

„Studierendenauswahl in der Zahnmedizin - Der Einfluss der Fähigkeiten räumliches  
Vorstellungsvermögen und manuelles Geschick auf den Erwerb zahnmedizinischer  
Fertigkeiten in den vorklinischen Laborkursen“

Dissertation zur Erlangung des Doktorgrades  
der Philosophischen Fakultät  
der Christian-Albrechts-Universität  
zu Kiel

vorgelegt von

Christian Kothe

Kiel, 30. Januar 2014

Erstgutachter: Prof. Dr. Udo Konradt

Zweitgutachter: Prof. Dr. Günter Köhnken

Tag der mündlichen Prüfung: 05. Mai 2014

durch den zweiten Prodekan Prof. Dr. Martin Krieger zum Druck genehmigt: 07. Mai 2014

## **Vorwort**

Eine Dissertation stellt zwar eine selbständig verfasste Abhandlung dar, die aber ohne Unterstützung vieler nicht umzusetzen ist. Herrn Prof. Dr. Udo Konradt möchte ich für die hervorragende Betreuung in allen Phasen der Arbeit danken, vor allem für das präzise Feedback, das mich zugleich forderte, aber auch förderte. Mein Dank gilt Herrn Prof. Dr. Günter Köhnken für das Erstellen des Zweitgutachtens. Frau Dr. Yvonne Garbers wusste auf meine zahlreichen Fragen immer eine hilfreiche Antwort, wofür ich ihr zu großem Dank verpflichtet bin.

Insbesondere möchte ich mich bei Herrn Prof. Dr. Hampe für die Unterstützung bedanken, Thema und Datenerhebung im Studierendenauswahlverfahren des Universitätsklinikums Hamburg-Eppendorf (UKE) implementieren zu können. Für Rat und Tat sei ebenfalls Frau Dipl.-Psych. Johanna Hissbach, Frau Dipl.-Psych. Anja Bath, Frau Dipl.-Psych. Mirjana Knorr und Herrn Dieter Münch-Harrach aus der Arbeitsgruppe Auswahlverfahren am UKE gedankt. Den Studierenden Frau Julia Weinberg, Frau Magdalena Sieversen, Herrn Johannes Marek und Herrn Alexander Vogelsang bin ich sehr dankbar für ihre tatkräftige Unterstützung bei der Testdurchführung und Datenerfassung.

Bei Herrn Dr. Levente Kriston möchte ich mich für die anregenden Diskussionen und Ratschläge bedanken sowie für die Möglichkeit, den Methoden-Workshop am UKE als Forum zu nutzen, um Fragestellungen aus ganz unterschiedlichen Perspektiven zu erörtern. Die Bewertung der Drahtbiegeproben übernahmen Frau Dr. Vahle-Hinz und Herr Knobloch aus dem Zentrum für Zahn-, Mund- und Kieferheilkunde am UKE, wofür ich mich ganz herzlich bedanken möchte.

Nicht nur für die Unterstützung bei der Datenakquise in den vorklinischen Laborkursen der Zahnmedizin am UKE bin ich Herrn Dr. Sascha Pieger sehr dankbar, sondern auch für die diversen Erklärungen der in den zahnmedizinischen Laborkursen anzufertigen Arbeiten.

Meinen Eltern und Freunden möchte ich ganz besonders für Unterstützung und Zuspruch danken. Spezieller Dank gilt an dieser Stelle meiner Mutter, Frau Dr. Gudrun Kothe, für das ausführliche Lektorat der Dissertationsschrift.

## **Zusammenfassung**

Bisher fehlt in der Literatur eine Antwort auf die Frage, welche Fähigkeiten Studierende in vorklinischen Laborkursen der Zahnmedizin benötigen, um praktische Untersuchungs- und Behandlungsmethoden zu erlernen. Im Anwendungsfeld der Studierendenauswahl ist die Fragestellung für den Einsatz von Verfahren relevant, um die in den praktischen Laborkursen benötigten Fähigkeiten in die Bewerberselektion einzubeziehen. Die Ergebnisse in der zahnmedizinischen Studieneignungsdiagnostik fallen inkonsistent im Hinblick auf die Frage aus, ob räumliches Vorstellungsvermögen als kognitive oder manuelles Geschick als nichtkognitive Fähigkeit ausschlaggebend für das Erlernen zahnmedizinischer Fertigkeiten sind (Spratley, 1990). Vor allem das Fehlen eines theoretischen Modells, in welches die widersprüchlichen Studienergebnisse integriert werden könnten, erweist sich als ein Hindernis im Forschungsprozess.

