum Einsatz kommt. Zum anderen zeigen *SJTs* keine bis geringe Korrelationen zur analytischen Intelligenz, die aktiv wird, wenn Personen auf abstrakte alltagsferne Probleme stoßen (Stemler & Sternberg, 2006). Die analytische Intelligenz spiegelt sich in fluiden Intelligenztests wieder (Sternberg, 1997). Demnach sollten *SJTs* nur gering mit fluiden Intelligenztests korrelieren. Auch aus inhaltlicher Perspektive sollte der *SJT* für soziale Kompetenzen gering mit dem Konstrukt der fluiden Intelligenz korrelieren. Methodisch bilden Intelligenztests die maximale Leistung eines Probanden ab, während verhaltensbasierte *SJTs* Verhaltenstendenzen erfassen (Cronbach, 1960). Somit sollten beide Verfahren auch methodisch gering miteinander korrelieren.

- H8: Der *SJT* korreliert nicht mit der fluiden Intelligenz gemessen über den CFT20-R.

Als weitere Verfahren der allgemeinen kognitiven Leistungsfähigkeit werden der Abiturnotendurchschnitt und der TMS verwendet. Diese sollten analog zum Intelligenztest weitgehend unabhängig vom *SJT* sein. Auf Subtestebene werden die TMS-Aufgabengruppen quantitative und formale Probleme, Textverständnis und konzentrier-

tes und sorgfältiges Arbeiten analysiert. Die Aufgabengruppe quantitative und formale Probleme erfasst die Fähigkeit mit Zahlen, Größen, Einheiten und Formeln umzugehen und sollte aufgrund der konstruktferne nicht mit dem *SJT* korrelieren. Die Aufgabengruppe Textverständnis misst die Fähigkeit, umfangreiches und komplexes Textmaterial zu bearbeiten. Komplex formulierte *SJT*-Items können die Anforderung an die verbale Kompetenz erhöhen und Einflüsse auf die Testfairness bezüglich ethischer Subgruppen hervorrufen (Sacco et al., 2000a). Der Einfluss kann reduziert werden, indem videobasierte statt schriftliche Situationen wie im vorliegenden Fall verwendet werden (Sacco et al., 2000a; 2000b). Daher sollten zwischen dem videobasierten *SJT* und dem Textverständnis nur geringe Zusammenhänge bestehen. Die TMS-Aufgabengruppe Konzentriertes und sorgfältiges Arbeiten erfasst die Konzentrationsfähigkeit einer Person. Konzentrationsfähigkeit ist die „Fähigkeit unter Bedingungen schnell und genau zu arbeiten, die das Erbringen einer kognitiven Leistung normalerweise erschweren“ (Schmidt-Atzert, Büttner & Bühner, 2004, S.9). Die Konzentrationsfähigkeit soll als diskriminantes Validierungskriterium für den *SJT* genutzt werden. Im Einzelnen wurde angenommen:

- H9: Der *SJT* korreliert nicht mit dem Abiturnotendurchschnitt.
- H10: Der *SJT* zeigt keinen Zusammenhang zum Gesamtwert des TMS.
- H11: Es besteht kein Zusammenhang zwischen dem *SJT* und dem TMS-Subtest quantitative und formale Probleme.
- H12: Der *SJT* zeigt nur geringe Korrelationen zum TMS-Subtest Textverständnis.
- H13: Der *SJT* korreliert nicht mit dem TMS-Subtest Konzentriertes und sorgfältiges Arbeiten.

Verfahren der Studierendenauswahl werden letztendlich daran gemessen, wie zuverlässig sie ein erfolgreiches Studium vorhersagen. Ein erfolgreiches Studium zeichnet sich u.a. durch gute Studienleistungen, durch die Zufriedenheit mit dem Studium und eine guten Passung zwischen dem Fähigkeitsprofil des Bewerbers und den Anforderungen des Studiums aus (Rindermann & Oubaid, 1999). Für die vorliegende Untersuchung wurden die Studienzufriedenheit und die Passung zum Studium als subjektives Maß über Selbsteinschätzungen erhoben. Der Studienerfolg wurde sowohl in subjektiver (Selbstauskunft) als auch in objektiver Form (Studiennoten) erfasst. Die Zufriedenheit korreliert gewöhnlich eher mit dem Arbeitsverhalten einer Person, während der Studienerfolg eher mit kognitiven Kriterien zusammenhängt (Giesen et al, 1986; Kersting, 2005). Konsistent dazu fand Kanning (2009) positive Zusammenhänge zwi-

schen der subjektiven Studienzufriedenheit und den sozialen Kompetenzen der Skala Offensivität und Selbststeuerung. Daher wurden folgende Hypothesen aufgestellt:

H14: Der *SJT* sagt die Studienzufriedenheit und die Passung zum Studium vorher.

H15: Der *SJT* sagt nicht den subjektiven Studienerfolg vorher.

Nicht-kognitive Kriterien sind besser für die Vorhersage von klinischen als von vorklinischen Studienleistungen geeignet (vgl. Kapitel 2.2.2), da in den klinischen Leistungen zwischenmenschliche und persönliche Kompetenzen eine höhere Rolle spielen. Daher wird auch für den vorliegenden *SJT* angenommen:

H16: Der *SJT* zeigt eine höhere prädiktive Validität für die klinischen als für die vorklinischen Studienleistungen.

8.2. Untersuchungsablauf und eingesetzte Verfahren

Die Erhebung fand im Oktober bis Dezember 2013 im PC-Pool der medizinischen Fakultät statt und dauerte ca. 120 Minuten pro Testung. Nach dem Lesen der Teilnehmerinformation und dem Ausfüllen der Einverständniserklärung bearbeiteten die Teilnehmer den Intelligenztest CFT 20-R. Anschließend wurden die Teilnehmer zu ihren soziodemografischen Angaben gefragt und füllten die Kurzversion des *Big Five Inventory (BFI-K)* aus. Danach bearbeiteten sie den videobasierten *SJT*, worauf eine Selbsteinschätzung bezüglich der fünf verbliebenen sozialen Kompetenzen des *SJTs* folgte. Nachdem die Teilnehmer das Inventar sozialer Kompetenzen (ISK) bearbeitet hatten, folgten Fragen zu Außenkriterien von sozialen Kompetenzen, zur Studienzufriedenheit, zum subjektiven Studienerfolg und zur Passung zum Studium. Der Test für Medizinische Studiengänge (TMS) und die Noten im Medizinstudium wurden im Nachhinein für die Teilnehmer abgerufen, die den Zugriff auf die universitäre Bewerber- und Studiennotendatenbank erlaubten. In einer Nacherhebung wurden die Teilnehmer um eine Fremdeinschätzung der sozialen Kompetenz von zwei Personen mit der Kurzform des Inventars sozialer Kompetenzen (ISK-K) gebeten. Tabelle 43 stellt den Untersuchungsablauf schematisch dar. Die einzelnen Verfahren werden im Folgenden in der Reihenfolge des Untersuchungsablaufs genauer beschrieben.

Tabelle 44: *Untersuchungsablauf der Validierungsstudie*

Verfahren	Minuten
Teilnehmerinformation, Einverständniserklärung	5
CFT20-R (1. Testteil)	26
soziodemografische Angaben	2
BFI-K	2
Videobasierter <i>SJT</i>	60
Selbsteinschätzungsskalen sozialer Kompetenz	2
ISK	20
Außenkriterien für soziale Kompetenzen	3
Studienzufriedenheit, -erfolg, -passung	2
gesamt:	122

CFT20-R. Der revidierte Grundintelligenztest CFT 20-R (Weiß, 2006) erfasst die *General Fluid Ability* im Sinne von Cattell (1963) mittels sprachfreien figuralen Darstellungen. Darunter versteht Cattell die Fähigkeit „komplexe Beziehungen in neuartigen Situationen wahrnehmen und erfassen zu können“ (1986, S.58). Der CFT 20-R besteht aus zwei gleichartigen Testteilen, die jeweils aus den Subtests Reihenfortsetzen, Klassifikationen, Matrizen und Topologien aufgebaut sind. In der vorliegenden Erhebung wird der als Kurzform geltende erste Testteil mit Testzeitverlängerung eingesetzt. Die reine Bearbeitungsdauer der Subtests umfasst 18 Minuten. Die interne Konsistenz des ersten Testteils liegt bei .92. Die Retest-Reliabilität für ein zweimonatiges Testintervall beträgt .85 und bei einem fünfmonatigen Testintervall .69. Alle Subtests zeigen hohe Ladungen auf dem General Fluid Ability Faktor. Der CFT 20-R korreliert zu .56 mit dem revidierten Prüfsystem für Schul- und Bildungsberatung für die vierte bis sechste Klasse (PSB-R 4-6; Lukesch, Kormann & Mayrhofer, 2002), welches die fluide Intelligenz und schulbezogene Wissensbereiche erfasst. Zudem wurden durchschnittliche Korrelationen von .50 mit der Mathematiknote festgestellt. Für den ersten Testteil liegen extrapolierte Altersnormen für Erwachsene von 20-60 Jahren vor. Das Verfahren differenziert gut im oberen Leistungsbereich. Der CFT 20-R wurde als standardisiertes Instrument zur Erfassung der allgemeinen kognitiven Leistungsfähigkeit in der vorliegenden Studie verwendet.

Soziodemografische Angaben. Die soziodemografischen Angaben umfassten das Alter, das Geschlecht, Deutsch als Erstsprache und das Fachsemester. Zudem wurde gefragt, ob die Teilnehmer eine mindestens einjährige medizinische Ausbildung (z.B.

Rettungssanitäter, Krankenpfleger) absolviert haben oder eine mindestens einjährige Erfahrung in der Arbeit mit Patienten mitbrachten. Diese Angaben gingen als Kontrollvariablen zur Prüfung potenzieller Einflüsse auf die Punktzahl im *SJT* ein.

BFI-K. Das *BFI-K* (Rammstedt & John, 2005) erfasst die fünf Persönlichkeitsfaktoren Extraversion, Neurotizismus, Gewissenhaftigkeit und Verträglichkeit mit jeweils vier Items sowie den Faktor Offenheit für Erfahrungen mit fünf Items. Mit einer Bearbeitungszeit von zwei Minuten ist das Verfahren sehr ökonomisch. Die Reliabilitäten und Validitäten liegen im ausreichenden bis sehr guten Bereich. Die internen Konsistenzen betragen zwischen .64 und .86 und die Retest-Reliabilitäten zwischen .76 und .86. Das *BFI-K* weist hohe Zusammenhänge zur Langversion des *BFI* ($r = .85$) und zum *Revised NEO Personality Inventory* ($r = .63-.82$; Costa & McCrae, 1992) auf. Das Verfahren wird für Gruppen- und Zusammenhangsanalysen, aber nicht für individuelle Testungen empfohlen. In der vorliegenden Untersuchung soll es zur Erhebung der fünf grundlegenden Persönlichkeitseigenschaften eingesetzt werden.

Videobasierte SJT. Der videobasierte *SJT* für soziale Kompetenzen umfasste 20 Items. Die Items wurden nach dem zeitlichen Ablauf des Studiums angeordnet, um der informierenden Funktion des *SJTs* gerecht zu werden. Die Filmsequenzen wurden durch eine kurze schriftliche Erklärung eingeleitet, wenn der Dialog nicht alle zum Verständnis notwendigen Informationen enthielt. An der jeweils erfolgskritischen Stelle stoppten die Filme und die Bewerber wurden aufgefordert, anzugeben wie sie an Stelle des dargestellten Studierenden reagieren würden. Dafür standen ihnen pro Item vier bis sechs Verhaltensalternativen zur Verfügung, die in zufälliger Reihenfolge angeordnet waren. Pro Verhaltensalternative sollten sie angeben, mit welcher Wahrscheinlichkeit sie diese ausführen würden. Insgesamt mussten 100 Prozent pro Item vergeben werden. Die bereits vergebenen Prozente wurden automatisch angezeigt. Die Filme konnten bei Bedarf erneut angeschaut werden. Eine zeitliche Limitation zur Beantwortung der Items bestand nicht.

Selbsteinschätzungsskalen sozialer Kompetenz. Die Selbsteinschätzungsskalen bestanden aus siebenstufigen Likert-Skalen, die die Teilnehmer aufforderten, sich in den fünf sozialen Kompetenzen des *SJTs* - verglichen mit Gleichaltrigen - einzuschätzen. Falls sich die Teilnehmer in einer Kompetenz nicht einzuschätzen wussten, stand ihnen die Option „kann ich nicht beurteilen“ zur Verfügung. Im Vorfeld wurde jeweils die Definition der sozialen Kompetenz gegeben. Die Selbsteinschätzungsskalen wurden als unstandardisiertes Instrument zur Erfassung von sozialen Kompetenzen eingesetzt.

ISK. Das Inventar sozialer Kompetenzen (Kanning, 2009) erfasst die vier Sekundärfaktoren Soziale Orientierung, Offensivität, Selbststeuerung und Reflexibilität mit insgesamt 17 untergeordneten Primärfaktoren (Tabelle 45).

Tabelle 45: *Primär- und Sekundärfaktoren des ISKs (Kanning, 2009, S. 28 ff.)*

Sekundärskalen	Definition	Primärskalen
Soziale Orientierung	Ausmaß, in dem eine Person anderen Menschen offen und mit positiver Grundhaltung gegenüber tritt.	Prosozialität Perspektivenübernahme Wertpluralismus Kompromissbereitschaft Zuhören
Offensivität	Fähigkeit, aus sich herauszugehen und im Kontakt mit anderen Menschen eigene Interessen aktiv verwirklichen zu können.	Durchsetzungsfähigkeit Konfliktbereitschaft Extraversion Entscheidungsfreudigkeit
Selbststeuerung	Fähigkeit eines Menschen, flexibel und rational zu handeln, wobei man sich selbst bewusst als Akteur begreift.	Selbstkontrolle Emotionale Stabilität Handlungsflexibilität Internalität
Reflexibilität	Ausmaß, in dem sich eine Person mit sich und ihren Interaktionspartnern aktiv auseinandersetzt.	Selbstdarstellung Direkte Selbstaufmerksamkeit Indirekte Selbstaufmerksamkeit Personenwahrnehmung

Für die Selbsteinschätzung der sozialen Kompetenz wurde die Langversion mit 108 Items eingesetzt. Die psychometrische Güte liegt mit internen Konsistenzen zwischen .64 und .90 und Retest-Reliabilitäten zwischen .80 und .87 im zufriedenstellenden bis guten Bereich. Faktorenanalysen bestätigten die vierfaktorielle Struktur. Zusammenhangsanalysen mit Persönlichkeitsfragebögen, Indikatoren sozialer Integration, der Arbeits-/Studienzufriedenheit und -leistung, sowie der Intelligenz stützen die Konstruktvalidität des Verfahrens. Für die Fremdeinschätzung der sozialen Kompetenz wurde die Kurzform bestehend aus 33 Items verwendet, die ausschließlich die vier Sekundärskalen erfasst. Die interne Konsistenz beträgt zwischen .69 und .75 und die Retest-Reliabilität zwischen .80 und .86.

Außenkriterien sozialer Kompetenzen. Außenkriterien sind Indikatoren für das Verhalten und Erleben von Personen. Die hier verwendeten Außenkriterien bilden die soziale Integration als Konsequenz sozial kompetenten Verhaltens ab. Zur Erhebung der sozialen Eingebundenheit im Studium wurden die drei nachfolgenden Items aus der fakultätsinternen Studierendenbefragung 2010 (Heidelberger Fragebogen zur Motivation und Interesse im Studium) entnommen, die eine interne Konsistenz von .90 (N = 370) aufwiesen und in Anlehnung an Müller (2001) entstanden sind: „In meiner Uni fühle ich mich sozial gut eingebunden“, „Mit meinen Kommilitonen verstehe ich mich insgesamt sehr gut“ und „Ich habe im Studium viele Leute, mit denen ich mich gut verstehe“. In Anlehnung an Kanning (2009) wurde als Außenkriterium für die Kontaktfähigkeit das Bemühen zu Netzwerken („Ich bemühe mich Freundschaften und Bekanntschaften zu bilden bzw. zu pflegen“) und die vorhandene soziale Unterstützung („Wenn ich im Studium oder privat Hilfe brauche, habe ich immer jemanden, den ich fragen kann“) aufgenommen. Die Beantwortung aller Items erfolgte auf einer fünfstufigen Ratingskala (1 = trifft voll zu bis 5 = trifft überhaupt nicht zu). Des Weiteren wurde ebenfalls orientiert an Kanning (2009) der Prozentsatz zufriedenstellend vs. aufbrausend gelöster Konflikte als Außenkriterium für die Kritikfähigkeit und Selbstkontrolle/Selbstdisziplin erfragt. Als Validierungskriterium für alle sozialen Kompetenzen wurde der prozentuale alleinige Medienkonsum in der Freizeit erhoben und nach dem prozentualen Anteil von sozialen Interaktionen gefragt, in denen sich die Teilnehmer unsicher fühlten.

Studienerfolgskriterien. Als subjektive Studienerfolgskriterien wurden die Studienzufriedenheit, der Studienerfolg und die Passung zum Studium erhoben. Alle drei Kriterien wurden durch ein allgemein formuliertes Item auf einer fünfstufigen Ratingskala (1 = trifft voll zu bis 5 = trifft überhaupt nicht zu) erfasst. Die Studienzufriedenheit wurde durch das Item „Insgesamt bin ich mit meinem Studium zufrieden“, der subjektive Studienerfolg über das Item „Ich bin mit meinen Leistungen im Studium zufrieden“ und die Passung zum Studium durch das Item „Die Anforderungen des Studiums passen gut zu meinen Fähigkeiten“ erfasst.

Bei vorhandener Einwilligung der Teilnehmer wurde darüber hinaus der objektive Studienerfolg anhand der Studienleistung erhoben. Dafür gingen die Leistungen aus dem ersten bis siebten Fachsemester ein. Da nur wenige fortgeschrittene Studierende an der Studie teilnahmen, lagen kaum Prüfungsleistungen aus dem achten bis zwölften Fachsemestern vor, sodass diese nicht in die Auswertung einfließen. Aus dem vorklinischen Studienabschnitt wurden die Leistungen verwendet, die eine Punktzahl, ein Ranking oder die prozentuale Leistung des Studierenden wiedergaben. Die vorklini-

schen Leistungen, die nur als bestanden vs. nicht bestanden im Notensystem der Fakultät vermerkt waren, wurden nicht berücksichtigt. Der Schwerpunkt der vorklinischen Prüfungen lag überwiegend in der Prüfung von naturwissenschaftlichem und medizinischem Grundlagenwissen. Die Leistungen der klinischen Studienabschnitte waren mit Noten von 1.0-5.0 beurteilt worden und waren analog zu Schulnoten zu interpretieren. Neben der Abfrage von krankheitsbezogenem Wissen, konzentrierten sich die klinischen Leistungsnachweise u.a. auch auf klinisch-praktische Fertigkeiten und den Umgang mit Patienten. Zum besseren Verständnis werden spezielle medizinische Prüfformate kurz erläutert. In der Inneren Medizin und Chirurgie wurden *Objective Structured Clinical Examinations (OSCE; Harden, Stevenson, Downie & Wilson, 1975; Jünger & Nikendei, 2005)* eingesetzt. OSCEs sind mündlich-praktische Prüfungen zur Beurteilung klinisch-praktischer Fertigkeiten. In dieser rotieren die Prüflinge durch einen Parcours von standardisierten Prüfungsstationen von je 5-9 Minuten Dauer. Die Prüfstationen spiegeln typische Situationen aus dem Arztalltag wieder. Zur Überprüfung kommunikativer Kompetenzen in der Arzt-Patienten-Beziehung werden Schauspieler eingesetzt, die eine standardisierte Patientenrolle einnehmen. In der Inneren Medizin und der Notfallmedizin kamen auch *Mini Clinical Evaluation Exercises (Mini-CEX; Norcini, Blank, Arnold & Kimball, 1995; Norcini & Burch, 2007)* und *Encounter Cards (Hatala & Norman, 2002; Paukert, Richards & Olney, 2002)* zum Einsatz. Mit diesen 5-15 minütigen Prüfungen am realen Patienten oder in Simulationssituationen werden klinisch-praktische Kompetenzen, der Umgang mit Patienten (Kommunikation, Empathie) und die Arbeitsorganisation (Entscheidungsfindung, Effizienz) anhand festgelegter Kriterien beurteilt. In einem Patientenbericht in der Inneren Medizin und Geriatrie stellen Studierende schriftlich einen konkreten Patientenfall dar mit einem Schwerpunkt auf Anamnese und körperlicher Untersuchung, Verdachtsdiagnosen und dem weiteren diagnostischen und therapeutischen Vorgehen unter Verwendung von aktueller Fachliteratur.

Allgemeine kognitive Leistungsfähigkeit. Als Maße der allgemeinen kognitiven Leistungsfähigkeit wurden neben dem IQ-Wert und dem Abiturnotendurchschnitt, der standardisierte Testwert des TMS, sowie drei TMS-Subskalen herangezogen. Die Subskala quantitative und formale Probleme erfasst die Fähigkeit mit abstrakten Formeln, Größen und Einheiten umzugehen. Der Subtest Textverständnis testet den Umgang mit komplexem Textmaterial, und die Subskala konzentriertes und sorgfältiges Arbeiten prüft die Konzentrationsfähigkeit einer Person. Eine Beschreibung des gesamten TMS wurde in Kapitel 2.1 gegeben.

8.3. Stichprobenrekrutierung und -beschreibung

An der Studie nahmen 47 (52%) weibliche und 43 (48%) männliche Humanmedizinstudierende der Medizinischen Fakultät Heidelberg teil. Die Rekrutierung erfolgte über den Email-Verteiler der Fakultät, über Aushänge in Lehrgebäuden und über Vorstellungen der Studie in Lehrveranstaltungen. Als Motivation zur Teilnahme wurden den Studierenden die Rückmeldung ihrer persönlichen Testergebnisse und die Teilnahme an einer Gutscheinerlosung in Aussicht gestellt. Es durften nur Personen teilnehmen, die den *SJT* zum ersten Mal bearbeiteten. An der Studie beteiligten sich Personen im Alter von 17 bis 29 Jahren mit einem Durchschnittsalter von 22.2 Jahren ($SD = 2.61$). Das Durchschnittsalter unterschied sich in der weiblichen und männlichen Stichprobe nicht ($M_w = 22.04$; $SD_w = 2.79$; $M_m = 22.44$; $SD_m = 2.61$). Die Mehrzahl der Studierenden befand sich zum Zeitpunkt der Untersuchung im 3.-4. Fachsemester, etwas weniger Studierende auch im 1.-2. und 7.-8. Fachsemester (Tabelle 46). Aus dem vorklinischen Studienabschnitt (1.-4. Fachsemester) nahmen 46 Studierende, aus dem klinischen Studienabschnitt (5.-12. Fachsemester) 42 Studierende teil. Zwei Personen hatten zum Untersuchungszeitpunkt gerade das Studium abgeschlossen. Getrennt nach Geschlecht war das 3.-4. Fachsemester wie in der Gesamtgruppe am häufigsten vertreten. Bei den Frauen folgte bezüglich der Häufigkeit das 1.-2. Fachsemester, bei den Männern das 7.-8. Fachsemester. In der weiblichen Stichprobe war das 11.-12. Fachsemester etwas stärker vertreten als in der männlichen Stichprobe. Es bestand wie in der Gesamtgruppe ein ausgewogenes Verhältnis zwischen der Anzahl von vorklinischen ($n_w = 24$, $n_m = 22$) und klinischen ($n_w = 21$, $n_m = 21$) Studierenden.

Tabelle 46: *Verteilung Fachsemester*

Fachsemester	gesamt		weiblich		männlich	
	N	%	N	%	N	%
1.-2.	18	20.0	11	23.4	7	16.3
3.-4.	28	31.1	13	27.7	15	34.9
5.-6.	9	10.0	5	10.6	4	9.3
7.-8.	17	18.9	6	12.8	11	25.6
9.-10.	4	4.4	2	4.3	2	4.7
11.-12.	10	11.1	7	14.9	3	7.0
>12.	2	2.2	1	2.1	1	2.3
Absolventen	2	2.2	2	4.3	0	0.0

Eine abgeschlossene medizinische Ausbildung oder eine mindestens einjährige Erfahrung in der Arbeit mit Patienten hatten 18% der Teilnehmer (Tabelle 47). Einundachtzig Prozent der Teilnehmer sprachen Deutsch als Erstsprache. Es lagen keine Unterschiede in der weiblichen und männlichen Stichprobe hinsichtlich einer abgeschlossenen medizinischen Ausbildung und der Erstsprache vor.

Tabelle 47: Verteilung medizinische Ausbildung und Erstsprache

	Medizinische Ausbildung				Erstsprache Deutsch			
	ja		nein		ja		nein	
	N	%	N	%	N	%	N	%
gesamt	16	17.8	74	82.2	73	81.1	17	18.9
weiblich	9	19.1	38	80.9	38	80.9	9	19.1
männlich	7	16.3	36	83.7	35	81.4	8	18.6

An der Fremdeinschätzung der sozialen Kompetenzen beteiligten sich 39 Personen im Alter von 17 bis 72 Jahren ($M = 31.4$; $SD = 14.9$). Darunter befanden sich 20 weibliche und 19 männliche Personen. Insgesamt lagen für 23 Teilnehmer ein bis zwei Fremdeinschätzungen vor. Diese Subgruppe setzte sich aus 15 weiblichen und acht männlichen Personen im Alter von 18-27 Jahren ($M = 23.1$; $SD = 2.3$) zusammen. Sieben Personen befanden sich im vorklinischen und 15 im klinischen Studienabschnitt. Drei hatten bereits vor dem Studium eine medizinische Ausbildung abgeschlossen. Zwei Personen sprachen Deutsch nicht als Erstsprache. Somit enthielt die Subgruppe einen höheren Anteil weiblicher Probanden als in der Gesamtstichprobe. Das Durchschnittsalter lag um ein Jahr höher als in der Gesamtgruppe. Es befanden sich mehr Studierende im klinischen als im vorklinischen Studienabschnitt. Die Stichproben waren vergleichbar hinsichtlich des Anteils an Personen mit medizinischer Ausbildung und Deutsch als Erstsprache.

8.4. Statistische Auswertung

Stichprobenumfang. Zur Berechnung des Stichprobenumfangs wurde das Programm *G* Power 3* (Faul et al., 2007) verwendet. Die statistische Power wurde nach Empfehlungen von Rost (2013) für alle Berechnungen auf .80 gesetzt. Für die bivariaten zweiseitigen Korrelationsanalysen wurden mit erwarteten mittleren Effektstärken 82 Testpersonen benötigt. Für die Unterschiedstestung zwischen dem Geschlecht bzw. dem vorklinischen vs. klinischen Abschnitt sollten bei einseitiger Testung

und einen erwarteten mittleren Effekt 102 Personen untersucht werden. Für die linearen multiplen Regressionsanalysen waren 68 Personen notwendig.

Deskriptive Statistik. Der Auswertungsschlüssel des *SJT*s wurde anhand des beschriebenen Vorgehens in Kapitel 6.4.6 erstellt. Zur Exploration der Daten wurden Minima, Maxima, Häufigkeiten, Mittelwerte und Standardabweichungen ermittelt. Bei den *SJT*-Items wurden zusätzlich die Schiefe und Kurtosis analysiert. Die Prüfung auf Normalverteilung der Daten fand mittels Kolmogorov-Smirnov-Test und der Betrachtung der Histogramme statt. Extremwerte und Ausreißer wurden anhand der *Boxplots* analysiert. Bei der Analyse der *SJT*-Items wurde zusätzlich zum arithmetischen Mittelwert der M-Schätzer nach Huber ermittelt. Fallen der arithmetische Mittelwert und der M-Schätzer ähnlich aus, kann dies als Hinweis gelten, dass sich keine gravierenden Ausreißer in der Stichprobe befinden. Lagen bei der Fremdeinschätzung sozialer Kompetenzen zwei Urteile pro Proband vor, wurde der Durchschnittswert beider gebildet. Für den Vergleich der Selbst- und Fremdeinschätzung wurden die unstandardisierten Skalenmittelwerte herangezogen, da keine standardisierten Werte für Fremdeinschätzungen im ISK-K vorlagen.

Unterschiedstestung. Der Unterschied in der *SJT*-Punktzahl zwischen Frauen und Männern wurde aufgrund der Normalverteilung mit dem T-test analysiert. Der Unterschiedstest zwischen den Persönlichkeitsfaktoren weiblicher und männlicher Probanden wurde mit dem nicht-parametrischen Kolmogorov-Smirnov-Test für metrische nicht normalverteilte Werte durchgeführt. Die Unterschiede in der Selbst- vs. Fremdbeurteilung sozialer Kompetenzen wurden mit der einfaktoriellen univariaten Varianzanalyse mit Messwiederholung geprüft. Als Messwiederholungsfaktor galt die Zeit mit zwei Faktorstufen. Die praktische Bedeutsamkeit der aufgetretenen Unterschiede wurde über das partielle Eta-Quadrat (η_p^2) bestimmt. Für dessen Interpretation wurde die Konvention von Cohen (1988) herangezogen. Dieser empfiehlt, Werte ab .01 als kleinen Effekt, .06 als mittleren Effekt und ab .14 als großen Effekt zu werten.

Korrelationsanalysen. Der Zusammenhang zwischen dem *SJT* und den Validierungskriterien wurde mit dem Korrelationskoeffizient nach Pearson und Spearman berechnet. Das parametrische Verfahren wurde gewählt, da die *SJT*-Skala normalverteilt war. Da jedoch die meisten Validierungskriterien nicht normalverteilt waren, wurde zusätzlich der nicht-parametrische Korrelationskoeffizient herangezogen. Die Anpassung des Signifikanzniveaus erfolgte mit der Bonferroni-Holm-Korrektur.

Regressionsanalysen. Für die Analyse der prädiktiven Validität wurde eine einfache lineare Regressionsanalyse mit dem *SJT* als alleinigem Prädiktor durchgeführt. Zur Ermittlung der inkrementellen Validität des *SJT*s wurden zusätzlich multiple hierarchische Regressionen berechnet. Die abhängigen Variablen waren jeweils die Studienzufriedenheit, der subjektive Studienerfolg, die Passung zum Studium und die soziale Eingebundenheit im Studium. In den hierarchischen Regressionen wurden die Prädiktoren in vier Blöcke aufgeteilt und sukzessiv mittels Einschluss-Methode in die Analyse einbezogen. Block 1 umfasste die Kontrollvariablen Geschlecht, Alter, Deutsch als Erstsprache und abgeschlossene medizinnahere Ausbildung. Im zweiten Schritt wurde der Abiturnotendurchschnitt (Block 2), im dritten der TMS (Block 3) und im vierten der *SJT* (Block 4) in die Analyse aufgenommen. Als Voraussetzungsprüfung wurde die gegenseitige Unabhängigkeit der Residuen anhand des Durbin-Watson-Koeffizienten geprüft. Diese Voraussetzung gilt als erfüllt, wenn dieser Werte zwischen 1.5-2.5 annimmt (Rudolf & Müller, 2004). Des Weiteren wurde die Homoskedastizität über Streudiagramme mit den z- standardisierten vorhergesagten Werten und den z- standardisierten Residuen kontrolliert. Die Punktwolken sollten dabei ein „(...) horizontal verlaufendes Band um den Mittelwert“ annehmen (S.56, Rudolf & Müller, 2004). Zudem sollten sowohl die Prädiktoren als auch abhängigen Variablen normalverteilt sein. Als wichtige Schätzgrößen bei den Regressionsanalysen dienen das standardisierte Beta-Gewicht (β), das quadrierte Bestimmtheitsmaß (R^2) und die Änderung im quadrierten Bestimmtheitsmaß (ΔR^2).

8.5. Ergebnisse

8.5.1. Deskriptive Statistik und vorausgegangene Analysen

Im folgenden Kapitel werden die deskriptiven Daten und die vorausgegangene Analysen für den *SJT* und die Validierungskriterien beschrieben. Für den *SJT* werden zunächst zentrale Statistiken auf Itemebene berichtet, bevor auf die *SJT*-Skala eingegangen wird.

SJT und dessen Items. Die Verteilung der *SJT*-Items bildet Tabelle 48 ab. Die Item-Mittelwerte lagen zwischen 3.1 und 4.4 von fünf erreichbaren Punkten. Die Standardabweichung betrug zwischen 0.5 und 1.1 Punkten. Der M-Schätzer nach Huber zeigte eine maximale Differenz von 0.2 Punkten zum arithmetischen Mittelwert. Eine zusätzliche Analyse erfolgte über die *Boxplots*, die für zehn Items keine (SBSR5, SKSD4, SKSD10, Kri1, Kri2, Kri12, SU1, SU12, SU13, KF6) und für die restlichen Items zwischen ein bis sechs Ausreißer anzeigten. Bei den Ausreißern handelte es

sich um Werte, die geringer als der Durchschnitt ausfielen. Für die Analysen wurden keine Werte ausgeschlossen, da keine inhaltlichen Gründe vorlagen und die Ausreißer gemäß des M-Schätzers einen geringen Verzerrungseffekt zu haben schienen.

Tabelle 48: *Deskriptive Statistik SJT-Items an Medizinstudierenden*

	M	M-Schätzer ^a	SD	Min	Max	Schiefe	Kurtosis
SBSR5	3.78	3.90	0.90	1.70	5.00	-0.56	-0.76
SBSR7	4.12	4.23	0.69	2.00	5.00	-1.00	0.71
SKSD4	3.03	3.00	0.75	1.50	5.00	0.20	-0.26
SKSD10	4.35	4.40	0.53	3.05	5.00	-0.49	-0.66
SKSD11	4.44	4.52	0.56	2.10	5.00	-1.60	3.51
Kri1	3.44	3.57	0.97	1.30	5.00	-0.49	-0.63
Kri2	4.26	4.35	0.69	2.00	5.00	-0.86	0.37
Kri4	3.92	3.96	0.61	1.80	5.00	-0.50	0.44
Kri7	4.19	4.27	0.67	1.50	5.00	-1.19	2.20
Kri10	4.05	4.12	0.81	1.00	5.00	-0.83	0.95
Kri12	4.37	4.40	0.48	3.30	5.00	-0.33	-0.91
SU1	4.03	4.05	0.45	2.85	5.00	-0.30	-0.08
SU8	4.34	4.43	0.57	2.80	5.00	-0.98	0.43
SU9	3.75	3.78	0.55	2.35	5.00	-0.32	0.09
SU10	4.42	4.48	0.45	2.75	5.00	-1.13	1.59
SU12	4.32	4.47	0.67	2.20	5.00	-0.90	0.03
SU13	3.17	3.19	0.90	1.10	4.90	-0.15	-0.53
SU14	4.04	4.13	0.62	2.20	5.00	-0.80	0.35
KF6	3.94	4.17	1.08	1.00	5.00	-0.95	-0.10
KF11	4.19	4.33	0.75	1.80	5.00	-1.06	0.64

Anmerkung. N = 90. ^a M-Schätzer nach Huber.

Itembezeichnung: SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Alle Items außer SKSD4 waren rechtssteil verteilt, wobei SKSD11 die höchste Abweichung aufwies. SKSD4 wies eine leichte Abweichung in linkssteiler Richtung auf. Die Kurtosis fiel für zehn Items steiler als die Normalverteilung aus (Werte > 0). Unter diesen zeigte SKSD11 gefolgt von Kri7 die höchste Kurtosis. Sieben Items wiesen eine leicht negative Kurtosis auf, sodass sie flacher als die Normalverteilung verliefen. Bei den Items SU1, SU9 und SU12 entsprach die Kurtosis fast der der Normalverteilung.

Der Kolmogorov-Smirnov-Test wies SKSD4, Kri12, SU9 und SU13 als normalverteilt aus. Die übrigen Items zeigten keine Normalverteilung ($p \leq .05$).

Die Interkorrelationen der *SJT*-Items zeigten überwiegend geringe bis mittlere Korrelationen zwischen den Items (Tabelle 49). Die höchsten Zusammenhänge bestanden zwischen KF11 und SU13 ($r(90) = .52$), Kri2 und Kri7 ($r(90) = .48$), KF6 und KF11 ($r(90) = .45$), Kri10 und SU14 ($r(90) = .41$), SU1 und SU12 ($r(90) = .41$), sowie SBSR5 und SU13 ($r(90) = .40$). Die meisten zu vernachlässigenden Korrelationen zu den anderen Testitems wiesen SU8 und SU9 auf.

Tabelle 49: Interkorrelation SJT-Items

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
SBSR5 (1)	-																			
SBSR7 (2)	.16	-																		
SKSD4 (3)	.37	.27	-																	
SKSD10 (4)	.36	.25	.32	-																
SKSD11 (5)	.21	.16	.11	.29	-															
Kri1 (6)	.31	.05	.21	.38	.04	-														
Kri2 (7)	.33	.09	.19	.19	.19	.25	-													
Kri4 (8)	.18	.10	.37	.28	.18	.24	.15	-												
Kri7 (9)	.21	.18	.18	.12	.11	.32	.48	.16	-											
Kri10 (10)	.38	.27	.39	.39	.20	.32	.51	.14	.29	-										
Kri12 (11)	.15	.17	.26	.23	.20	.09	.25	.20	.23	.38	-									
SU1 (12)	.34	.15	.21	.33	.32	.35	.27	.13	.23	.31	.18	-								
SU8 (13)	.12	.29	.02	-.01	.19	-.02	.10	.08	.13	.09	.08	.27	-							
SU9 (14)	.22	.16	.08	.09	.19	.07	.02	.10	.05	.07	.13	.08	.21	-						
SU10 (15)	.27	.07	.15	.20	.24	.08	.12	.27	.06	.12	.18	.21	.01	.30	-					
SU12 (16)	.19	.28	.06	.27	.09	.38	.32	.14	.29	.27	.13	.41	.22	.05	.23	-				
SU13 (17)	.40	.01	.31	.20	.16	.28	.16	.14	.24	.31	.08	.33	.04	.31	.31	.24	-			
SU14 (18)	-.10	.10	-.03	.16	.15	.01	.21	.06	.15	.41	.28	.03	.08	.11	.14	.28	.22	-		
KF6 (19)	.38	.15	.32	.28	.19	.19	.22	.31	.22	.29	.13	.11	.06	.28	.35	.18	.38	.01	-	
KF11 (20)	.39	.27	.25	.25	.11	.29	.12	.27	.33	.37	.10	.28	.09	.24*	.31	.13	.52	.17	.45	-

Anmerkung. Fett markierte Werte zeigen die höchsten Interkorrelationen zwischen den Items an.