In der vorliegenden Arbeit wurde Ackermans (1988) Theorie des Fertigkeitserwerbs angewandt, um den Einfluss des räumlichen Vorstellungsvermögens und des manuellen Geschicks auf die Aufgabenleistungen beim Erlernen zahnmedizinischer Fertigkeiten in den vorklinischen Laborkursen zu definieren. Entgegen der Untersuchung von Gray und Deem (2002a; b) wurden die vorklinischen Laborkursaufgaben in der vorliegenden Untersuchung in die Typen konsistent-einfach und inkonsistent-komplex unterteilt und in separate Prädiktionsmodelle als Kriterien eingebunden. Die Datenerhebung erfolgte im Studierendenauswahlverfahren der Zahnmedizin an der Universität Hamburg, um Messverzerrungen zu vermeiden, die beispielsweise vorliegen könnten, falls die Testverfahren nicht mit der Motivation absolviert würden, ein bestmögliches Ergebnis zu erzielen. Die Operationalisierung der räumlichen Fähigkeit und der manuellen als instrumentelles, manuelles Geschick wiesen insgesamt eine akzeptable Güte auf, wie aus den Ergebnissen der konfirmatorischen Faktorenanalyse hervorging. Zur Auswertung der postulierten Beziehungen zwischen den Fähigkeitsarten und der Aufgabenleistung in den vorklinischen Laborkursen wurde das „partial least squares“ Verfahren eingesetzt, dessen Parameterschätzungen komplementär mittels robuster Regressionsanalyse überprüft wurden.

Das instrumentelle, manuelle Geschick besaß konform zu den Erwartungen einen moderaten Einfluss auf die Leistung konsistenter, einfacher Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten in der Vorklinik, zu der die kognitive Fähigkeit entgegen den postulierten Erwartungen keinen (schwachen) Zusammenhang aufwies.

Zur Leistung inkonsistenter, komplexer Aufgaben wiesen beide Fähigkeiten wie postuliert einen moderaten Einfluss auf. Die Untersuchungsergebnisse bestätigten räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick als in den vorklinischen Laborkursen benötigte Fähigkeiten zum Erlernen zahnmedizinischer Fertigkeiten, deren Einfluss in Abhängigkeit der Aufgabentypen konsistent-einfach und inkonsistent-komplex ausfiel. Ackermans (1988) Theorie des Fertigkeitserwerbs stellt ein bewährtes Modell dar, das als theoretisches Fundament in der Studieneignungsdiagnostik der Zahnmedizin eingesetzt werden kann, um die Beziehungen zwischen Fähigkeiten und dem Erlernen von Fertigkeiten zu charakterisieren.

Das quasiexperimentelle Untersuchungsdesign, die Stichprobenbeschaffenheit und die nur bedingt zu akzeptierende Messungsgüte des räumlichen Vorstellungsvermögens stellen Limitationen der vorliegenden Untersuchungsergebnisse dar. Die wechselseitige Beziehung der räumlichen und manuellen Fähigkeiten, der Einfluss der Motivation auf die Aufgabenleistungen beim Erwerb zahnmedizinischer Fertigkeiten in den Laborkursen sowie der Phasenverlauf des Fertigkeitserwerbs in Ackermans (1988) Modell stellen mögliche Fragestellungen nachfolgender Studien dar.

## **Abstract**

In literature, an answer to the question, which abilities are needed by students to acquire practical examination and treatment methods in preclinical laboratory courses in dentistry, is still missing. This question is important for the use of tests in the field of student selection to ensure, that in practical laboratory courses required abilities are included in the selection of applicants. In dental study aptitude diagnostic, inconsistent findings are reported to the question, whether spatial thinking as cognitive ability or manual dexterity as noncognitive ability are determining for dental skill acquisition (Spratley, 1990). Especially the lack of a theoretical model, that could be used to integrate contradictory results, proves to be an obstacle in research progress.