Itembezeichnung SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Die Skalenbildung erfolgte genauso wie in der vorausgegangenen Studie. Die Teilnehmer erreichten im Durchschnitt 75.8 Punkte im *SJT*. Aufgrund der Hypothese H1 wurde der Gesamtwert im *SJT* auch nach Geschlechtern getrennt untersucht. Die männliche Stichprobe zeigte verglichen mit der weiblichen Stichprobe einen um 0.7 Punkte geringeren Punktwert und eine höhere Standardabweichung im *SJT* (Tabelle 50). Der Mittelwertsunterschied war nicht signifikant ($t(88) = 0.52$; $p = .61$) und zeigte eine zu vernachlässigende Effektstärke ($d = .11$), sodass H1 zurückgewiesen werden muss. Von der theoretisch möglichen Punktspanne von 80 Punkten wurde in der Gesamtgruppe ein Spanne von 31.8 Punkten ausgenutzt, wobei die Frauen eine geringe Punktspanne aufwiesen als die Männer (28.7 vs. 30.2). Die Subgruppe, für die eine Fremdeinschätzung der sozialen Kompetenz vorlag, zeigte einen höheren Mittelwert und ein höheres Minima als die anderen Gruppen.

Tabelle 50: *Deskriptive Statistik SJT Medizinstudierende*

	M	M-Schätzer ^a	SD	Min	Max	Schiefe	Kurtosis
gesamt	75.79	75.77	6.68	59.45	91.25	-0.04	-0.40
weiblich	76.13	76.38	6.19	59.45	88.10	-0.36	-0.14
männlich	75.40	74.90	7.23	61.10	91.25	0.22	-0.49
fremd	77.99	77.59	6.68	68.20	91.25	0.33	-0.96

Anmerkung. $N_{\text{gesamt}} = 90$; $n_w = 47$; $n_m = 43$; $n_{\text{fremd}} =$ Anzahl der Personen mit Fremdeinschätzung sozialer Kompetenzen = 23. ^a M-Schätzer nach Huber.

Der M-Schätzer unterschied sich in allen Gruppen nur minimal von dem arithmetischen Mittelwert, was darauf hindeutete, dass keine Ausreißer in den Stichproben vorlagen. Die Boxplots zeigten keine Ausreißer in der Gesamtgruppe, der männlichen Stichprobe und für die Personen mit Fremdeinschätzung an. Unter den Frauen wurde eine Studentin mit der geringsten Punktzahl im *SJT* als Ausreißer markiert. Die Schiefe entsprach in der Gesamtstichprobe fast der Normalverteilung. In der weiblichen Stichprobe war die Verteilung leicht rechtssteil. In der männlichen Gruppe, sowie bei den Personen mit Fremdeinschätzung war sie leicht linkssteil ausgeprägt. Die Kurtosis fiel in allen Gruppen flacher als die Normalverteilung aus, wobei die weibliche Stichprobe die geringste Abweichung von null zeigte. Der Kolmogorov-Smirnov-Test gab für alle Gruppen an, dass sie normalverteilt waren ($Z_{\text{ges}} = 0.07$, $p = .20$; $Z_w = 0.08$, $p = .20$; $Z_m = 0.10$, $p = .20$; $Z_{\text{fremd}} = 0.13$; $p = .20$). Die Histogramme bildeten der Normalverteilung ähnliche Häufigkeitsverteilungen ab (Abbildung 25).

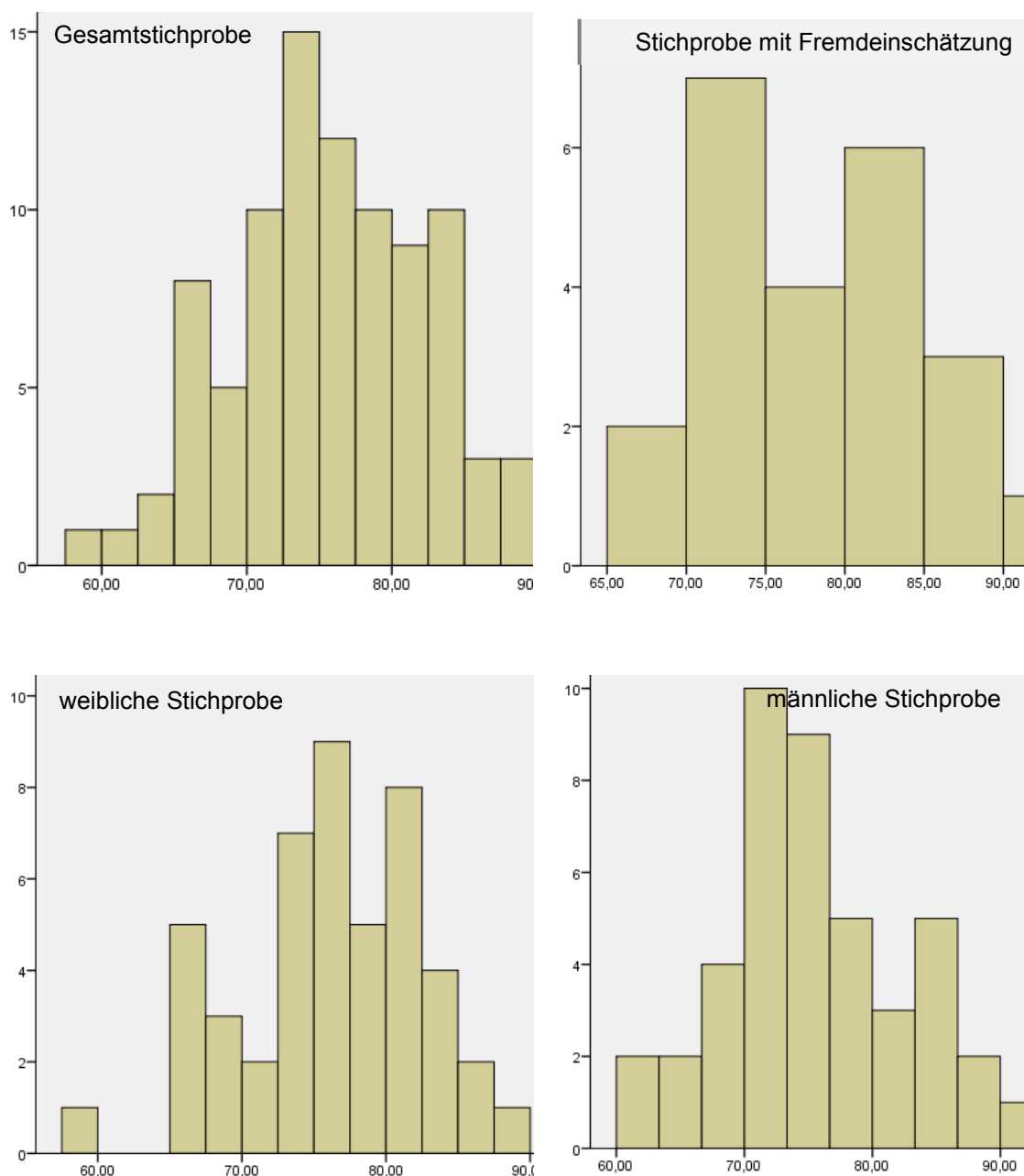


Abbildung 25. Verteilung SJT Validierungsstudie Medizinstudierende.

Abszisse = Häufigkeit; Ordinate = Punktzahl im SJT.

Die nachfolgenden Analysen wurden aufgrund der nicht vorhandenen Unterschiede zwischen Frauen und Männern nur für die Gesamtstichprobe vorgenommen. Auf die Darstellung der weiteren Item-Kennwerte für die Subgruppe der fremdeingeschätzten Probanden wurde verzichtet, da sie nur für die Gesamtgruppe von Interesse war.

Die interne Konsistenz des SJTs betrug .84. Die Trennschärfen lagen zwischen .20 und .60 (Tabelle 51). Item SU8, SU9 und SU14 wiesen die geringsten Trennschärfen auf und fielen unter das geforderte Mindestmaß von .30 (Backhaus et al., 2008).

Die interne Konsistenz würde jedoch nur marginal steigen bzw. sinken, wenn die drei Items entnommen werden würden. Die durchschnittliche Trennschärfe lag bei .41. Der Schwierigkeitsgrad lag zwischen .61 und .88 mit einer durchschnittlichen Schwierigkeit von .80. Sieben Items lagen im empfohlenen Bereich von .20-.80. Die anderen Items waren mit Schwierigkeiten > .80 sehr leicht zu lösen. Die Itemschwierigkeit unterschied sich nicht in Abhängigkeit der Antwortanzahl pro Item ($F(2,17) = 1.097$; $p = .36$). So zeigten Items mit vier Antworten ($N = 4$) eine mittlere Schwierigkeit von .78, Items mit fünf Antworten ($N = 9$) eine von .78 und Items mit sechs Antworten ($N = 7$) eine von .84.

Tabelle 51: *Trennschärfe und Itemschwierigkeit der SJT-Items Medizinstudierende*

	r_{it}	α , falls Item entfernt	p_i
SBSR5	.54	.82	.76
SBSR7	.32	.83	.82
SKSD4	.45	.83	.61
SKSD10	.49	.83	.87
SKSD11	.32	.83	.89
Kri1	.41	.83	.69
Kri2	.45	.83	.85
Kri4	.37	.83	.78
Kri7	.43	.83	.84
Kri10	.60	.82	.81
Kri12	.35	.83	.87
SU1	.48	.83	.81
SU8	.20	.84	.87
SU9	.29	.83	.75
SU10	.38	.83	.88
SU12	.43	.83	.86
SU13	.51	.82	.63
SU14	.23	.84	.81
KF6	.49	.83	.79
KF11	.55	.82	.84

Anmerkung. r_{it} = korrigierte Trennschärfe; p_i = Itemschwierigkeit.

Itembezeichnung: SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/Selbstdisziplin; Kri = Kritikfähigkeit; R = Respektfähigkeit; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

ISK. Die Medizinstudierenden beschrieben ihre sozialen Fähigkeiten als durchschnittlich (Tabelle 52), wobei die individuellen Werte zwischen weit unterdurchschnittlich ($M < 80$) bis weit überdurchschnittlich ($M > 120$) variierten. Die Primärskalen emotionale Stabilität und Internalität, sowie die Sekundärskalen Offensivität und Selbststeuerung folgten der Normalverteilung ($p \geq .20$). Alle anderen Skalen waren nicht normalverteilt ($p \leq .05$). Die internen Konsistenzen nahmen Werte zwischen .62 und .86 bei den Primärskalen und zwischen .83 und .88 bei den Sekundärskalen an. Die Skalen Prosozialität, Wertepluralismus, Selbstkontrolle, Handlungsflexibilität und Selbstdarstellung wiesen Werte $< .70$ auf.

Tabelle 52: *Deskriptive Statistik ISK*

	M	SD	Min	Max	α
soziale Orientierung	103.87	10.13	74	126	.88
Prosozialität	104.07	10.09	78	121	.68
Perspektivenübernahme	104.26	9.95	77	120	.75
Wertepluralismus	104.03	10.12	77	124	.64
Kompromissbereitschaft	102.31	11.34	70	124	.78
Zuhören	99.91	10.03	73	120	.78
Offensivität	101.84	10.75	70	127	.88
Durchsetzungsfähigkeit	101.71	10.97	73	122	.79
Konfliktbereitschaft	99.28	10.53	75	123	.82
Extraversion	102.16	10.84	74	124	.86
Entscheidungsfreudigkeit	101.76	10.75	80	121	.82
Selbststeuerung	106.27	9.00	82	122	.84
Selbstkontrolle	105.47	8.80	86	123	.64
Emotionale Stabilität	102.07	9.81	77	124	.76
Handlungsflexibilität	105.56	9.00	80	126	.62
Internalität	105.16	10.58	80	129	.76
Reflexibilität	104.29	9.10	73	126	.83
Selbstdarstellung	104.67	8.84	83	124	.66
Direkte Selbstaufmerksamkeit	102.20	10.49	70	121	.80
Indirekte Selbstaufmerksamkeit	102.74	9.43	70	122	.76
Personenwahrnehmung	102.90	10.19	71	122	.80

Anmerkung. $N = 90$. Mittelwerte entsprachen T-Werte. Sekundärskalen sind fett gedruckt.

Für die Teilnehmer, für die eine Fremdeinschätzung der sozialen Kompetenz vorlag, wurden die Mittelwerte des ISK-K verglichen (Tabelle 53). Die Fremdeinschätzung der Offensivität und Selbststeuerung fiel signifikant höher aus als die Selbsteinschätzung.

Die Beurteilung der sozialen Orientierung unterschied sich nicht in der Selbst- und Fremdeinschätzung. Die Reflexibilität wurde in der Fremdeinschätzung signifikant geringer eingeschätzt als in der Selbsteinschätzung. Die Unterschiede waren von hoher praktischer Bedeutsamkeit (Cohen, 1988). Die Werte der Sekundärskalen waren in der Selbst- und Fremdeinschätzung normalverteilt ($p \geq .20$).

Tabelle 53: *Selbst- vs. Fremdeinschätzung ISK-K*

	Selbsteinschätzung		Fremdeinschätzung		Unterschiedstestung		
	M	SD	M	SD	F	p	η^2
soziale Orientierung	3.08	0.23	2.93	0.42	2.84	.11	.11
Offensivität	2.51	0.46	2.98	0.44	29.28	.00	.57
Selbststeuerung	2.80	0.34	3.09	0.38	12.44	.00	.37
Reflexibilität	3.06	0.32	2.75	0.41	11.12	.00	.34

Anmerkung. N = 23. Skala von 1 = trifft nicht zu bis 4 = trifft zu. F = Prüfgröße der Varianzanalyse mit Messwiederholung; η^2 = Partielles Eta-Quadrat.

Es bestand ein hoher Zusammenhang in der Selbst- und Fremdeinschätzung der Offensivität ($r(23) = .56, p = .01$), ein mittlerer in der Selbststeuerung ($r(23) = .38; p = .08$) und ein niedriger in der sozialen Orientierung ($r(23) = .27; p = .21$) sowie der Reflexibilität ($r(23) = .28; p = .19$).

Selbsteinschätzungsskalen sozialer Kompetenzen. Die Medizinstudierenden beschrieben ihre sozialen Kompetenzen als höher ausgeprägt als bei Gleichaltrigen (Tabelle 54). Am höchsten schätzten sie ihre Selbstreflexionsfähigkeit ein, am niedrigsten die Fähigkeit, soziale Unterstützung aufzusuchen. Der Skalenrange wurde ausgenutzt. Die Skalenwerte waren nicht normalverteilt ($p < .05$).

Tabelle 54: *Deskriptive Statistik Selbsteinschätzungsskalen Medizinstudierende*

	M	SD	Min	Max
Selbstreflexion	5.34	1.04	2	7
Selbstkontrolle	4.74	1.18	2	7
Kritikfähigkeit	4.84	1.04	2	7
Soziale Unterstützung	3.97	1.39	2	6
Kontaktfähigkeit	4.56	1.34	1	7
Allg. soziale Kompetenz	4.77	1.13	1	7

Anmerkung. N = 85. Bewertung von 1 = geringer als bei Gleichaltrigen bis 7 = höher als bei Gleichaltrigen.

Außenkriterien sozialer Kompetenzen. Die drei Items der sozialen Eingebundenheit wurden zu einer Skala zusammengefasst, welche eine interne Konsistenz von .80 aufwies. Die anderen Außenkriterien standen für sich allein und wurden nicht zu einer Skala zusammengefasst. Die Medizinstudierenden fühlten sich gut eingebunden, konnten bei Bedarf auf soziale Unterstützung zurückgreifen und bemühten sich, soziale Netzwerke zu bilden und zu erhalten (Tabelle 55). Die Studierenden lösten 77% ihrer Konflikte in den letzten Jahren zufriedenstellend und 14% aufbrausend. In 22% der sozialen Interaktionen fühlten sie sich unsicher im Handeln. Sie verbrachten im Durchschnitt 25% ihrer Freizeit mit alleinigem Medienkonsum. Die Ausprägung aller Kriterien war normalverteilt ($p < .05$).

Tabelle 55: *Deskriptive Statistik Außenkriterien sozialer Kompetenzen*

	M	SD	Min	Max
Soziale Eingebundenheit ^a	4.17	0.79	2.00	5.00
Erlebte soziale Unterstützung ^a	4.51	0.82	1.00	5.00
Bemühen zu Netzwerken ^a	4.36	0.78	2.00	5.00
zufriedenstellend gelöste Konflikte ^b	76.90	14.79	10.00	100.00
aufbrausend gelöste Konflikte ^b	14.32	14.19	0.00	60.00
unsicher gefühlt ^b	22.45	17.32	0.00	80.00
Medienkonsum allein ^b	24.94	19.57	0.00	85.00

Anmerkung. ^a Bewertung auf einer Skala von 1 = trifft überhaupt nicht zu bis 5 = trifft voll zu.

^b Angabe in Prozent.

BFI-K. Die Medizinstudierenden beschrieben sich als durchschnittlich bis überdurchschnittlich offen gegenüber neue Erfahrungen, gewissenhaft, extravertiert und verträglich (Tabelle 56). In Neurotizismus beschrieben sie sich eher als durchschnittlich bis unterdurchschnittlich. Die Varianz der Werte fiel mit einer Streuung von maximal 0.9 Punkten gering aus. Die internen Konsistenzen der Skalen betrug .50-.80. Die Konsistenz der Verträglichkeitsskala zeigte mit .50 keinen zufriedenstellenden Wert (vgl. Hossiep et al., 2000). Bis auf Neurotizismus war die Ausprägung aller Skalen normalverteilt ($p < .05$).

Tabelle 56: *Deskriptive Statistik BFI-K*

	M	SD	Min	Max	α
Extraversion	3.67	0.86	1.25	5.00	.80
Verträglichkeit	3.19	0.73	1.25	5.00	.50
Gewissenhaftigkeit	3.77	0.74	1.50	5.00	.74
Neurotizismus	2.94	0.85	1.25	4.75	.78
Offenheit	4.07	0.69	2.00	5.00	.73

Anmerkung. Bewertung von 1 = sehr unzutreffend bis 5 = sehr zutreffend.

Die Persönlichkeitsfaktoren Verträglichkeit und Gewissenhaftigkeit wurden zusätzlich für Männer und Frauen geprüft, da sie in Zusammenhang mit Geschlechtsunterschieden im *SJT* stehen (Whetzel et al., 2008). Frauen und Männer unterschieden sich nicht in der Verträglichkeit ($M_w = 3.31$ vs. $M_m = 3.05$, $p = .10$) und Gewissenhaftigkeit ($M_w = 3.87$ vs. $M_m = 3.66$, $p = .12$).

Kognitive Kriterien. Der Abiturnotendurchschnitt lag mit 1.4 im sehr guten Bereich (Tabelle 57). Mit einem *Range* von 1.0-2.9 wurde die theoretisch mögliche Notenspanne für Studienplatzbewerber von 1.0-4.0 nicht ausgenutzt. Der mittlere Intelligenzquotient betrug 121.7, der als überdurchschnittlich bis weit überdurchschnittlich einzuordnen ist. Der TMS-Gesamtwert und dessen Subskalen lagen im Mittel im durchschnittlichen bis überdurchschnittlichen Bereich. Bis bei den TMS-Subskalen quantitative und formale Probleme, sowie konzentriertes und sorgfältiges Arbeiten, war die Ausprägung aller Skalen nicht normalverteilt ($p < .05$).

Tabelle 57: *Deskriptive Statistik kognitive Kriterien*

	N	M	SD	Min	Max
Abiturnotendurchschnitt	89	1.41	0.45	1.00	2.90
IQ (IQ-Skala)	90	121.71	12.39	95	148
TMS (T-Skala)	50	115.86	6.81	96	131
Quantitative Probleme	50	114.72	8.20	91	130
Textverständnis	50	112.48	6.12	97	125
Konzentration	49	108.51	9.08	92	124

Subjektive Studienerfolgskriterien. Die Studierenden schätzten sich als zufrieden mit dem Studium ein (Tabelle 58). Sie gaben an, dass ihre Fähigkeiten überwiegend zu den Studienanforderungen passten und waren größtenteils mit ihren Leistungen zufrieden. Die maximale Bewertungsspanne von 1-5 wurde ausgenutzt. Alle drei subjektiven Studienerfolgskriterien waren nicht normalverteilt ($p < .05$).

Tabelle 58: *Deskriptive Statistik subjektive Studienerfolgskriterien*

	M	SD	Min	Max
Studienzufriedenheit	4.30	0.91	1	5
subjektive Studienerfolg	3.65	1.19	1	5
Passung	3.90	0.95	1	5

Anmerkung. N = 90. Bewertung mit 1 = sehr unzutreffend bis 5 = sehr zutreffend.

Objektive Studienerfolgskriterien. Die Mehrheit der vorklinischen Leistungen befand sich im oberen Leistungsdrittel (Tabelle 59). Die Leistungen waren, mit Ausnahme der Zellbiologie-Prüfung, nicht normalverteilt ($p < .05$).

Tabelle 59: *Deskriptive Statistik vorklinische Studienleistung*

Studienleistung	N	M	SD	Min	Max
Osteologie (%)	71	84.83	11.59	57.10	100.00
Makroskopische Anatomie (%)	67	74.21	14.22	44.40	100.00
Humangenetik (Rangplatz)	54	21.91	2.92	13.00	26.00
Mikroskopische Anatomie, Biochemie- Molekularbiologie, Physiologie 1 (max. 90 Punkte)	56	65.36	9.36	42.00	78.00
Mikroskopische Anatomie, Biochemie- Molekularbiologie, Physiologie 2 (max. 90 Punkte)	36	71.19	8.25	54.00	84.00
Zellbiologie (max. 88 Punkte)	56	65.54	10.36	31.00	81.00

Die Mehrheit der klinischen Studienleistungen lag im sehr guten bis guten Bereich (Tabelle 60). Die Streuung der Leistungen war bei den *Encounter Cards*, den Patientenbericht der Inneren Medizin, die Präsentationsnoten der klinisch-pathologischen Konferenz, sowie die Noten der Notfall- und Palliativmedizin gering. Die Leistungen, die in Punkten angegeben wurden, waren normalverteilt ($p \leq .05$). Die Leistungen, die als Noten angegeben wurden, waren nicht normalverteilt ($p \geq .05$), mit Ausnahme der Mitarbeit in der Notfallmedizin.

Tabelle 60: *Deskriptive Statistik klinische Studienleistung*

	Studienleistung	N	M	SD	Min	Max
Propädeutik (5. FS)	Blockpraktikum Chirurgie (Note)	38	1.76	0.54	1.00	3.00
	Blockpraktikum Innere Medizin (Note)	38	1.76	0.54	1.00	3.00
	Hygiene, Mikrobiologie, Virologie (Note)	37	2.16	0.83	1.00	4.00
	Pathologie (Note)	38	1.82	0.69	1.00	3.00
	Bildgebende Verfahren (Note)	38	1.63	0.75	1.00	3.00
	Gesundheitsökonomie (Note)	38	1.58	0.76	1.00	4.00
	Prävention, Gesundheitsförderung (Note)	38	1.87	0.96	1.00	4.00
	Pharmakologie, Toxikologie (Note)	38	2.05	1.16	1.00	5.00
Block I (6. oder 7. FS)	Allgemeinmedizin (Note)	27	2.52	0.58	2.00	4.00
	Innere Medizin (Note)	27	2.52	0.58	2.00	4.00
	MC-Klausur (max. 45 Punkte)	16	36.03	5.53	21.50	43.00
	OSCE (max. 250 Punkte)	16	209.47	19.15	175.00	250.0
	Mini-CEX (Note)	15	1.50	0.62	1.00	3.00
	Encounter Cards (Note)	15	1.44	0.22	1.17	1.81
	Patientenbericht (Note)	15	1.53	0.36	1.00	2.50
	Kl. Chemie, Labordiagnostik (Note)	27	2.52	0.58	2.00	4.00
	Kl. Pharmakologie, -therapie (Note)	27	1.93	1.07	1.00	5.00
	Gerontologie (Note)	26	1.54	0.65	1.00	3.00
Block II (6. oder 7. FS)	Klausur (max. 29 Punkte)	15	25.00	2.13	21.50	29.00
	Fallbericht	15	1.50	0.64	1.00	3.00
	Anästhesiologie (Note)	29	2.31	0.60	1.00	3.00
	Chirurgie (Note)	29	2.31	0.60	1.00	3.00
	Orthopädie (Note)	29	2.31	0.60	1.00	3.00
	MC-Klausur (max. 65 Punkte)	23	54.30	3.73	46.50	61.00
	OSCE (max. 325 Punkte)	23	272.70	14.30	249.00	306.00
	Klinisch-pathologische Konferenz (Note)	29	2.31	0.79	1.00	4.00
	Klausur (max. 66 Punkte)	28	46.71	11.47	2.40	61.5
	Präsentation (Note)	28	1.06	0.57	0.50	2.30

Fortsetzung Tabelle 60:

Studienleistung		N	M	SD	Min	Max
Block II (6. oder 7. FS)	Notfallmedizin (Note)	28	1.58	0.46	1.00	2.31
	Mitarbeit (Note)	23	1.64	0.40	1.00	2.50
	Mini-CEX (Note)	23	1.32	0.41	1.00	2.00
	Klausur (max. 87.5 Punkte)	28	76.48	4.58	66.50	83.50
	Palliativmedizin (Note)	21	1.57	0.50	1.00	2.00
	Rehabilitation, Naturheilverfahren (Note)	28	2.36	0.95	0.90	5.00
	Urologie (Note)	29	2.31	0.60	1.00	3.00

Anmerkung. Note von 1.0-5.0. MC-Klausur = Multiple Choice Klausur; Kl. = klinische.

8.5.2. konvergente und diskriminante Validität des SJTs

Zusammenhang ISK und SJT. Der SJT zeigte signifikante Korrelationen mit den Sekundärskalen soziale Orientierung und Offensivität in der Gesamtgruppe, aber nicht mit den Sekundärskalen Selbststeuerung und Reflexibilität (Tabelle 61).

Tabelle 61: *Zusammenhang ISK und SJT*

	Selbsteinschätzung Gesamtgruppe		Selbsteinschätzung Subgruppe		Fremdeinschätzung Subgruppe	
	r	r_s	r	r_s	r	r_s
	soziale Orientierung	.40**	.41**	.38	.41	.10
Prosozialität	.28**	.30**				
Perspektivenübernahme	.32**	.30**				
Wertpluralismus	.33**	.27*				
Kompromissbereitschaft	.23*	.17				
Zuhören	.32**	.34**				
Offensivität	.30**	.31**	.43*	.48*	.33	.31
Durchsetzungsfähigkeit	.15	.14				
Konfliktbereitschaft	.07	.10				
Extraversion	.36**	.37**				
Entscheidungsfreudigkeit	.23*	.25*				
Selbststeuerung	.09	.11	.17	.21	.20	.25
Selbstkontrolle	-.04	-.05				
Emotionale Stabilität	-.03	-.05				
Handlungsflexibilität	.07	.13				
Internalität	.22*	.25*				

Fortsetzung Tabelle 61

	Selbsteinschätzung		Selbsteinschätzung		Fremdeinschätzung	
	Gesamtgruppe		Subgruppe		Subgruppe	
	r	r _s	r	r _s	r	r _s
Reflexibilität	.01	.04	.09	.14	.31	.35
Selbstdarstellung	-.16	-.15				
Direkte Selbstaufmerksamkeit	-.10	-.14				
Indirekte Selbstaufmerksamkeit	.07	.11				
Personenwahrnehmung	.22*	.18				

Anmerkung. N_{Gesamtgruppe} = 90; N_{Subgruppe} = 23. r_s = Korrelationskoeffizient nach Spearman; * p ≤ .05; ** p ≤ .01. Sekundärskalen sind fett gedruckt.

Auf Primärskalenniveau korrelierte der *SJT* in der Gesamtgruppe signifikant mit der Prosozialität, der Perspektivenübernahme, dem Wertepluralismus, der Kompromissbereitschaft, dem Zuhören, der Extraversion, der Entscheidungsfreudigkeit, der Internalität und der Personenwahrnehmung. Nach der Alpha-Fehler-Korrektur blieben die Zusammenhänge zu den vier Primärskalen Perspektivenübernahme, Wertepluralismus, Zuhören und Extraversion signifikant. Die höchste Korrelation auf Primärskalenniveau bestand zwischen dem *SJT* und Extraversion. Die meisten bedeutsamen Korrelationen zeigte der *SJT* zu den Primärskalen der sozialen Orientierung. Zu den Primärskalen der Selbststeuerung und der Reflexibilität bestanden keine Zusammenhänge. In der Subgruppe trat eine signifikante Korrelation mittlerer Höhe zwischen dem *SJT* und der selbsteingeschätzten Offensivität auf. Die Zusammenhänge zu den anderen selbsteingeschätzten Sekundärskalen waren nicht signifikant. Zwischen der Fremdeinschätzung sozialer Kompetenzen und dem *SJT* traten keine bedeutsamen Korrelationen auf. Die Korrelationshöhe fiel jedoch verglichen zu den Selbsteinschätzungsdaten beider Gruppen für die soziale Orientierung niedriger und für die Reflexibilität höher aus. Die Korrelationshöhe des *SJT*s mit der fremdeingeschätzten Offensivität-Skala war vergleichbar mit der Korrelation in der Gesamtgruppe. Somit kann H2 nur für die zwei Sekundärskalen soziale Orientierung und Offensivität, sowie den Primärskalen Perspektivenübernahme, Wertepluralismus, Zuhören und Extraversion in der Selbsteinschätzung bestätigt werden. Für die Sekundärskalen Selbststeuerung und Reflexibilität, sowie deren Subskalen in der Selbsteinschätzung, muss die Annahme widerlegt werden.

Ebenfalls konnte der Zusammenhang zwischen der Fremdeinschätzung sozialer Kompetenzen und dem *SJT* nicht nachgewiesen werden.

Zusammenhang Selbsteinschätzungsskalen sozialer Kompetenzen und SJT.

Der *SJT* korrelierte niedrig mit der selbsteingeschätzten Kontaktfähigkeit und der Suche nach sozialer Unterstützung (Tabelle 62). Die Zusammenhänge verloren nach der Bonferroni-Holm-Korrektur ihre Signifikanz. Der *SJT* zeigte positive Tendenzen im Zusammenhang mit der Kritikfähigkeit, der Selbstkontrolle und der allgemeinen sozialen Kompetenz. Es bestand eine Nullkorrelation zur Fähigkeit zur Selbstreflexion. Das parametrische und das nicht-parametrische Verfahren zeigten vergleichbare Ergebnisse. Somit kann die Annahme H3 nicht bestätigt werden.

Tabelle 62: *Zusammenhang Selbsteinschätzungsskalen und SJT*

	r	r _s
Selbstreflexion	.05	.02
Selbstkontrolle	.16	.15
Kritikfähigkeit	.13	.10
soziale Unterstützung	.24*	.23*
Kontaktfähigkeit	.24*	.23*
allg. soziale Kompetenz	.18	.10

Anmerkung. r_s = Korrelationskoeffizient nach Spearman; * p ≤ .05.

Zusammenhang Außenkriterien sozialer Kompetenzen und SJT. Der *SJT* korrelierte in mittlerer Höhe mit der sozialen Eingebundenheit und der erlebten sozialen Unterstützung im Studium (Tabelle 63). Diese Zusammenhänge blieben auch nach der Alpha-Fehler-Korrektur bestehen. Mit dem Bemühen zu Netzwerken zeigte der *SJT* einen geringen Zusammenhang, der nach der Alpha-Fehler-Korrektur nicht mehr signifikant war. Der *SJT* zeigte eine positive Tendenz mit dem Prozentsatz an zufriedenstellend gelösten sozialen Konflikten. Negative Tendenzen traten zwischen dem *SJT* und dem Anteil an sozialen Interaktionen, in denen sich die Person unsicher gefühlt hat, sowie dem alleinigen Medienkonsum auf. Ein Nullzusammenhang zeigte der *SJT* zu dem Prozentsatz an aufbrausend gelösten sozialen Konflikten. Das parametrische und das nicht-parametrische Verfahren wiesen vergleichbare Ergebnisse auf.

Tabelle 63: *Zusammenhang Außenkriterien und SJT*

	r	r_s
Soziale Eingebundenheit	.39**	.38**
Erlebte soziale Unterstützung	.30**	.31**
Bemühen zu Netzwerken	.23*	.24*
zufriedenstellend gelöster Konflikte	.15	.11
aufbrausend gelöster Konflikte	.01	.01
unsicher gefühlt	-.20	-.16
Medienkonsum allein	-.12	-.08

Anmerkung. r_s = Korrelationskoeffizient nach Spearman; * $p \leq .05$; ** $p \leq .01$.

Fett gedruckte Werte blieben auch nach der Bonferroni-Holm-Korrektur signifikant.

Auf Basis der Ergebnisse kann H4 für die erlebte soziale Unterstützung und die soziale Eingebundenheit im Studium, aber nicht für das Bemühen zu Netzwerken sowie für das Ausmaß an friedfertig gelösten Konflikten bestätigt werden. H5 muss aufgrund nicht signifikanter negativer Korrelationen zum alleinigen Medienkonsum, dem Ausmaß an aufbrausend gelösten Konflikten und der Unsicherheit in sozialen Interaktionen abgelehnt werden.

Zusammenhang BFI-K und SJT. Der *SJT* zeigte niedrige Korrelationen zu Extraversion und Gewissenhaftigkeit, wobei letztere nach der Bonferroni-Holm-Korrektur keine Bedeutsamkeit mehr aufwies (Tabelle 64). Der *SJT* korrelierte nicht mit Verträglichkeit, Offenheit für neue Erfahrungen und Neurotizismus. Das parametrische und das nicht-parametrische Verfahren zeigten identische Ergebnisse. Somit musste bis auf Ausnahme der Extraversion H6 und H7 abgelehnt werden.

Tabelle 64: *Zusammenhang BFI-K und SJT*

	r	r_s
Extraversion	.29**	.29**
Verträglichkeit	.04	.04
Gewissenhaftigkeit	.25*	.25*
Neurotizismus	-.10	-.10
Offenheit	-.08	-.08

Anmerkung. r_s = Korrelationskoeffizient nach Spearman; * $p \leq .05$; ** $p \leq .01$.

Fett gedruckte Werte blieben auch nach der Bonferroni-Holm-Korrektur signifikant.

Zusammenhang kognitive Kriterien und SJT. Der *SJT* zeigte eine positive Tendenz zum Abiturnotendurchschnitt (Tabelle 65). Nullkorrelationen traten zum Intelligenzquotienten, zum TMS-Gesamtwert, sowie zu den drei TMS-Subskalen auf. Das parametrische und das nicht-parametrische Verfahren zeigten vergleichbare Ergebnisse. Somit konnte bestätigt werden, dass der *SJT* nicht mit den kognitiven Kriterien korrelierte (H8-H13).

Tabelle 65: *Zusammenhang kognitive Kriterien und SJT*

	r	r _s
Abiturnotendurchschnitt	.16	.20
IQ	.03	.03
TMS	.04	.02
Quantitative Probleme	-.07	-.10
Textverständnis	.07	.07
Konzentration	-.11	-.08

Anmerkung. r_s = Korrelationskoeffizient nach Spearman.

Zusammenhang Studienerfolgskriterien und SJT. Der *SJT* korrelierte in mittlerer Höhe mit der Studienzufriedenheit und der Passung zum Studium (Tabelle 66), sodass H14 bestätigt werden konnte. Mit dem subjektiven Studienerfolg bestanden keine Zusammenhänge, sodass auch H15 bestätigt wurde. Das parametrische und das nicht-parametrische Verfahren zeigten vergleichbare Ergebnisse.

Tabelle 66: *Zusammenhang subjektive Studienerfolgskriterien und SJT*

	N	r	r _s
Studienzufriedenheit	90	.31**	.30**
subjektive Studienerfolg	90	.05	.06
Passung	90	.28**	.30**

Anmerkung. r_s = Korrelationskoeffizient nach Spearman; ** p ≤ .01.

Aus Tabelle 67 kann entnommen werden, dass der *SJT* keinen Zusammenhang mit den vorklinischen Studienleistungen zeigte.

Tabelle 67: *Zusammenhang vorklinische Studienleistungen und SJT*

Studienleistung	N	r	r _s
Osteologie (%)	71	-.15	-.14
Makroskopische Anatomie (%)	67	.19	.19
Humangenetik (Rangplatz)	54	.12	.11
Mikroskopische Anatomie, Biochemie- Molekularbiologie, Physiologie 1 (Punkte)	56	.10	.13
Mikroskopische Anatomie, Biochemie- Molekularbiologie, Physiologie 2 (Punkte)	36	.12	.20
Zellbiologie (Punkte)	56	.04	.09

Anmerkung. r_s = Korrelationskoeffizient nach Spearman.

Bei den klinischen Studienleistungen zeigten sich negative Tendenzen zu den meisten Leistungen, die in Noten angegeben und positive Tendenzen zu den meisten Leistungen, die in Punkten angegeben wurden (Tabelle 68). Es lag eine signifikant hohe Korrelation zwischen der *SJT*-Skala und der Note im *Mini-CEX* vor. Die mittleren bis hohen Tendenzen zwischen dem *OSCE* und den *Encounter Cards* erreichten keine Signifikanz.

Aufgrund der geringen Stichprobengrößen von 15-38 an Stelle der poweranalytisch notwendigen 68 und den nur wenigen signifikanten Zusammenhängen wurde auf Regressionsanalysen zur Untersuchung der prädiktiven Validität verzichtet. Insgesamt konnte H16 nur insoweit bestätigt werden, dass der *SJT* keine signifikanten Zusammenhänge zu den vorklinischen Noten zeigte. Für die klinischen Noten konnte nur ein signifikanter Zusammenhang gefunden werden, sodass die Hypothese nicht bestätigt werden konnte.

Tabelle 68: Zusammenhang klinische Studienleistungen und SJT

	Studienleistung	N	r	r _s
Propädeutik (5. FS)	Blockpraktikum Chirurgie (Note)	38	-.05	-.05
	Blockpraktikum Innere Medizin (Note)	38	-.05	-.05
	Hygiene, Mikrobiologie, Virologie (Note)	37	-.20	-.21
	Pathologie (Note)	38	-.25	-.26
	Bildgebende Verfahren (Note)	38	-.22	-.24
	Gesundheitsökonomie (Note)	38	-.10	-.13
	Prävention, Gesundheitsförderung (Note)	38	-.20	-.07
	Pharmakologie, Toxikologie (Note)	38	-.01	-.10
Block I (6. oder 7. FS)	Allgemeinmedizin (Note)	27	-.04	-.10
	Innere Medizin (Note)	27	-.04	-.10
	MC-Klausur (Punkte)	16	.15	.24
	OSCE (Punkte)	15	.31	.44
	Mini-CEX (Note)	15	-.78**	-.77**
	Encounter Cards (Note)	15	-.50	-.40
	Patientenbericht (Note)	15	-.28	-.21
	Kl. Chemie, Labordiagnostik (Note)	27	-.04	-.10
	Kl. Pharmakologie, -therapie (Note)	27	.21	.20
	Gerontologie (Note)	26	-.29	-.35
	Klausur (Punkte)	15	-.05	.05
	Fallbericht	15	-.04	-.08
	Block II (6. oder 7. FS)	Anästhesiologie (Note)	29	-.08
Chirurgie (Note)		29	-.08	-.17
Orthopädie (Note)		29	-.08	-.17
MC-Klausur (Punkte)		23	-.04	-.03
OSCE (Punkte)		23	.21	.29
Klinisch-pathologische Konferenz (Note)		29	.06	-.09
Klausur (Punkte)		28	-.01	-.01
Präsentation (Note)		28	-.06	-.12
Notfallmedizin (Note)		28	.12	-.03
Mitarbeit (Note)		23	.26	.30
Mini-CEX (Note)		23	.26	.14
Klausur (Punkte)		28	-.26	-.19
Palliativmedizin (Note)		21	-.12	-.14
Rehabilitation, Naturheilverfahren (Note)		28	-.06	-.17
Urologie (Note)		29	-.08	-.17

Anmerkung. r_s = Korrelationskoeffizient nach Spearman; **p ≤ .01.

Zusammenhang Kontrollvariablen und SJT. Der *SJT* korrelierte nicht mit dem Alter ($r(90) = .05$), Deutsch als Erstsprache ($r(90) = -.08$), der medizinischen Vorbildung ($r(90) = -.07$) und dem Fachsemester ($r(90) = -.17$). Wurden die Studierenden anhand des Fachsemesters den vorklinischen oder klinischen Studienabschnitt zugeteilt, zeigten sich keine signifikanten Unterschiede in der Punktzahl im *SJT* ($M_{\text{Vorklinik}} = 76.60$ vs. $M_{\text{Klinik}} = 74.94$; $t(86) = 1.17$; $p = .25$). Die Effektstärke fiel gering aus ($d = .25$).

8.5.3. Prädiktive und inkrementelle Validität des SJTs

Vor der Ergebnisdarstellung der Regressionsanalysen werden die Ergebnisse der Voraussetzungsprüfung berichtet. Die statistische Unabhängigkeit der Residuen war gegeben, da die Durbin-Watson-Statistik bei der Studienzufriedenheit (1.97), dem subjektiven Studienerfolg (1.82) und der Passung zum Studium (2.12) Werte zwischen 1.5 und 2.5 annahm. Die Homoskedastizität der Prädiktoren war für die lineare und multiple Regression erfüllt. Die Streuung der Schätzfehler verlief konstant mit steigenden Kriterien (Anhang H-1, H-2). Die Normalverteilung war für den *SJT* und den TMS, aber nicht für die Kontrollvariablen, dem Abiturnotendurchschnitt und den abhängigen Variablen erfüllt. Da geringfügige Verletzungen der bivariaten Normalverteilung zu tolerierbaren Verzerrungen in den Ergebnissen führen (Rudolf & Müller, 2004), konnte die Regressionsanalyse durchgeführt werden.

Zur Analyse der prädiktiven Validität wurde der *SJT* als alleiniger Prädiktor in die Regression aufgenommen. Der *SJT* klärte 10% der Varianz der Studienzufriedenheit und 8% der Varianz der Passung zum Studium auf (Tabelle 69). Für die Zufriedenheit mit der eigenen Leistung im Studium (subjektiver Studienerfolg) lieferte der *SJT* keinen Vorhersagewert. Die Schätzfehler waren gering.

Tabelle 69: *Prädiktive Validität des SJTs für den subjektiven Studienerfolg*

Abhängige Variable	β	SE	R^2	korr. R^2
Studienzufriedenheit	.31**	0.01	.10**	.09
subjektiver Studienerfolg	.05	0.02	.00	-.01
Passung	.28**	0.02	.08**	.07

Anmerkung. $N = 90$; β = Beta-Gewicht entspricht hier der Produkt-Moment-Korrelation; SE = Standardfehler des Schätzers; R^2 = Bestimmtheitsmaß; korr. R^2 = korrigiertes Bestimmtheitsmaß.

Zur Untersuchung der inkrementellen Validität wurde analysiert, inwieweit der *SJT* neben soziodemografischen Variablen, dem Abiturnotendurchschnitt und dem TMS einen zusätzlichen Erklärungsbeitrag für die Studienzufriedenheit, dem subjektiven Studienerfolg und der Passung zum Studium leistet (Tabelle 70).

Tabelle 70: *Inkrementelle Validität des SJTs für den subjektiven Studienerfolg*

AV	Prädiktoren	B	SE	β	R^2	korr. R^2	ΔR^2
Studienzufriedenheit	Block 1				.05	-.03	.05
	Block 2				.07	-.04	.02
	Abi	.39	0.40	.19			
	Block 3				.08	-.05	.01
	TMS	-.01	0.02	-.08			
	Block 4				.19	.05	.11*
	<i>SJT</i>	.05	0.02	.36*			
subjektiver Studienerfolg	Block 1				.02	-.07	.02
	Block 2				.14	.04	.12*
	Abi	-1.40	0.56	-.47*			
	Block 3				.15	.03	.01
	TMS	.01	0.03	.08			
	Block 4				.19	.05	.04
	<i>SJT</i>	.04	0.03	.21			
Passung	Block 1				.12	.05	.12
	Block 2				.14	.04	.01
	Abi	.40	0.47	.16			
	Block 3				.14	.02	.00
	TMS	.00	0.02	.02			
	Block 4				.27*	.15	.13**
	<i>SJT</i>	.06	0.02	.39**			

Anmerkung. N = 50. B = nicht standardisierter Regressionskoeffizient; SE = Standardfehler des Schätzers; β = standardisierter Regressionskoeffizient; R^2 = Bestimmtheitsmaß; ΔR^2 = Änderung im Bestimmtheitsmaß; R^2 korr. = korrigiertes Bestimmtheitsmaß.

* $p < .05$, ** $p < .01$. Signifikante Werte wurden fett markiert.

Block 1 = Kontrollvariablen Alter, Geschlecht, Deutsch als Erstsprache, medizinische Ausbildung;

Block 2 = Kontrollvariablen, Abiturnotendurchschnitt (Abi);

Block 3 = Kontrollvariablen, Abi, Test für Medizinische Studiengänge (TMS);

Block 4 = Kontrollvariablen, Abi, TMS, *SJT*.

Die Studienzufriedenheit konnte durch keinen der Prädiktoren signifikant vorhergesagt werden. Zwar waren das Beta-Gewicht ($\beta = .36$, $p \leq .05$) und die Änderung im Bestimmtheitsmaß ($\Delta R^2 = .11$, $p \leq .05$) bei Aufnahme des *SJT*s signifikant, aber das Gesamtmodell zeigte keine Signifikanz. Beim subjektiven Studienerfolg führte die Aufnahme des *SJT*s zu keinen signifikanten Kenngrößen. Die Passung zum Medizinstudium konnte durch die Kombination der Kontrollvariablen, dem Abiturnotendurchschnitt, dem TMS und dem *SJT* zu 27% vorhergesagt werden, wobei der *SJT* einen inkrementellen Beitrag von 13% lieferte. Der *SJT* wies bei den Regressionsanalysen einen geringen Schätzfehler auf.

Für die Interpretation der Ergebnisse wurde zudem berücksichtigt, ob die Prädiktoren untereinander hoch korrelieren (Multikollinearität; Tabelle 71), um mögliche Redundanz- und Subpressionseffekte zu erkennen. Der *SJT*, der TMS und das Geschlecht korrelierten mit keinen der anderen Prädiktoren. Der Abiturnotendurchschnitt korrelierte positiv mit dem Alter und negativ mit einer der abgeschlossenen medizinischen Ausbildung und Deutsch als Erstsprache. Das Alter korrelierte außer mit dem Abiturnotendurchschnitt mit der medizinischen Ausbildung negativ.

Tabelle 71: *Interkorrelation der Prädiktoren*

	1	2	3	4	5	6	7
Abi (1)	-						
TMS (2)	.08	-					
<i>SJT</i> (3)	.16	.04	-				
Alter (4)	.53**	.03	.06	-			
Geschlecht (5)	.16	.01	-.06	.08	-		
Erstsprache (6)	-.27*	-.14	-.08	-.11	-.01	-	
Ausbildung (7)	-.62**	.11	-.03	-.40**	.04	.08	-

Die Kollinearitätsstatistik zeigte für alle Prädiktoren außer dem Abiturnotendurchschnitt und der medizinischen Ausbildung hohe Toleranzwerte und niedrige Varianzinflationsfaktoren (Tabelle 72), was darauf hindeutet, dass keine Multikollinearität zwischen den Prädiktoren vorliegt. Zwischen dem Abiturnotendurchschnitt und der medizinischen Ausbildung könnten Suppressionseffekte bestehen, was jedoch für die Untersuchung der inkrementellen Validität des *SJT*s keine Bedeutung hat und deshalb nicht tiefer analysiert wurde.

Tabelle 72: *Kollinearitätsstatistik*

Modell		Toleranz	Varianzinflationsfaktoren
Block 1	Geschlecht	.92	1.08
	Alter	.91	1.10
	Ausbildung	.89	1.12
	Erstsprache	.97	1.03
Block 2	Geschlecht	.92	1.08
	Alter	.79	1.27
	Ausbildung	.64	1.56
	Erstsprache	.95	1.05
	Abinote	.55	1.81
Block 3	Geschlecht	.92	1.08
	Alter	.79	1.27
	Ausbildung	.62	1.62
	Erstsprache	.93	1.07
	Abinote	.54	1.85
	TMS	.94	1.07
Block 4	Geschlecht	.89	1.12
	Alter	.79	1.27
	Ausbildung	.61	1.63
	Erstsprache	.87	1.15
	Abinote	.54	1.86
	TMS	.93	1.08
	SJT	.88	1.13

8.6. Diskussion

In den folgenden Abschnitten wird zuerst auf Einschränkungen hinsichtlich der Stichprobe eingegangen, bevor die Deskriptiva und Reliabilitäten des *SJT*s diskutiert werden. Darauf folgt die Auseinandersetzung mit den Ergebnissen zur Validität des *SJT*s. Diese werden in der Reihenfolge der Forschungshypothesen kommentiert.

Stichprobe. Einschränkungen hinsichtlich der Stichprobe könnten in der Selektivität liegen. So könnten sich bei der Rekrutierung insbesondere die Studierenden angesprochen gefühlt haben, die ein höheres Bewusstsein für soziale Kompetenzen zeigten und über höher ausgeprägte soziale Kompetenzen verfügten. Ein Indiz dafür könnten die hohen Werte auf den Selbsteinschätzungsskalen sein. Dagegen sprechen jedoch

die im Mittel durchschnittlich ausfallenden Werte des ISKs. Das Geschlechterverhältnis der Stichprobe spiegelt gut das Geschlechterverhältnis der Heidelberger Studierenden wieder. Der Stichprobenumfang entsprach den in den Power-Analysen berechneten Stichprobenumfängen für Korrelations- und Regressionsanalysen. Eine Ausnahme bildeten die objektiven Studienerfolgskriterien operationalisiert über die Studiennoten, die für viele Teilnehmer insbesondere aus den höheren Semestern nicht vorlagen. Für die Unterschiedstestungen war die tatsächliche Stichprobengröße etwas unterhalb der notwendigen Stichprobengröße von 102 Personen einzuordnen. Für die Fremdeinschätzung der sozialen Kompetenzen konnten nur 39 Personen trotz wiederholter Aufforderung gewonnen werden. Somit lag nur für einen kleinen Teilnehmerkreis eine Fremdeinschätzung vor. Die Ergebnisse sollten daher mit Vorsicht interpretiert werden.

Deskriptiva und Reliabilität des SJTs. Die Werte der Medizinstudierenden befanden sich eher in der oberen Hälfte der *SJT*-Skala. Der maximale Punktwert wurde jedoch nicht erreicht, sodass die Differenzierungsfähigkeit im höheren Bereich nicht zwangsläufig eingeschränkt sein muss. Die Medizinstudierenden erzielten verglichen mit den Studienplatzbewerbern und -interessenten durchschnittlich fünf Punkte weniger im *SJT*. Dies entspricht dem Forschungsbefund, dass Bewerber eine höhere Punktzahl erreichen als Stelleninhaber (McDaniel & Nguyen, 2001). Die interne Konsistenz, die Trennschärfen und Itemschwierigkeiten des *SJTs* fielen vergleichbar zu den Werten der Studienplatzbewerber-Stichproben aus. Auch konnte bestätigt werden, dass die Itemschwierigkeit unabhängig von der Antwortanzahl pro Item war. Für eine Diskussion der Reliabilität wird daher auf Kapitel 7.6 verwiesen.

Validität des SJTs. Zunächst wird das Vorgehen der Validierung diskutiert, bevor der Fokus auf die spezifischen Hypothesen gelenkt wird. Bei der Hypothesentestung wurde im Allgemeinen streng vorgegangen. Neben den parametrischen Verfahren wurden nicht-parametrische Tests zur statistischen Absicherung durchgeführt. Zudem wurden Alpha-Fehler-Korrekturen vorgenommen. Somit wurden mit höherer Wahrscheinlichkeit Alternativhypothesen zu Gunsten der Nullhypothesen abgelehnt. Ein hoher β -Fehler kann jedoch für die Untersuchung als angemessener beurteilt werden als ein hoher α -Fehler.

Die Prüfung der konvergenten Validität beruhte ausschließlich auf Selbst- und Fremdeinschätzungsverfahren. Inwieweit diese fähig sind, Kompetenzen zu erheben, kann angezweifelt werden. Streng genommen wird mit diesen Verfahren nicht eine Kompetenz erhoben, sondern nur die Einschätzung einer Kompetenz. Soziale Kompetenzen äußern sich letztendlich erst im konkret beobachtbaren Verhalten (Erpenbeck,

1997; Kanning, 2003; Riemann & Allgöwer, 1993). Zudem sind Einschätzungsverfahren subjektabhängig. Es ist unklar, ob die Personen ihre Einstellungen, Handlungsweisen oder Fähigkeiten beurteilt haben (Seyfried, 1995). Auffallend war, dass sich die Studieninteressenten und -bewerber besser in den sozialen Kompetenzen via *SJT* und Selbsteinschätzungsskalen einschätzten als die Studierenden. Vermutlich sind dafür positive Selbstdarstellungstendenzen verantwortlich (Sonntag & Schäfer-Rauser, 1993). Die Selbstdarstellungstendenzen sind nach Kanning (2009) nicht als grundlegend schlecht zu bewerten. Vielmehr integriert er sie als eine Facette der sozialen Kompetenz. Darüber hinaus kann das „Problem“ umgangen werden, indem als Normierungsstichprobe Bewerber statt Studierende herangezogen werden. Eine andere mögliche Erklärung für die höher eingeschätzten sozialen Kompetenzen der Studieninteressenten und -bewerber ist, dass sich Personen mit schwächeren Leistungen oft besser einschätzen als Personen mit hohen Leistungen (Kruger & Dunning, 1999). Hier kann nur spekuliert werden, dass die Studierenden über höhere soziale Kompetenzen aufgrund der im Studium integrierten Kommunikationstrainings etc. verfügten als die Studienplatzbewerber. Bergmann (2007) argumentiert, dass Kompetenzmessungen durch Selbsteinschätzung durchaus valide sind, wenn sie anonym, dimensions- und verhaltensorientiert erhoben werden und die Person Erfahrung in der Selbsteinschätzung mitbringt. Die Anonymität wurde in der vorliegenden Stichprobe gewährleistet. Die Dimensions- und Verhaltensorientierung war beim ISK, aber nicht bei den globalen Selbsteinschätzungsskalen und den Außenkriterien sozialer Kompetenzen gegeben. Ob Studierende erfahren mit Selbsteinschätzungen waren, kann nicht belegt werden. Vermutlich waren die Studierenden aus höheren Fachsemestern vertrauter in der Selbsteinschätzung als die aus frühen Fachsemestern, da sie sich in der medizinischen Lehre wiederholt in simulierten klinischen Settings einschätzen müssen. Trotz der positiven Argumentationen für die Validität von Selbsteinschätzungsverfahren (Bergmann, 2007), ist dennoch die Verhaltensbeobachtung die adäquateste Methode zur Erhebung von Kompetenzen (Goldfried & D’Zurilla, 1969; Kauffeld et al., 2002). Diese sollte in zukünftigen Studien Verwendung finden.

Die Prüfung der diskriminanten Validität erfolgte über subjektive und objektive Kriterien mit unterschiedlichem Aggregationsniveau. Während der Abiturnotendurchschnitt ein Leistungsaggregat der zwei letzten Schuljahre darstellt, sind der Intelligenztest und der TMS punktuelle Leistungsmessungen. Die Studienleistungen setzten sich aus punktuellen und aus aggregierten Leistungen über ein Semester zusammen. Darüber hinaus wurde die Zufriedenheit mit den eigenen Studienleistungen als subjektives Kriterium erhoben. Die diskriminante Validierung war demnach methodisch vielfältig gestaltet.

Geschlechtereffekt. Frauen erzielten einen ca. einen Punkt höheren *SJT*-Wert als Männer. Der Unterschied war jedoch nicht signifikant. Dieses Ergebnis stützt das Resultat der vorherigen Studie an Studienplatzbewerbern und -interessenten. Somit scheint der entwickelte *SJT* genderfair zu sein. Der nicht aufgetretene Geschlechtsunterschied könnte darauf zurückgeführt werden, dass der *SJT* keinen Zusammenhang zu den Persönlichkeitseigenschaften Gewissenhaftigkeit und Verträglichkeit aufwies. Diese Persönlichkeitseigenschaften moderieren gewöhnlich den Geschlechtereffekt im *SJT*, da Frauen meist eine höhere Ausprägung in diesen Eigenschaften aufweisen als Männer (Whetzel et al., 2008). Die Medizinstudierenden zeigten keine Geschlechterunterschiede für Verträglichkeit und Gewissenhaftigkeit, sodass der Moderator-Effekt vermutlich nicht zum Tragen gekommen ist.

Zusammenhang ISK und SJT. Der *SJT* zeigte signifikante Zusammenhänge mit der Selbsteinschätzung der Sekundärskala soziale Orientierung ($r(90) = .40$) und deren drei von fünf Subskalen Perspektivenübernahme ($r(90) = .32$), Wertepluralismus ($r(90) = .33$) und Zuhören ($r(90) = .32$) in der Gesamtgruppe. Zu den zwei übrigen Subskalen Prosozialität und Kompromissbereitschaft lagen positive Tendenzen vor. Somit scheint der *SJT* eher das Ausmaß zu messen, inwieweit eine Person anderen offen und mit positiver Grundhaltung gegenüber tritt. Dies beinhaltet nach den Skalenbeschreibungen von Kanning (2009) die Fähigkeiten, sich für andere Menschen einzusetzen, aufmerksam zuzuhören, andere Perspektiven einzunehmen, andere Meinungen zu tolerieren und Kompromisse anzustreben. Des Weiteren erfasst der *SJT* vermutlich die Offensivität einer Person ($r(90) = .30$), die im ISK als Fähigkeit beschrieben wird, aus sich heraus zu gehen und im Kontakt mit anderen Menschen eigene Interessen aktiv verwirklichen zu können (Kanning, 2009). Dabei korrelierte der *SJT* am höchsten mit der Primärskala Extraversion ($r(90) = .36$). Dies lässt schlussfolgern, dass der *SJT* von den Facetten der Offensivität am meisten wiedergibt, wie aktiv eine Person auf andere zugeht und soziale Kontakte knüpft. Die Skalen Extraversion und Offensivität ähneln den Konstrukten Kontaktfähigkeit und Suche nach sozialer Unterstützung, die bei der *SJT*-Entwicklung zu Grunde lagen und mit elf Items in den *SJT* einfließen. Zur Entscheidungsfreudigkeit lag eine positive Tendenz vor. Dies ist naheliegend, da der *SJT* von den Probanden verlangte, zwischen vier bis sechs Verhaltensalternativen zu priorisieren. Die situative Entscheidungsfähigkeit ist auch nach den Theorien des Tacit Knowledge (Sternberg, 1997, 1999) und des prozeduralen Wissens (Motowidlo et al., 2006) ein Bestandteil von *SJT*s. Mit der Sekundärskala Selbststeuerung bestand kein Zusammenhang zum *SJT*, obwohl die Definition der untergeordneten Primärskalen Selbstkontrolle und emotionale Stabilität dem bei der *SJT*-Entwicklung zugrunde lie-

genden Konstrukt Selbstkontrolle/Selbstdisziplin ähnelte. Vermutlich ist das Konstrukt mit nur drei Items unterrepräsentiert oder die Items sind nicht in der Lage Selbstkontrolle/Selbstdisziplin adäquat wiederzugeben. Der *SJT* scheint zudem keine sozialreflexiven Eigenschaften zu erfassen, da er keine Korrelation zu der Sekundärskala Reflexibilität aufwies. Die zwei Items, die ursprünglich auf Basis des Konstrukts Selbstbeobachtung/Selbstreflexion entwickelt wurden, können das Konstrukt vermutlich nicht ausreichend abbilden. Zwischen der Primärskala Personenwahrnehmung und dem *SJT* bestand eine positive Tendenz. Dies ist insoweit erklärbar, dass die Teilnehmer die Personen im Video beobachten mussten, um dem Handlungsstrang zu folgen und das Item beantworten zu können. Die negative Tendenz zu der Primärskalen Selbstdarstellung könnte einen Hinweis darauf geben, dass es den Teilnehmern nicht daran gelegen war, einen positiven Eindruck zu erzeugen. Zusammengefasst misst der *SJT* eher soziale Kompetenzen, die das Zugehen auf andere Menschen abbilden, aber keine reflexiven sozialen Kompetenzen. Verglichen mit dem Arbeitsmodell sozialer Kompetenzen (Kapitel 3.4) scheint der *SJT* eher die inter-, aber nicht die intrapersonellen sozialen Kompetenzen zu erfassen. Einschränkungen in den Korrelationshöhen könnten darauf zurückzuführen sein, dass der *SJT* eine eingeschränkte Punktspanne aufwies. Die ISK-Skalen Prosozialität, Wertepluralismus, Selbstkontrolle, Handlungsflexibilität und Selbstdarstellung zeigten zudem Einschränkungen in der internen Konsistenz. Verglichen mit zwei Studien, die den Zusammenhang eines videobasierten *SJT*s mit dem ISK analysierten, fielen die Korrelationshöhen in der vorliegenden Untersuchung erfreulich aus. In der Studie von Möller (2010) korrelierte der *SJT* für Kundenorientierung lediglich mit der Sekundärskala Offensivität ($r = .21$), aber nicht mit den übrigen Sekundärskalen des ISKs. Ähnlich verhielt es sich bei einem videobasierten *SJT* für berufsspezifische soziale Kompetenzen von Polizeibeamten (Kanning, 2005). Der *SJT* wies nur eine geringe Korrelation zur Skala Selbstbehauptung auf ($r = .13$; heutige Primärskala Durchsetzungsfähigkeit), aber keine zu den übrigen ISK-Skalen.

Bezüglich der Selbst- und Fremdeinschätzung im ISK-K trat eine hohe Übereinstimmung für die Offensivität und mittlere Übereinstimmungen für die soziale Orientierung, Selbststeuerung und Reflexibilität auf. Gewöhnlich werden mittlere Übereinstimmungen in der Selbst- und Fremdeinschätzung und hohe zwischen verschiedenen Fremdeinschätzungen festgestellt (Harris & Schaubroeck, 1988). Dabei fallen die Selbsteinschätzungen positiver als die Fremdeinschätzungen aus, was selbige Autoren mit der Attributionstheorie (DeVader, Bateson & Lord, 1986; Jones & Nisbett, 1972) begründen. Diese besagt, dass gute Leistungen der eigenen Person zugeschrieben, während schlechte Leistungen eher auf Umweltbedingungen zurückgeführt werden. In der vorliegenden Studie war lediglich die Beurteilung der Reflexibilität positiver in der

Selbst- als in der Fremdeinschätzung. Für die soziale Orientierung unterschied sich die Selbsteinschätzung nicht von der Fremdeinschätzung. Für die Offensivität und Selbststeuerung fiel die Fremdeinschätzung sogar positiver als die Selbsteinschätzung aus. Vermutlich wirkte hier die Attributionstheorie weniger stark, da es sich um Stelleninhaber statt Bewerber handelte (vgl. McDaniel & Nguyen, 2001). Die Fremdeinschätzung mit dem ISK-K bestätigte nicht die Zusammenhänge, die zwischen der Selbsteinschätzung des ISKs und dem *SJT* gefunden wurden. Zwischen der Fremdeinschätzung sozialer Kompetenzen und dem *SJT* lagen keine bedeutsamen Korrelationen vor. Die Korrelationshöhe zur sozialen Orientierung war deutlich niedriger und die der Reflexibilität deutlich höher als in der Selbsteinschätzung. Einzig die fremdeingeschätzte Offensivität zeigte eine ähnliche Korrelationshöhe zum *SJT* wie deren Selbsteinschätzung. Auch wenn nur die Selbsteinschätzungen der Teilnehmer mit dem *SJT* korreliert wurden, für die eine Fremdeinschätzung vorlag, fanden sich keine signifikanten Zusammenhänge, mit Ausnahme der Offensivität, zum *SJT*. Die selbsteingeschätzte Offensivität korrelierte sogar höher mit dem *SJT* als in der Gesamtgruppe. Eine Ursache für die Ergebnisdifferenzen kann in der geringen Subgruppe von 23 Personen liegen, mit der signifikante Korrelationen schwer nachzuweisen waren. Auch könnten die Ergebnisdifferenzen auf die Abweichungen zwischen der Selbst- und Fremdeinschätzung der sozialen Kompetenzen zurückgeführt werden. Eine Fremdeinschätzung sozialer Kompetenzen sollte in zukünftigen Studien an einer größeren Stichprobe wiederholt werden, um aussagekräftigere Ergebnisse zu erhalten.

Zusammenhang Selbsteinschätzungsskalen sozialer Kompetenzen und SJT. Der *SJT* zeigte lediglich positive Tendenzen mit der Kontaktfähigkeit, der Suche nach sozialer Unterstützung, der allgemeinen sozialen Kompetenzen, der Selbstkontrolle und der Kritikfähigkeit. Die zwei erstgenannten Selbsteinschätzungsskalen wiesen die höchsten Tendenzen zum *SJT* auf. Dies kann von der Tendenz konsistent zum Ergebnis des ISKs gewertet werden, in dem die inhaltlich verwandten Skalen Extraversion und Offensivität mittlere Zusammenhänge zum *SJT* zeigten. Ein Nullzusammenhang bestand zwischen der Selbstreflexion und dem *SJT*, was konsistent zu den Nullkorrelationen mit der Reflexibilitäts-Skala des ISKs zu werten ist. Die Ergebnisse der Selbsteinschätzungsskalen waren insgesamt wenig aussagekräftig, sodass keine weiteren Schlussfolgerungen abgeleitet werden. Mögliche Gründe wurden bereits in der Diskussion der vorherigen Studie beschrieben (Kapitel 7.6).

Zusammenhang Außenkriterien und SJT. *SJT* korrelierte signifikant mit der sozialen Eingebundenheit im Studium ($r(90) = .39$) und der vorhandenen sozialen Unterstützung ($r(90) = .30$). Die Skala der sozialen Eingebundenheit wies dabei eine gute interne Konsistenz auf ($\alpha = .80$). Zum Bemühen Netzwerke zu bilden/zu erhalten und dem Ausmaß an friedfertig gelösten Konflikten lagen positive Tendenzen vor. Die Außenkriterien bilden die soziale Integration als Konsequenz sozial kompetenten Verhaltens ab. Die gefundenen Zusammenhänge stützen die Vermutung, dass der *SJT* soziale Kompetenzen erfasst. Der umgekehrte Ansatz, dass der *SJT* negativ mit den Konsequenzen für inkompetentes Sozialverhalten korreliert, gelang weniger gut. Hier zeigten sich lediglich negative Tendenzen des *SJT*s zu der Unsicherheit in sozialen Interaktionen und dem alleinigen Medienkonsum in der Freizeit. Zum Prozentsatz an aufbrausend gelösten sozialen Konflikten bestand eine Nullkorrelation. Bezüglich des alleinigen Medienkonsums kann spekuliert werden, dass das Medienverhalten sich verändert hat. Oft sitzen Personen zwar allein vor dem Computer, aber sie kommunizieren häufig gleichzeitig mit vielen Menschen über soziale Netzwerke. Der alleinige Medienkonsum muss somit nicht mit weniger sozial kompetentem Verhalten einhergehen.

Zusammenhang Persönlichkeitseigenschaften und SJT. Der *SJT* korrelierte signifikant mit der Dimension Extraversion des *BFI-K* ($r(90) = .29$). Dieses Ergebnis ist konsistent zum Zusammenhang des *SJT*s mit der Extraversionsskala des *ISKs*. Verglichen mit den Metaanalysen von McDaniel et al. (2001, 2007) korrelierte der vorliegende *SJT* stärker mit Extraversion als der Durchschnitt der *SJT*s ($r = .13$). Dies stützt die Annahme, dass der *SJT* Aspekte des aktiven Zugehens auf andere Menschen erfasst. Der Zusammenhang zwischen Gewissenhaftigkeit und dem *SJT* wies vergleichbare Korrelationshöhen zu denselben Metaanalysen auf (.25 vs. .23), verlor aber nach der Alpha-Fehler-Korrektur die Signifikanz. Der negative Trend zur emotionalen Instabilität kann im Sinne Kanning (2005) so ausgelegt werden, dass Neurotizismus sich negativ auf soziale Kompetenzen auswirkt. Verwunderlich ist, dass der *SJT* Nullkorrelationen zur Verträglichkeit und Offenheit für neue Erfahrungen zeigte. Erwartungsgemäß sollten Persönlichkeitsskalen, die sich positiv auf soziale Interaktionen auswirken, mit sozialen Kompetenzen korrelieren (Asendorpf & Neyer, 2012; Kanning, 2005). Im Mittel zeigen *SJT*s Korrelationen zu .22 mit Verträglichkeit und zu .11 mit Offenheit für Erfahrungen (McDaniel et al., 2001; 2007). Eine Erklärung für die Nullkorrelationen bzw. niedrigen Zusammenhänge könnte in den geringen Streuungen der Persönlichkeitseinschätzungen gelegen haben. Auch zeigte die Verträglichkeitsskala eine niedrige interne Konsistenz. Andererseits ist es nicht ungewöhnlich, dass *SJT*s nicht mit Verträglichkeit und Offenheit für Erfahrungen korrelieren (McDaniel & Nguyen, 2001; Smith & McDaniel,

1998). Auch der videobasierte *SJT* von Möller (2010) zeigte zu keiner der fünf Persönlichkeitseigenschaften signifikante Zusammenhänge. Patterson et al. (2012) weisen darauf hin, dass die exakten Korrelationshöhen zwischen *SJT*s und Persönlichkeitseigenschaften unklar sind. Insgesamt kann daher nicht eindeutig interpretiert werden, ob die Ergebnisse für oder gegen den *SJT* sprechen. Es kann nur geschlussfolgert werden, dass der *SJT* eher extravertiertes Verhalten zu erfassen vermag.

Zusammenhänge kognitive Kriterien und SJT. Der *SJT* zeigte Nullkorrelationen zur fluiden Intelligenz, zum TMS-Gesamtwert, sowie zu den TMS-Subskalen quantitative und formale Probleme, Textverständnis und konzentriertes und sorgfältiges Arbeiten. Zum Abiturnotendurchschnitt lag eine positive Tendenz vor. Somit scheint der *SJT* relativ unabhängig von der allgemeinen kognitiven Leistungsfähigkeit zu sein. Jedoch muss berücksichtigt werden, dass die Medizinstudierenden durch das kognitiv orientierte Auswahlverfahren der Fakultät einen eingeschränkten Wertebereich in den Abiturdurchschnittsnoten und dem TMS aufwiesen, was zu geringeren Korrelationen geführt haben könnte. In der vorliegenden Studie spiegelten die kognitiven Kriterien eher den durchschnittlichen bis weit überdurchschnittlichen Leistungsbereich wieder. Aus derzeitiger Forschungslage können keine eindeutigen Schlussfolgerungen gezogen werden, wie hoch *SJT*s mit kognitiven Kriterien zusammenhängen sollten. In der Metaanalyse von McDaniel et al. (2001) wurde ein mittlerer Zusammenhang zu .42 gefunden, jedoch variierten die Korrelationen fast über den gesamten Korrelationsbereich von null bis eins. Auch für videobasierte *SJT*s zeigt sich ein ähnliches Bild. Der Zusammenhang zwischen videobasierten *SJT*s und der allgemeinen kognitiven Leistungsfähigkeit nimmt Nullkorrelationen (Kanning, 2005) bis moderate Zusammenhänge an (Smiderle et al., 1994; Weekley & Jones, 1997). Clevenger et al. (2001) führen die starke Varianz der Korrelationen auf die zugrunde liegenden Konstrukte der *SJT*s zurück. So sollten *SJT*s, die sich aus analytischen Problemlösesituationen zusammensetzen, höher mit kognitiven Kriterien korrelieren, als *SJT*s, die soziale Interaktionen abbilden. Letztere sollten höher mit Persönlichkeitseigenschaften korrelieren. Dies trifft auf den vorliegenden *SJT* zu. Dieser korrelierte nicht mit den kognitiven, aber mit den sozialen Kriterien und bedingt mit den Persönlichkeitseigenschaften. Die Ergebnisse fallen konform zur Theorie des *Tacit Knowledge* aus (Sternberg, 1997, 1999; Stemler & Sternberg, 2006). Nach dieser verlangen *SJT*s situationsspezifische, aber keine abstrakte Problemlösungsfähigkeit, die anhand der analytischen Intelligenz erfasst werden kann. Demnach sollten *SJT*s geringer mit fluiden Intelligenztests korrelieren. Zusätzlich könnte die verhaltensbasierte Instruktion des *SJT*s dazu beigetragen haben, dass keine Korrelationen zur allgemeinen kognitiven Leistungsfähigkeit vorlagen. *SJT*s mit ver-

haltensbasierten Instruktionen korrelieren im Allgemeinen geringer mit kognitiven Fähigkeiten als *SJTs* mit wissensbasierten Instruktionen (McDaniel & Nguyen, 2001; Whetzel et al., 2008). Dass kein Zusammenhang zwischen Textverständnis und dem *SJT* vorlag, kann ebenfalls positiv gewertet werden. Somit scheint der videobasierte *SJT* keine starken Anforderung an die verbale Kompetenz von angehenden Akademikern zu stellen (vgl. Sacco et al., 2000a; 2000b). Auch scheint er testfair gegenüber ethnischen Minderheiten zu sein (Clevenger et al., 2001; Motowidlo et al., 1990).

Vorhersage des Studienerfolgs. Der *SJT*klärte 10% der Varianz der Studienzufriedenheit und 8% der Varianz der Passung zum Studium auf. Der subjektive Studien-erfolg konnte durch den *SJT* nicht vorhergesagt werden. Diese Ergebnisse stehen im Einklang mit dem Befund, dass die Studienzufriedenheit eher mit dem Arbeitsverhalten korreliert, während der Studienerfolg eher Zusammenhänge mit kognitiven Kriterien aufweist (Giesen et al., 1986; Kersting, 2005). So zeigte der *SJT*, der eher das Studierverhalten einer Person wiedergibt, einen Zusammenhang mit der Studienzufriedenheit und Passung zum Studium, aber nicht mit dem stärker kognitiv geprägten Kriterium des subjektiven Studienerfolgs.

Zu den Studiennoten der Vorklinik lagen keine signifikanten Zusammenhänge für den *SJT* vor. Dies ist konsistent zu dem Befund, dass nicht-kognitive Kriterien wenig geeignet sind, den vorklinischen Studienerfolg vorherzusagen (Eva et al., 2004; Kulatunga-Moruzi & Norman, 2002; Lievens, 2013; Lievens et al., 2009a). Im Gegensatz dazu belegten dieselben Autoren, dass nicht-kognitive Kriterien durchaus in der Lage sind, die klinischen Leistungen vorherzusagen. In der vorliegenden Untersuchung trat nur eine signifikante Korrelation des *SJTs* mit der Note im *Mini-CEX* auf. Des Weiteren lagen mittlere bis hohe Tendenzen zwischen dem *SJT*, dem *OSCE*, sowie den *Encounter Cards* der Inneren Medizin vor. Diese drei Prüfungsformate spiegeln einen höheren Anteil an nicht-kognitiven Aspekten wie den Umgang mit Patienten und klinisch-praktischen Fertigkeiten wieder. Dies kann als Indiz für die Validität des *SJTs* gewertet werden, der soziale Kompetenzen erfassen soll. Andererseits bestanden zu der Mehrheit der klinischen Studienleistungen keine hypothesenkonformen Zusammenhänge. Zum einen könnte dies auf die sehr geringen Stichprobengrößen zurückgeführt werden, zum anderen schienen die spezifischen Teilnoten besser für die Analyse geeignet zu sein als die zahlreicher eingegangenen Fachnoten. Fachnoten besitzen meist eine geringere Spezifität als die Teilnoten, was nach dem *Bandwidth-fidelity dilemma* (Ones & Viswesvaran, 1996) den Zusammenhang zwischen dem *SJT* und den objektiven Studienerfolgskriterien beeinträchtigt haben könnte. Der Propädeutische Block, der die klinisch-theoretischen Fächer umfasst, wird im Heidelberger Curriculum

dem klinischen Studienabschnitt zugeordnet, enthält aber außer den Blockpraktika der Inneren Medizin und Chirurgie keine praktischen Prüfungsformate mit Patientenkontakt. Dies könnte ein Grund dafür sein, dass keine Korrelationen zwischen dem *SJT* und den Leistungen im Propädeutischen Block nachweisbar waren. Um eine höhere Aussagekraft zu erzielen, sollten insbesondere die Teilnoten aus höheren Semestern von einer größeren Studierendenanzahl mit Regressionsanalysen analysiert werden. Zudem empfehlen sich prospektive statt retrospektive Untersuchungen.