In the present thesis, Ackerman's (1988) theory of skill acquisition was used to identify the influence of spatial thinking and manual dexterity on task performance during dental skill acquisition in preclinical laboratory courses. Different to the research by Gray and Deem (2002a; b), the preclinical laboratory tasks were divided into the categories "consistent-simple" and "inconsistent-complex" in the present study and included in separate prediction models as criteria. The data survey took place during student selection procedure in dentistry of the Universität Hamburg to avoid a measurement bias, e.g. if the abilities tests are not completed with the motivation to achieve the best possible results. The operationalization of spatial ability and manual as instrumental, manual dexterity proved to be of acceptable goodness, shown by confirmatory factor analysis' results. The partial least squares-method was used for analysis of postulated relationships between the types of abilities and task performance in preclinical laboratory courses; its parameter estimations were complementary controlled through robust regression method.

As expected, the instrumental, manual dexterity showed a moderate influence on performance of consistent, simple tasks during skill acquisition in pre-clinic, but the cognitive ability, contrary to the postulated expectation, had no (weak) relationship to those. In agreement with expectations, both abilities revealed a moderate influence on performance of inconsistent, complex tasks. The present research findings confirmed spatial thinking and instrumental, manual dexterity as required abilities for acquisition of dental skills in preclinical laboratory courses, which are depending on the task categories consistent-simple and inconsistent-complex. Ackerman's (1988) theory of skill acquisition represented a valid model, which can be used as a theoretical basis in

dental study aptitude diagnostics for characterizing the relationships between abilities and skill acquisition.

The quasi experimental design, the sample quality and the measurement quality restrictions of spatial ability are limitations of the present thesis. The recursive relationship of spatial and manual abilities, the influence on task performance at dental skill acquisition in laboratory courses through motivation and the phase progress of skill acquisition in Ackerman's (1988) model are feasible research questions in further studies.

# Inhaltsverzeichnis

Vorwort.....	III
Zusammenfassung .....	IV
Abstract.....	VI
Inhaltsverzeichnis .....	VIII
Tabellenverzeichnis .....	XIII
Abbildungsverzeichnis .....	XV
Abkürzungsverzeichnis .....	XVI
1. Einleitung .....	1
2. Theoretische und empirische Herleitung .....	4
2.1 Studierendenauswahl .....	4
2.1.1 Ziele hochschuleigener Auswahlverfahren .....	4
2.1.2 Studierfähigkeit und Studienerfolg .....	5
2.1.3 Vorhersage des Studienerfolgs.....	6
2.2 Vorhersage des Studienerfolgs in der Zahnmedizin.....	8
2.2.1 Ablauf des Zahnmedizinstudiums.....	8
2.2.2 Schulabschlussnoten .....	8
2.2.3 Studierfähigkeitstests .....	10
2.2.3.1 Test für medizinische Studiengänge.....	10
2.2.3.2 Besonderer Auswahltest Zahnmedizin .....	11
2.2.3.3 Dental Admission Test .....	12
2.2.4 Interviews.....	13
2.2.5 Nichtkognitive Merkmale .....	13
2.2.6 Situational Judgment Tests .....	14
2.2.7 Universitäre Auswahlverfahren .....	15
2.2.8 Zusammenfassung und Resümee .....	16
2.3 Vorhersage der praktischen Studienleistung in der Zahnmedizin.....	18
2.3.1 Definitionen .....	18
2.3.1.1 Fähigkeit.....	18
2.3.1.2 Fertigkeit .....	19
2.3.1.3 Manuelles Geschick .....	19
2.3.1.4 Räumliches Vorstellungsvermögen.....	20
2.3.1.5 Psychomotorik.....	21
2.3.2 Frühe Studien .....	22
2.3.3 Dental Admission Test.....	22
2.3.3.1 Subtests „Spatial Relations“ und „Carving Dexterity“ .....	22
2.3.3.2 Dimensionalität der Subtests „Carving Dexterity“ und „Spatial Relations“ .....	24
2.3.3.3 Entnahme des Subtests „Carving Dexterity“ .....	25
2.3.3.4 Die Einführung des „Perceptual Ability Tests“ .....	27
2.3.4 Fähigkeiten als Prädiktoren.....	28
2.3.4.1 Räumliches Vorstellungsvermögen.....	28
2.3.4.1.1 Frühere Studien.....	28

2.3.4.1.2	Aktuellere Studien .....	30
2.3.4.2	Manuelles Geschick .....	33
2.3.4.2.1	Frühere Studien.....	33
2.3.4.2.2	Aktuellere Studien .....	35
2.3.4.3	Kognitive Stile .....	37
2.3.4.4	Psychomotorik.....	38
2.3.4.4.1	Frühere Studien.....	38
2.3.4.4.2	Aktuellere Studien .....	41
2.3.5	Methodische Kritik .....	42
2.3.5.1	Psychometrische Güte des praktischen Studienerfolgs .....	42
2.3.5.1.1	Noten als Kriterien der praktischen Studienleistung.....	42
2.3.5.1.2	Faktoren der zahnmedizinischen Studienleistung.....	43
2.3.5.2	Statistische Auswertungsverfahren .....	43
2.3.5.3	Stichprobenbeschaffenheit .....	45
2.3.5.4	Validität der Testinstrumente .....	46
2.3.5.4.1	Gütekriterien der Klassischen Testtheorie .....	46
2.3.5.4.2	Inhaltliche Güte .....	47
2.3.6	Fertigkeitserwerb in den vorklinischen Laborkursen.....	48
2.3.6.1	Vielfalt zahnmedizinischer Fertigkeiten.....	48
2.3.6.2	Frühe Phase des Erlernens zahnmedizinischer Fertigkeiten.....	49
2.3.6.3	Lernphasen zahnmedizinischer Fertigkeiten .....	50
2.3.7	Zusammenfassung und Resümee .....	51
2.3.7.1	Dental Admission Test .....	51
2.3.7.2	Fähigkeiten als Prädiktoren.....	52
2.3.7.3	Methodische Kritik.....	55
2.3.7.4	Fertigkeitserwerb in den vorklinischen Laborkursen .....	56
2.4	Vorhersage erlernter Leistung: Ackermans Theorie des Fertigkeitserwerbs .....	57
2.4.1	Definition allgemeine Intelligenz.....	57
2.4.2	Ackermans Theorie des Fertigkeitserwerbs .....	57
2.4.2.1	Komplexität und Konsistenz von Aufgaben.....	59
2.4.2.2	Prädiktion der Leistung beim Fertigkeitserwerb .....	59
2.4.2.3	Resource Allocation Model.....	60
2.4.2.4	Prädiktion praktischer Leistung in der Zahnmedizin.....	61
2.4.3	Vorhersage beruflicher Leistung.....	63
2.4.3.1	Drei-Phasen-Modell von Ackerman.....	63
2.4.3.2	Der „whole-Person“ Ansatz als Prädiktor von Leistung .....	65
2.4.4	Zusammenfassung und Resümee .....	66
3.	Fragestellung und Hypothesen .....	68
3.1	Vorhersage der praktischen Studienleistung .....	68
3.2	Ziele der vorliegenden Arbeit .....	70
3.3	Hypothesen .....	72
3.3.1	Diskrimination der Fähigkeitsarten.....	72
3.3.2	Vorhersage der Leistung in den vorklinischen Laborkursen .....	74
3.3.2.1	Konsistente, einfache Aufgaben.....	76
3.3.2.2	Inkonsistente, komplexe Aufgaben .....	79
3.3.3	Graphische und tabellarische Darstellung der Hypothesen.....	83