Zusammenhang Kontrollvariablen und SJT. Der *SJT* korrelierte nicht mit dem Alter, Deutsch als Erstsprache, der medizinischen Vorbildung und dem Fachsemester. Auch zwischen Studierenden der Vorklinik vs. Klinik, bestand kein Unterschied in der *SJT*-Punktzahl. Die Altersspanne war verglichen mit der vorherigen Untersuchung geringer (15-54 vs. 17-29 Jahre), sodass ein fehlender Zusammenhang zwischen dem Alter und dem *SJT* zu erwarten war. Schon in der Stichprobe der Studienplatzbewerber zeigte sich zwar ein signifikanter, aber kein praktisch bedeutsamer Zusammenhang zwischen dem Alter und dem *SJT*. Somit scheint das Alter keinen bedeutsamen Einfluss auf den *SJT*-Wert der Studienbewerber, -interessenten und Medizinstudierenden zu haben

Die Unabhängigkeit des *SJTs* von der Erstsprache unterstützt das Ergebnis, dass kein Zusammenhang zwischen dem Textverständnis und dem *SJT* festgestellt wurde. Somit wird die Schlussfolgerung gestärkt, dass der videobasierte *SJT*, keine starken Anforderung an die verbale Kompetenz von angehenden Akademikern stellt (vgl. Sacco et al., 2000a, 2000b). Eine Einschränkung bestand jedoch darin, dass nur 17 (19%) Teilnehmer Nicht-Muttersprachler waren. Verglichen mit dem durchschnittlichen Anteil von 4% Nicht-EU-Bürger und EU-Bürger aus nicht-deutschsprachigen Ländern im Bewerberpool des AdHs (Kohorten 2010-2013) war der Anteil Nicht-Muttersprachler allerdings hoch ausgeprägt. Auch im Bezug zum durchschnittlichen Anteil von 11% Nicht-EU-Bürgern und EU-Bürgern aus nicht-deutschsprachigen Ländern pro immatrikulierter Studierendenkohorte (2007, 2008, 2010, 2011) war der Anteil nicht-deutschsprachiger Muttersprachler in der Stichprobe hoch. Anhand der Näherungswerte kann geschlussfolgert werden, dass die Erstsprache keinen gravierenden Einfluss auf die Leistung im *SJT* für die Zielgruppen der medizinischen Studienplatzbewerber und -interessenten sowie Medizinstudierenden zu haben scheint.

Die Unabhängigkeit des *SJTs* von der medizinischen Vorbildung kann wie bei der Erstsprache dadurch angezweifelt werden, dass nur 16 Personen der Stichprobe (18%) eine medizinische Ausbildung abgeschlossen oder mindestens ein Jahr mit Patienten zusammengearbeitet haben. Verglichen mit der gewöhnlichen Verteilung der

Heidelberger Studienplatzbewerber und Medizinstudierenden befand sich ein hoher Anteil von Personen mit Berufsausbildung in der Stichprobe. Im Durchschnitt weisen 12% der Bewerber im hochschuleigenen Auswahlverfahren (Kohorten 2010-2013) und 5% der in Heidelberg immatrikulierten Medizinstudierenden eine medizinische Ausbildung vor (Kohorten 2007, 2008, 2010, 2011). Daher kann für die gewöhnliche Zusammensetzung der Heidelberger Studienplatzbewerber und Medizinstudierenden geschlossen werden, dass eine medizinische Ausbildung keinen Einfluss auf die *SJT*-Ergebnisse hervorruft.

Der Nullzusammenhang zwischen dem Fachsemester und dem *SJT* kann durch die wenig vertretenen höheren Fachsemester (9.-12.) begünstigt worden sein. Für aussagekräftigere Ergebnisse sollte die Studierendenzahl pro Fachsemester erhöht werden.

Beurteilung des Auswertungsschlüssels. Da der Auswertungsschlüssel direkt die Punktzahl im *SJT* bestimmt und somit Auswirkungen auf die Validität des *SJT*s haben kann, soll er nach den Kriterien von Bergman et al. (2006) beurteilt werden. Nach ihnen sollte ein Auswertungsschlüssel eine hohe Kriteriumsvalidität und inkrementelle Validität aufweisen. Er sollte mit verwandten Konstrukten höher korrelieren als mit inhaltlich fremden Konstrukten und keine Subgruppeneffekte hervorrufen. Der in dieser Studie verwendete Auswertungsschlüssel zeigte Zusammenhänge zu prosozialem und offensivem Verhalten, der Extraversion als Persönlichkeitseigenschaft und den Konsequenzen sozial kompetenten Verhaltens. Die Kriteriumsvaliditäten können verglichen mit den in Metaanalysen berichteten Korrelationen um .30 zu Persönlichkeitsfaktoren und sozialen Fähigkeiten (z.B. Christian et al., 2010; McDaniel & Nguyen, 2001) als gut eingeschätzt werden. Dies traf aber nicht für die unstandardisierten Selbsteinschätzungsskalen zu. Der *SJT* zeigt Potenzial zur Vorhersage der Studienzufriedenheit und der Passung zum Studium, sowie eine inkrementelle Validität für die Passung zum Studium über das Abitur und dem TMS hinaus. Hier sollte jedoch noch stärker die prädiktive Validität von klinischen Studienleistungen erforscht werden, um einen Rückschluss auf die Qualität des Auswertungsschlüssels und des *SJT*s zu ziehen. Der *SJT*-Punktwert korrelierte mit den verwandten Konstrukten der sozialen Kompetenz und Persönlichkeit höher als mit den inhaltlichen fremden Konstrukten der allgemeinen kognitiven Leistungsfähigkeit. Der Auswertungsschlüssel rief keine Subgruppeneffekte hervor. Die *SJT*-Punktzahl war unabhängig vom Geschlecht, dem Alter, der Erstsprache und einer abgeschlossenen medizinischen Ausbildung. Mit Vorsicht bezüglich der inkrementellen Validität folgt der Auswertungsschlüssel allen Empfehlungen und kann daher als gut beurteilt werden.

9. Studienübergreifende Zusammenfassung der Ergebnisse

Tabelle 73 fasst die Ergebnisse beider Validierungsstudien anhand der Forschungshypothesen zusammen.

Tabelle 73: *Ergebnisse der Hypothesentestung*

Aspekt	Hypothese	Bestätigt?
Faktorenstruktur	- Dem videobasierten <i>SJT</i> liegen zwei Faktoren zugrunde. Ein Faktor setzt sich aus den zwei sozialen Kompetenzen Selbstkontrolle/ Selbstdisziplin und Selbstbeobachtung/ Selbstreflexion zusammen. Der zweite Faktor bildet die vier sozialen Kompetenzen Respektfähigkeit, Kontaktfähigkeit, Kritikfähigkeit und Suche nach sozialer Unterstützung ab (2).	nein
Akzeptanz/ Nützlichkeit	- Der <i>SJT</i> wird von den Studienplatzbewerbern akzeptiert (2).	ja
	- Der <i>SJT</i> ist in der Lage, die Teilnehmer über die Anforderungen zu informieren, die das Medizinstudium an die sozialen Kompetenzen der Studierenden stellt (2).	ja
	- Der <i>SJT</i> regt die Teilnehmer zum Reflektieren der eigenen sozialen Kompetenzen an (2).	ja
	- Personen, mit hohen <i>SJT</i> -Werten, fühlen sich bestärkt Medizin zu studieren, während Personen mit niedrigen <i>SJT</i> -Werten von einem Medizinstudium abgeschreckt werden (2).	nein
soziale Kompetenzen	- Der <i>SJT</i> korreliert signifikant mit den Selbsteinschätzungs-skalen der sozialen Kompetenzen (3)	teils
	- Der <i>SJT</i> korreliert mit den Skalen des ISKs (1).	teils
	- Der <i>SJT</i> korreliert positiv mit dem Bemühen Netzwerke zu bilden/erhalten, der sozialen Unterstützung bei Problemen, der sozialen Eingebundenheit und dem Ausmaß an friedfertig gelösten Konflikten (1).	teils
	- Der <i>SJT</i> korreliert negativ mit dem alleinigen Medienkonsum in der Freizeit, dem Ausmaß an aufbrausend gelösten Konflikten und der Unsicherheit in sozialen Interaktionen (1).	nein
Persönlichkeit	- Der <i>SJT</i> zeigt positive Zusammenhänge zu den Persönlichkeitseigenschaften Verträglichkeit, Gewissenhaftigkeit, emotionale Stabilität und Offenheit für Erfahrungen (1).	teils
	- Es besteht ein negativer Zusammenhang zwischen dem <i>SJT</i> und der Persönlichkeitseigenschaft Neurotizismus (1).	nein

Fortsetzung Tabelle 73:

Aspekt	Hypothese	Bestätigt?
Kognition	- Der <i>SJT</i> korreliert nicht mit der fluiden Intelligenz gemessen über den CFT20-R (1).	ja
	- Der <i>SJT</i> korreliert nicht mit dem Abiturnotendurchschnitt (3)	ja
	- Der <i>SJT</i> korreliert nicht mit dem Gesamtwert des TMS (3).	ja
	- Es besteht kein Zusammenhang zwischen dem <i>SJT</i> und dem TMS-Subtest quantitative und formale Probleme (1).	ja
	- Der <i>SJT</i> zeigt nur geringe Korrelationen zum TMS-Subtest Textverständnis (1).	ja
	- Der <i>SJT</i> korreliert nicht mit dem TMS-Subtest Konzentriertes und sorgfältiges Arbeiten (1).	ja
Geschlechtereffekt	- Frauen erreichen höhere Punktzahlen im <i>SJT</i> als Männer (3).	nein
Studienzufriedenheit	- Der <i>SJT</i> sagt die Studienzufriedenheit und die Passung zum Studium vorher (1).	teils
	- Der <i>SJT</i> sagt nicht den subjektiven Studienerfolg vorher (1).	ja
	- Der <i>SJT</i> zeigt eine höhere prädiktive Validität für die klinischen als für die vorklinischen Studienleistungen (1).	Tendenz ja

Anmerkung. Anzahl der Stichprobentestung in Klammern.

Aus Tabelle 73 kann entnommen werden, dass die Abgrenzung des vorliegenden *SJT*s zu kognitiven Kriterien (fluide Intelligenz, Abiturnotendurchschnitt, TMS) konsistenter gelang als der Nachweis der konvergenten Validität. Bei letzterem zeigten sich überwiegend Zusammenhänge mit Skalen, die eine hohe Kontaktfähigkeit, Offenheit und soziale Orientierung einer Person widerspiegeln bzw. Folge dieser sind. Korrelationen mit sozialen Kompetenzen der Selbststeuerung und Reflexion konnten nicht nachgewiesen werden. Dies kann auch ein Indiz dafür sein, dass die angenommene zweifaktorielle Struktur des *SJT*s nicht bestätigt werden konnte. Die Ergebnisse der Faktorenanalyse legten eine einfaktorielle Struktur nahe. Zudem konnte der *SJT* die Studienzufriedenheit und Passung zum Studium vorhersagen. Der *SJT* war wie vermutet nicht in der Lage, die Zufriedenheit mit den eigenen Studienleistungen vorherzusagen. Bezüglich der inkrementellen Validität leistete er einen Beitrag für die Passung zum Studium, aber nicht für die Studienzufriedenheit über den Abiturnotendurchschnitt und den TMS hinaus. Der *SJT* zeigte tendenziell höhere Zusammenhänge mit klinischen Studienleistungen, deren Schwerpunkt auf sozial kompetentem Verhalten liegen, als mit den vorklinischen Leistungen. Außerdem konnte festgestellt werden, dass die Zielgruppe den

SJT akzeptierte und als nützlich erlebte. Die aus den Forschungsergebnissen resultierenden Implikationen stellt das nächste Kapitel dar.

10. Implikationen für Forschung und Praxis

Mit dem videobasierten *SJT* wurde ein ansprechendes Selbstselektionsinstrument für die Medizinische Fakultät Heidelberg geschaffen, das Studienplatzbewerbern und -interessenten über Anforderungen informiert, die das Medizinstudium an soziale Kompetenzen stellt. Neben dem videobasierten *SJT* von Lievens und Sackett (2006), der beansprucht interpersonelle Fähigkeiten zu erfassen, existiert nun ein zweiter videobasierter *SJT* für die medizinische Studierendenauswahl. Der *SJT* steht allen Interessenten auf der Homepage der Fakultät zur Verfügung und kann als Ressourcenschonendes Breitbandverfahren eingesetzt werden. Er fördert die Auseinandersetzung der Teilnehmer mit den eigenen sozialen Kompetenzen und unterstützt die Entscheidung für oder gegen ein Medizinstudium. Die Bewerber akzeptieren den *SJT* und loben dessen Realitätsnähe, die Anschaulichkeit und Qualität der Filme. Somit erfüllt der *SJT* den Anspruch, Auswahlverfahren zu erschaffen, die die Bewerber über allgemeine als auch fachspezifische Komponenten des Studiums in Form von Selbstselektionsinstrumenten und beratenden *Self-Assessments* informieren (Benbassat & Baumal, 2007; Bore et al., 2009; Deutsche Gesellschaft für Psychologie e. V., 2004). Ein Ziel dieser Verfahren ist die Gewährleistung einer hinreichenden Passung zwischen dem Studienangprofil und dem Bewerber (Wenzel, 2008; Wissenschaftsrat, 2004). Diesem Ziel scheint der vorliegende *SJT* gerecht zu werden. So kann der *SJT* einen Anteil der Studienzufriedenheit und der Passung zum Studium vorhersagen. Zudem leistet er einen inkrementellen Beitrag für die Passung zum Studium über den Abiturnotendurchschnitt und den TMS hinaus. Ob der *SJT* auch den objektiven Studienerfolg via Noten vorhersagen kann, ist aufgrund der geringen Anzahl von fortgeschrittenen Medizinstudierenden in der Untersuchungsgruppe noch ungewiss. Hierfür sollte insbesondere die prädiktive Validität des *SJTs* für Noten des klinischen Studienabschnitts analysiert werden, da die Stärke nicht-kognitiver Verfahren in der Vorhersage des klinischen Studienerfolgs liegt (Eva et al., 2009; Eva et al., 2012; Hojat et al., 1993; Kulatunga-Moruzi & Norman, 2002; Lievens, 2013). Die Weiterverfolgung der Teilnehmer bis zum Studienabschluss ist geplant. Um die Stabilität der Ergebnisse zu prüfen, soll auch der Zusammenhang zwischen dem *SJT*-Ergebnis und den Studiennoten weiterer Studierendengruppen analysiert werden. Auf Basis der vorhandenen Abiturnoten, TMS- und *SJT*-Ergebnisse ist darüber hinaus vorgesehen, den Einsatz des *SJTs* als kompensatorisches vs. konsekutives Auswahlkriterium zu simulieren. Auch wird in Erwägung gezogen, die Teilnahme am *SJT* als Bestandteil des AdHs allein zu Forschungszwecken verpflichtend zu gestalten, um der Selektivität der Stichproben entgegen zu wir-

ken. In diesem Falle sollte eine wissensbasierte Instruktionsform eingesetzt werden, um der sozialen Erwünschtheit vorzubeugen.

Verglichen zu anderen *SJT*s besitzt der vorliegende *SJT* eine hohe interne Konsistenz, die eine Auswertung auf Individualebene erlaubt. Die Ermittlung der Retest- und Paralleltestreliabilität nach Empfehlung von Whetzel und McDaniel (2009) stehen indes noch aus. Die Erhebung der Paralleltestreliabilität gestaltet sich allerdings schwierig, weil dafür zusätzliche videobasierte Items entwickelt werden müssten.

Die Faktorenstruktur des *SJT*s setzt sich aus einer Art Generalfaktor, gefolgt von Faktoren mit geringen spezifischen Varianzbeiträgen konsistent zu den Schilderungen von Gillespie et al. (2002) und Schmitt und Chan (2006) zusammen. Der „Generalfaktor“ des vorliegenden *SJT*s repräsentiert dabei das Ausmaß, inwieweit eine Person auf andere zugeht, um eigene Interessen aktiv zu verwirklichen und wie offen und positiv sie anderen begegnet. Der *SJT* grenzt sich zudem von kognitiven Kriterien wie der fluiden Intelligenz, dem Abiturnotendurchschnitt und dem TMS ab. Aufgrund des derzeitig kognitiv geprägten AdHs der Medizinischen Fakultät Heidelberg ist die Abgrenzung des *SJT*s zu kognitiven Leistungskriterien zu befürworten.

Der *SJT* scheint unabhängig von Geschlecht, Alter, Deutsch als Erstsprache, einer abgeschlossenen medizinischen Ausbildung und dem Fachsemester zu sein, sodass auf eine getrennte Testnormierung hinsichtlich dieser Variablen verzichtet werden kann. Die Analyse der Testfairness von *SJT*s konzentrierte sich bisher nur auf das Geschlecht, die Rasse und die verbalen Fähigkeiten als potenzielle Einflussfaktoren. Somit liefert diese Arbeit erste Forschungsbeiträge zum Einfluss des Alters, einer abgeschlossenen medizinischen Ausbildung und dem Fachsemester auf die Punktzahl im *SJT*. Der Einfluss des Fachsemesters auf die *SJT*-Leistung sollte dabei detaillierter mit Medizinstudierenden aus unterschiedlichen Studienabschnitten analysiert werden. Hierfür bietet sich ein Vergleich der *SJT*-Werte von Studienanfänger vs. Studierende im praktischen Jahr oder mit Absolventen an.

Auch für andere Forscher ist es von Interesse, das Konstrukt von *SJT*s weiter zu untersuchen. Aufbauend auf die *Tacit Knowledge* Theorie (Sternberg, 1997, 1999; Stemler & Sternberg, 2006) könnte der vorliegende *SJT* als auch andere *SJT*s anhand kontextspezifischer Problemlösungsverfahren analysiert werden. Ebenfalls bietet sich die Erforschung der *SJT*-Struktur mit Hilfe der *ITPs* (Motowidlo et al., 2006) an. Dazu müssten die Antwortoptionen des *SJT*s nach dem Ausmaß an Persönlichkeitseigenschaften bewertet und mit der Angemessenheit der Antwortoptionen und dem *SJT*-Punktwert in Beziehung gesetzt werden.

Für die weitere Untersuchung der konvergenten Validität des *SJT*s sollten zusätzliche soziale Kompetenzkriterien einbezogen werden, die im Optimalfall über eine Ver-

haltensbeobachtung erhoben werden (Goldfried & D’Zurilla, 1969; Kauffeld et al., 2002). Da sowohl der klinische Alltag, als auch die Übungssituationen in der medizinischen Lehre eine hohe Variabilität aufweisen, empfiehlt es sich dafür, künstliche Settings zu entwerfen. Ein bereits bestehendes standardisiertes Setting bieten die *OSCE*-Prüfungen. In diesen könnten die sozialen Kompetenzen von Studierenden während eines Arzt-Patienten-Gesprächs beobachtet und beurteilt werden. Auch die *Mini-CEX*-Prüfungen, in denen Kommunikations- und Teamfähigkeit während einer Simulation bewertet werden, scheinen zur Erhebung sozialer Kompetenzen geeignet. Ein anderer Ansatz der Konstruktvalidierung besteht darin, Personen mit überdurchschnittlichen, durchschnittlichen und unterdurchschnittlichen sozialen Kompetenzen den *SJT* bearbeiten zu lassen. So bietet es sich an, sozial ängstliche und autistische Menschen - als Personen mit gewöhnlich unterdurchschnittlichen sozialen Kompetenzen - als Stichprobe zu verwenden und mit Personen mit durchschnittlicher Ausprägung zu vergleichen. Auch könnten mit der Sozialkompetenz verwandte Konstrukte zur weiteren Validierung des *SJT*s verwendet werden. Statt dem eingesetzten *BFI-K* könnten die Persönlichkeitsfaktoren mit dem umfassenderen *NEO-FFI* (Borkenau & Ostendorf, 1993) oder *NEO-PIR* (Ostendorf & Angleitner, 2004) erhoben werden. Aus dem Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (Hossiep, Paschen & Mühlhaus, 2003) würden sich die Skalen Sensitivität, Kontaktfähigkeit und Soziabilität zur Validierung des *SJT*s eignen, da sie den bei der *SJT*-Entwicklung eingegangenen Konstrukten ähneln. Des Weiteren kämen soziale und emotionale Intelligenztests als angrenzende Konstrukte für die Validierung des *SJT*s in Frage.

Von Interesse für andere Testentwickler könnten die neu entwickelte Instruktionsform und der Auswertungsschlüssel des vorliegenden *SJT*s sein. Der Auswertungsschlüssel berücksichtigt mehrere Antwortoptionen gleicher Effektivität und weist nach Bergman et al. (2006) bislang gute Eigenschaften auf. Auch wurden die Erfahrungen, die während der *SJT*-Entwicklung gemacht wurden, als Empfehlung für zukünftige Testkonstrukteure festgehalten. Der *SJT* eignet sich außerdem für den Einsatz in der Studierendenauswahl an anderen Fakultäten. Hierfür sind multizentrische Studien geplant.

Einige Autoren kritisieren, dass kaum vergleichende experimentelle Studien für *SJT*s vorliegen (Arthur & Villado, 2008; McDaniel & Nguyen, 2001; Ployhart & MacKenzie, 2011). Als Reaktion auf die Kritik wurde der *SJT* an Studienplatzbewerbern und Medizinstudierenden vergleichend analysiert. Die Arbeit liefert einen zusätzlichen Beleg dafür, dass Bewerber höhere Punktwerte im *SJT* erzielen als Stelleninhaber (McDaniel & Nguyen, 2001). Zukünftig könnte zudem ein Vergleich zwischen dem *SJT* in Schrift-, Video- und Drehbuchformat durchgeführt werden. Auch bietet es sich an,

verschiedene kontinuierliche Instruktionsformen und Auswertungsschlüssel nebeneinander zu analysieren (vgl. Bergman et al., 2006; Polyhart & Ehrhart, 2003), wobei insbesondere die Betrachtung von wissensbasierten Instruktionen oder freien Antwortformaten aufschlussreich wäre. Freie Antwortformate bieten zudem die Möglichkeit, neue Antwortoptionen für den *SJT* zu gewinnen, in dem z.B. Studienbewerber oder Absolventen befragt werden.

Ein weiteres Einsatz- und Forschungsgebiet ist die Implementierung des *SJTs* als Trainingsinstrument in die medizinische Lehre. Dafür wurde eine Anleitung für Lehrende erstellt, die zusammen mit den *SJT*-Videos auf der fakultätsinternen Lehrplattform zur Verfügung gestellt werden soll. Der *SJT* ließe sich für die Initiierung von Rollenspielen verwenden. Nach der videobasierten Situationsschilderung würden die Medizinstudierenden die Rollen der Filmakteure übernehmen und verschiedene Handlungsalternativen durchspielen. Der Einsatz sollte wissenschaftlich begleitet werden, in dem die Effektivität des *SJTs* als Trainingsinstrument analysiert wird. Alternativ könnte der *SJT* auch als Progress-Test für soziale Kompetenzen über die gesamte Studiendauer an verschiedenen Fakultäten eingesetzt und erforscht werden.

Literaturverzeichnis

- Albrecht, G. (1997). Neue Anforderungen an Ermittlung und Bewertung von beruflicher Kompetenz. In Arbeitsgemeinschaft Qualifikation–Entwicklung–Management (Hrsg.), *Kompetenzentwicklung '97: Berufliche Weiterbildung in der Transformation – Fakten und Visionen* (S. 85-140). Münster: Waxmann.
- Albrecht, T. L., & Adelman, M. B. (1987). Communicating social support: A theoretical perspective. In T. L. Albrecht & M. B. Adelman (Eds.), *Communicating social support* (pp. 18-39). Newbury Park: Sage.
- Althoff, K. (1986). Zur Aussagekraft von Schulzeugnissen im Rahmen der Eignungsdiagnostik. *Psychologie und Praxis*, 30, 77-85.
- Akkreditierungsrat (2013). *Regeln für die Akkreditierung von Studiengängen und für die Systemakkreditierung*. http://www.akkreditierungsrat.de/fileadmin/Seiteninhalte/AR/Beschluesse/AR_Regeln_Studiengaenge_aktuell.pdf (Zugriff: 17.03.2014).
- Amelang, M. (1987). *An investigation of the factorial structure and external validity of tests of social intelligence* (Tech. Rep. No. 59). Heidelberg: Universität Heidelberg, Psychologisches Institut.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Asendorpf, J. & Neyer, F. J. (2012). *Psychologie der Persönlichkeit* (5. vollständig überarbeitete Aufl.). Berlin: Springer.
- Asendorpf, J. & Wallbott, H. G. (1979). Maße der Beobachterübereinstimmung: Ein systematischer Vergleich. *Zeitschrift für Sozialpsychologie*, 10, 243-252.
- Backhaus, K., Erichson, B., Plinke, P. & Weiber, R. (2008). *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*. Berlin: Springer.
- Baron-Boldt, J., Schuler, H. & Funke, U. (1988). Prädiktive Validität von Schulabschlußnoten: Eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 2 (2), 79-90.
- Barr, D.A. (2010). Science as superstition: Selection medical students. *Lancet*, 376, 678-679.
- Bass, B. M. & Barrett, G. V. (1972). *Man, work and organizations: An introduction to industrial and organizational psychology*. Oxford: Allyn & Bacon.

- Bastians, F. & Runde, B. (2002). Instrumente zur Messung sozialer Kompetenzen. *Zeitschrift für Psychologie*, 210 (4), 186-196.
- Baumert, J. & Watermann, R. (2000). Institutionelle und regionale Variabilität und die Sicherung der gemeinsamen Standards in der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Band 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (S. 317-372). Opladen: Leske & Budrich.
- Beach, M. C., Duggan, P. S., Cassel, C. K. & Geller, G. (2007). What does 'respect mean'? Exploring the moral obligation of health professionals to respect patients. *Society of General Internal Medicine*, 22, 692-695.
- Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12 (2), 155-187.
- Bean, J. P. (1985). Interaction effects based on class level in an explanatory model of college student dropout syndrome. *American Educational Research Journal*, 22 (1), 35-64.
- Beauducel, A. & Wittmann, W. W. (2005). Simulation study on fit indices in confirmatory factor analysis based on data with slightly distorted simple structure. *Structural Equation Modeling*, 12, 41-75.
- Becker, R. E. & Heimberg, R. G. (1988). Assessment of social skills. In A. Bellack & M. Hersen (Eds.), *Behavioral assessment* (pp. 365-395). Oxford: Pergamon Press.
- Benbassat, J. & Baumal, R. (2007). Uncertainties in the selection of applicants for medical school. *Advance Health Science Education Theory and Practice*, 12, 509-521.
- Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring situational judgement tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment*, 14, 223-235.
- Bergman, M. E., Drasgow, F., Donovan, M. A. & Juraska, S. E. (2003). *Scoring Situational Judgement Tests*. Paper presented at the 18th annual meeting of the Society for Industrial and Organizational Psychology, Orlando.
- Bergman, A. V., Susarla, M., Howell, H. & Karimbu, N. Y. (2006). Dental admission test scores and performance on NBDE Part I, Revisited, *Journal of Dental Education*, 70 (3), 258-262.

- Bergmann, B. (2007). Selbstkonzept beruflicher Kompetenz. In J. Erpenbeck & L. von Rosenstiel, (Hrsg.), *Handbuch Kompetenzmessung - Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* (2. Aufl., S. 94-223). Stuttgart: Schäffer-Poeschel.
- Bore, M., Munro, D. & Powis, D. (2009). A comprehensive model for the selection of medical students. *Medical Teacher*, 31, 1066-1072.
- Borkenau, P. & Ostendorf, F. (1993): *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae*. Göttingen: Hogrefe.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6. Aufl.). Berlin: Springer.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer.
- Bortz, J., Lienert, G. A. & Boehnke, K. (2000). *Verteilungsfreie Methoden in der Biostatistik* (2. Aufl.). Heidelberg: Springer.
- Brooks, M. E. & Highhouse, S. (2006). Can good judgement be measured? In J. A. Weekly & R. E. Polyhart (Eds.), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 39-55). Mahwah: Lawrence Erlbaum Associates.
- Bucksch-Beudt, C. Büchel, A., Berkhoff, S., Janko, S., Kirchhoff, A., Kompatscher, J., Kraft, H.G., Kujumdshiev, S., Nürnberger, F., Ochsendorf, F., Rehner, M. & Schulze, J. (2006). Möglichkeiten und Grenzen der Fragebogen-gestützten Erhebung von Soft skills als Zulassungskriterien zum Medizinstudium. *GMS Zeitschrift für Medizinische Ausbildung*, 23 (4), o.S.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Diplom-Psychologen und Sozialwissenschaftler*. München: Pearson Studium.
- Buhrmester, D. Furman, W. Wittenberg, T. & Reis, H. T. (1988). Five domains of interpersonal competence in peer relationships. *Journal of Personality and Social Psychology*, 55 (6), 991-1008.
- Bundestag (2004). *Siebtes Gesetz zur Änderung des Hochschulrahmengesetzes (7. HRGÄndG)*. Bonn: Bundesgesetzblatt (Teil I Nr. 47).

- Bundestag (2012). Grundgesetz für die Bundesrepublik Deutschland.
<http://www.gesetze-im-internet.de/bundesrecht/gg/gesamt.pdf> (Zugriff: 20.03.2014).
- Caldarella, P. & Merrell, K. W. (1997). Common dimensions of social skills of children and adolescents: A taxonomy of positive behaviours. *School Psychology Review*, 26, 264-278.
- Campbell, D. T. (1963). Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science* (pp. 94-172). New York: McGraw-Hill.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait multimethod matrix. *Psychological Bulletin*, 56 (2), 81-105.
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 687-732). Palo Alto: Consulting Psychologists Press.
- Cantor, N. & Harlow, R. E. (1994). Social intelligence & personality: Flexible life task pursuit. In R.J. Sternberg & P. Ruzgis (Eds.), *Personality and Intelligence* (pp. 137-168). New York: Cambridge University Press.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence; a critical experiment. *Educational Psychology*, 54, 1-22.
- Chan, D. & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgement tests: Subgroup differences in test performance and face validity perception. *Journal of Applied Psychology*, 82 (1), 143-159.
- Christian, M. S., Edwards, B. D. & Bradley, J. C. (2010). Situational Judgement Tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63, 83-117.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, S. & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, 98, 310-357.

- Costa, P. T. & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five factor Inventory. Professional Manual*. Odessa: Psychological Assessment Resources.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.
- Cronbach, L.J. (1960). *Essentials of Psychological Testing*. New York: Harper & Row.
- Crushing, A. M., Ker, J. S., Kinnersley, P., McKeown, P., Silverman, J., Patterson, J. & Westwood, O. M. R. (2014). Patient safety and communication: A new assessment for doctors trained in countries where language differs from that of the host country: Results of a pilot using a domain-based assessment. *Patient Education and Counseling*, *95* (3), 332-339.
- Cureton, E. E. (1950). Validity, reliability, and baloney. *Educational and Psychological Measurement*, *10*, 94-96.
- Darden, C. & Gintner, E. (1996). Life skills development scale-Adolescent Form: The theoretical and therapeutic relevance of life skills. *Journal of Mental Health Counseling*, *18* (2), 142-164.
- De Fruyte, F. & Mervielde, I. (1998). Personality and interests as predictors of educational streaming and achievement. *European Journal of Personality*, *10* (5), 405-425.
- Deidesheimer, Kreis (1993). *Feststellung der Studieneignung im Rahmen der Hochschulzulassung*. Bonn: Bericht für das Bundesministerium für Bildung und Wissenschaft.
- Deidesheimer Kreis (1997). Hochschulzulassung und Studieneignungstests. Göttingen: Vandenhoeck und Ruprecht.
- DeRossier, M., Patterson, C., Kuperschmidt, J. B. (1994). Children's academic and behavioral adjustment as a function of the chronicity and proximity of peer rejection. *Child Development*, *65* (6), 1799-1813.
- Deutsche Gesellschaft für Psychologie e.V. (2004). Stellungnahme der Deutschen Gesellschaft für Psychologie e.V. (DGPs) zur Auswahl von Studierenden durch die Hochschulen (vom 22. November 2004). *Psychologische Rundschau*, *56* (2), 153-154.

- DeVader, C. L., Bateson, A. G. & Lord, R. G. (1986). Attribution theory: A meta-analysis of attributional hypotheses. In E. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 63-81). Lexington: Lexington Books.
- Dlugosch, S. (2005). *Prognose von Studienerfolg*. Aachen: Shaker.
- Dodge, K. A. (1993). Social-cognitive mechanisms in the development of conduct disorder and depression. *Annual Review of Psychology*, 44, 559-584.
- Dodson, M., Crotty, B., Prideaux, D., Carne, R., Ward, A. & de Leeuw, E. (2009). The multiple mini-interview: How long is long enough? *Medical Education*, 43, 168-174.
- Döpfner, M., Schlüter, S. & Rey, E. R. (1981). Evaluation eines sozialen Kompetenztrainings für selbstunsichere Kinder im Alter von neun bis zwölf Jahren: Ein Therapievergleich. *Zeitschrift für Kinder- und Jugendpsychiatrie*, 9, 233-252.
- Donnon, T., Paolucci, E. O. & Violato, C. (2007). The predictive validity of the MCAT for medical school performance and medical board licensing examinations: A meta-analysis of the published research. *Academic Medicine*, 82, 100-106.
- Eisenberg, N. & Harris, J. D. (1984). Social competence: A developmental perspective. *School Psychology Review*, 13, 267-277.
- Ellis, S. Carette, B., Anseel, F. & Lievens, F. (2014). Systematic reflection: Implications for learning from failures and successes. *Current Directions in Psychological Science*, 23 (1), 67-72.
- Ellis, S. & Davidi, I. (2005). After-event reviews: Drawing lessons from successful and failed experience. *Journal of Applied Psychology*, 90, 857-871.
- Erpenbeck, J. (1997). Selbstgesteuertes, selbstorganisiertes Lernen. In Arbeitsgemeinschaft Qualifikations-Entwicklungs-Management (Hrsg.), *Kompetenzentwicklung '97: Berufliche Weiterbildung in der Transformation – Fakten und Visionen*. Münster: Waxmann.
- Euler, D. (1997). Sozialkompetenz als didaktische Kategorie – vom «didaktischen Impressionsmanagement» zu einem Forschungsprogramm. In R. Dubs & R. Luzzi (Hrsg.), *25 Jahre IWP: Tagungsbeiträge: Schule in Wissenschaft, Politik und Praxis* (S. 279-318). St. Gallen: Hochschule St. Gallen, Institut für Wirtschaftspädagogik.

- Eva, K. W., Reiter, H. I., Rosenfeld, J. & Norman, G. R. (2004). The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Academic Medicine*, 79 (10), 40-42.
- Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J. & Norman, G. R. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education*, 43, 767-775.
- Eva, K. W., Reiter, H. I., Rosenfeld, J., Trinh, K., Wood, T. J. & Norman, G. R. (2012). Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *Journal of the American Medical Association*, 308 (21), 2233-2240.
- 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.
- Ferguson, E., James, D., O'Hehir, F. & Sanders, A. (2003). Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *British Medical Journal*, 326, 429-432.
- File, Q. W. & Remmers, H. H. (1971). *How supervise? Manual 1971 revision*. Cleveland: Psychological Cooperation.
- Fishbach, A., & Finkelstein, S. R. (2012). How feedback influences persistence, disengagement, and change in goal pursuit. In H. Aarts & A. Elliot (Eds.), *goal-directed behavior* (pp. 203-230). New York: Psychology Press.
- Fisseni, H. J. (2004). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Fisseni, H. J., Olbrich, E., Halsig, N., Mailahn, J. & Ittner, E. (1993). *Auswahlgespräche mit Medizinstudenten*. Göttingen: Hogrefe.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 41, 237-358.
- Fleischer, T. (2007). Soziale Kompetenz entwickeln. In T. Fleischer, N. Grewe, B. Jötten, K. Seifried & B. Sieland (Hrsg.), *Handbuch Schulpsychologie: Psychologie für die Schule* (S. 210-220). Stuttgart: Kohlhammer.

- Flexner, A. (1910). *Medical Education in the United States and Canada: A report to the Carnegie Foundation for the advancement of teaching*. New York: The Carnegie Foundation for the Advancement of Teaching.
- Frank, J. R., Snell, L. et al. (2014). *Draft CanMEDS 2015 Physician Competency Framework – Series I*. Ottawa: The Royal College of Physicians and Surgeons of Canada.
- Frey, A. & Balzer, L. (2003). Soziale und methodische Kompetenzen - der Beurteilungsbogen smk: Ein Messverfahren für die Diagnose von sozialen und methodischen Kompetenzen. *Empirische Pädagogik*, 17 (2), 148-175.
- Freyer, K. (2013). *Zum Einfluss von Studieneingangsvoraussetzungen auf den Studienerfolg Erstsemesterstudierender im Fach Chemie*. Berlin: Logos.
- Funke, U. & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection and Assessment*, 6, 115-123.
- Gaba, D. M., (2000). Anaesthesiology as a model for patient safety in health care. *British Medical Journal*, 18, 785-788.
- Gessner, T. M. & Klimoski, R. J. (2006). Making Sense of Situations. In J. A. Weekly & R. E. Polyhart (Eds.), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 13-38). Mahwah: Lawrence Erlbaum Associates.
- Giesen, H., Gold, A., Hummer, A. & Jansen, R. (1986). *Prognose des Studienerfolgs. Ergebnisse aus Längsschnittuntersuchungen*. Frankfurt a. M.: Institut für Pädagogische Psychologie.
- Gillespie, M. A., Oswald, F. L., Schmitt, N., Manheim, L. & Kim, B. (2002). *Construct validation of situational Judgement test of college students success*. Paper presented at the 17th annual convention of the Society for Industrial and Organizational Psychology, Toronto.
- Gold, A. & Souvignier, E. (1997). Examensleistung und Studiererleben bei Hochschulabsolventen. *Zeitschrift für Pädagogische Psychologie*, 11 (1), 53-63.
- Gold, A. & Souvignier, E. (2005). Prognose der Studierfähigkeit. Ergebnisse aus Längsschnittanalysen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37, 214-222.