4. Methodik.....	85
4.1 Operationalisierung der Untersuchungsvariablen.....	85
4.1.1 Instrumentelles, manuelles Geschick.....	86
4.1.1.1 Beschreibung des HAM-Man.....	86
4.1.1.2 Auswertung des HAM-Man.....	87
4.1.1.3 Güte des HAM-Man.....	87
4.1.1.4 Operationalisierung als latentes Konstrukt.....	88
4.1.2 Räumliches Vorstellungsvermögen.....	90
4.1.2.1 Testbeschreibung des LPS.....	90
4.1.2.2 Auswertung des LPS.....	90
4.1.2.3 Gütekriterien des LPS.....	90
4.1.2.4 Operationalisierung als latentes Konstrukt.....	91
4.1.2.5 Neukonzeption des LPS.....	92
4.1.3 Schulabschlussnote.....	92
4.1.3.1 Beschreibung der Abiturdurchschnittsnote.....	92
4.1.3.2 Güte der Schulabschlussnote.....	93
4.1.3.3 Operationalisierung als latentes Konstrukt.....	93
4.1.4 Naturwissenschaftliches Verständnis.....	94
4.1.4.1 Beschreibung des HAM-Nat.....	94
4.1.4.2 Auswertung des HAM-Nat.....	95
4.1.4.3 Güte des HAM-Nat.....	95
4.1.4.4 Operationalisierung als latentes Konstrukt.....	95
4.1.5 Praktische Studienleistung.....	96
4.1.5.1 Beschreibung der Laborkurse.....	96
4.1.5.2 Operationalisierung als latentes Konstrukt.....	97
4.2 Untersuchungsdurchführung.....	99
4.2.1 Datenerhebung.....	99
4.2.2 Stichprobe.....	99
4.2.3 Kontrolle von Störvariablen.....	100
4.3 Auswertung der Untersuchungsfragestellung.....	102
4.3.1 Exkurs: Güte des HAM-Man nach Klassischer Testtheorie.....	103
4.3.2 Konfirmatorische Faktorenanalyse.....	104
4.3.2.1 Modellidentifikation.....	104
4.3.2.2 Parameterrestriktionen.....	105
4.3.2.3 Parameterschätzungen.....	106
4.3.2.4 Modellevaluation.....	107
4.3.2.5 Graphische Darstellung des Messmodells.....	111
4.3.3 Partial Least Squares-Pfadanalyse.....	113
4.3.3.1 Vor- und Nachteile.....	114
4.3.3.2 Modellevaluation.....	116
4.3.3.3 Evaluation von Single-Item-Konstrukten.....	119
4.3.3.4 Inkrementelle Validität.....	121
4.3.4 Robuste Regressionsanalyse.....	122
4.3.4.1 MM-Schätzer.....	123
4.3.5 Prädiktionsmodelle der praktischen Studienleistung.....	124
4.3.6 Software.....	126

5. Ergebnisse.....	127
5.1 Vorbereitende Auswertungsanalysen .....	127
5.1.1 Deskriptive Statistik.....	127
5.1.2 Exkurs: Testgüte des HAM-Man nach Klassischer Testtheorie .....	130
5.2. Diskrimination der Fähigkeitsarten .....	132
5.2.1 Voraussetzungen der konfirmatorischen Faktorenanalyse.....	132
5.2.2 Konkurrenzmodelle.....	133
5.2.3 Modellevaluation .....	135
5.2.3.1 Reliabilitäten der manuellen und räumlichen Fähigkeit.....	137
5.2.3.2 Diskrimination der manuellen und räumlichen Fähigkeit .....	138
5.3 Prädiktion der vorklinischen Laborkursleistung.....	140
5.3.1 Evaluation der Prädiktionsmodelle .....	140
5.3.1.1 Analyse der Messmodelle .....	140
5.3.1.2 Analyse der Strukturmodelle.....	146
5.3.2 Hypothesenprüfung .....	150
5.3.2.1 Konsistente, einfache Aufgaben.....	151
5.3.2.2 Inkonsistente, komplexe Aufgaben .....	153
5.3.2.3 Zusammenfassung .....	154
6. Diskussion .....	157
6.1 Ziele der vorliegenden Untersuchung.....	157
6.2 Diskriminanz der Fähigkeitsart.....	159
6.2.1 Reliabilität und diskriminante Validität der Fähigkeitsarten .....	159
6.2.2 Operationalisierungsgüte des räumlichen Vorstellungsvermögens .....	160
6.2.2.1 Statistischer Erklärungsansatz.....	160
6.2.2.2 Psychometrische Güte des LPS .....	160
6.2.2.3 Zusammenfassende Bewertung der Messungsgüte .....	162
6.2.3 Operationalisierungsgüte des instrumentellen, manuellen Geschicks .....	162
6.2.3.1 Gütekriterien der Klassischen Testtheorie.....	162
6.2.3.2 Bi-Faktor-Modell der manuellen Fähigkeit.....	162
6.2.3.3 Konkurrente Validität.....	164
6.2.4 Beziehung zwischen räumlicher und manueller Fähigkeit.....	165
6.3 Vorhersage der Leistung in den vorklinischen Laborkursen.....	167
6.3.1 Prädiktion der Leistung konsistenter, einfacher Aufgaben .....	167
6.3.1.1 Einfluss der räumlichen und manuellen Fähigkeit .....	167
6.3.1.2 Einfluss der komplementären Prädiktoren .....	168
6.3.1.3 Inkrementelle Validitäten der räumlichen und manuellen Fähigkeit.....	169
6.3.2 Prädiktion der Leistung inkonsistenter, komplexer Aufgaben.....	170
6.3.2.1 Einfluss der räumlichen und manuellen Fähigkeit .....	171
6.3.2.2 Einfluss komplementärer Prädiktoren .....	172
6.3.2.3 Inkrementelle Validitäten der räumlichen und manuellen Fähigkeit.....	172
6.3.3 Resümee .....	173
6.4 Applikation der Untersuchungsergebnisse .....	174
6.4.1 Prädiktion vorklinischer Laborkursleistungen .....	174
6.4.1.1 Prädiktion konsistenter-einfacher Laborkursaufgaben (Vorklinik).....	174

6.4.1.2 Prädiktion inkonsistenter-komplexer Laborkursaufgaben (Vorklinik).....	175
6.4.2 Prädiktive Stärke der räumlichen und manuellen Fähigkeit .....	176
6.4.2.1 „Silo“-Problem.....	177
6.4.2.2 Beurteilung der Einflussstärken in der vorliegenden Untersuchung .....	178
6.4.2.3 Geschlechtsspezifität des räumlichen Vorstellungsvermögens .....	179
6.4.3 Operationalisierung des manuellen Geschicks.....	180
6.4.3.1 Weiterentwicklung des Drahtbiegetests HAM-Man .....	181
6.4.3.2 Trainierbarkeit manueller Fähigkeiten .....	182
6.4.3.3 Anforderungen an Aufgaben manueller Tests.....	183
6.4.4 Operationalisierung der praktischen Studienleistung.....	184
6.4.5 Motivation als Störvariable .....	185
6.4.5.1 Einfluss der Motivation auf die Fähigkeitsmessungen.....	186
6.5 Limitationen.....	188
6.5.1 Untersuchungsstichprobe .....	188
6.5.2 Messzeitpunkt der praktischen Studienleistung .....	189
6.5.3 Güte der latenten Variable räumliches Vorstellungsvermögen.....	189
6.5.4 Interkulturelle Disparität .....	190
6.6 Ausblick und zukünftige Fragestellungen .....	191
6.6.1 Empfehlungen für zukünftige Studien .....	191
6.6.1.1 Operationalisierung von Fähigkeiten .....	191
6.6.1.2 Nachverfolgung von Studienabbrüchen .....	192
6.6.1.3 Erhöhung der Stichprobengröße.....	193
6.6.1.4 Klassifikation der Aufgabencharakteristik .....	193
6.6.1.5 Operationalisierung der praktischen Studienleistung .....	194
6.6.2 Zukünftige Fragestellungen .....	196
6.6.2.1 Phasenverlauf des Fertigkeitserwerbs in Ackermans Modell.....	196
6.6.2.2 Einfluss der Motivation auf den Erwerb zahnmedizinischer Fertigkeiten.....	196
7. Literatur .....	198
8. Anhang .....	220