- Goldfried, M. R. & D'Zurilla, T. J. (1969). A behavior-analytic model for assessing competence. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (pp.151-196). New York: Academic Press.
- Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. New York: Bantam.
- Gough, H. G. & Hall, W. B. (1975). An attempt to predict graduation from medical school. *Journal of Medical Education*, 50, 940-950.
- Gresham, F., Sugai, G., & Horner, R. (2001). Interpreting outcomes of social skills training for students with high-incidence disabilities. *Exceptional Children*, 67, 331-344.
- Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, 62, 255-262.
- Hampe, W., Hissbach, J., Kadmon, M., Kadmon, G., Klusmann, D. & Scheutzel, P. (2009). Wer wird ein guter Arzt? Verfahren zur Auswahl von Studierenden der Human- und Zahnmedizin. *Bundesgesundheitsblatt*, 52, 821-830.
- Hampe, W., Klusmann, D., Buhk, H., Münch-Harrach, D. & Harendza, S. (2008). Reduzierbarkeit der Abbrecherquote im Humanmedizinstudium durch das Hamburger Auswahlverfahren für Medizinische Studiengänge-Naturwissenschaftsteil (HAM-Nat). *GMS Zeitschrift für Medizinische Ausbildung*, 25 (2), 1-11.
- Harden, R. M., Stevenson, M., Downie, W. W., Wilson, G. M. (1975). Assessment of clinical competence using objective structured examination. *British Medical Journal*, 1, 447-51.
- Hargie, O., Saunders, C. & Dickson, D. (1981). *Social Skills in Interpersonal Communication*. London: Croom Helm.
- Harris, M. M. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer and peer-supervisor ratings. *Personnel Psychology*, 41, 43-62.
- Hatala, R. & Norman, G. (2002): Adapting the key features examination for a clinical clerkship. *Medical Education*, 36 (2), 160-165.
- Hattie, J. (1958). Methodology review : Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.

- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77 (1), 81-112.
- Hauenstein, N. M. A., Findlay, R. A. & McDonald, D. P. (2010). Using situational judgement tests to assess training effectiveness: Lessons learned evaluating military equal opportunity advisor trainees. *Military Psychology*, 22, 262-281.
- Havighurst, R. J. (1982). *Developmental tasks and education*. New York: Longman.
- Hell, B., Trapmann, S. & Schuler, H. (2007). Eine Metaanalyse der Prognosekraft von Studierfähigkeitstests. *Empirische Pädagogik*, 21 (3), 251-270.
- Heublein, U., Spangenberg, H. & Sommer, D. (2003). *Ursachen des Studienabbruchs (Analyse 2002, Hochschulplanung Band 163)*. Hannover: Hochschul-Informationssystem.
- Hinsch, R. & Pfungsten, U. (2007). *Gruppentraining sozialer Kompetenzen (5. völlig veränderte Aufl.)*. Weinheim: Psychologie Verlags Union.
- Höschl, C. & Kozeny, J. (1997). Predicting academic performance of medical students: the first three years. *American Journal of Psychiatry*, 154 (6), 87-92.
- Hofmann, M., Rieger, M. A. & Ostermann, T. (2007). Schulische und nicht-schulische Prädiktoren für die Studienplatzzusage an der Universität Witten / Herdecke - Ergebnisse einer QUEST-Analyse. *GMS Zeitschrift für Medizinische Ausbildung*, 24 (4), o.S.
- Hogan, J. B. (1994). Empirical keying of background data measures. In G. S. Stokes, M. D. Mumford & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 69-107). Palo Alto: Consulting Psychologists Press.
- Hogan, J., Barrett, P. & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92, 1270-1285.
- Hojat, M., Robeson, M., Damjanov, I., Veloski, J. J., Glaser, K. & Gonnella, J.S. (1993). Students' characteristics as predictors of academic performance in Medical School. *Academic Medicine*, 68, 635-637.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.

- Holtz, K. L. (1994). *Geistige Behinderung und soziale Kompetenz. Analyse und Integration psychologischer Konstrukte*. Heidelberg: HVA-Edition Schindele.
- Hooper, A. C., Cullen, M. J. & Sackett, P. R. (2006). Operational Threats to the Use of SJTs: Faking, Coaching and Retesting Issues. In J. A. Weekly & R. E. Polyhart (Eds.), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 253-278). Mahwah: Lawrence Erlbaum Associates.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Hossiep, R., Paschen, M. & Mühlhaus, O. (2000). *Persönlichkeitstests im Personalmanagement*. Göttingen: Verlag für Angewandte Psychologie.
- Hossiep, R., Paschen, M. & Mühlhaus, O. (2003). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung* (2. vollständig überarbeitete Aufl.). Göttingen: Hogrefe.
- Hough, L. & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G. S. Stokes, M. D. Mumford & W. A. Owens (Eds.), *Biodata handbook: Theory research, and use of biographical information in selection and performance prediction* (pp. 109-145). Palo Alto: Consulting Psychologists Press.
- Humphreys, L. G. (1960). Investigations of simplex. *Psychometrika*, 25, 313-323.
- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96 (1), 72-98.
- Ihme, J. M., Jia, S., Penzold, M. & Zimmermann, R. (2008). *Die Macht der Sterne - wie Sie sehen, sehen Sie nichts*. Posterpräsentation online: <http://www2.uni-jena.de/svw/allgpsy2/emprakong3/poster/9.pdf> (Zugriff: 20.06.2011).
- Ittner, E. & Halsig, N. (1993). Prognostische Relevanz des Auswahlgesprächs als qualitativ neue Methode beim Zugang zum Studium der Medizin. In F. Baumgärtel (Hrsg.), *Klinische Psychologie im Spiegel ihrer Praxis* (S. 113-119). Bonn: Deutscher Psychologenv Verlag.
- James, D., Yates, J. & Nicholson, S. (2010). Comparison of A level and UKCAT performance in students applying to UK medical and dental schools in 2006: cohort study. *British Medical Journal*, 340, 1-7.

- Jones, A. C., Courts, F. J., Sandow, P. L. & Watson, R. E. (1997). Myers-Briggs Type Indicator and dental school performance. *Journal of Dental Education*, 61 (12), 928-933.
- Jones, E. E. & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. A. Kanouse, H. H. Kelly, R. E. Nisbett, S. Valines & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79-94). Morristown: General Learning Press.
- Jünger, J. & Nikendei, C. (Hrsg.). (2005). OSCE Prüfungsvorbereitung Innere Medizin. Stuttgart: Thieme.
- Julian, E. R. (2005). Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine*, 80 (10), 910-917.
- Kadmon, G., Möltner, A., Kahmann, J. & Kadmon, M. (zur Publikation eingereicht). Evidence based weighted compensatory admission procedure: different competencies predict similar levels of success.
- Kanning, U. P. (2003). *Diagnostik sozialer Kompetenzen*. Göttingen: Hogrefe.
- Kanning, U. P. (2005). Computergestützte Personalauswahl von Polizeibeamten. *Zeitschrift für Arbeits- und Organisationspsychologie*, 49, 140-146.
- Kanning, U. P. (2009). *Inventar sozialer Kompetenzen (ISK)*. Manual. Göttingen: Hogrefe.
- Kanning, U. P., Grewe, K., Hollenberg, S. & Hadouch, M. (2006). From the subject's point of view: Reaction to different types of situational judgement items. *European Journal of Psychological Assessment*, 22, 168-176.
- Kauffeld, S., Frieling, E. & Grote, S. (2002). Soziale, personale, methodische oder fachliche: Welche Kompetenzen zählen bei der Bewältigung von Optimierungsaufgaben in betrieblichen Gruppen? *Zeitschrift für Psychologie*, 210 (4), 197-208.
- Kohn L. T., Corrigan J. M. & Donaldson, M. S. (2000). *To err is human*. Washington: National Academy Press.
- Kulatunga-Moruzi, C. & Norman, G. R. (2002). Validity of admissions measures in predicting performance outcomes: the contribution of cognitive and non-cognitive dimensions. *Teaching and Learning in Medicine*, 14 (1), 34-42.

- Kelava, A. & Moosbrugger, H. (2008). Deskriptivstatistische Evaluation von Items (Itemanalysen) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Testkonstruktion* (S. 73-98). Heidelberg: Springer.
- Kersting, M. (2005). Beratung und Auswahl von Studienplatzbewerbern: Ziele und Methoden. *Psychologische Rundschau*, 56 (2), 149-150.
- Kiessling, C., Dieterich, A., Fabry, G. Hölzer, H., Wolf, L. Mühlinghaus, I., Pruskil, S., Scheffer, S. & Schubert, S. (2010). Communication and social competencies in medical education in German-speaking countries: The Basel Consensus Statement. Results of a Delphi Survey. *Patient Education and Counseling* 81, 259-266.
- Konegen-Grenier, C. (2001). *Studierfähigkeit und Hochschulzugang*. Köln: Deutscher Instituts-Verlag.
- Krokos, K. J., Meade, A. W., Cantwell, A. R., Pond, S. B., & Wilson, M. A. (2004). *Empirical keying of situational judgement tests: Rationale and some examples*. Paper presented at the 19th annual meeting of the Society for Industrial and Organizational Psychology, Chicago.
- Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it. How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*. 77 (6), 1121-1134.
- Kuh, G., Kinzie, J., Buckley, J.A., Bridges, B.K. & Hayek, C. (2007). *Piecing Together the Student Success Puzzle: Research, Propositions and Recommendations* (ASHE Higher Education Report, Volume 32, Number 5). San Francisco: Jossey-Bass.
- Lakey, B., & Cohen, S. (2000). Social support theory and measurement. In S. Cohen, L. Underwood, & B. Gottlieb (Eds.), *Measuring and intervening in social support* (pp. 29-52). New York: Oxford University Press.
- The Lancet (2010). What makes a good doctor? *The Lancet*, 376 (28), 658.
- Lang, D. (2008). *Soziale Kompetenz und Persönlichkeit - Zusammenhänge zwischen sozialer Kompetenz und den Big Five der Persönlichkeit bei jungen Erwachsenen*. Unveröff. Diss., Universität Koblenz-Landau.
- Latham, G. P., Saari, L. M., Pursell, E. D. & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.

- Law, K., Wong, C. & Song, L. (2004). The construct and criterion validity of emotional intelligence and its potential utility for management studies. *Journal of Applied Psychology, 89* (3), 483-496.
- Legree, P. (1995). Evidence for an oblique social intelligence factor established with a likert-based testing procedure. *Intelligence, 21*, 247-266.
- Libbrecht, N., Lievens, F., Carette, B. & Côté, S. (2014). Emotional intelligence predicts success in medical school. *Emotion, 14* (1), 64-73.
- Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests. *Medical Education, 47*, 182-189.
- Lievens, F. & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact and construct validity. *International Journal of Selection and Assessment, 10*, 245-257.
- Lievens, F., Ones, D. S. & Dilchert, S. (2009a). Personality Scale validities increase throughout medical school. *Journal of Applied Psychology, 94*, 1514-1535.
- Lievens, F. & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91* (5), 1181-1188.
- Lievens, F., Sackett, P. R. & Buyse, T. (2009b). The effects of response instructions on situational judgement test performance and validity in a high-stakes context. *Journal of Applied Psychology, 94*, 1095-1101.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Psychologie Verlags Union.
- Lindgren, A. & Heikkinen, A. (Eds.). (2004). *Social Competences in Vocational and Continuing Education*. Bern: Lang.
- Lukesch, H., Kormann, A. & Mayrhofer, S. (2002). *Revidierte Fassung des Prüfsystems für Schul- und Bildungsberatung von Horn für 4. Bis 6. Klassen (PSB-R 4-6), Manual*. Göttingen: Hogrefe.
- MacCallum, R. C., Widaman, K. F., Zhang, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4* (1), 87-99.

- Mace, F. C., Belfiore, P., & Shea, M. (1989). Operant theory and research in self-regulation. In B.J. Zimmerman & D.H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 27-50). New York: Springer.
- Magnus, S. A. & Mick, S. S. (2000). Medical schools, affirmative action, and the neglected role of social class. *American Journal of Public Health, 90* (8), 1197-1201.
- Makoul G. (2001). Essential elements of communication in medical encounters: The Kalamazoo consensus statement. *Academic Medicine, 76* (4), 390-393.
- Makoul, G. & Schofield, T. (1999). Communication teaching and assessment in medical education: an international consensus statement. Netherlands Institute of Primary Health Care. *Patient Education Counseling, 37* (2), 191-195.
- Malti, T. & Perren, S. (Hrsg.). (2008). *Soziale Kompetenz bei Kindern und Jugendlichen. Entwicklungsprozesse und Förderungsmöglichkeiten*. Stuttgart: Kohlhammer.
- Manhal, S., Neges, H., Reibnegger, G., Ithaler, G. & Smolle, J. (2012). *Situational Judgement Test – eine wesentliche Erweiterung des Grazer Auswahlverfahrens für Medizinstudierende*. Poster auf der 15. Grazer Konferenz – Qualität der Lehre: Teaching and Learning – Expanding our Resources 28. - 30. April 2011 Wien.
- Mathers, J., Sitch, A., Marsh, J.L. & Parry, J. (2002). Widening access to medical education for underrepresented socioeconomic groups: population-based cross-sectional analysis of UK data. *British Medical Journal, 342*.o.S.
- Matthews, G., Zeidner, M., & Roberts, R. (2002). *Emotional Intelligence: Science and myth*. Boston: MIT Press.
- Mayer, J. D., Salovey, P. & Caruso, D. R. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (2nd ed., pp. 396-422). New York: Cambridge University Press.
- Mayer, J.D., Salovey, P., Caruso, D. R. & Sitarenios, G. (2003). Modeling and measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*, 97-105.
- Mayring, P. (1990). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Deutscher Studienverlag.

- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgement tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60* (1), 63-91.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgement tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86* (4), 730-740.
- McDaniel, M. A. & Nguyen, N. T. (2001). Situational judgement tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103-113.
- McFall, R. (1982). A review and reformulation of the concept of social skills. *Behavioral Assessment, 4*, 1-33.
- MacLane, C. N., Barton, M. G., Holloway-Lundy, A. E. & Nickles, B. J. (2001). *Keeping score: Expert weights on situational judgement responses*. Paper presented at the 16th annual meeting of the Society for Industrial and Organizational Psychology, San Diego.
- McManus, I. C., Powis, D. A., Wakeford, R., Ferguson, E., James, D. & Richard, P. (2005). Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *British Medical Journal, 331*, 555-559.
- Mercer, A. & Puddey, I.B. (2011). Admission selection criteria as predictors of outcomes in an undergraduate medical course: A prospective study. *Medical Teacher, 33*, 997-1004.
- Möller, J. H. (2010). *Kundenorientierung im Hotelfach. Die Entwicklung und Validierung eines Situational Judgement Tests*. Unveröff. Diss., Westfälische Wilhelms-Universität, Münster.
- Montague, W. & Odds, F. C. (1990). Academic selection criteria and subsequent performance. *Medical Education, 24* (2), 151-157.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*, 683-729.
- Moss, F. A. (1926). Do you know how to get along with people? Why some people get ahead in the world while others do not. *Scientific American, 135*, 26-27.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640-647.

- Motowidlo, S. J., Hooper, A. C. & Jackson, H. L. (2006). A Theoretical Basis for Situational Judgement Tests. In J. A. Weekly & R. E. Polyhart (Eds.), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 57-81). Mahwah: Lawrence Erlbaum Associates.
- Motowidlo, S. J. & Tippins, N. (1993). Further studies of the low fidelity simulation in the form of situational inventory. *Journal of Occupational and Organizational Psychology*, 66, 337-344.
- Müller, F. H. (2001). *Studium und Interesse. Eine empirische Untersuchung bei Studierenden*. Münster: Waxmann.
- Mumford, M. D. & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement*, 11, 1-31.
- Nguyen, N. T., Biderman, M. D. & McDaniel, M. A. (2005). Effects of response instructions on faking in a situational judgement test. *International Journal of Selection and Assessment*, 13, 250-260.
- Norcini, J. J., Blank, L. L., Arnold, G. K. & Kimball, H. R. (1995). The mini-CEX (clinical evaluation exercise): A preliminary investigation. *Annals of Internal Medicine*, 123 (10), 795-799.
- Norcini, J. J. & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher*, 29 (9), 855-871.
- Northrop, L. C. (1989). *The psychometric history of selected ability constructs*. Washington: U.S. Office of Personnel Management.
- O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. L. & Lawrence, A. (2007). Incremental validity of situational judgement tests for task and contextual job performance. *International Journal of Selection Assessment*, 15 (1), 19-29.
- O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, 43, 971-990.
- Olson-Buchanan, J. B. & Drasgow, F. (2006). Multimedia Situational Judgement Tests: The Medium Creates the Message. In J. A. Weekly & R. E. Polyhart (Eds.), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 253-278). Mahwah: Lawrence Erlbaum Associates.

- Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A. & Donovan, M. A. (1998). An interactive video assessment of conflict resolution skills. *Personnel Psychology, 51*, 1-24.
- Ones, D. S. & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17* (6), 609-626.
- Organisation for Economic Cooperation and Development. (2004). *Career Guidance and Public Policy - Bridging the Gap*. Paris: Author.
- Ostendorf, F. & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae: Revidierte Fassung*. Göttingen: Hogrefe.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college effects students: Findings and insights from twenty years of search*. San Francisco: Jossey-Bass.
- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.
- Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M. & O'Neill, P. (2012). Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education, 46*, 850-868.
- Patterson, F., Baron, H., Carr, V., Plint, S. & Lane, P. (2009). Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. *Medical Education, 43* (1), 50-57.
- Paullin, C. & Hanson, M. A. (2001). *Comparing the validity of rationally-derived and empirically-derived scoring keys for a situational Judgement inventory*. Paper presented at the 16th annual meeting of the Society for Industrial and Organizational Psychology, San Diego.
- Paukert, J. L., Richards, M. L. & Olney, C. (2002). An encounter card system for increasing feedback to students. *The American Journal of Surgery, 183*, 300-304.
- Payton, J. W., Wardlaw, D. M., Graczyk, P. A., Bloodworth, M. R., Tompsett, C. J. & Weissberg, R. P. (2000). Social and emotional learning: A framework for promoting mental health and reducing risk behaviors in children and youth. *Journal of School Health, 70*, 179-185.
- Petermann, F. (2002). Klinische Kinderpsychologie: Das Konzept der sozialen Kompetenz. *Zeitschrift für Psychologie, 210* (4), 175-185.

- Petermann, U. & Petermann, F. (1989). *Training mit sozial unsicheren Kindern*. München: Psychologie Verlags Union.
- Petermann, U. & Petermann, F. (2003). Ängste im Kindesalter. Therapeutische Ansätze und Entspannungsverfahren. *Praxis der Psychomotorik*, 28, 88-92.
- Pereira, G. & Harvey, V. S. (1999). Situational Judgement tests: Do they measure ability, personality, or both? In J. P. Clevenger (chair), *The Construct Validity of the Situational Judgement Inventory*. Symposium presented at the 14th annual conference of the Society for Industrial and Organizational Psychology, Atlanta.
- Polyhart, R. E. (2006). The Predictor Response Process Model. In J. A. Weekly & R. E. Polyhart (Eds.), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 83-105). Mahwah: Lawrence Erlbaum Associates.
- Polyhart, R. E. & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of SJT. *International Journal of Selection and Assessment*, 11, 1-16.
- Polyhart, R. E. & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racio-ethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153-172.
- Polyhart, R. E. & MacKenzie, W. I. (2011). Situational Judgement Tests: A Critical Review and Agenda for the Future. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization* (pp. 237-252). Washington: American Psychological Association.
- Polyhart, R. E. & Ryan, A. M. (2000). *Integrating personality tests with situational judgement tests for the prediction of customer service performance*. Paper presented at the 15th annual conference of the Society for Industrial and Organizational Psychology, New Orleans.
- Powis, D. (2010). Improving the selection of medical students. *British Medical Journal*, 340, 708.
- Prideaux, D., Roberts, C., Eva, K., Centeno, A., Mccrorie, P., McManus, C., Patterson, F., Powis, D., Tekian, A. & Wilkinson, D. (2011). Assessment for selection for the health care professions and specialty training: Consensus statement and recommendations from the Ottawa 2010 Conference, *Medical Teacher*, 33, 215-223.

- Rammstedt, B. & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K): Entwicklung und Validierung eines ökonomischen Inventars zur Erfassung der fünf Faktoren der Persönlichkeit. *Diagnostica*, 51 (4), 195-206.
- Rasch, B. Friese, M. Hofmann, W. J. & Naumann, E. (2006). *Quantitative Methoden* (Band 2, 2. Aufl.). Heidelberg: Springer.
- Renk, K. & Phares, V. (2004). Cross informant ratings of social competencies in children and adults. *Clinical Psychology Review*, 24, 239-254.
- Reschke, K. (1995). Soziale Kompetenz entwickeln – Ressourcen entdecken helfen. Interventive Forschung auf der Basis des Kompetenzmodells von Vorweg & Schröder (1980). In: J. Margraf & K. Rudolf (Hrsg.), *Training sozialer Kompetenz* (S. 205-228). Baltmannsweiler: Röttger-Schneider.
- Rey, G. D. & Wender, K. F. (2008). *Neuronale Netze. Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung*. Bern: Huber.
- Reynolds, D. H., Sydell, E. J., Scott, D. R. & Winter, J. L. (2000). *Factors affecting situational judgement test characteristics*. Paper presented at the 15th annual conference of the Society for Industrial and Organizational Psychology, New Orleans.
- Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, 85 (6), 880-887.
- Riemann, R. & Allgöwer, A. (1993). Eine deutschsprachige Fassung des „Interpersonal Competence Questionnaire“ (ICQ). *Zeitschrift für Differentielle und Diagnostische Psychologie*, 14, 153-163.
- Rindermann, H. (2009). *Emotionale-Kompetenz-Fragebogen (EKF)*. Göttingen: Hogrefe.
- Rindermann, H. & Oubaid, V. (1999). Auswahl von Studienanfängern durch Universitäten – Kriterien, Verfahren und Prognostizierbarkeit des Studienerfolgs. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 20 (3), 172-191.
- Robins, L. S., Braddock, C. H. & Fryer-Edwards, K. A. (2002). Using the American Board of Internal Medicine's "Elements of Professionalism" for undergraduate ethics education. *Academic Medicine*, 77 (6), 523-31.

- Robbins, S. B., Lauver, K. Le. H., Davis, D., Langley, H. & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130 (2), 261-288.
- Rolf, E. R. & Powis, D. A. (1997). Selecting Australian doctors of the future. *Medical Journal Australia*, 166, 229-230.
- Rosenthal, R., Hall, J. A., Archer, D., DiMatteo, M. R. & Rogers, P. L. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore: John Hopkins University Press.
- Rost, D. H. (2013). *Interpretation und Bewertung pädagogisch-psychologischer Studien. Eine Einführung* (3. Aufl.). Stuttgart: UTB.
- Rychen, D. S. & Salganik, L. H. (2003). *Key competencies for a successful life and a well-functioning society*. Göttingen: Hogrefe & Huber.
- Sacco, J. M., Scheu, C., Ryan, A. M. & Schmitt, N. W. (2000a). *Understanding race differences on situational judgement tests using readability statistics*. Paper presented at the 15th annual convention of the Society for Industrial and Organizational Psychology, New Orleans.
- Sacco, J. M., Schmidt, D. B. & Rogg, K. L. (2000b). *Using readability statistics and reading comprehension scores to predict situational judgement tests performance, Black-White differences and validity*. Paper presented at the 15th annual convention of the Society for Industrial and Organizational Psychology, New Orleans.
- Salovey, P. & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality*, 9, 185-211.
- Schiefele, U. & Pekrun, R. (1996). Psychologische Modelle des selbstgesteuerten und fremdgesteuerten Lernens. In F.E. Weinert (Hrsg.), *Psychologie des Lernens und der Instruktion* (Enzyklopädie der Psychologie, Serie Pädagogische Psychologie, Bd. 2, S. 249-278). Göttingen: Hogrefe.
- Schmidt, J. U. (1995). Psychologische Messverfahren für soziale Kompetenzen. In B. Seyfried (Hrsg.), *Stolperstein Sozialkompetenz. Was macht es so schwierig sie zu erfassen, zu fördern und zu beurteilen* (Berichte zur Beruflichen Bildung Bd. 179, S.117-135). Bielefeld: Bertelsmann.
- Schmidt, F. L. & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability and jobknowledge. *Current Directions in Psychological Science*, 2, 8-9.

- Schmidt, F. & Hunter, J. (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.
- Schmidt-Atzert, L., Büttner, G. & Bühner, M. (2004). Theoretische Aspekte der Aufmerksamkeits-/Konzentrationsdiagnostik. In G. Büttner & L. Schmidt-Atzert (Hrsg.), *Diagnostik von Konzentration und Aufmerksamkeit. Jahrbuch der pädagogisch-psychologischen Diagnostik* (Bd. 3, S. 3-22). Göttingen: Hogrefe.
- Schmitt, N. & Chan, D. (2006). Situational Judgement Tests: Method or Construct? In J. A. Weekly & R. E. Polyhart (Eds.), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 135-155). Mahwah: Lawrence Erlbaum Associates.
- Schneider, R., Ackerman, P. & Kanfer, R. (1996). "To act wisely in human relations": Exploring the dimensions of social competence. *Personality and Individual Differences*, 4, 469-481.
- Schneider, B. H. & Byrne, B. M. (1985). Children's social skills training: A meta-analysis. In B.H. Schneider, K. Rubin, & J.E. Ledingham (Eds.), *Children's Peer relations: Issues in assessment and intervention* (pp. 175-190). New York: Springer.
- Schumann, S. (2008). Förderung von Sozialkompetenzen im Unterricht – aber wie? Die Bedeutung und Wirkungsweise kooperativer Lernformen. *Netzwerk Wissenschaft*, 2, 10-23.
- Segrin, F. (2000). Poor social skills are a vulnerability factor in the development of psychosocial problems. *Human Communication Research*, 26 (3), 489-514.
- Seyfried, B. (1995). Soziales Verhalten: Die Illusion „objektiver Beurteilung“. In B. Seyfried (Hrsg.), *Stolperstein Sozialkompetenz* (S. 137-152). Bielefeld: Bertelsmann.
- Sheffer, C., Penn, D. & Cassisi, J. (2001). The effects of impression management demands on heart rate, self-reported social anxiety, and social competence in undergraduate males. *Journal of Anxiety Disorders*, 15, 171-182.
- Simmons-Morton, B. & Crump, A. (2003). Association of parental involvement and social competence with school adjustment and engagement among sixth graders. *Journal of School Health*, 73 (3), 121-126.
- Simpson, M., Buckman, R., Stewart, M., Maguire, P., Lipkin, M., Novack, D. & Till, J. (1991). Doctor-patient communication: the Toronto consensus statement. *British Medical Journal*, 303, 1385-1387.

- Siu, E. & Reiter, H. I. (2009). Overview: What's worked and what hasn't as a guide towards predictive admissions tool development. *Advances in Health Sciences Education*, 14 (5), 759-775.
- Smith, K. G. & McDaniel, M. A. (1998). Criterion and construct validity evidence for a situational judgement measure. Paper presented at the 13. Annual Conference of the Society of Industrial and Organizational Psychology, Dallas.
- Sonntag, K. H. & Schäfer-Rauser, U. (1993). Selbsteinschätzung beruflicher Kompetenzen bei der Evaluation von Bildungsmaßnahmen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37 (4), 163-171.
- Spitzberg, B. H. & Cupach, W. R. (1989). *Handbook of interpersonal competence research*. New York: Springer.
- Stemler, S. E., Elliott, J., Grigorenko, E. L. & Sternberg, R. J. (2006). There's more to teaching than instruction: Seven strategies for dealing with the social side of teaching. *Educational Studies*, 31 (1), 85-102.
- Stemler, S. E. & Sternberg, R. J. (2006). In J. A. Weekly & R. E. Polyhart (Eds.), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 107-131). Mahwah: Lawrence Erlbaum Associates.
- Sternberg, R. J. (1988). *The Triarchic Mind. A new theory of human intelligence*. New York: Viking Penguin.
- Sternberg, R. J. (1997). *Successful intelligence: How practical and creative intelligence determine success in life*. New York: Plume.
- Sternberg, R. J. (1999). The theory of successful intelligence. *Review of General Psychology*, 3, 292-316.
- Stumpf, H. & Nauels, H. U. (1990). Zur prognostischen Validität des «Tests für medizinische Studiengänge» (TMS) im Studiengang Humanmedizin. *Diagnostica*, 36 (1), 16-32.
- Such, M. J. & Schmidt, D. B. (2004). *Examining the effectiveness of empirical keying: A cross-cultural perspective*. Paper presented at the 19th annual conference of the Society for Industrial and Organizational Psychology, Chicago.

- Tashiro, J.; Shimpuku, Y., Naruse, K., Matsutani, M. & Matsutani, M. (2013). Concept analysis of reflection in nursing professional development. *Japan Journal of Nursing Science*, 10 (2), 170-179.
- Thiel, F., Veit, S., Blüthmann, I., Lepa, S. & Ficzkó, M. (2008). *Ergebnisse der Befragung der Studierenden in den Bachelorstudiengängen an der Freien Universität Berlin - Sommersemester 2008*. <http://www.fu-berlin.de/universitaet/entwicklung/qualitaetsmanagement/bachelorbefragung/bachelorbefragung-2008.pdf?1304061426> (Zugriff: 20.03.2014).
- Thoits, P. A. (1995). Stress, coping, and social support processes: Where are we? What next? *American Sociological Association*, 35, 53-79.
- Thorndike, R. L. (1920). Intelligence and its use. *Harper's Magazin*, 140, 227-235.
- Thorndike, R. L. (1936). Factor analysis of social and abstract intelligence. *Journal of Educational Psychology*, 27, 231-233.
- Thorndike, R. L. & Stein, S. (1937). An evaluation of attempts to measure social intelligence. *Psychological Bulletin*, 34 (5), 275-285.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45, 89-125.
- Tinto, V. (1993). *Leaving college: Rethinking the cause and cures of student attrition* (2nd ed.). Chicago: University of Chicago.
- Tripe, M. D. & Foti, R. J. (2003). *An evaluation of the construct validity of situational judgement tests*. Paper presented at the 18th annual conference of the Society for Industrial and Organizational Psychology, Orlando.
- Trost, G., Hrsg. (1994). *Test für medizinische Studiengänge (TMS): Studien zur Evaluation*. Bonn: Institut für Test- und Begabungsforschung (18. Arbeitsbericht).
- Trost, G., Hrsg. (1996). *Test für medizinische Studiengänge (TMS): Studien zur Evaluation*. Bonn: Institut für Test- und Begabungsforschung (20. Arbeitsbericht).
- Trost, G., Klieme, E. & Nauels, H. U. (1997). Prognostische Validität des Tests für medizinische Studiengänge (TMS). In T. Hermann (Hrsg.), *Hochschulentwicklung - Aufgaben und Chancen* (S. 57-78). Heidelberg: Asanger.

- Trost, G., Flum, F. & Fay, E., Klieme, E., Maichle, U., Meyer, M. & Nauels, H. U. (1998). *Evaluation des Tests für Medizinische Studiengänge (TMS): Synopse der Ergebnisse*. Bonn: ITB.
- Tsay, C. H.-H. (2013). Understanding students' adaptation to graduate school: An integration of social support theory and social learning theory. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 74, o.S.
- Tutton P. J. (1996). Psychometric test results associated with high achievement in basic science components of a medical curriculum. *Academic Medicine*, 71 (2), 181-186.
- Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Ullrich, R. & Ullrich de Muynck, R. (Hrsg.) (1978). *Soziale Kompetenz. Experimentelle Ergebnisse zum Assertiveness-Trainingsprogramm ATP* (Band 1: Meßmittel und Grundlagen). München: Pfeiffer.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
- Venezia, A., Callan, P. M., Finney, J.E., Kirst, M. W. & Usdan, M. D. (2005). *The governance divide: A report on a four-state study on improving college readiness and success*. San Jose: The Institute for Educational Leadership, the National Center for Public Policy and Higher Education, and the Stanford Institute for Higher Education Research.
- Von Fragstein, M., Silverman, J., Cushing, A., Quilligan, S., Salisbury, H. & Wiskin, C. (2008). UK consensus statement on the content of communication curricula in undergraduate medical education. *Medical Education*, 42, 1100-1107.
- Walker, D. & Goldenberg, R. (2004). *Bringing selection into the 21st century: A look at video based testing within U.S. Customs and Border Protection*. Paper presented at the annual conference of the International Public Management Association – Assessment Council, Seattle.
- Waters, E. & Soufre, L. A. (1983). Social competence as a developmental construct. *Development Review*, 3, 79-97.
- Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology*, 50, 25-49.

- Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, 52 (3), 679-700.
- Weekly, J.A. & Polyhart, R.E. (Eds.). (2006). *Situational judgement tests: Theory, Measurement and Application*. Mahwah: Lawrence Erlbaum.
- Weekley, J. A., Polyhart, R. E. & Holtz, B. C. (2006). On the Development of Situational Judgement Tests: Issues in Item Development, Scaling and Scoring. In J. A. Weekly & R. E. Polyhart (Eds.), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 57-81). Mahwah: Lawrence Erlbaum Associates.
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2- Revision. CFT 20-R*. Göttingen: Hogrefe.
- Wentzel, K. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, 62, 1066-1078.
- West, S.G., Finch, J. F. & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks: Sage.
- Westermann, R., Heise, E., Spies, K. & Trautwein, U. (1996). Identifikation und Erfassung von Komponenten der Studienzufriedenheit. *Psychologie in Erziehung und Unterricht*, 43, 1-22.
- Whetzel, D. L. & McDaniel, M. A. (2009). Situational judgement tests: An overview of current research. *Human Research Management Review*, 19, 188-202.
- Whetzel, D. L., McDaniel, M. A. & Nguyen, N. (2008). Subgroup differences in situational judgement test performance: a meta-analysis. *Human Performance*, 21, 291-309.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wissenschaftsrat (2004). *Empfehlungen zur Reform des Hochschulzugangs*. (Drucksache 5920/04). Berlin: Wissenschaftsrat. <http://www.wissenschaftsrat.de/texte/5920-04.pdf> (Zugriff: 12.06.2012).
- Wottawa, H. & Hossiep, R. (1987). *Grundlagen psychologischer Diagnostik. Eine Einführung*. Göttingen: Hogrefe.

Zhi-Feng, E. & Lee, C. Y. (2013). Using peer feedback to improve learning via online peer assessment. *The Turkish Online Journal of Educational Technology*, 12 (1), 187-199.

Ziv, A., Rubin, O., Moshinsky, A., Gafni, N., Kotler, M., Dagan, Y., Lichtenberg, D., Mekori, Y. A. & Mittelman, M. (2008). MOR: a simulation-based assessment centre for evaluating the personal and interpersonal qualities of medical school candidates. *Medical Education*, 42 (10), 991-998.

Tabellenverzeichnis

Tabelle 1: Aufgabengruppen des TMS	13
Tabelle 2: Prädiktive Validität der Hochschulzulassungsberechtigung	17
Tabelle 3: Prädiktive Validität Einzelfachnoten	19
Tabelle 4: Vorhersageleistung von Studierfähigkeitstests	19
Tabelle 5: Entwicklung des AdH in Heidelberg	25
Tabelle 6: Dimensionen sozialer Kompetenzen	34
Tabelle 7: Framework of Social and Emotional Learning	34
Tabelle 8: Instruktionsformen von SJTs	53
Tabelle 9: Reliabilitäten in Abhängigkeit von der Instruktionsform des SJTs	58
Tabelle 10: Fragebögen der Anforderungsanalyse am Beispiel der Verantwortungsübernahme	68
Tabelle 11: Demografie der Teilnehmer der Anforderungsanalyse	69
Tabelle 12: Tätigkeitsfelder der Teilnehmer der Anforderungsanalyse	70
Tabelle 13: Wichtigkeit sozialer Kompetenzen in Studium und Beruf (Humanmediziner)	72
Tabelle 14: Wichtigkeit sozialer Kompetenzen in Studium und Beruf (Zahnmediziner)	73
Tabelle 15: Wichtigkeit sozialer Kompetenzen in Studium vs. Beruf (Human- vs. Zahnmediziner)	75
Tabelle 16: Wichtigkeit sozialer Kompetenzen in Studium vs. Beruf (Gesamtgruppe) .	76
Tabelle 17: Paarweiser Vergleich der Wichtigkeit sozialer Kompetenzen in Studium und Beruf	78
Tabelle 18: Zusammenhang Wichtigkeit sozialer Kompetenzen mit Kontrollvariablen .	79
Tabelle 19: Verständlichkeit der SJT-Items	90
Tabelle 20: Teilnehmerdemografie Auswertungsschlüssel	95
Tabelle 21: Realitätsnähe und Beurteilerübereinstimmung	98
Tabelle 22: Punktbildung	102
Tabelle 23: Altersverteilung der Stichproben	113
Tabelle 24: Deskriptive Statistik auf Itemebene	118
Tabelle 25: Multivariate Normalverteilung	120
Tabelle 26: Modelfit-Indizes	121
Tabelle 27: Regressionskoeffizienten und Kommunalitäten der KFA	122
Tabelle 28: Interkorrelationen und MSA-Koeffizienten der SJT-Items	124
Tabelle 29: Erklärte Gesamtvarianz nach dem Kaiser-Guttman-Kriterium	127
Tabelle 30: Ergebnisse Parallelanalyse	128
Tabelle 31: Ergebnisse MAP-Test	129
Tabelle 32: Prozent der nicht reproduzierten Residuen	130

Tabelle 33: zweifaktorielle Lösungen Stichprobe 1	131
Tabelle 34: Unrotierte vierfaktorielle Lösung Stichprobe 2.....	132
Tabelle 35: Promaxrotierte vierfaktorielle Lösung Stichprobe 2	133
Tabelle 36: Interkorrelation vierfaktorielle Lösung Stichprobe 2.....	134
Tabelle 37: Einfaktorielle Lösungen Stichprobe 1 und 2	135
Tabelle 38: Deskriptive Daten der SJT-Skala	137
Tabelle 39: Trennschärfe und Itemschwierigkeit der SJT-Skala	139
Tabelle 40: Deskriptive Statistik der Validitätskriterien.....	140
Tabelle 41: Korrelationen der SJT-Skala mit Validierungskriterien und Kontrollvariablen	141
Tabelle 42: SJT-Punktzahl abhängig vom Alter in Stichprobe 1	142
Tabelle 43: Qualitative Rückmeldungen zum SJT	144
Tabelle 44: Untersuchungsablauf der Validierungsstudie	157
Tabelle 45: Primär- und Sekundärfaktoren des ISKs	159
Tabelle 46: Verteilung Fachsemester	162
Tabelle 47: Verteilung medizinische Ausbildung und Erstsprache	163
Tabelle 48: Deskriptive Statistik SJT-Items an Medizinstudierenden	166
Tabelle 49: Interkorrelation SJT-Items.....	168
Tabelle 50: Deskriptive Statistik SJT Medizinstudierende	169
Tabelle 51: Trennschärfe und Itemschwierigkeit der SJT-Items Medizinstudierende .	171
Tabelle 52: Deskriptive Statistik ISK.....	172
Tabelle 53: Selbst- vs. Fremdeinschätzung ISK-K.....	173
Tabelle 54: Deskriptive Statistik Selbsteinschätzungsskalen Medizinstudierende	174
Tabelle 55: Deskriptive Statistik Außenkriterien sozialer Kompetenzen.....	174
Tabelle 56: Deskriptive Statistik BFI-K.....	175
Tabelle 57: Deskriptive Statistik kognitive Kriterien.....	176
Tabelle 58: Deskriptive Statistik subjektive Studienerfolgskriterien	176
Tabelle 59: Deskriptive Statistik vorklinische Studienleistung	176
Tabelle 60: Deskriptive Statistik klinische Studienleistung	177
Tabelle 61: Zusammenhang ISK und SJT	178
Tabelle 62: Zusammenhang Selbsteinschätzungsskalen und SJT	180
Tabelle 63: Zusammenhang Außenkriterien und SJT	181
Tabelle 64: Zusammenhang BFI-K und SJT	181
Tabelle 65: Zusammenhang kognitive Kriterien und SJT.....	182
Tabelle 66: Zusammenhang subjektive Studienerfolgskriterien und SJT	182
Tabelle 67: Zusammenhang vorklinische Studienleistungen und SJT	183
Tabelle 68: Zusammenhang klinische Studienleistungen und SJT	184

Tabelle 69: Prädiktive Validität des SJTs für den subjektiven Studienerfolg	185
Tabelle 70: Inkrementelle Validität des SJTs für den subjektiven Studienerfolg.....	186
Tabelle 71: Interkorrelation der Prädiktoren.....	187
Tabelle 72: Kollinearitätsstatistik	188
Tabelle 73: Ergebnisse der Hypothesentestung	199

Abbildungsverzeichnis

Abbildung 1. Bedingungsmodell des Studienerfolgs.....	15
Abbildung 2. Varianzaufklärung des Studienerfolgs durch psychosoziale Faktoren.....	23
Abbildung 3. Studierendenauswahlmodell.....	27
Abbildung 4. Definition von sozialer Kompetenz und sozial kompetentem Verhalten...	31
Abbildung 5. Verwandte Konstrukte sozialer Kompetenzen.....	32
Abbildung 6. Modell sozialer Kompetenzen nach Holtz	36
Abbildung 7. Drei-Ebenen Modell sozialer Kompetenzen	37
Abbildung 8. Draft CanMed Framework.....	38
Abbildung 9. Eigenes Arbeitsmodell sozialer Kompetenzen.	41
Abbildung 10. SJT-Item aus dem Management Bereich.....	45
Abbildung 11. Tacit Knowledge Theory	47
Abbildung 12. Theory of SJT procedural knowledge.....	48
Abbildung 13. Schritte und Aspekte der SJT-Entwicklung	50
Abbildung 14. Konstruktbezogene Antworten durch hohe und niedrige Ausprägung von Verträglichkeit.	51
Abbildung 15. Konstrukt bezogene Antworten durch Problemlösestrategien.	52
Abbildung 16. SJT-Item für die Facharztauswahl.	63
Abbildung 17. Entwicklungsablauf des SJTs für soziale Kompetenzen.....	66
Abbildung 18. Checkliste zur Entwicklung der Antwortoptionen des SJTs.	85
Abbildung 19. Instruktion des SJTs in der Vortestung.	88
Abbildung 20. Drehbuchauszug des SJTs.....	106
Abbildung 21. Angenommene zweifaktorielle Struktur des SJTs.....	107
Abbildung 22. Kritische Filmstelle und Instruktion des SJTs.	111
Abbildung 23. Verteilung der SJT-Skala.....	138
Abbildung 24. Nützlichkeit und Akzeptanz des SJTs.....	143
Abbildung 25. Verteilung SJT Validierungsstudie Medizinstudierende.....	170

Abkürzungsverzeichnis

Abb.	Abbildung
AdH	Auswahlverfahren der Hochschulen
A	Cronbachs Alpha
Art.	Artikel
AV	Abhängige Variable
Bsp.	Beispiel
BVerfGE	Bundesverfassungsgericht
bzw.	beziehungsweise
CI90	Konfidenzintervall auf dem 90 Prozentniveau
D	Effektstärke d
DAT	Dental Admission Test
etc.	ecetera
EMS	Eignungstest für das Medizinstudium in der Schweiz
HRG	Hochschulrahmengesetz
GMA	Gesellschaft für Medizinische Ausbildung
GG	Grundgesetz
ggf.	gegebenenfalls
ges.	gesamt
GPA	Grade Point Average
H	Hypothese
HAM-Nat	Hamburger Auswahlverfahren für Medizinische Studiengänge-Naturwissenschaftsteil
HeiCuMed	Heidelberger Curriculum Medicinale
HRG	Hochschulrahmengesetz
HZB	Hochschulzulassungsberechtigung
ISK	Inventar sozialer Kompetenzen
KV	Kontrollvariable
Hyp.	Hypothese
ITP	Implicit Trait Policy
KF als Itembezeichnung	Kontaktfähigkeit
Kri als Itembezeichnung	Kritikfähigkeit
Max	Maximum
MCAT	Medical College Admission Test
MediKIT	Medizinische Kommunikations- und Interaktions- training

Min	Minimum
MMI	Multiple Mini Interview
M	Arithmetischer Mittelwert
MSA	Measure of Sample Adequacy
N	Anzahl der Studienteilnehmer
NEO-FFI	NEO Fünf Faktoren Inventar
OSCE	Objective Structured Clinical Examination
p	Irrtumswahrscheinlichkeit
POL	Problemorientierten Lernen
r	Korrelationskoeffizient nach Pearson
r_s	Korrelationskoeffizient nach Spearman
R als Itembezeichnung	Respektfähigkeit
S.	Seite
SBSR als Itembezeichnung	Selbstbeobachtung/Selbstreflexion
SD	Standardabweichung
SJT	Situational Judgement Test
SKSD als Itembezeichnung	Selbstkontrolle/Selbstdisziplin
SU als Itembezeichnung	Suche nach sozialer Unterstützung
TMS	Test für Medizinische Studiengänge
UV	Unabhängige Variable
u.a.	unter anderem
vgl.	Vergleich
vs.	versus
z.B.	zum Beispiel
η_p^2	Effektstarke Partielles Eta Quadrat
%	Prozent
&	und

Erklärung gemäß § 8 Abs. 1 Buchst. b) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe.

Erklärung gemäß § 8 Abs. 1 Buchst. c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.

Vorname, Nachname

Datum, Unterschrift

Anhang

Anhang A Definition sozialer Kompetenzen des Arbeitsmodells

Ebene 0	
Selbstkonzept:	Kennen und Werten der eigenen Kompetenzen, Einstellungen, Vorlieben und des Aussehens.
Selbstwertgefühl:	Affektive Komponente des Selbstkonzepts, die eine Wertung des eigenen Selbst enthält.
Selbstwirksamkeitserwartung	Überzeugung einer Person, kompetent handeln zu können und damit die gewünschte Wirkung zu erzielen.
Selbstbeobachtung/ Selbstreflexion:	Bereitschaft und Fähigkeit eigene Gefühle, Gedanken und Verhalten zu beobachten, eigenen Stärken, Schwächen und Grenzen zu erkennen und einzugestehen und Wirkung des eigenen Verhaltens auf andere wahrzunehmen.
Selbstkontrolle/ Selbstdisziplin:	Selbstregulierende Prozesse und Maßnahmen entgegen der Impulse und Gewohnheiten zur Erreichung eines Ziels. Bereitschaft und Fähigkeit eigene Bedürfnisse, Gedanken, Gefühle und Verhalten zu kontrollieren und in Belastungssituationen (z.B. Stress, Ärger, Frustration) aufgabenorientiert und zuverlässig zu handeln und diese ggf. situationsangemessen ausdrücken zu können.
Soziale Wahrnehmungsfähigkeit:	Fähigkeit, Situationen und Personen angemessen und richtig wahrzunehmen und relevante Signale korrekt zu interpretieren.
Ebene 1	
Kommunikationsfähigkeit:	Bereitschaft und Fähigkeit sich verbal und nonverbal verständlich auszudrücken, aktiv zuzuhören und Signale des Gesprächspartners angemessen zu interpretieren.
Durchsetzungsfähigkeit:	Bereitschaft und Fähigkeit eigene Wünsche, Forderungen und Erwartungen an andere Personen zu stellen, um ein Ziel zu erreichen.
Einfühlungsvermögen/ Empathie:	Bereitschaft und Fähigkeit sich in die Lage eines anderen hineinzuversetzen und Verständnis für dessen Perspektive oder Befindlichkeit zu entwickeln.
Hilfsbereitschaft:	Bereitschaft und Fähigkeit andere zu unterstützen
Kompromissfähigkeit:	Bereitschaft und Fähigkeit verschiedene Interessen zusammenzutragen und Vereinbarungen zu treffen, die unter den gegebenen Bedingungen die günstigste Strategie darstellen.
Kritikfähigkeit:	Bereitschaft und Fähigkeit konstruktive Kritik zu äußern und mit berechtigter und unberechtigter Kritik an der eigenen Person angemessen umzugehen.

Fortsetzung Anhang A Definition sozialer Kompetenzen des Arbeitsmodells

Ebene 1	
Persönliche und soziale Verantwortungsübernahme:	Bereitschaft und Fähigkeit verantwortungsbewusst gegenüber der eigenen Person und der Gesellschaft zu handeln.
Respektfähigkeit:	Bereitschaft und Fähigkeit Unterschiede von Personen zu akzeptieren und würdigen, sowie die Autonomie und Rechte aller Personen wertzuschätzen.
Suche nach sozialer Unterstützung:	Bereitschaft und Fähigkeit Unterstützungsbedarf zu erkennen, sowie auf angemessene und erreichbare Personen zurückgreifen.
Ebene 2	
Konfliktfähigkeit:	Bereitschaft und Fähigkeit Konflikte zu erkennen, zu analysieren, zuzulassen und zu lösen und diese auch als Chance zu verstehen.
Ebene 3	
Integrationsfähigkeit:	Bereitschaft und Fähigkeit sich in soziale Gruppen einzufügen und die eigene Rolle anzunehmen bzw. andere Personen einzugliedern, um effektiv auf ein gemeinsames Ziel hin zu arbeiten.
Kooperationsfähigkeit:	Kooperationsfähigkeit ist ein Unteraspekt der Teamfähigkeit. Es ist die Bereitschaft und Fähigkeit, mit anderen Personen effektiv ein gemeinsames Ziel zu erreichen. Es beinhaltet aufeinander einzugehen, eigene Ideen einzubringen, sich aber auch zurücknehmen zu können.
Ebene 4	
Teamfähigkeit:	Bereitschaft und Fähigkeit einer Person zur effektiven, zielgerichteten und positiv erlebbaren Zusammenarbeit mit anderen.
Führungsfähigkeit:	Bereitschaft und Fähigkeit Personen und Gruppen konstruktiv und produktiv anzuleiten, zu betreuen und das Teamklima zu pflegen.

Anhang B Anforderungsanalyse

Anhang B-1 Online-Fragebogen

Liebe Human- und Zahnmediziner,

soziale Kompetenzen werden neben fachlichen Fähigkeiten in unserer Gesellschaft als zunehmend wichtig für Studium und Beruf angesehen. Die Medizin sieht sich daher mit Fragen konfrontiert wie:

Welche sozialen Kompetenzen sind für die medizinische Ausbildung relevant?

Wie trainierbar sind soziale Kompetenzen im Studium?

Wie hoch sollten sozialen Kompetenzen bei medizinischen Studienbewerbern mindestens ausgeprägt sein?

Sollten soziale Kompetenzen bei der medizinischen Studierendenauswahl mit berücksichtigt werden?

Wie ist Ihre Meinung als Arzt zu diesen Fragen?

Zur vollständigen Beantwortung des Fragebogens benötigen Sie maximal 20 Minuten. Falls Sie eine Frage nicht beantworten möchten oder können, lassen Sie diese einfach aus. Die Beantwortung des Fragebogens erfolgt anonym.

Wenn Sie an einer Rückmeldung der Studienergebnisse interessiert sind, können Sie im Anschluss Ihre Email-Adresse hinterlassen. Diese wird unabhängig von Ihren Antworten gespeichert.

Danke für Ihr Interesse und Ihre Teilnahme!

Dipl.-Psych. Janine Kahmann
janine.kahmann@med.uni-heidelberg.de

Sind Sie als Human- oder Zahnmediziner tätig?

- Als Humanmediziner
- Als Zahnmediziner

In welchem Monat sind Sie geboren?

Diese Frage dient lediglich der zufälligen Verteilung auf verschiedene Fragebogenversionen.

- Januar bis April
- Mai bis August
- September bis Dezember

Anmerkung: Wenn Sie als Humanmediziner tätig sind, füllen Sie den Fragebogen bitte im Hinblick auf das Humanmedizinstudium aus. Wenn Sie als Zahnmediziner tätig sind, füllen Sie den Fragebogen bitte im Hinblick auf das Zahnmedizinstudium aus.

Bei Fragen zum Arztberuf stellen Sie sich bitte einen Human- bzw. Zahnmediziner mit Patienten- und Kollegenkontakt vor.

Im Folgenden werden Ihnen 12 soziale Kompetenzen vorgestellt. Beginnend mit der...

1. Ausdrucksfähigkeit/Kommunikationsfähigkeit:

Definition (Version 1-3): Bereitschaft und Fähigkeit, sich auf verbaler und nonverbaler Ebene (Gestik, Mimik, Körperhaltung) verständlich auszudrücken, aktiv zuzuhören und Signale des Gesprächspartners angemessen zu interpretieren.

Beispiel aus dem Studium (Version 2):

- sich mit Kommilitonen und Dozenten über medizinische Inhalte austauschen, Wissen in Vorträgen, Diskussionen oder schriftlichen Ausarbeitungen in verständlicher Art und Weise weitergeben und erklären können

Beispiel aus dem Beruf (Version 3):

- medizinisches Wissen in verständlicher Art und Weise während interdisziplinärer Fallbesprechungen weitergeben und erklären können

Wie wichtig schätzen Sie die Ausdrucksfähigkeit/Kommunikationsfähigkeit für das Human-/Zahnmedizinstudium bzw. für den Beruf als Human-/Zahnmediziner ein?

1 = sehr unwichtig

7 = sehr wichtig

	1	2	3	4	5	6	7	Weiß nicht
für das Studium	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
für den Beruf	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Hier folgten die weiteren 11 sozialen Kompetenzen des Arbeitsmodells...

Zur besseren Interpretierbarkeit der Daten, benötigen wir noch Angaben zu Ihrer Person.

In welchem Facharztbereich sind Sie tätig? (freie Antwort)

Wie viele Jahre arbeiten Sie bereits als Arzt?

Bitte geben Sie eine ganzzahlige Ziffer an: ____

Wie viele Jahre lehren Sie bereits Studierende?

Bitte geben Sie eine ganzzahlige Ziffer an: ____

Ihr Geschlecht:

männlich

weiblich

Gibt es noch etwas, was Sie ergänzend mitteilen möchten? Hier ist Platz für Anregungen, Kritik und Meinungen: (freie Antwort)

Falls Sie an einer Rückmeldung zu den Fragebogenergebnissen interessiert sind, können Sie hier Ihre Email-Adresse hinterlassen. Die Ergebnisse werden Ihnen dann nach Abschluss der Erhebung zugesandt.

Anhang B-2: Tätigkeitsbereich der Teilnehmer

Facharzttrichtung Humanmedizin	Version 1		Version 2		Version 3	
	N	%	N	%	N	%
Allgemeinmedizin	1	7.7	2	9.1	3	16.7
Anästhesie	0	0.0	2	9.1	1	5.6
Gynäkologie und Urologie	1	7.7	2	9.1	2	11.1
Innere Medizin	2	15.4	4	18.2	5	27.8
Psychiatrie und Psychotherapie	2	15.4	4	18.2	2	11.1
Neurologie	0	0.0	1	4.5	1	5.6
sonstige mit Patientenkontakt	3	23.1	5	22.7	2	11.1
sonstige ohne Patientenkontakt	1	7.7	2	9.1	2	11.1
gültig	10	77.0	22	100.0	18	100.0
Facharzttrichtung Zahnmedizin						
allgemeine Zahnmedizin	8	53.3	13	68.4	9	40.9
sonstige mit Patientenkontakt	2	13.3	0	0.0	3	13.6
gültig	10	66.6	13	68.4	12	54.5

Anmerkung. Die Prozentwerte beziehen sich auf die Teilnehmer, die Ihr Tätigkeitsfeld angegeben haben.

Anhang C Teilnehmer und Materialien der Workshops

Anhang C-1 Teilnehmer des ersten Workshops

Person	Geschlecht	Berufsgruppe	Fachgebiet	Position
1	weiblich	Humanmedizinerin	Viszeralchirurgie	Oberärztin
2	männlich	Humanmediziner	Kinder- und Jugend- psychiatrie	Oberarzt
3	weiblich	Zahnmedizinerin	Zahnerhaltungskunde	Zahnärztin
4	männlich	Zahnmediziner	Zahnerhaltungskunde	Zahnarzt
5	männlich	Zahnmediziner	Prothetik	Zahnarzt
6	männlich	Psychologe	Pädagogische Psychologie	Wissenschaft- licher Mitarbeiter
7	weiblich	Erziehungs- wissenschaftlerin	Allgemeinmedizin und Versorgungsforschung	Wissenschaft- licher Mitarbeiter
8	weiblich	Biochemikerin	Biochemie	Wissenschaft- licher Mitarbeiter
9	männlich	Physiologe	Physiologie und Pathophysiologie	Wissenschaftli- cher Mitarbeiter
10	männlich	Krankenpfleger	Psychiatrie	Krankenpfleger
11	männlich	Krankenpfleger	Viszeralchirurgie	Krankenpfleger

Anhang C-2 Teilnehmer des zweiten Workshops

Person	Geschlecht	Berufsgruppe	Fachgebiet	Position
1	weiblich	Humanmedizinerin	Viszeralchirurgie	Oberärztin
2	weiblich	Humanmedizinerin	Pädiatrie	Assistenzärztin
3	männlich	Humanmediziner	Kinder- und Jugendpsychiatrie	Oberarzt
4	weiblich	Humanmedizinerin	Viszeralchirurgie	Assistenzärztin
5	männlich	Humanmediziner	Psychosomatik	Facharzt
6	weiblich	Zahnmedizinerin	Allgemeine Zahnmedizin	Zahnärztin
7	männlich	Zahnmediziner	Kieferorthopädie	Zahnarzt
8	männlich	Zahnmediziner	Prothetik	Zahnarzt
9	männlich	Zahnmediziner	Zahnerhaltungskunde	Zahnarzt
10	männlich	Psychologe	Pädagogische Psychologie	Wissenschaftlicher Mitarbeiter
11	weiblich	Erziehungswissenschaftlerin	Allgemeinmedizin und Versorgungsforschung	wissenschaftlicher Mitarbeiter
12	weiblich	Biochemikerin	Biochemie	wissenschaftlicher Mitarbeiter
13	männlich	Physiologe	Physiologie und Pathophysiologie	wissenschaftlicher Mitarbeiter
14	männlich	Krankenpfleger	Psychiatrie	Krankenpfleger
15	männlich	Krankenpfleger	Viszeralchirurgie	Krankenpfleger

Anhang C-3 Vorlage zur Entwicklung der Antwortoptionen

Situationsbeschreibung:

--

Handlungsoptionen:

Ausformulierte Handlungsoption (Einzelarbeit)	Qualitätsprüfung (Arbeit im Tandem)	
	Ausprägung der sozialen Kompetenz: ++ sehr hoch + hoch 0 neutral - gering -- sehr gering	Änderungsvorschläge
		...

Anhang D Itempool SJT

* In Endversion des SJTs mit 20 Items enthalten

Items Selbstbeobachtung/Selbstreflexion (SBSR)

SBSR1: In einer mündlichen Prüfung merken Sie, dass Sie trotz guter Vorbereitung zwei kleine Wissenslücken haben. Sie geben Ihr Bestes und versuchen, möglichst kompetent aufzutreten. Nach der Prüfung werden Sie kurz gebeten den Raum zu verlassen, damit sich Ihr Prüfer und der Beisitzer beraten können. Derweil haben Sie Zeit, die Prüfung noch einmal im Kopf Revue passieren zu lassen und Ihre Prüfungsleistung selbst einzuschätzen.

- a) Ich schätze meine Prüfungsleistung aufgrund der Wissenslücken als durchschnittlich ein.
- b) Ich schätze meine Prüfungsleistung trotz Wissenslücken als sehr gut ein.
- c) Ich bin mir unsicher, wie ich meine Prüfungsleistung aufgrund der Wissenslücken einschätzen soll.
- d) Ich schätze meine Prüfungsleistung als gut ein und mache mir meine Wissenslücken klar.
- e) Ich schätze meine Prüfungsleistung aufgrund der Wissenslücken als sehr schlecht ein.
- f) Ich schätze meine Prüfungsleistung aufgrund der Wissenslücken als schlecht ein.

SBSR2: Sie absolvieren Ihr erstes Praktikum im klinischen Studienabschnitt in einer Ambulanz und erhalten von einer Ärztin den Auftrag bei mehreren Patienten Blut abzunehmen. Sie haben zwar schon im einen vorklinischen Praktikum die Blutabnahme an einem Modell geübt, fühlen sich aber unsicher, diese an realen Patienten durchzuführen. Alle anderen anwesenden Praktikanten/Innen beherrschen die Blutabnahme bereits. Sie haben Sorge, dass die Ärztin den Auftrag zukünftig immer an einen anderen Praktikanten weitergibt, wenn Sie die Unsicherheit zeigen.

- a) Ich führe die Blutentnahme durch, weil die Prozedur für die Patienten ungefährlich ist.
- b) Ich bitte einen anderen Praktikanten, mir die Blutentnahme an einem Patienten zu zeigen. Bei den folgenden Patienten führe ich dann selbstständig die Blutentnahmen durch.
- c) Ich bitte die Ärztin, mir die Blutentnahme an einem Patienten zu zeigen. Bei den folgenden Patienten führe ich dann selbstständig die Blutentnahmen durch.
- d) Ich bitte einen anderen Praktikanten die Blutentnahme zu übernehmen, da ich mich noch unsicher fühle.
- e) Ich bitte die Ärztin, die Blutentnahme zu übernehmen, da ich mich noch unsicher fühle.

SBSR3: Ihr erster Patientenkontakt im vorklinischen Studienabschnitt findet zur Übung an einem Schauspielpatienten statt. Bei diesem sollen Sie Krankheiten in der Familie erheben. Dabei werden Sie von sieben Mitstudenten und einem Dozenten beobachtet. Das Gespräch gelingt Ihnen gut. Sie fragen die meisten relevanten Inhalte ab. Sie erkennen auch, dass Sie ein paar Details nicht erfragt haben. Im Anschluss an das simulierte Patientengespräch sollen zunächst Sie selbst das Gespräch bewerten bevor Sie Feedback vom Schauspielpatienten, Ihren Mitstudenten und dem Dozenten erhalten.

- a) Ich beobachte die Gesichtsausdrücke der Anwesenden und gebe je nach Ausdruck eine positive oder negative Bewertung meiner eigenen Person.
- b) Ich sage, dass ich alle relevanten Fragen gestellt habe. Sollten mich die anderen auf die Detailfragen ansprechen, würde ich mich einsichtig zeigen.
- c) Ich sage, dass ich alle relevanten Fragen bis auf ein paar Details gestellt habe.
- d) Ich sage, dass mir das Gespräch gut gelungen ist.
- e) Ich erkläre selbstkritisch, dass ich einige Details vernachlässigt habe.

SBSR4: Sie unterhalten sich mit einer neuen Mitstudentin über Ihre Studiensituation und haben den Eindruck, dass sich während des Gesprächs die Stimmung zwischen Ihnen beiden verschlechtert. Plötzlich verabschiedet sich die Mitstudentin knapp und beendet das Gespräch.

- a) Ich mache mir keine Gedanken über das Gespräch, da sie vermutlich einfach nur schnell los musste.
- b) Ob ich mir Gedanken über das Gespräch mache, hängt davon ab wie sympathisch mir die Mitstudentin ist.
- c) Ich spreche sie beim nächsten Treffen konkret darauf an, um den Grund für ihr Verhalten zu verstehen.
- d) Ich mache mir Gedanken über das Gespräch, ob ich irgendetwas Beleidigendes oder Unangemessenes gesagt habe.
- e) Ich vermute, dass sie nicht über dieses Thema sprechen möchte und vermeide es in Zukunft.

SBSR5*: Während eines vierwöchigen hausärztlichen Praktikums haben Sie nach einer Woche den Eindruck, dass der Hausarzt Ihre Arbeitsweise gut findet. Die Arzthelferinnen scheinen Ihrer Arbeitsweise aber kritisch gegenüber zu stehen. Sie haben den Eindruck nicht sehr herzlich aufgenommen worden zu sein.

- a) Ich mache mir nichts weiter aus dem Verhalten der Arzthelferinnen. Wichtiger ist, dass die Hausärztin mit mir zufrieden ist.
- b) Ich verhalte mich den Arzthelferinnen gegenüber genauso wie sie es mir gegenüber tun.
- c) Ich suche ein klärendes Gespräch mit den Arzthelferinnen.

- d) Ich schildere der Hausärztin die Situation und bitte sie um ihre Einschätzung.
- e) Ich suche nicht das Gespräch. Vermutlich geht die wenig herzliche Aufnahme der Arzthelferinnen auf andere Ursachen zurück, die nichts mit meiner Arbeitsweise zu tun haben.

SBSR6: Sie stehen kurz vor der Biochemieprüfung im vorklinischen Studienabschnitt und müssen nur noch diese Prüfung bestehen, um zum ersten Staatsexamen (sehr umfassende Prüfung zum Abschluss des vorklinischen Studienabschnitts, die jedes halbe Jahr absolviert werden kann) zugelassen zu werden. Nach der Biochemieprüfung haben Sie 4-6 Wochen Zeit, um sich auf das erste Staatsexamen vorzubereiten. Trotz aller Anstrengung merken Sie, dass Sie die Stoffmenge nicht bewältigen können. Sie spüren eine Woche vor der Biochemieprüfung einen zunehmenden Erschöpfungszustand. Sie essen aus Appetitmangel kaum noch etwas und wachen morgens immer wieder unerholt auf.

- a) Ich versuche die Prüfungsbelastung zu reduzieren, indem ich z.B. einen befreundeten Mitstudenten bitte, dass wir uns gemeinsam auf die Prüfungen vorbereiten.
- b) Ich hole mir fachkundigen Rat (z.B. beim Arzt oder einer psychologischen Beratungsstelle), wie ich mit meiner Erschöpfung umgehen sollte.
- c) Ich erkenne die kritische Situation für mich selbst und verschiebe beide Prüfungen.
- d) Ich versuche trotz Anstrengung für beide Prüfungen zu lernen, da ich die Prüfungen nicht verschieben möchte.
- e) Ich nehme homöopathische Medikamente, um besser schlafen zu können, sodass ich besser lernen kann.
- f) Ich versuche die Woche noch durchzustehen, da ich die Biochemieprüfung

SBSR7*: Bei Ihrem ersten Praktikum auf einer Station des Uniklinikums haben Sie mehreren Patientin den Blutdruck gemessen. Sie sind sich nicht sicher, ob Sie die Werte richtig erhoben haben. Abends telefonieren Sie mit einer befreundeten Mitstudentin, die gerade in einem anderen Krankenhaus im Praktikum ist und bereits mehrere Praktika absolviert hat. Sie fragt Sie, wie Sie mit Blutdruck messen zurechtgekommen sind.

- a) Ich erzähle ihr von meiner Unsicherheit und lasse mir von ihr erklären, wie sie beim Blutdruck messen vorgegangen ist.
- b) Ich schlage ihr vor über was anderes zu erzählen, um vom Tag abzuschalten.
- c) Ich erzähle ihr, dass ich den Blutdruck falsch gemessen habe.
- d) Ich erzähle ihr, dass ich mehreren Patienten den Blutdruck gemessen habe und dabei kein Problem hatte.
- e) Ich erzähle ihr, dass ich noch lernen muss, wie der Blutdruck genau gemessen wird.

Items Selbstkontrolle/Selbstdisziplin (SKSD)

SKSD1: Sie sind Human-/Zahnmedizinstudierende(r) im vierten Semester. In vier Wochen findet die letzte Klausur statt, die Sie unbedingt bestehen müssen, um im Herbst am ersten Staatsexamen teilnehmen zu können. Eigentlich benötigen Sie diese Zeit für die Prüfungsvorbereitung. Dazwischen liegen die zweiwöchigen Pfingstferien. Ihre Freunde haben für die gesamten Pfingstferien einen sehr attraktiven Urlaub geplant und fragen, ob Sie mitkommen.

- a) Ich fahre eine Woche mit, ohne mir etwas zu Lernen mitzunehmen. Nach dem Urlaub arbeite ich umso mehr, um die verlorene Lernzeit wieder aufzuholen.
- b) Ich fahre nicht mit, sondern bleibe zu Hause, um zu Lernen. Ab und zu gönne ich mir dafür eine Belohnung vor Ort.
- c) Ich fahre die gesamten zwei Wochen mit, ohne mir etwas zu Lernen mitzunehmen. Nach dem Urlaub arbeite ich umso mehr, um die verlorene Lernzeit wieder aufzuholen.
- d) Ich fahre nicht mit, sondern bleibe zu Hause, um jede Minute für die Klausurvorbereitung zu nutzen.
- e) Ich fahre die gesamten zwei Wochen mit, nehme mir aber etwas zu Lernen mit.
- f) Ich fahre eine Woche mit und nehme mir etwas zu Lernen mit.

SKSD2: Sie möchten sich um ein Stipendium für ein Auslandspraktikum bewerben. Dazu benötigen Sie ein Gutachten. Ein(e) Dozent/In hat eingewilligt und versprochen, dass Sie das Gutachten zu einem bestimmten Termin erhalten. Der Termin ist bereits seit letzter Woche verstrichen. Sie warten immer noch auf das Gutachten und sind zunehmend verärgert. Die Frist zur Einreichung der Unterlagen für das Stipendium endet in 3 Tagen.

- a) Um mich nicht weiter zu ärgern, frage ich eine(n) andere(n) Dozenten/Dozentin, ob er/sie mir ein Gutachten schreibt.
- b) Ich warte einfach noch 2 Tage. Sich Ärgern macht die Situation nicht besser.
- c) Ich lasse mir meinen Ärger nicht anmerken und schreibe ihm/ihr eine höfliche Email, um ihn/sie an das Gutachten zu erinnern.
- d) Ich lasse mir meinen Ärger nicht anmerken und rufe ihn/sie an, um höflich zu erfragen wie weit er/sie mit dem Gutachten vorangeschritten ist.
- e) Ich gehe verärgert zu dem Dozenten/der Dozentin und frage ihn/sie, warum ich das versprochene Gutachten noch nicht erhalten habe.

SKSD3: Sie sind schlecht auf eine Klausur in der Vorklinik vorbereitet, haben sich aber bereits für diese Klausur angemeldet. Wenn Sie nicht teilnehmen, gilt der Leistungsnachweis als „nicht bestanden“. Insgesamt können Sie ein Mal durchfallen. Am Abend vor der Klausur werden Sie zunehmend nervöser, da Sie merken, dass Sie die Klausur mit Ihrem derzeitigen Wissensstand wahrscheinlich nicht bestehen werden.

- a) Ich lerne die gesamte Nacht durch und schreibe die Klausur mit.
- b) Ich gestehe mir ein, dass ich die Klausur nicht bestehen kann und schreibe sie nicht mit.
- c) Ich lasse mir ein ärztliches Attest ausstellen, sodass ich mir beide Chancen zum Bestehen der Klausur sichere.
- d) Ich höre mit dem Lernen auf, um mich nicht noch nervöser zu machen und schreibe die Klausur mit.

SKSD4*: Sie sind Praktikant/In auf einer Station und sollen bei einer Patientin Blut abnehmen, da Sie es schon gut beherrschen. Sie stellen sich bei der Patientin als „Student/In“ vor, worauf die Patientin mit heftiger Ablehnung reagiert und sagt: „Dann holen Sie bitte einen Arzt. Ich lasse keinen Studenten an meinen Körper“.

- a) Ich frage sie, ob sie schon schlechte Erfahrungen gemacht hat, um die Situation zu verstehen.
- b) Ich sage höflich, dass ich ihr Verhalten übertrieben finde.
- c) Ich sage, dass sie sich keine Sorgen machen muss, da ich das Blutabnehmen beherrsche.
- d) Ich akzeptiere ihre Meinung und gebe der Ärztin Bescheid, dass die Patientin von einem Arzt behandelt werden möchte.

SKSD5: Sie absolvieren ein Praktikum in einer chirurgischen Abteilung und haben nach einigen Tagen das Gefühl, immer wieder für Routinetätigkeiten wie die Blutentnahme eingeteilt zu werden. Die Ärzte nehmen sich keine Zeit, Ihnen etwas beizubringen. Auch Patienten, die Sie aufgenommen und untersucht haben, besprechen sie nicht gemeinsam. In der zweiten Woche werden Sie von der Visite (Begutachtung des Patienten am Krankenbett durch einen oder mehrere Ärzte)weggeschickt, um sämtliche Blutentnahmen der Station durchzuführen.

- a) Ich bitte einen anderen Praktikanten, die Blutentnahmen durchzuführen, damit ich die Visite begleiten kann.
- b) Ich sage, dass ich die Visite gern begleiten möchte, um etwas Neues zu lernen und anschließend die Blutentnahme mache.
- c) Ich erledige die Blutentnahmen und bitte die Ärzte nach der Visite um ein Gespräch zur Ausbildungssituation.
- d) Ich versuche, auf eine andere Station zu wechseln.
- e) Ich erledige die Blutentnahmen ohne Diskussion.

SKSD6: Sie möchten Humanmedizin studieren und müssen dafür ein Pflegepraktikum in einem Krankenhaus absolvieren. Sie sind auf einer unfallchirurgischen Station mit einem/einer Studierenden eingeteilt, der/die schon seit einem Monat auf der Station im Praktikum ist. Sie helfen ihm/ihr beim Aufsägen eines Gipsverbandes, wobei diese(r) den Patienten mit der Säge leicht

an der Haut verletzt. Er/sie beschuldigt Sie, den Fuß falsch gehalten zu haben, obwohl Sie ihn genauso gehalten haben, wie er/sie es vorher beschrieben hat. Das ärgert Sie.

- a) Ich sage nichts dazu.
- b) Ich bitte eine(n) Arzt/In zur Hilfe, um die Situation zu klären.
- c) Ich entschuldige mich bei ihm/ihr, dass ich den Fuß vielleicht doch nicht richtig gehalten habe.
- d) Ich sage ihm/ihr im einen 4-Augen-Gespräch, dass er/sie so nicht mit mir umgehen kann.
- e) Ich sage ihm/ihr im einen 4-Augen-Gespräch, dass er mich gern korrigieren kann, aber bitte in freundlicher Form.
- f) Ich gehe ihm/ihr in den nächsten Tagen aus dem Weg.

SKSD7: Die Sekretärin aus dem Studentensekretariat schickt Ihnen per Email aus Versehen ein falsches Dokument zu, in denen die vorläufigen Klausurnoten Ihres Semesters stehen. Die Sekretärin bittet Sie daraufhin, die Noten auf keinen Fall weiterzugeben, da sie noch nicht offiziell bestätigt wurden.

- a) Ich behalte die Klausurnoten für mich.
- b) Ich lösche das Dokument.
- c) Ich teile nur meine eigene Klausurnote meinen engsten Mitstudenten mit und bitte sie, es für sich zu behalten.
- d) Ich leite alle Klausurnoten an meine Mitstudenten weiter.
- e) Ich teile meinen engsten Mitstudenten meine und ihre Klausurnoten mit und bitte sie, es für sich zu behalten.

SKSD8: Sie sind Teil eines studentischen Teams, das einen Patientenfall aufbereitet und in einer Unterrichtseinheit vorstellen soll. Für den Vortrag erhalten Sie als Team eine Gesamtnote. Ein Mitglied des Teams, mit dem Sie bisher noch nicht zusammengearbeitet haben, beteiligt sich nicht und hat bis zum Tag vor der Vorstellung seinen Anteil der Arbeit nicht erfüllt. Da er nicht zu erreichen ist, sehen Sie ihn erst direkt zur Patientenvorstellung wieder.

- a) Ich bereite mit den anderen aktiven Teammitgliedern den fehlenden Arbeitsanteil vor und wir sprechen den inaktiven Mitstudenten am Vorstellungstag darauf an.
- b) Ich erledige den Arbeitsteil des inaktiven Mitstudenten und spreche den inaktiven Mitstudenten am Vorstellungstag darauf an.
- c) Ich verständige mich mit den aktiven Teammitgliedern, dass wir den Vortrag so belassen und abwarten wie der inaktive Mitstudent am Vorstellungstag handelt.
- d) Ich verständige mich mit den aktiven Teammitgliedern, dass wir den Vortrag so belassen und die Dozentin über den inaktiven Mitstudenten informieren.

- e) Ich verständige mich mit den aktiven Teammitgliedern, dass wir den Vortrag so belassen und den inaktiven Mitstudenten am Vorstellungstag darauf ansprechen.