## Tabellenverzeichnis

Tabelle 2.01: Trimodaler Ansatz der Berufseignungsdiagnostik in der Studierendenauswahl.....	7
Tabelle 2.02: Testaufgaben der BATZ und erfasste Merkmale .....	11
Tabelle 3.01: Übersicht zu den Untersuchungshypothesen.....	84
Tabelle 4.01: Skalen des LPS (2. Auflage, 1983) .....	90
Tabelle 4.02: Aufgaben in den vorklinischen Laborkursen .....	97
Tabelle 4.03: Gütekriterien der konfirmatorischen Faktorenanalyse .....	108
Tabelle 4.04: Gütekriterien varianzbasierter Strukturgleichungsmodelle.....	116
Tabelle 5.01: Deskriptive Statistik .....	127
Tabelle 5.02: Interkorrelationen der Untersuchungsvariablen .....	129
Tabelle 5.03: Globale Gütemaße der KFA-Modelle .....	134
Tabelle 5.04: Deskriptive Statistik der Indikatoren.....	135
Tabelle 5.05: Parameterschätzungen der Indikatoren.....	136
Tabelle 5.06: Faktorenkennwerte .....	137
Tabelle 5.07: Diskriminante Validitäten .....	138
Tabelle 5.08: Faktorladungen im TPK-Messmodell .....	141
Tabelle 5.09: Faktorladungen im PhaI-Messmodell .....	142
Tabelle 5.10: Gütekriterien des TPK-Messmodells .....	143
Tabelle 5.11: Gütekriterien des PhaI-Messmodells.....	144
Tabelle 5.12: Kreuzkorrelationen im TPK-Prädiktionsmodell .....	145
Tabelle 5.13: Kreuzkorrelationen im PhaI-Prädiktionsmodell.....	146
Tabelle 5.14: Gütekriterien des TPK-Strukturmodells.....	147
Tabelle 5.15: Gütekriterien des PhaI-Strukturmodells .....	147
Tabelle 5.16: Korrelationen zwischen den latenten Variablen im TPK-Strukturmodell.....	148
Tabelle 5.17: Korrelationen zwischen den latenten Variablen im PhaI-Strukturmodell .....	148
Tabelle 5.18: Regressionskoeffizienten der robusten Regressionsanalyse .....	149
Tabelle 5.19: Inkrementelle Validitäten im TPK-Prädiktionsmodell.....	152
Tabelle 5.20: Inkrementelle Validitäten im PhaI-Prädiktionsmodell.....	154
Tabelle 5.21: Resultate der Hypothesenprüfung .....	155
Tabelle A01: Deskriptive Statistik der Erhebungsinstrumente .....	222
Tabelle A02-1: Mittelwerte, Standardabweichung, Reliabilitäten und Interkorrelationen der Erhebungsinstrumente.....	223
Tabelle A02-2: Mittelwerte, Standardabweichung, Reliabilitäten und Interkorrelationen der Erhebungsinstrumente (Fortsetzung) .....	224

Tabelle A02-3: Mittelwerte, Standardabweichung, Reliabilitäten und Interkorrelationen der Erhebungsinstrumente (Fortsetzung) .....	225
Tabelle A03: Interrater-Reliabilität des HAM-Man .....	226
Tabelle A04: Retest-Reliabilität des HAM-Man .....	226
Tabelle A05: Interkorrelationen HAM-Man (OE) und HAM-Man (AdH) .....	227
Tabelle A06: HAM-Man OE - Mittelwerte, Standardabweichung, Reliabilitäten und Korrelationen .....	228
Tabelle A07: HAM-Man OE - Deskriptive Statistik des HAM-Man in der Orientierungseinheit .....	229
Tabelle A08: HAM-Man OE - Interrater-Reliabilitäten ( $ICC_{unjustiert}$ ) .....	230
Tabelle A09: Parameterschätzungen der Indikatoren im KFA-Modell (Ein-Faktor-Modell) .....	230
Tabelle A10: Faktorenkennwerte im KFA-Modell (Ein-Faktor-Modell) .....	230
Tabelle A11: Korrelationen und Validitäten im KFA-Modell (Ein-Faktor-Modell) ...	231
Tabelle A12: Parameterschätzungen der Indikatoren im KFA-Modell (Drei-Faktoren-Modell) .....	231
Tabelle A13: Faktorenkennwerte im KFA-Modell (Drei-Faktoren-Modell) .....	231
Tabelle A14: Korrelationen und Validitäten im KFA-Modell (Drei-Faktoren-Modell) .....	232
Tabelle A15: Faktorladungen im PhaII-Messmodell .....	233
Tabelle A16: Gütekriterien des PhaII-Messmodells .....	234
Tabelle A17: Kreuzkorrelationen im PhaII-Prädiktionsmodell.....	235
Tabelle A18: Korrelationen zwischen den latenten Variablen im PhaII-Strukturmodell .....	235
Tabelle A19: Gütekriterien des PhaII-Strukturmodells.....	235
Tabelle A20: Inkrementelle Validitäten im PhaII-Prädiktionsmodell.....	236
Tabelle A21: Koeffizienten der robusten Regressionsanalyse .....	236
Tabelle A22: Bewertungskriterien der vorklinischen Laborkursaufgaben.....	236

## **Abbildungsverzeichnis**

Abbildung 2.01: Jährliche Kosten in Euro pro Studierenden an deutschen Hochschulen in 2008 .....	5
Abbildung 2.02: Ackermans Modell des Fertigkeitserwerbs .....	58
Abbildung 3.01: Untersuchungsmodell zur Vorhersage der praktischen Aufgabenleistungen.....	83
Abbildung 4.01: Figurenvorlagen des HAM-Nat.....	86
Abbildung 4.02: Bi-Faktor-Modell des instrumentellen, manuellen Geschicks .....	89
Abbildung 4.03: Messmodell im Prädiktionsmodell der praktischen Studienleistung .....	112
Abbildung 4.04: Konzeptionelles PLS-Prädiktionsmodell .....	124
Abbildung A01: Vorlage der zu biegenden Figur in der DBP .....	220
Abbildung A02: Operationalisierung des instrumentellen, manuellen Geschick als Ein-Faktor-Modell .....	220
Abbildung A03: Operationalisierung des instrumentellen, manuellen Geschicks als Drei-Faktoren-Modell .....	221