SKSD9: Sie fühlen sich im mündlichen ersten Staatsexamen ungerecht bewertet. Sie sind überzeugt, dass Sie genau so viel wussten wie Ihre Mitstudentin, die mit Ihnen gemeinsam die Prüfung absolviert hat. Jedoch hat diese eine Note besser erhalten als Sie.

- a) Ich bitte um eine Wiederholung der Prüfung, um eine objektivere Bewertung zu bekommen.
- b) Ich spreche die Prüfer darauf an und bitte sie, mir die Bewertung zu erklären.
- c) Ich zeige, dass ich mich ungerecht bewertet fühle, unternehme aber nichts weiter.
- d) Ich frage meine Mitstudentin, ob sie die Sache genauso sieht und bitte gegebenenfalls um ein Gespräch mit den Prüfern.
- e) Ich möchte nicht mit den Prüfern diskutieren und akzeptiere die Note

SKSD10*: Sie befinden sich im letzten Jahr Ihres Humanmedizinstudiums im sogenannten Praktischen Jahr (PJ), in dem die Ausbildung am Patienten im Mittelpunkt steht und sie auf verschiedenen Stationen arbeiten. Derzeit sind Sie auf einer chirurgischen Station. Obwohl Sie Interesse und Arbeitsbereitschaft signalisieren, erhalten Sie ständig Routineaufgaben, durch die Sie nichts Neues dazu lernen. Sie fühlen sich schlecht betreut, weil die Ärzte kaum Zeit für Sie haben und sind zunehmend frustrierter. Sie sind noch vier Wochen auf dieser Station eingeteilt, bevor Sie auf eine andere Station wechseln.

- a) Ich frage die Ärzte, ob ich sie bei ihrer Arbeit begleiten kann, um etwas dazuzulernen.
- b) Ich melde dem Stationsarzt nach meinem Einsatz zurück, dass ich nicht viel gelernt habe.
- c) Ich halte durch und freue mich auf den Einsatz auf der nächsten Station.
- d) Ich frage die Ärzte, was ich außer den Routineaufgaben noch übernehmen kann.
- e) Ich rede mit dem Stationsarzt über meine Ausbildungssituation und bitte um Lösungsmöglichkeiten.
- f) Ich sage den Ärzten, dass ich keine Routineaufgaben übernehmen möchte, da ich dadurch nichts dazu lerne.

SKSD11*: Sie befinden sich am Ende Ihres Humanmedizinstudiums im Praktischen Jahr, in dem die Ausbildung am Patienten im Mittelpunkt steht und sie auf verschiedenen Stationen arbeiten. Auf Ihrer derzeitigen Station hat die Stationsärztin den Angehörigen eines an Krebs erkrankten Patienten die schlechte Nachricht überbracht, dass eine OP wenig aussichtsreich ist. Nun fragen die Angehörigen verzweifelt bei Ihnen nach, ob das wahr sein kann? Sie fühlen sich mit der Situation fachlich und menschlich überfordert.

- a) Ich sage dem Angehörigen, dass ich als Student/In die Situation noch nicht umfassend beurteilen kann und die Stationsärztin informiere. Bis sie kommt, höre ich mir die Sorgen der Angehörigen aufmerksam an.
- b) Ich sage dem Angehörigen, dass ich als Student/In die Situation noch nicht umfassend beurteilen kann und höre mir seine Sorgen aufmerksam an. Seine Fragen versuche ich meines Wissenstands entsprechend zu beantworten.
- c) Ich sage dem Angehörigen, dass ich als Student/In die Situation noch nicht umfassend beurteilen kann und versuche den Angehörigen zu beruhigen, indem ich ihm Hoffnung mache, dass die Situation nicht ausweglos ist.
- d) Ich sage dem Angehörigen, dass ich als Student/In die Situation noch nicht umfassend beurteilen kann und der Stationsärztin Bescheid gebe.

Items Kritikfähigkeit (Kri)

Kri1*: Sie möchten Humanmedizin studieren und müssen dafür ein Pflegepraktikum in einem Krankenhaus absolvieren. Sie sind auf einer unfallchirurgischen Station mit einer Studentin eingeteilt, die schon seit einem Monat auf der Station im Praktikum ist. Sie helfen ihr beim Aufsagen eines Gipsverbandes, wobei diese den Patienten mit der Säge leicht an der Haut verletzt. Sie beschuldigt Sie hinterher, den Fuß falsch gehalten zu haben, obwohl Sie ihn genauso gehalten haben, wie sie es vorher beschrieben hat. Das ärgert Sie. Ein Pfleger hat zufällig die Situation beobachtet.

- a) Ich sage ihm unter vier Augen, dass er mich gern korrigieren kann, aber bitte in freundlicher Form.
- b) Ich bitte den Studenten, mir den Fehler konkret zu zeigen, da ich ihn nicht erkennen kann.
- c) Ich sage nichts dazu und meide zukünftig die Zusammenarbeit mit ihm.
- d) Ich entschuldige mich bei ihm, dass ich den Fuß vielleicht doch nicht richtig gehalten habe.
- e) Ich bitte die Pflegerin in Anwesenheit des Studenten um ihre Meinung.
- f) Ich sage nichts dazu und arbeite normal weiter.

Kri2*: Im Rahmen eines Seminars hat Ihre Sitznachbarin, die sie bisher nicht gut kennen, ein offensichtlich inhaltlich schlechtes Referat gehalten. Ihre Sitznachbarin scheint dennoch von der Qualität ihres Referats überzeugt und fragt Sie nach dem Seminar erwartungsvoll nach Ihrer Einschätzung.

- a) Ich sage ihr, was ich an ihrem Vortrag gut fand und was sie noch verbessern könnte.
- b) Ich sage ihr, dass niemand Ihren Vortrag verstanden hat und ihn alle schlecht fanden.
- c) Ich sage ihr, dass ich ihren Vortrag gut fand, da ich sie nicht verletzen möchte.
- d) Ich sage ihr, was ich an ihren Vortrag nicht gut fand.

- e) Ich sage ihr, dass ich es nicht richtig einschätzen kann und bitte sie, für eine aussagekräftige Rückmeldung die Dozentin zu fragen.
- f) Ich versuche einen weiteren Mitstudenten einzubeziehen, damit er eine Rückmeldung gibt.

Kri3: Nach Ihrer ersten zahnmedizinischen Patientenbehandlung während eines Behandlungskurses teilt Ihnen Ihre 53-jährige energisch auftretende Patientin mit, dass sie sich nicht mehr von Ihnen behandeln lassen möchte, weil Sie zu unsicher auf sie wirkten.

- a) Ich versuche gemeinsam mit der Patientin und meinem Kursbetreuer eine Lösung zu finden.
- b) Ich versuche, die Patientin von meinen fachlichen Fähigkeiten zu überzeugen.
- c) Ich bitte meinen Kursleiter, mir einen anderen Patienten zuzuordnen.
- d) Ich akzeptiere ihren Wunsch und behandle sie nicht weiter.
- e) Ich erkläre der Patientin, dass sie in einem Lehrkrankenhaus damit rechnen muss, dass einfache medizinische Maßnahmen durch Studierende durchgeführt werden.
- f) Ich erkläre der Patientin, dass sie meine erste Patientin ist und meine Unsicherheit entschuldigen möchte.

Kri4*: Sie halten ein Referat, auf das Sie sich gut vorbereitet haben und das in Ihre Semesterendnote einfließt. Sie haben den Eindruck, das Referat gut gemeistert zu haben, obwohl in der anschließenden Diskussion leichte Unsicherheiten auftraten. Sie erhalten für das Referat 20 von 25 Punkten mit konkreter Begründung des Dozenten. Sie haben dennoch das Gefühl nicht schlechter gewesen zu sein als eine Mitstudentin, die mit 22 Punkten bewertet wurde.

- a) Ich versuche die Bewertung zu akzeptieren, obwohl ich anderer Meinung bin.
- b) Ich frage die Mitstudentin in Gegenwart des Dozenten wie gerechtfertigt sie die Bewertung findet.
- c) Ich sage dem Dozenten sachlich, dass ich die Bewertung nicht gerecht finde.
- d) Ich bitte dem Dozenten die Bewertung im Vergleich zur Mitstudentin zu erklären, da ich sie nicht nachvollziehen kann.
- e) Ich beschwere mich bei dem Dozenten über die ungerechte Bewertung.

Kri5: Sie sind Teil eines studentischen Teams, das einen Patientenfall aufbereiten und in einem Seminar vorstellen soll. Sie hatten mit Ihrem Team vereinbart, dass Sie alle Arbeitsanteile zu einem Gesamtvortrag zusammenstellen. Ein Teammitglied beteiligt sich nicht und hat bis zum Vorabend des Präsentationstags seinen/ihren Anteil der Arbeit nicht erfüllt, obwohl Sie ihn/sie per Email noch einmal daran erinnert haben.

- a) Ich rufe am Vorabend das betreffende Teammitglied an und bitte ihn/sie mir zu erklären, warum er/sie sich nicht beteiligt.
- b) Ich arbeite seinen/ihren Anteil am Vorabend aus. Am Tag der Vorstellung bitte ich das betreffende Teammitglied, dem Dozenten zu sagen, dass er/sie sich nicht beteiligt hat.
- c) Ich berate mich am Vorabend mit den anderen Teammitgliedern per Telefon, um abzustimmen, wie wir weiter vorgehen.
- d) Ich arbeite seinen/ihren Anteil am Vorabend aus. Während der Präsentation nenne ich dann den Namen der beteiligten und unbeteiligten Gruppenmitglieder vor allen Studierenden und vor dem/der Lehrenden.
- e) Ich rufe am Vorabend das betreffende Teammitglied an und bitte ihn/sie erneut darum für morgen seinen/ihren Anteil noch auszuarbeiten.
- f) Ich arbeite seinen/ihren Anteil am Vorabend aus. Am Tag der Vorstellung kritisiere ich das betreffende Teammitglied nicht, um mich nicht noch mehr unnötig zu ärgern.
- g) Ich arbeite vorsichtshalber seinen/ihren Anteil am Vorabend aus. Morgens am Tag der Vorstellung kontaktiere ich den/die Lehrende(n) und frage ihn, wie wir vorgehen sollen.
- h) Ich arbeite seinen/ihren Anteil am Vorabend nicht aus, sondern warte ab, wie er/sie seinen/ihren Anteil während der Vorstellung meistert. Danach kann ich ihn/sie auch noch kritisieren.

Kri6: Sie und eine Mitstudentin, die Sie noch nicht gut kennen, sollen gemeinsam ein Referat halten, wofür Sie eine gemeinsame Note erhalten. In der Vorbereitung auf das Referat sprechen sie genau ab, für welche Inhalte Sie oder die Referatspartnerin zuständig ist. Während des Referats nimmt die Mitstudentin im ersten Teil viele Inhalte vorweg, für die Sie eigentlich zuständig sind.

- a) Ich ergänze meine Referatspartnerin geschickt während des Vortrags, wenn sie Inhalte vorwegnimmt, indem ich den Zuhörern sage, dass ich Ihnen gleich darüber mehr berichte.
- b) Ich halte mein Referatteil so gut es geht. Hinterher sage ich meiner Referatspartnerin, dass sie mich etwas durcheinander gebracht hat, weil sie einige Inhalte vorweg genommen hat.
- c) Ich sage während des Referats den Zuhörern, dass es vorkommen kann, dass ich etwas durcheinanderkomme, weil meine Referatspartnerin einige Inhalte vorweg genommen hat.
- d) Ich lasse mir während des Referats nichts anmerken. Hinterher sage ich dem Dozenten, dass mich meine Referatspartnerin etwas durcheinander gebracht hat, da sie einige Inhalte vorweg genommen hat.
- e) Ich lasse mir während des Referats nichts anmerken, vermeide aber in Zukunft die Zusammenarbeit mit der Mitstudentin.

Kri7*: Zum Bestehen des Biochemiepraktikums müssen Sie mit zwei Mitstudentinnen ein Protokoll ausarbeiten, für das sie eine gemeinsame Note erhalten. Sie besprechen untereinander, wer für welche Abschnitte des Protokolls zuständig ist. Sie haben sich bereit erklärt, die einzelnen Abschnitte zum Gesamtprotokoll zusammenzufügen. Sie bekommen wie besprochen die ausgearbeiteten Abschnitte von beiden Mitstudenten per Email zugeschickt. Während Sie die Abschnitte in das Gesamtprotokoll einarbeiten, merken Sie, dass der Beitrag der Mitstudentinnen von schlechter Qualität ist. Der Abgabetermin des Protokolls ist in einer Woche.

- a) Ich lasse ihre Beiträge so wie sie sind und akzeptiere die schlechtere Gesamtnote für alle.
- b) Ich sende ihnen ihre Beiträge mit Verbesserungsvorschlägen zurück und bitte sie, wenn sie diese für sinnvoll halten einzuarbeiten.
- c) Ich verbessere ihre Beiträge, sage aber nichts weiter dazu.
- d) Ich verbessere ihre Beiträge und melde ihnen zurück, was ich an ihren Anteilen verändert habe.

Kri8: Bei Ihrem ersten Praktikum auf einer Station des Uniklinikums haben Sie einem Patienten/einer Patientin Blut abgenommen. Obwohl Sie sich sicher sind, Ihre Arbeit gut durchgeführt zu haben, beklagt sich der Patient/die Patientin bei Ihnen, von einem Studierenden behandelt zu werden, der/die doch offensichtlich noch ungeübt und noch kein richtiger Arzt ist.

- a) Ich ärgere mich, aber spreche den Patienten/die Patientin nicht darauf an.
- b) Ich frage den Patienten/die Patientin, ob sie sich durch mich nicht gut behandelt gefühlt hat.
- c) Ich bitte einem Arzt/eine Ärztin dem Patienten/der Patientin zu versichern, dass ich Blutabnehmen kann.
- d) Ich erkläre dem Patienten/der Patientin, dass sie an Universitätskliniken bei einfachen Verfahren mit der Behandlung durch Studierende rechnen müsste.
- e) Ich erwidere darauf nichts und tue so als ob ich es nicht gehört habe.

Kri9: Im Rahmen eines Seminars halten Sie den ersten Vortrag zu einem schwierigen Thema. Nach dem Vortrag werden Sie von der Dozentin vor allen Studierenden in Ihren Augen unangemessen deutlich kritisiert, da Sie ein offensichtlich inhaltlich schlechtes Referat gehalten haben.

- a) Ich sage vor der Gesamtgruppe, dass ich die Kritik als unangemessen empfinde.
- b) Ich versuche die Kritik etwas abzuschwächen, indem ich sage, dass es ein schwieriges Thema war und ich als erste Person im Seminar drankam.
- c) Ich wende mich nach dem Seminar an die Dozentin und erkläre ihr freundlich, dass ich ihre Kritik als unangemessen empfinde.
- d) Ich denke mir meinen Teil und schweige.

- e) Ich spreche mit meinem Banknachbarn über die unangemessene Kritik der Dozentin.

Kri10*: Nach jedem Semester haben Sie als Studierende(r) die Möglichkeit, anonym Lehrveranstaltungen und die dafür verantwortlichen Lehrenden schriftlich zu beurteilen. Dafür können Sie Noten und freie Anmerkungen geben. Einen Dozenten fanden Sie dieses Semester dermaßen schlecht, dass Sie es kundgeben wollen.

- a) Ich gebe dem Dozenten eine schlechte Note und mache keine weiteren Anmerkungen.
- b) Ich nehme nicht an der schriftlichen Bewertung teil, sondern rate meinen Mitstudenten, nicht an Lehrveranstaltungen des Dozenten teilzunehmen.
- c) Ich gebe dem Dozenten eine schlechte Note und beschreibe sachlich, was ich nicht gut fand.
- d) Ich gebe dem Dozenten eine schlechte Note und mache in den freien Anmerkungen meinem Ärger Luft.
- e) Ich gebe dem Dozenten eine mittelmäßige Note, da ich ihn trotz der schlechten Veranstaltungen nicht verletzen möchte.

Kri11: Als Humanmedizinstudierende(r) machen Sie Ihr Pflegepraktikum auf einer chirurgischen Station. Bei einer gemeinsamen Pause mit einem „alten Pflege-Hasen“ spricht dieser in diskriminierender Weise über Patienten mit Migrationshintergrund, da diese immer Extrawünsche bei der Pflege haben.

- a) Ich frage den Pfleger, welche konkreten Erfahrungen er mit den Patienten gemacht hat.
- b) Ich signalisiere meine gegenteilige Ansicht durch Schweigen.
- c) Ich sage, dass ich seine Meinung nicht gut finde.
- d) Ich bringe meinen negativen Erfahrungen mit Emigranten ein.
- e) Ich sage nichts dazu und berichte es anschließend der Stationsleitung.
- f) Ich akzeptiere seine Meinung aufgrund seiner beruflichen Erfahrung.

Kri12*: Sie fertigen als Studierende(r) der Zahnmedizin, im Rahmen eines praktischen klinischen Kurses eine Präparation an. Sie sind stolz auf Ihre Arbeit und haben auch schon äußerst positives Feedback von Ihren Mitstudenten erhalten. Nun haben Sie Ihre Präparation bei Ihrem Dozenten eingereicht. Anhand von festgelegten Bewertungskriterien wird Ihre Arbeit mit der Note 3 (befriedigend) bewertet.

- a) Ich versuche, den Dozenten von der Qualität meiner Arbeit zu überzeugen.
- b) Ich zeige meine Arbeit einem anderen Dozenten und bitte um eine Zweitmeinung.
- c) Ich informiere den Vorgesetzten des Dozenten über die ungerechte Bewertung.
- d) Ich frage den Dozenten, ob er mir erläutern kann wie meine Note zustande gekommen ist.
- e) Ich versuche die Bewertung zu akzeptieren, obwohl ich anderer Meinung bin.

- f) Ich orientiere mich zukünftig stärker an den Bewertungskriterien und bitte Dozenten um Rückmeldungen während der Arbeitsphasen.

Items Respekt gegenüber anderen (R)

R1: Sie und ein(e) Mitstudent/In sollen gemeinsam ein Referat halten. Sie haben sich vorher über die Verteilung der Inhalte abgesprochen. Während des Referats nimmt der/die Mitstudent/In im ersten Teil viele Inhalte vorweg, für die Sie eigentlich zuständig sind.

- a) Ich halte meinen Referatsteil wie geplant und sage meinem/meiner Mitstudenten/Mitstudentin nichts.
- b) Ich halte meinen Referatsteil und versuche die bereits erwähnten Inhalte wegzulassen. Im Anschluss an das Referat spreche ich meine(n) Mitstudenten/Mitstudentin auf die nicht eingehaltene Absprache an.
- c) Ich halte meinen Referatsteil wie geplant und erwähne zu Beginn, dass die Zuhörer mit einigen Wiederholungen rechnen müssen, da mein(e) Mitstudent/In Inhalte vorweggenommen hat.
- d) Ich empfehle meinen anderen Mitstudenten/Mitstudentinnen, nicht mit meinem Referatspartner zusammen zu arbeiten, da er/sie sich nicht an Absprachen hält.
- e) Während des Referats unterbreche ich meine(n) Mitstudenten/Mitstudentin und bitte ihn/sie, sich auf die abgesprochenen Inhalte zu beschränken.

R2: Sie gehen zum Studentensekretariat, da Sie wegen einer großen Familienfeier einen Seminartermin mit Anwesenheitspflicht tauschen möchten. Obwohl sich die Sekretärin bemüht einen Tauschtermin zu finden, ist dies nicht möglich, da die anderen Seminare bereits voll sind. Ihre Mitstudenten haben Sie bereits vergeblich befragt, ob sie mit Ihnen tauschen würden. Und der Praktikumsleiter hat Sie an das Sekretariat verwiesen.

- a) Ich frage die Sekretärin erneut, ob es nicht doch noch eine Möglichkeit gibt.
- b) Ich sage ihr, dass ich mir wünschen würde, dass die Anliegen von Studierenden mehr berücksichtigt werden.
- c) Ich bedanke mich für ihre Mühe, obwohl es nicht geklappt hat.
- d) Ich gehe morgen noch einmal in das Sekretariat und hoffe, dass eine andere Sekretärin anwesend ist, die ich fragen kann.

R3: Sie sind Studierende(r) der Human-/Zahnmedizin und nehmen den ersten Tag an einem zweiwöchigen Anatomie-Seminar teil. Ihr bis dahin unbekannter Sitznachbar ist ungepflegt und hat einen unangenehmen Körpergeruch.

- a) Ich versuche den anderen Mitstudenten zu signalisieren, dass ich es nicht bin, die/der riecht.
- b) Ich versuche mit ihm ins Gespräch zu kommen und weise ihn unter vier Augen auf seinen Hygienemangel hin.
- c) Ich setze mich wo anders hin, um den Mitstudenten nicht bloßzustellen.
- d) Ich sage ihm beiläufig, dass heute ein unangenehmer Geruch in der Luft liegt.
- e) Ich bleibe auf meinem Platz sitzen und versuche, mich nicht von dem Geruch stören zu lassen.

R4: Sie sind Studierende(r) der Human-/Zahnmedizin. Innerhalb Ihrer regelmäßig stattfindenden Arbeitsgruppe lästern einige Mitstudenten/Innen stark über den Sprachfehler eines Gruppenmitglieds, das noch nicht eingetroffen ist.

- a) Ich beteilige mich, da ich kein Spielverderber sein möchte.
- b) Ich empfehle dem/der betroffene(n) Mitstudenten/Mitstudentin nach dem Treffen unter vier Augen seinen Sprachfehler behandeln zu lassen.
- c) Ich halte mich raus, da ich es unfair finde.
- d) Ich spreche den/die betroffene(n) Mitstudenten/Mitstudentin während des Treffens offen auf seinen/ihren Sprachfehler an.
- e) Ich bitte meine Mitstudenten/Innen, mit dem Lästern aufzuhören.

R5: Nach jedem Semester haben Sie als Studierende(r) die Möglichkeit, anonym Lehrveranstaltungen und die dafür verantwortlichen Lehrenden schriftlich zu beurteilen. Dafür können Sie Noten und freie Anmerkungen geben. Eine(n) Lehrende(n) fanden Sie dieses Semester dermaßen schlecht, dass Sie es kundgeben wollen.

- a) Ich gebe dem/der Lehrenden eine schlechte Note und beschreibe sachlich, was ich nicht gut fand.
- b) Ich nehme nicht an der schriftlichen Bewertung teil, da ich es für wirksamer halte meine Meinung im Gespräch mit Mitstudenten/Mitstudentinnen kund zu tun.
- c) Ich gebe dem/der Lehrenden eine schlechte Note und mache die Anmerkung, dass der/die Lehrende wenig geeignet ist, zu lehren.
- d) Ich gebe dem/der Lehrenden eine mittelmäßige Note, da ich ihn/sie trotz der schlechten Veranstaltung nicht persönlich verletzen möchte.
- e) Ich nehme nicht an der schriftlichen Bewertung teil, sondern spreche den/die Lehrende persönlich darauf an, da ich es für respektvoller halte.
- f) Ich gebe dem/der Lehrenden eine schlechte Note. Zusätzlich bitte ich meine Mitstudenten/Mitstudentinnen, ihn/sie auch schlecht zu bewerten.
- g) Ich gebe dem/der Lehrenden eine schlechte Note, um ihm/ihr zu signalisieren, dass die Veranstaltung nicht gut war.

R6: Sie sind Studierende(r) der Human-/Zahnmedizin. Zu Beginn des Semesters wird Ihnen für ein Referat ein Austauschstudent als Partner zugeteilt. Die sprachliche Verständigung erweist sich als schwierig, da der Austauschstudent nur gebrochen deutsch und englisch spricht. Sonst beherrschen Sie beide keine andere gemeinsame Sprache.

- a) Ich schlage ihm vor, dass ich das Referat allein ausarbeite und ihm dann seinen fertigen Teil gebe.
- b) Ich gehe zu dem Dozenten, schildere ihm das Problem und bitte um eine Lösung.
- c) Ich frage ihn, ob er jemanden kennt, der Deutsch und seine Erstsprache spricht, damit dieser für die gemeinsame Ausarbeitung übersetzen kann.
- d) Ich sage dem Austauschstudenten, dass wir mit seinem derzeitigen Sprachniveau leider nicht miteinander arbeiten können.
- e) Ich versuche das Nötigste auf Deutsch und Englisch mit ihm abzustimmen, sodass jeder seinen Referatsteil allein ausarbeiten kann.

R7: Sie hospitieren im OP, um sich einen Eindruck von der Tätigkeit eines Chirurgen zu machen. Während der OP redet das ärztliche Personal abfällig über die Figur des Patienten/der Patientin, der/die relativ dick ist.

- a) Ich sage dazu nichts, weil die Situation im OP für mich neu ist und ich unsicher bin.
- b) Ich warte weitere OPs ab, um einzuschätzen, ob das Verhalten die Regel darstellt.
- c) Ich stimme mit ein, ein bisschen Spaß bei der Arbeit muss sein.
- d) Ich frage während der OP, ob die Figur des Patienten/der Patientin eine Bedeutung für das Ergebnis der Operation hat.
- e) Ich frage nach dem Eingriff, ob das Gespräch ernst gemeint war.
- f) Ich frage andere Hospitanten nach ihren Erfahrungen im OP.

R8: In einer studentischen Gruppe sollen Sie über das Semester hinweg ein Projekt durchführen. Gerade treffen Sie sich alle das erste Mal, um Termine für Ihre Arbeitsgruppe abzustimmen. Einer Ihrer Mitstudenten ist zeitlich sehr unflexibel, da er mehrere Nebenjobs zur Finanzierung seines Studiums zu bewältigen hat. Der erste Eindruck von ihm wirkt auf Sie nicht sehr sympathisch.

- a) Ich bitte den Mitstudenten freundlich darum, sich eine andere Arbeitsgruppe zu suchen.
- b) Ich bespreche mit den anderen Gruppenmitgliedern, ob wir den Mitstudenten finanziell unterstützen können.
- c) Ich schlage vor, dass wir ihm etwas Arbeit abnehmen, damit er sein Studium finanzieren kann.
- d) Ich diskutiere mit allen Gruppenmitgliedern, um eine Arbeitsweise zu vereinbaren.
- e) Ich suche mir eine andere Arbeitsgruppe, in der es leichter ist, Termine zu vereinbaren.
- f) Ich und die anderen Gruppenmitglieder bitten den zuständigen Dozenten um Rat.

Items Suche nach sozialer Unterstützung (SU)

SU1*: Sie beginnen den ersten Tag Ihres Pflegepraktikums (in diesem lernen Sie die Arbeit von Pflegekräften und den Stationsablauf kennen) auf einer Krankenhausstation. Nach einer kurzen Begrüßung vom Arzt- und Pflegepersonal gehen alle hektisch Ihrer Arbeit nach. Nur Sie wissen gar nicht, was sie eigentlich tun sollen. Ein Ansprechpartner wurde Ihnen nicht zugewiesen.

- a) Ich warte ab bis ein Pfleger Zeit hat und auf mich zukommt.
- b) Ich bitte einen Pfleger darum, dass ich einen Ansprechpartner zugewiesen bekomme.
- c) Ich setze mich in das Stationszimmer und lese die Patientenakten, um mir einen Überblick über die derzeitigen Patienten zu verschaffen.
- d) Ich frage das Pflegepersonal was ich tun kann.
- e) Ich beobachte die Abläufe und versuche mich bestmöglich einzubringen.

SU2: Sie haben große finanzielle Probleme während Ihres vorklinischen Studiums. Das Geld reicht nicht mal mehr für das Nötigste. Auch Ihre Eltern können Sie nicht finanziell unterstützen. Andererseits haben Sie wenig Zeit, da das Studium Sie zeitlich voll beansprucht.

- a) Ich lasse mich bei einer Bank über Kredite beraten.
- b) Ich gehe nebenbei arbeiten und nehme eine eventuelle Verlängerung des Studiums in Kauf.
- c) Ich breche das Studium ab, da ich es mir nicht leisten kann.
- d) Ich frage Mitstudenten, die auch finanzielle Probleme haben, wie sie vorgegangen sind.
- e) Ich erkundige mich nach studentischen Beratungsstellen, um mich dort über Finanzierungsmöglichkeiten beraten zu lassen.

SU3: Eigentlich hatten Sie den Eindruck, ganz gut mit Ihren Mitstudenten zu Recht zu kommen. Seit einiger Zeit haben Sie jedoch das Gefühl, dass Sie von den Mitstudenten ausgeschlossen werden. Einen Anlass dafür ist Ihnen nicht bewusst. Sie fühlen sich sozial nicht gut eingebunden.

- a) Ich verzichte auf den Kontakt mit den entsprechenden Mitstudenten und versuche Kontakt zu anderen Mitstudenten aufzubauen.
- b) Ich warte eine Weile ab, ob sich die Situation wieder verbessert.
- c) Ich frage die entsprechenden Mitstudenten, ob mein Eindruck stimmt und ob etwas vorgefallen ist.
- d) Ich frage einen anderen Mitstudenten, ob ihm aufgefallen ist, dass ich mich unangemessen gegenüber den anderen verhalten habe.
- e) Ich wende mich an eine studentische Beratungsstelle.

SU4: Als Studierender der Humanmedizin werden Sie im Rahmen eines Praktikums im Krankenhaus von einer Ärztin gebeten, bei einer Patientin eine Untersuchung durchzuführen, die Sie noch nicht beherrschen.

- a) Ich erkläre der Stationsärztin, dass ich die Untersuchung noch nie durchgeführt habe. Ich bitte sie mir zu zeigen, damit ich sie anschließend durchführen kann.
- b) Ich schaue mir den Ablauf der Untersuchung im Lehrbuch/Internet an und führe sie anschließend bei der Patientin durch.
- c) Ich nutze die Gelegenheit und probiere die Untersuchung aus.
- d) Ich bitte einen erfahrenen Praktikanten, mir die Untersuchung zu zeigen, damit ich sie anschließend durchführen kann.

SU5: Sie befinden sich im ersten Studienjahr des Human-/Zahnmedizinstudiums. In zwei Wochen beginnt der Präparierkurs in der Anatomie. Sie machen sich ernsthafte Sorgen, den Präparierkurs in der Anatomie emotional zu bewältigen, da Sie noch nie eine Leiche gesehen haben und Sie bisher wenig mit dem Thema Tod in Kontakt gekommen sind.

- a) Ich spreche mit meinen Mitstudenten über die Situation und wie es ihnen damit geht.
- b) Ich spreche mir Mut zu, um den Kurs zu durchstehen, da ich nicht der/die erste Medizinstudierende sein werde, der/die sich Sorgen macht.
- c) Ich bitte meine Hausärztin, mir für den Kurszeitraum ein beruhigendes Medikament zu verschreiben.
- d) Ich breche das Studium ab.
- e) Ich bitte einen Betreuer des Präparierkurses um einen Termin und erzähle ihm von meinen Sorgen.
- f) Ich wende mich an eine studentische Beratungsstelle der Universität.

SU6: Sie stehen am Anfang Ihres Human-/Zahnmedizinstudiums. In den ersten Tagen kommt einiges an Verwaltungsaufwand auf Sie zu. Sie müssen sich in einige Kurse einschreiben, sich im PC-Pool und für die Druckzentrale anmelden, eine Karte für die Mensa organisieren und auch noch einige Unterlagen wegen Ihres Umzugs besorgen. Sie verlieren den Überblick und sind sich nicht wirklich sicher wo und wie Sie dies erledigen sollen.

- a) In der nächsten Lehrveranstaltung spreche ich einen Mitstudenten an und erkundige mich, wo und wie er/sie bei allen Dingen vorgegangen ist.
- b) Ich schaue auf der Homepage der Universität nach möglichen Ansprechpartnern und Tipps.
- c) Ich lasse erstmal alles auf mich zukommen. Das anfängliche Chaos wird sich im Laufe des Semesters auflösen.
- d) In der nächsten Lehrveranstaltung spreche ich einen Mitstudenten an und frage, ob wir die Dinge zusammen erledigen.

- e) In der nächsten Vorlesung spreche ich einen Lehrenden an, wohin ich mich bezüglich meiner Fragen wenden kann.
- f) Ich nehme mir die Zeit, alle Dinge aufzulisten und einzeln für mich selbst zu organisieren.

SU7: Sie stecken mitten im Lernen auf eine mündliche Prüfung. Gleichzeitig müssen Sie noch einen Vortrag vorbereiten und das Protokoll vom letzten Biochemie-Praktikum muss auch rechtzeitig abgegeben werden. Und obendrein zeigt der Putzplan Ihrer Wohngemeinschaft, dass Sie an der Reihe sind.

- a) Ich erledige alle Dinge, die anstehen, jedoch mit etwas weniger Sorgfalt als sonst.
- b) Ich frage eine(n) Mitstudenten/Mitstudentin, ob er/sie mir sein Protokoll zuschickt, damit ich schon mal eine Vorlage habe. Dann kann ich die Uni-Sachen schneller erledigen und es bleibt noch Zeit zum Putzen.
- c) Ich bitte meine Mitbewohner/Innen, ob sie diese Woche das Putzen für mich übernehmen könnten. Ich konzentriere mich auf alle Sachen, die ich für die Uni erledigen muss.
- d) Ich konzentriere mich auf meine Prüfung. Bei allen anderen Sachen schau ich, was ich noch zusätzlich schaffe.

SU8*: Als Studierende(r) der Zahnmedizin haben Sie größere Schwierigkeiten mit dem Herstellen einer zahntechnischen Arbeit. Sie bemerken, dass eine Mitstudentin von Ihnen deutlich besser mit der Aufgabe zurechtkommt als Sie.

- a) Ich schaue mir die Arbeit der zwei Mitstudenten genau an und lasse mir Tipps geben.
- b) Ich lasse mich dadurch in meiner Arbeit nicht beeinflussen.
- c) Ich tausche mich mit meiner Mitstudentin aus, der die Arbeit auch nicht so leicht fällt.
- d) Ich bitte die Dozentin, sich meine Arbeit anzuschauen und mir Verbesserungshinweise zu geben.

SU9*: Sie sind als Studierende(r) der Human-/Zahnmedizin zum zweiten Mal durch eine wichtige Prüfung gefallen, die Sie für die Anmeldung zur Zwischenprüfung (Physikum) benötigen. Beide Male haben Sie sich intensiv auf die Prüfung vorbereitet. Wenn Sie erneut durchfallen, können Sie Ihr Studium nicht mehr fortsetzen.

- a) Ich frage meine Mitstudenten, die sich in einer ähnlichen Situation befinden, um gemeinsam auf die Prüfung zu lernen.
- b) Ich lerne genauso konzentriert wie zuvor mit der Hoffnung, dass ich beim 3. Mal die Prüfung bestehe.
- c) Ich breche das Studium ab, da ich die Prüfung bestimmt wieder nicht schaffe.
- d) Ich schildere dem Prüfer meine Situation und bitte ihn um Tipps zur Prüfungsvorbereitung.

- e) Ich wende mich an eine studentische Beratungsstelle, um mögliche Strategien zur Prüfungsvorbereitung zu besprechen.

SU10*: Sie befinden sich am Ende des vorklinischen Studienabschnitts der Human-/Zahnmedizin. In ein paar Monaten stehen Ihnen wichtige mündliche Zwischenprüfungen (Physikum) bevor. In den letzten zwei Prüfungen des vergangenen Semesters waren Sie sehr aufgereggt und hatten sogar Blackouts.

- a) Ich spreche mit einem Dozenten über meine Befürchtungen.
- b) Ich frage befreundete Mitstudenten wie sie mit Prüfungssituationen umgehen.
- c) Ich bereite mich so vor wie bei den Prüfungen zuvor.
- d) Ich übe die Prüfungssituation mit befreundeten Mitstudenten ein.
- e) Ich verschiebe die mündliche Prüfung auf das nächste Semester.
- f) Ich suche professionelle Hilfe auf (z.B. beim Psychologen, beim Studentenwerk).

SU11: Sie stehen kurz vor einer mündlichen Prüfung und geraten kurz vorher in eine Konfliktsituation wegen einer Klausurbewertung mit Ihrem/Ihrer Dozenten/Dozentin, der/die Sie auch mündlich prüfen wird.

- a) Ich stelle einen Antrag von einem anderen Prüfer aufgrund Befangenheit geprüft zu werden
- b) Ich lasse mir nichts anmerken und bereite mich so gut ich kann auf die Prüfung vor.
- c) Ich spreche den Prüfer darauf an und versuche den Konflikt zu lösen.
- d) Ich hole mir ein Attest, um die Prüfung nicht machen zu müssen. Bei dem neuen Prüfungstermin hoffe ich einen anderen Prüfer zu haben.

SU12*: Sie absolvieren Ihr erstes Praktikum im klinischen Studienabschnitt in einer Ambulanz und erhalten von einer Ärztin den Auftrag bei mehreren Patienten Blut abzunehmen. In einem vorklinischen Praktikum haben Sie die Blutabnahme an einem Modell geübt und von Ihrem Dozenten die Rückmeldung erhalten, dass es schon gut klappt. Sie fühlen sich dennoch unsicher am realen Patienten Blut abzunehmen. Alle anderen anwesenden Praktikanten/Innen beherrschen die Blutabnahme bereits. Sie haben Sorge, dass die Ärztin den Auftrag zukünftig immer an einen anderen Praktikanten weitergibt, wenn Sie die Unsicherheit zeigen.

- a) Ich bitte einen anderen Praktikanten, mir bei der ersten Blutabnahme über die Schulter zu schauen. Bei den folgenden Patienten führe ich die Blutentnahmen dann selbstständig durch.
- b) Ich bitte einen Arzt, mir bei der ersten Blutabnahme über die Schulter zu schauen. Bei den folgenden Patienten führe ich die Blutentnahmen dann selbstständig durch.
- c) Ich bitte einen anderen Praktikanten die Blutentnahme zu übernehmen, da ich mich noch unsicher fühle.

- d) Ich bitte die Ärztin, die Blutentnahme zu übernehmen, da ich mich noch unsicher fühle.
- e) Ich führe die Blutentnahme durch.

SU13*: Sie sind Human-/Zahnmedizinstudierende(r) im zweiten Semester, in dem Klausuren in Chemie und Physik anstehen. Sie haben in den Seminaren und Praktika vieles nicht verstanden und haben auch nicht genügend Vorkenntnisse aus der Schule. An hilfreichen Unterlagen (Mitschriften, Zusammenfassungen, alte Klausuren) haben Sie bereits alles zusammengetragen.

- a) Ich konzentriere mich auf die Inhalte, die ich verstehe und nehme die restlichen Wissenslücken in Kauf.
- b) Ich versuche mir dennoch den Lernstoff durch erhöhte Anstrengung allein anzueignen.
- c) Ich frage Mitstudenten, ob sie mit mir eine Lerngruppe gründen möchten.
- d) Ich spreche Studierende aus älteren Semestern an, ob sie mir bei meinen Fragen helfen können.
- e) Ich fasse meine wichtigsten Fragen in einer Email zusammen und bitte die zuständigen Lehrenden um Beantwortung.

SU14*: Sie haben krankheitsbedingt einige human-/zahnmedizinische Vorlesungen versäumt. Als Sie wieder gesund sind, fragen Sie Ihren befreundeten Mitstudenten, ob sie etwas verpasst haben. Dieser antwortet, dass die Dozentin meinte, die Vorlesungen seien prüfungsrelevant. Er fand sie aber nicht wirklich spannend.

- a) Ich besorge mir die Vorlesungsunterlagen nicht, sondern konzentriere mich umso mehr auf die nächsten Vorlesungen.
- b) Ich schließe mich zur Klausurvorbereitung einer Lerngruppe an, um das Versäumte aufzuarbeiten.
- c) Ich bitte meinen befreundeten Mitstudenten, seine Unterlagen kopieren zu dürfen und arbeite sie dann eigenständig nach.
- d) Ich besorge mir die Vorlesungsunterlagen nicht, sondern erarbeite mir die verpassten Inhalte aus dem Lehrbuch.
- e) Ich warte bis mir mein befreundeter Mitstudent die Unterlagen von sich aus anbietet.
- f) Ich bitte die Dozentin, mir die Vorlesungsunterlagen zuzuschicken, um sie dann selbst anhand des Lehrbuchs nachzuarbeiten.

Items Kontaktfähigkeit (KF)

KF1: Sie ziehen aus Ihrer Heimatstadt in Ihre Studienstadt zum Human-/Zahnmedizinstudium ins erste Semester und kennen niemanden. Sie haben ein Einzelzimmer im Wohnheim zugeteilt bekommen, in dem Sie nun Ihr erstes Wochenende verbringen. Sie fühlen sich etwas einsam, da Sie noch niemanden in Ihrer Studienstadt kennen.

- a) Ich klopfe an die Türen der anderen Mitbewohner, um mich vorzustellen.
- b) Ich warte ab, bis mich die anderen Mitbewohner ansprechen.
- c) Ich halte mich in Gemeinschaftsräumen (z.B. Küche, Clubraum) auf, um mit den Mitbewohnern ins Gespräch zu kommen.
- d) Ich bleibe erst einmal auf meinem Zimmer, um mich selbst einzugewöhnen. Ich werde noch genügend Leute im Studium kennen lernen.
- e) Ich stelle mich den Mitbewohnern vor, denen ich auf dem Gang begegne.
- f) Ich frage Freunde aus der Heimat, ob sie mich das Wochenende besuchen möchten.

KF2: Sie sind Human-/Zahnmedizinstudierende(r) im zweiten Semester, in dem Klausuren in Chemie und Physik anstehen. Sie haben in den Seminaren und Praktika vieles nicht verstanden und haben auch nicht genügend Vorkenntnisse aus der Schule. An hilfreichen Unterlagen (Mitschriften, Zusammenfassungen, alte Klausuren) haben Sie bereits alles zusammengetragen.

- a) Ich fasse meine wichtigsten Fragen in einer Email zusammen und bitte die zuständigen Lehrenden um Beantwortung.
- b) Ich spreche Studierende aus älteren Semestern an, ob sie mir bei meinen Fragen helfen können.
- c) Ich frage Mitstudenten, ob sie mit mir eine Lerngruppe gründen möchten.
- d) Ich verschiebe die Klausur und versuche erst einmal die schulischen Wissenslücken zu schließen.
- e) Ich warte ab, ob sich die offenen Fragen bei der Vorbereitung noch klären.
- f) Ich versuche mir den Lernstoff durch erhöhte Anstrengung allein anzueignen.

KF3: Sie sind Human-/Zahnmedizinstudierende(r) im ersten Semester. Sie kennen zwar viele Mitstudenten, haben aber den Eindruck, sozial nicht gut eingebunden zu sein. Nach einer Vorlesung hören Sie wie sich Mitstudenten zu einem Treffen verabreden. Sie werden jedoch nicht gefragt.

- a) Ich spreche die Mitstudenten nicht an, da ich mich nicht aufdrängen möchte.
- b) Ich spreche die Mitstudenten dieses Mal nicht an, aber bei der nächsten Gelegenheit.
- c) Ich frage die Mitstudenten, ob ich mich anschließen darf.
- d) Ich gehe zur verabredeten Zeit "zufällig" zu dem Treffpunkt und hoffe, dass ich mich anschließen darf.
- e) Ich frage die Mitstudenten nach ihrem Treffen wie es war und hoffe, dass sie mich dadurch das nächste Mal auch einbeziehen.
- f) Ich spreche die Mitstudenten an, ob es einen Grund gibt, warum ich nicht gefragt werde.

KF4: Sie haben krankheitsbedingt einige human-/zahnmedizinische Vorlesungen versäumt. Der Inhalt dieser Vorlesungen ist jedoch klausurrelevant.

- a) Ich bitte einen befreundeten Mitstudenten, seine/ihre Unterlagen kopieren zu dürfen und arbeite sie dann eigenständig nach.
- b) Ich besorge mir die Vorlesungsunterlagen nicht, sondern erarbeite mir die Inhalte aus dem Lehrbuch.
- c) Ich warte bis mir meine befreundeten Mitstudenten die Unterlagen von sich aus anbieten.
- d) Ich schließe mich zur Klausurvorbereitung einer Lerngruppe an, um das Versäumte aufzuarbeiten.
- e) Ich bitte den Dozenten, mir die Vorlesungsunterlagen zuzuschicken, um sie dann selbst anhand des Lehrbuchs nachzuarbeiten.
- f) Ich hoffe, dass die Inhalte doch nicht in der Klausur drankommen und konzentriere mich umso mehr auf die nächsten Vorlesungen.

KF5: Sie sind Zahnmedizinstudierende(r) im achten Semester und bekommen eine Patientin zugewiesen, die Sie das erste Mal während eines praktischen Kurses behandeln sollen. Bei ihr soll eine unkomplizierte Füllung vorgenommen werden. Während Sie ihr die Hand zur Begrüßung reichen, wendet sie sich etwas von Ihnen ab. Sie haben den Eindruck, dass sie ängstlich ist.

- a) Ich schildere der Patientin meinen Eindruck und frage vor was sie Angst hat.
- b) Ich thematisiere die Ängstlichkeit der Patientin nicht, da ich sie nicht noch mehr verunsichern möchte.
- c) Ich sage der Patientin, dass die Behandlung nicht schlimm ist und beginne mit der Behandlung.
- d) Ich bitte meinen Dozenten, um einen anderen Patienten, da ich mir unsicher bin, wie ich mit ihr umgehen soll.
- e) Ich bitte meinen Dozenten um Rat.

KF6*: Sie befinden sich in der „Ersti- Woche“ (Einführungswoche, die von fortgeschrittenen Studierenden für Studienanfänger unmittelbar vor Beginn des ersten Semesters durchgeführt wird). Weil Sie krank sind, können Sie erst ab dem dritten Tag teilnehmen. Während Ihrer Abwesenheit haben sich schon Grüppchen gebildet, die sich gemeinsam auf den Weg in die Mensa zum Mittagessen begeben. Sie konnten sich durch Ihre Abwesenheit noch nicht bekannt machen.

- a) Ich stelle mich in die Nähe einer Gruppe und warte, dass mir angeboten wird mit in die Mensa zu kommen.
- b) Ich gehe in die Mensa und hoffe, dass sich einer meiner Mitstudenten mit zu mir an den Tisch setzt.

- c) Ich gehe in die Mensa und frage eine Gruppe von Mitstudenten, ob ich mich zu ihnen setzen darf.
- d) Ich gehe nach Hause, um dort zu essen.

KF7: Sie haben als Human-/Zahnmedizinstudierende(r) in einem Fach sehr gute Fähigkeiten. Der Stoff und das Lernen fallen Ihnen nicht schwer. In einem zugehörigen Seminar bekommen Sie Aufgaben gestellt, nach deren Bearbeitung Sie gehen dürfen. Sie merken, dass Ihr Banknachbar sich schwer tut und nur sehr langsam die Aufgaben löst während Sie schon fertig sind.

- a) Ich schiebe meinem Banknachbarn meine Lösungen zu.
- b) Ich frage meinen Banknachbarn, ob ich ihm helfen kann.
- c) Ich warte kurz ab, ob mich mein Banknachbar um Hilfe bittet. Vielleicht möchte er gar keine Unterstützung von mir.
- d) Ich gehe und nutze die gewonnene Zeit, um andere Dinge zu erledigen.

KF8: Sie arbeiten gerade einen medizinischen Vortrag aus. Dabei haben Sie einige Inhalte noch nicht ganz verstanden. Sie haben noch eine Woche bis zum Vortrag Zeit. Der Dozent hat im Seminar angeboten, dass Sie sich bei Fragen an ihn wenden können.

- a) Ich versuche, es mir selbst zu erarbeiten, auch wenn es mir schwer fällt.
- b) Ich frage ältere Studierende, ob sie sich in dieser Thematik auskennen.
- c) Ich kontaktiere den Dozenten, um meine Fragen zu klären.
- d) Ich frage meine Mitstudenten, ob sie sich in dieser Thematik auskennen.

KF9: Sie sind Human-/Zahnmedizinstudierende(r) im dritten Semester und haben sich bereits im ersten Semester mit einem Mitstudenten angefreundet. Sie verbringen einen großen Teil Ihrer Freizeit miteinander. Nun wird Ihnen die Freundschaft zu eng.

- a) Ich breche die Freundschaft ab.
- b) Ich mache mich rar und sage Treffen ab.
- c) Ich versuche andere Mitstudenten in die Freundschaft einzubeziehen.
- d) Ich warte ab, ob sich das "Engegefühl" bei mir wieder legt.
- e) Ich rede mit ihm offen darüber, dass ich mehr Zeit für mich brauche.

KF10: Sie sind Human-/Zahnmedizinstudierende(r) des 3. Semesters und sollen mit einem/einer Mitstudenten/Mitstudentin ein gemeinsames Referat ausarbeiten. Sie haben ihm/ihr bereits eine Email geschickt, die er/sie jedoch nicht beantwortet hat. Die Telefonnummer des/der Mitstudenten/Mitstudentin kennen Sie nicht und in sozialen Netzwerken ist sie auch nicht vertreten. Jetzt ist schon Freitagabend. Das Referat sollen Sie am Dienstagmorgen halten.

- a) Ich bereite meinen Referatsteil vor und schicke eine E-Mail an den/die verantwortliche(n) Dozenten/Dozentin, um ihn/ihr die Situation zu schildern und frage um Rat.
- b) Ich bereite das gesamte Referat allein vor, da mir die Situation zu heikel wird.
- c) Ich bereite meinen Referatsteil vor und warte ab, ob sich der/die Mitstudent/Mitstudentin am Wochenende meldet, vielleicht ist ihm/ihr einfach etwas dazwischen gekommen.
- d) Ich bereite meinen Referatsteil vor und spreche den/die Mitstudenten/Mitstudentin in der Vorlesung am Montagmorgen an.

KF11*: Sie sind Human-/Zahnmedizinstudierende(r) im ersten Semester und neu in der Stadt. Sie wohnen alleine in einer Wohnung und Ihre Mitstudenten kennen Sie bisher nur flüchtig. Es ist Freitagnachmittag und eigentlich hätten Sie Lust heute Abend zu einem Konzert zu gehen, das Sie im Internet entdeckt haben.

- a) Ich gehe allein los, vielleicht treffe ich dort ein bekanntes Gesicht.
- b) Ich fahre in eine andere Stadt, um dort meine alten Freunde zu treffen.
- c) Ich bleibe in meiner Wohnung, da ich noch niemanden richtig kenne.
- d) Ich schreibe über den E-Mail-Verteiler (oder ähnliches) meines Semesters, ob jemand Lust hätte, heute Abend mitzukommen.
- e) Ich melde mich bei Mitstudenten, die ich schon flüchtig kennen gelernt habe.

KF12: Sie absolvieren zusammen mit einem/einer Mitstudenten/Mitstudentin einen praktischen zahnmedizinischen Kurs. In diesem ist es Vorschrift, dass einer den/die Patienten/Patientin behandelt und der/die andere Ihnen als "Zahnarzhelfer/In" assistiert. Ihr(e) Tandempartner/In ruft Sie am Morgen des Kurstages an und sagt, dass er/sie nicht kommen kann, weil er/sie krank ist.

- a) Ich gehe zum Kurs und hoffe, dass ein(e) weitere(r) Mitstudent/In, keine(n) Partner/In hat.
- b) Ich gehe zum Kurs und frage den/die Kursleiter/In welche Lösungsmöglichkeiten es gibt.
- c) Ich gehe nicht zum Kurs, da ich ohne Partner/In keinen Patienten behandeln kann.
- d) Ich informiere den/die Kursleiter/In im Anschluss an das Telefonat und frage wie ich vorgehen soll.
- e) Ich frage unter den Mitstudenten/Mitstudentinnen am Morgen, ob jemand auch keine(n) Partner/In hat.

Anhang E Vortestung an Gymnasiasten

Anhang E-1 Beispielttestbogen

Drei kurze Fragen zu Ihnen:

Ihr Alter?

___Jahre

Ihr Geschlecht?

- weiblich
 männlich

Sind Sie an einem Human- oder Zahnmedizinstudium interessiert?

- Ja
 Nein
 weiß nicht

Nun werden Ihnen 5 Situationen vorgestellt, die Ihre Fähigkeit erfassen sollen, soziale Unterstützung zu suchen. Darunter ist zu verstehen, ob Sie in schwierigen oder problematischen Situationen erkennen, ob Sie Unterstützung von anderen benötigen und diese von geeigneten Personen anfordern.

Situation SU1: Sie beginnen den ersten Tag Ihres Praktikums auf einer Krankenhausstation. Nach einer kurzen Begrüßung vom Arzt- und Pflegepersonal gehen alle hektisch Ihrer Arbeit nach. Nur Sie wissen gar nicht, was sie eigentlich tun sollen. Ein(e) Ansprechpartner/In wurde Ihnen nicht zugewiesen.

Wie wahrscheinlich würden Sie wie folgt handeln?

Die Prozentangaben müssen insgesamt 100% ergeben.

- | | |
|---|------|
| Ich frage das Pflegepersonal was ich tun kann. | ___% |
| Ich frage eine(n) Arzt/Ärztin, ob ich ihn/sie bei der Arbeit begleiten darf. | ___% |
| Ich setze mich in das Stationszimmer und lese die Patientenakten, um mir einen Überblick über die derzeitigen Patienten zu verschaffen. | ___% |
| Ich gehe zu dem/der Klinikchef/In und bitte ihn/sie, mir einen Ansprechpartner zuzuteilen. | ___% |
| Ich bitte eine(n) Arzt/Ärztin darum, dass ich eine(n) Ansprechpartner/In zugewiesen bekomme. | ___% |

Wie würden Sie in dieser Situation vielleicht noch reagieren?

Wie verständlich ist die Situation?

sehr leicht verständlich sehr schwer verständlich

Was ist an der Situation nicht verständlich?

Ist für die Situation medizinisches Vorwissen notwendig?

- ja
- nein
- weiß nicht

Es folgten Situationen 2-5.

Platz für weitere Kommentare:

Anhang E-2 Mittlere Häufigkeiten pro Antwortoption

Item	N	Häufigkeiten pro Antwortoption															
		A		B		C		D		E		F		G		H	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
SBSR1	3	6.67	11.55	70.00	20.00	0.00	0.00	8.33	15.00	15.00	28.33						
SBSR2	3	3.33	5.77	18.33	27.54	26.67	25.17	23.33	25.17	25.66	15.30						
SBSR3	3	30.00	43.59	31.67	16.07	5.00	5.00	30.00	36.06	3.33	5.77						
SBSR4	3	15.00	13.23	40.00	26.46	13.33	15.28	16.67	28.87	13.33	5.77	1.67	2.89	0.00	0.00		
SBSR5	2	5.00	7.07	45.00	21.21	10.00	14.14	.00	.00	15.00	21.21	25.00	7.07				
SBSR6	2	32.50	10.61	20.00	28.28	5.00	7.07	22.50	17.68	20.00	28.28						
SBSR7	2	75.00	35.36	0.00	0.00	25.00	35.36	0.00	0.00								
SKSD1	3	38.33	29.30	13.00	7.55	5.33	4.51	30.00	34.64	13.33	10.41						
SKSD2	3	10.00	10.00	8.33	2.89	45.00	37.75	16.67	11.55	20.00	34.64						
SKSD3	3	46.67	35.12	10.00	10.00	36.67	28.87	6.67	5.77								
SKSD4	2	40.00	42.43	30.00	28.28	10.00	0.00	20.00	14.14	0.00	0.00						
SKSD5	3	10.00	10.00	56.67	20.82	3.33	5.77	16.67	5.77	13.33	11.55						
SKSD6	4	2.50	5.00	15.00	23.80	15.00	23.80	22.50	26.30	45.00	5.77	0.00	0.00				
SKSD7	4	25.00	50.00	47.50	49.92	22.50	33.04	5.00	10.00								
SKSD8	3	0.00	0.00	0.00	0.00	3.33	5.77	36.67	55.08	60.00	51.96						
SKSD9	3	26.67	20.82	0.00	0.00	0.00	0.00	63.33	15.28	10.00	10.00						
SKSD10	3	1.67	2.89	6.67	11.55	31.67	7.64	5.00	8.66	43.33	20.82	11.67	2.89				
SKSD11	3	15.00	13.23	0.00	0.00	25.00	25.00	31.67	16.07	25.00	25.00	3.33	5.77				
Kri1	3	13.33	11.55	16.67	5.77	3.33	5.77	3.33	5.77	63.33	11.55						
Kri2	3	0.00	0.00	1.67	2.89	21.67	17.56	0.00	0.00	73.33	11.55	3.33	5.77				

Fortsetzung Anhang E-2 Mittlere Häufigkeiten pro Antwortoption

Item	N	Häufigkeiten pro Antwortoption															
		A		B		C		D		E		F		G		H	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Kri3	4	7.50	15.00	2.50	5.00	18.75	11.81	41.25	21.75	17.50	5.00	12.50	15.00				
Kri4	3	1.67	2.89	0.00	0.00	30.00	26.46	21.67	7.64	10.00	17.32	36.67	5.77				
Kri5	4	21.25	16.52	2.50	5.00	20.00	27.08	2.50	5.00	23.75	9.46	5.00	10.00	0.00	0.00	25.00	26.46
Kri6	3	6.67	11.55	33.33	32.15	30.00	10.00	10.00	10.00	20.00	20.00						
Kri7	2	30.00	28.28	5.00	7.07	15.00	21.21	5.00	7.07	5.00	7.07	40.00	14.14				
Kri8	3	6.67	5.77	26.67	5.77	20.00	10.00	40.00	10.00	6.67	5.77						
Kri9	3	43.33	25.17	10.00	10.00	33.33	5.77	3.33	5.77	10.00	17.32						
Kri10	4	15.00	12.91	0.00	0.00	5.00	5.77	17.50	17.08	2.50	5.00	15.00	23.81	45.00	5.77		
Kri11	3	0.00	0.00	60.00	45.83	33.33	35.12	0.00	0.00	6.67	11.55						
Kri12	3	1.67	2.89	0.00	0.00	63.33	23.09	35.00	21.79								
R1	4	23.25	19.72	43.75	20.56	26.75	18.14	2.50	5.00	3.75	4.79						
R2	4	7.50	8.66	1.25	2.50	46.25	34.00	10.00	20.00	5.00	10.00	6.25	9.46	23.75	14.93		
R3	4	10.00	20.00	10.00	11.55	25.00	43.59	17.50	22.17	3.75	4.79	18.75	28.39	15.00	12.91		
R4	4	0.00	0.00	15.00	19.15	25.00	26.46	0.00	0.00	60.00	27.08						
R5	4	38.75	19.31	0.00	0.00	7.50	9.57	20.00	14.14	23.75	20.56	0.00	0.00	10.00	11.55		
R6	4	22.75	26.02	12.50	25.00	57.25	28.23	0.00	0.00	7.50	9.57						
R7	4	27.50	32.02	27.50	25.00	0.00	0.00	2.50	5.00	18.75	21.75	23.75	22.87				
R8	4	1.25	2.50	17.50	23.63	13.75	17.97	57.50	17.08	10.00	11.55						
SU1	3	10.00	17.32	33.33	15.28	10.00	10.00	26.67	29.30	20.00	13.23						
SU2	3	23.33	20.82	6.67	5.77	36.67	15.28	0.00	0.00	11.67	2.89	21.67	16.07				

Fortsetzung Anhang E-2 Mittlere Häufigkeiten pro Antwortoption

Item	N	Häufigkeiten pro Antwortoption															
		A		B		C		D		E		F		G		H	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
KF11	3	15.00	21.79	16.67	11.55	8.33	2.89	26.67	23.09	33.33	11.55	0.00	0.00				
KF12	3	13.33	5.77	33.33	5.77	0.00	0.00	35.00	8.66	18.33	7.64						

Anhang F Entwicklung des Auswertungsschlüssels

Anhang F-1 Beispielbogen

Mit der nun folgenden Situation soll Kritikfähigkeit erfasst werden. Dies ist die Bereitschaft und Fähigkeit konstruktive Kritik zu äußern und mit berechtigter und unberechtigter Kritik an der eigenen Person angemessen umzugehen.

Situation (Kri2):

Im Rahmen eines Seminars hat Ihr Sitznachbar, den sie bisher nicht gut kennen, ein offensichtlich inhaltlich schlechtes Referat gehalten. Ihr Sitznachbar scheint dennoch von der Qualität seines Referats überzeugt und fragt Sie nach dem Seminar erwartungsvoll nach Ihrer Einschätzung.

Wie realistisch schätzen Sie die Situation für ein Humanstudium ein?

vollkommen realistisch überhaupt nicht realistisch

Wie angemessen schätzen Sie die folgenden Handlungen bezüglich der gezeigten Kritikfähigkeit ein?

Das Ausmaß der Kritikfähigkeit ist für diese Situation ...						
	sehr an- gemessen	ange- messen	teils angemessen / teils <u>un</u> angemessen	<u>un</u> ange- messen	sehr <u>un</u> an- gemessen	Kein Zusam- menhang mit Kritikfähigkeit
a) Ich sage ihm, dass ich sein Referat gut fand, da ich ihn nicht verletzen möchte.						
b) Ich sage ihm, was ich an seinem Referat nicht gut fand.						
c) Ich sage ihm, dass ich es nicht richtig einschätzen kann und bitte ihn, für eine aussagekräftige Rückmeldung den Dozenten zu fragen.						
d) Ich sage ihm, was ich an seinem Referat gut fand und was er noch verbessern könnte.						
e) Ich versuche einen weiteren Mitstudenten einzubeziehen, damit er eine Rückmeldung gibt.						
f) Ich sage ihm vorsichtig, dass ich sein Referat nicht gut fand.						

Anmerkungen zu der Situation oder den Handlungsmöglichkeiten:

Anhang G Validierung an Studienbewerbern und Studieninteressierten

Anhang G-1 Immer oder nie gewählte Antwortoptionen

		Stichprobe 1				Stichprobe 2			
		zu 0% gewählt		zu 100% gewählt		zu 0% gewählt		zu 100% gewählt	
		N	%	N	%	N	%	N	%
SBSR5	a	358	46.6	4	0.5	70	42.9	0	0.0
	b	637	82.8	0	0.0	139	85.3	1	0.6
	c	42	5.5	74	9.6	13	8.0	20	12.3
	d	199	25.9	3	0.4	44	27.0	3	1.8
	e	395	51.4	4	0.5	91	55.8	0	0.0
SBSR7	a	22	2.9	121	15.7	7	4.3	29	17.8
	b	519	67.5	1	0.1	121	74.2	0	0.0
	c	330	42.9	2	0.3	68	41.7	1	0.6
	d	624	81.1	4	0.5	134	82.2	0	0.0
	e	195	25.4	4	0.5	47	28.8	2	1.2
SKSD1	a	408	53.1	13	1.7	89	54.6	6	3.7
	b	137	17.8	49	6.4	35	21.5	8	4.9
	c	683	88.8	2	0.3	150	92.0	0	0.0
	d	421	54.7	2	0.3	93	57.1	2	1.2
	e	433	56.3	5	0.7	102	62.6	0	0.0
	f	152	19.8	31	4.0	33	20.2	8	4.9
SKSD4	a	100	13.0	2	0.3	23	14.1	3	1.8
	b	509	66.2	0	0.0	124	76.1	0	0.0
	c	118	15.3	8	1.0	28	17.2	2	1.2
	d	70	9.1	25	3.3	27	16.6	10	6.1
SKSD10	a	42	5.5	1	0.1	8	4.9	2	1.2
	b	450	58.5	0	0.0	113	69.3	0	0.0
	c	356	46.3	2	0.3	93	57.1	1	0.6
	d	67	8.7	3	0.4	21	12.9	1	0.6
	e	88	11.4	4	0.5	25	15.3	1	0.6
	f	559	72.7	0	0.0	115	70.6	0	0.0
SKSD11	a	52	6.8	102	13.3	17	10.4	25	15.3
	b	206	26.8	19	2.5	55	33.7	6	3.7
	c	410	53.3	2	0.3	94	57.7	0	0.0
	d	390	50.7	4	0.5	90	55.2	0	0.0

Fortsetzung Anhang G-1 Immer oder nie gewählte Antwortoptionen

		Stichprobe 1				Stichprobe 2			
		zu 0% gewählt		zu 100% gewählt		zu 0% gewählt		zu 100% ge- wählt	
		N	%	N	%	N	%	N	%
Kri1	a	148	19.2	3	0.4	44	27.0	0	0.0
	b	47	6.1	14	1.8	2	1.2	2	1.2
	c	496	64.5	1	0.1	114	69.9	1	0.6
	d	231	30.0	1	0.1	46	28.2	0	0.0
	e	470	61.1	0	0.0	101	62.0	1	0.6
	f	330	42.9	4	0.5	82	50.3	0	0.0
Kri2	a	16	2.1	129	16.8	3	1.8	31	19.0
	b	661	86.0	0	0.0	140	85.9	0	0.0
	c	521	67.8	2	0.3	117	71.8	0	0.0
	d	383	49.8	7	0.9	82	50.3	0	0.0
	e	270	35.1	0	0.0	67	41.1	2	1.2
	f	420	54.6	0	0.0	108	66.3	0	0.0
Kri4	a	208	27.0	23	3.0	59	36.2	8	4.9
	b	617	80.2	0	0.0	129	79.1	0	0.0
	c	169	22.0	13	1.7	51	31.3	3	1.8
	d	98	12.7	47	6.1	18	11	15	9.2
	e	609	79.2	0	0.0	140	85.9	0	0.0
Kri7	a	629	81.8	0	0.0	137	84	1	0.6
	b	50	6.5	80	10.4	10	6.1	19	11.7
	c	471	61.2	2	0.3	118	72.4	0	0.0
	d	103	13.4	25	3.3	24	14.7	8	4.9
Kri10	a	559	72.7	2	0.3	120	73.6	0	0.0
	b	575	74.8	2	0.3	127	77.9	0	0.0
	c	12	1.6	270	35.1	3	1.8	77	47.2
	d	628	81.7	1	0.1	142	87.1	1	0.6
	e	391	50.8	1	0.1	98	60.1	1	0.6
Kri12	a	408	53.1	0	0.0	91	55.8	0	0.0
	b	307	39.9	0	0.0	72	44.2	0	0.0
	c	687	89.3	0	0.0	142	87.1	0	0.0
	d	17	2.2	22	2.9	6	3.7	6	3.7
	e	398	51.8	0	0.0	98	60.1	1	0.6
	f	73	9.5	8	1.0	20	12.3	3	1.8
R2	a	237	30.8	8	1.0	53	32.5	3	1.8
	b	586	76.2	0	0.0	133	81.6	0	0.0
	c	17	2.2	187	24.3	5	3.1	42	25.8
	d	527	68.5	0	0.0	114	69.9	0	0.0

Fortsetzung Anhang G-1 Immer oder nie gewählte Antwortoptionen

		Stichprobe 1				Stichprobe 2			
		zu 0% gewählt		zu 100% gewählt		zu 0% gewählt		zu 100% gewählt	
		N	%	N	%	N	%	N	%
SU1	a	416	54.1	1	0.1	103	63.2	0	0.0
	b	167	21.7	2	0.3	44	27	0	0.0
	c	421	54.7	0	0.0	91	55.8	0	0.0
	d	12	1.6	11	1.4	2	1.2	2	1.2
	e	57	7.4	1	0.1	11	6.7	0	0.0
SU2	a	444	57.7	1	0.1	88	54	0	0.0
	b	79	10.3	7	0.9	20	12.3	3	1.8
	c	710	92.3	0	0.0	157	96.3	1	0.6
	d	62	8.1	3	0.4	21	12.9	0	0.0
	e	34	4.4	2	0.3	7	4.3	0	0.0
SU5	a	29	3.8	8	1.0	9	5.5	1	0.6
	b	31	4.0	17	2.2	5	3.1	4	2.5
	c	693	90.1	0	0.0	148	90.8	0	0.0
	d	740	96.2	0	0.0	156	95.7	0	0.0
	e	275	35.8	0	0.0	63	38.7	0	0.0
	f	403	52.4	1	0.1	81	49.7	0	0.0
SU8	a	25	3.3	7	0.9	4	2.5	1	0.6
	b	445	57.9	1	0.1	109	66.9	0	0.0
	c	182	23.7	0	0.0	52	31.9	0	0.0
	d	18	2.3	5	0.7	7	4.3	2	1.2
SU9	a	60	7.8	3	0.4	18	11.0	0	0.0
	b	262	34.1	1	0.1	60	36.8	0	0.0
	c	734	95.4	0	0.0	158	96.9	0	0.0
	d	73	9.5	1	0.1	19	11.7	1	0.6
	e	78	10.1	0	0.0	14	8.6	1	0.6
SU10	a	172	22.4	1	0.1	33	20.2	0	0.0
	b	61	7.9	1	0.1	17	10.4	0	0.0
	c	394	51.2	2	0.3	84	51.5	3	1.8
	d	43	5.6	4	0.5	13	8.0	0	0.0
	e	616	80.1	3	0.4	128	78.5	0	0.0
	f	225	29.3	3	0.4	45	27.6	2	1.2
SU12	a	155	20.2	9	1.2	30	18.4	3	1.8
	b	55	7.2	71	9.2	9	5.5	14	8.6
	c	554	72.0	1	0.1	128	78.5	0	0.0
	d	538	70.0	0	0.0	119	73.0	0	0.0
	e	282	36.7	11	1.4	78	47.9	1	0.6
SU13	a	426	55.4	0	0.0	91	55.8	0	0.0
	b	169	22.0	2	0.3	37	22.7	3	1.8
	c	29	3.8	11	1.4	10	6.1	0	0.0
	d	71	9.2	0	0.0	16	9.8	0	0.0

e	249	32.4	0	0.0	55	33.7	0	0.0
---	-----	------	---	-----	----	------	---	-----

Fortsetzung Anhang G-1 Immer oder nie gewählte Antwortoptionen

		Stichprobe 1				Stichprobe 2			
		zu 0% gewählt		zu 100% gewählt		zu 0% gewählt		zu 100% gewählt	
		N	%	N	%	N	%	N	%
SU14	a	724	94.1	0	0.0	156	95.7	0	0.0
	b	83	10.8	3	0.4	29	17.8	0	0.0
	c	54	7.0	10	1.3	8	4.9	8	4.9
	d	559	72.7	0	0.0	124	76.1	0	0.0
	e	719	93.5	0	0.0	150	92.0	0	0.0
	f	84	10.9	10	1.3	37	22.7	3	1.8
KF6	a	421	54.7	2	0.3	96	58.9	0	0.0
	b	334	43.4	2	0.3	87	53.4	0	0.0
	c	16	2.1	253	32.9	4	2.5	62	38.0
	d	598	77.8	0	0.0	127	77.9	1	0.6
KF11	a	356	46.3	5	0.7	74	45.4	0	0.0
	b	287	37.3	1	0.1	81	49.7	0	0.0
	c	377	49.0	4	0.5	93	57.1	0	0.0
	d	252	32.8	4	0.5	60	36.8	1	0.6
	e	30	3.9	22	2.9	4	2.5	9	5.5

Anmerkung. N_{SP1} = 769; N_{SP2} = 163; SBSR = Selbstbeobachtung/Selbstreflexion; SKSD = Selbstkontrolle/Selbstdisziplin; Kri = Kritikfähigkeit; R = Respekt gegenüber anderen; SU = Suche nach sozialer Unterstützung; KF = Kontaktfähigkeit.

Anhang G-2 Häufigkeiten Nützlichkeit und Akzeptanz des SJTs

		Stichprobe 1		Stichprobe 2	
		N	%	N	%
Anforderungen	trifft zu	273	35	55	33.7
bewusst	trifft eher zu	324	41.5	62	38.0
	teils/teils	105	13.4	22	13.5
	trifft eher nicht zu	55	7	16	9.8
	trifft nicht zu	23	2.9	8	4.9
Nachdenken	trifft zu	162	20.7	44	27.0
angeregt	trifft eher zu	286	36.6	50	30.7
	teils/teils	139	17.8	29	17.8
	trifft eher nicht zu	127	16.3	28	17.2
	trifft nicht zu	61	7.8	12	7.4
verstärkt zu	trifft zu	263	33.7	45	27.6
studieren	trifft eher zu	210	26.9	53	32.5
	teils/teils	183	23.4	39	23.9
	trifft eher nicht zu	63	8.1	10	6.1
	trifft nicht zu	59	7.6	16	9.8
abgeschreckt	trifft zu	36	4.6	18	11.0
zu studieren	trifft eher zu	33	4.2	4	2.5
	teils/teils	82	10.5	18	11.0
	trifft eher nicht zu	134	17.2	34	20.9
	trifft nicht zu	491	62.9	89	54.6
SJT weiter-	trifft zu	266	34.1	74	45.4
empfehlen	trifft eher zu	268	34.3	41	25.2
	teils/teils	147	18.8	23	14.1
	trifft eher nicht zu	60	7.7	8	4.9
	trifft nicht zu	37	4.7	16	9.8

Anmerkung. $N_{SP1} = 767$; $N_{SP2} = 163$.

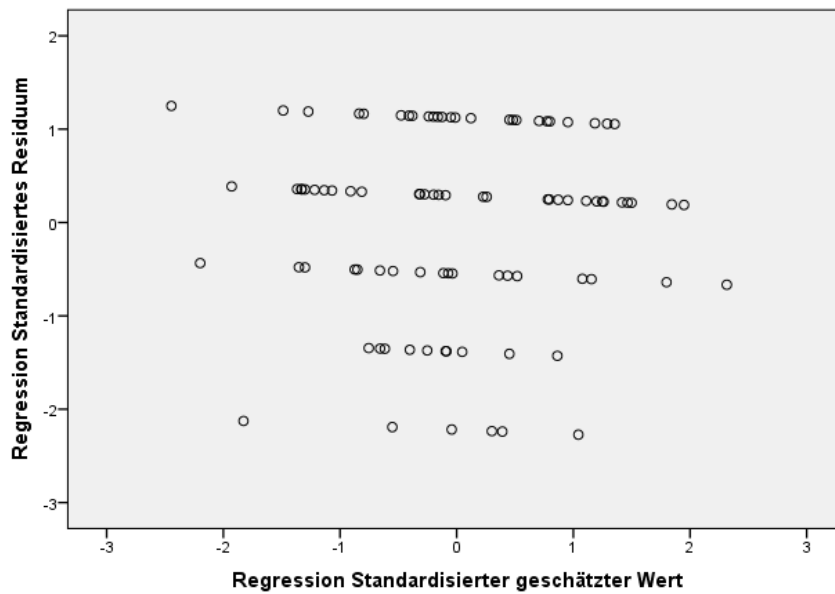
Anhang H Validierung an Medizinstudierenden

Anhang H-1 Prüfung der Homoskedastizität Lineare Regression

AV = Studienzufriedenheit



AV = subjektiver Studienerfolg

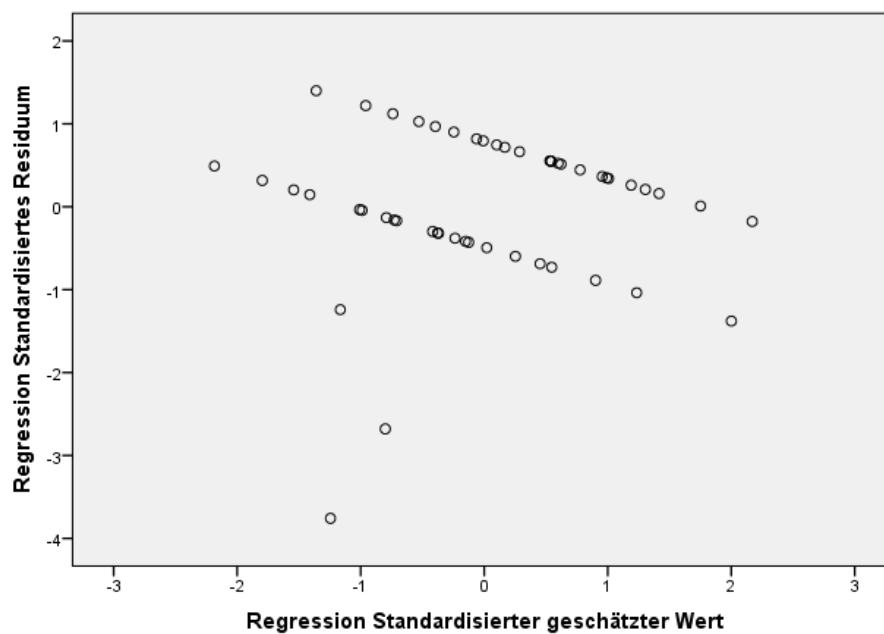


AV = Passung zum Studium



Anhang H-2 Prüfung der Homoskedastizität hierarchische Regression

AV = Studienzufriedenheit



AV = subjektiver Studienerfolg



AV = Passung zum Studium