## Abkürzungsverzeichnis

$\alpha$	Signifikanzniveau
$\beta$	Regressionskoeffizient
$\hat{\beta}^{(0)}$	S-Schätzer (RR)
$\hat{\beta}^{MM}$	MM-Schätzer (RR)
$\delta$	Differenz
$\delta_i$	Fehlerterm
$\delta_{hj}$	Residualvariable
$\zeta_j$	Residualvariable
$\varepsilon_{hj}$	Residualvariable
$\eta$	latente Variable (Konstrukt)
$\theta_{ii}$	Messfehlervarianzen, Kovarianzmatrix der Fehlerterme
$\lambda_{ij}$	standardisierte Faktorladung (KFA)
$\lambda_{ijn}$	nichtstandardisierte Faktorladung (KFA)
$\lambda_i^2$	Ladungskoeffizient
$\Lambda$	Matrix der Faktorladungen
$\Sigma$	Varianz-Kovarianz-Matrix
$\pi_{hj}$	Gewichte
$\xi_j$	latente Variable (Faktor)
$\rho$	mittlere gewichtete korrigierte Validität
$\rho_c$	Faktorreliabilität
$\sigma^2$	Varianz
$\phi_{jj}$	Faktorvarianz
$\Phi$	Kovarianzmatrix der Faktorwerte
$\chi^2$	Chi-Quadrat-Koeffizient
$\omega_{hj}$	Ladungskoeffizient
&	und
$\sum$	Summe
$\int_j$	Skalar
$\pi_{hj}$	Gewichte (PLS)
Abb.	Abbildung
ACT	Adaptive Control of Thought
ADA	American Dental Association
AGFI	Adjusted Goodness of Fit Index

b	Regressionskoeffizient (RR)
BATZ	Besonderer Auswahltest Zahnmedizin
beta	standardisierter Regressionskoeffizient
bzw.	beziehungsweise
C-Werte	Centile
CFI	Comparative Fit Index
cov	Covariance Matrix
C.R.	Critical Ratio
Cronbachs $\alpha$	Cronbachs Alpha
d	Cohens d
DAT	Dental Admission Test
DBP	Drahtbiegeprobe
DBP (F)	Drahtbiegeprobe, Skala Form
DBP (S)	Drahtbiegeprobe, Skala Sauberkeit
DD	Dritte Dimension
DDT	Dental Dexterity Test
DE	Dreieck
DEV	Durchschnittlich erfasste Varianz
df	Freiheitsgrad
DG	Deckungsgleichheit
Dim.	Dimension
et al.	et alii
etc.	et cetera
Exz	Exzess
F	Testprüfgröße des F-Tests
$F_{ML}$	ML-Funktionsgleichung
$f^2$	Effektstärke (PLS)
f.	und folgende Seite
f.f.	und folgende Seiten
FKE	Fertigkeitserwerb
FLK	Fornell-Larcker-Kriterium
g	Faktor der allgemeinen Intelligenz
GFI	Goodness of fit index
GoF	Goodness of Fit

---

GPA	grade point average
H	Hypothese
$h^2$	Kommunalität (KFA)
$H_j^2$	Kreuzvalidierungsmaß Kommunalität (PLS)
HAM-Man	Hamburger Auswahlverfahren für Medizinische Studiengänge - Manueller Test
HAM-Nat	Hamburger Auswahlverfahren für Medizinische Studiengänge - Naturwissenschaftstest
ICC	Intra-Klassen-Korrelationen
instrument.	instrumentelles
k	Anzahl der Prädiktoren
KFA	konfirmatorische Faktorenanalyse
KI	Konfidenzintervall
log	Logarithmus
LPS	Leistungsprüfsystem
LPS-2	Leistungsprüfsystem 2
LV	latente Variable
M	Mittelwert
Man	instrumentelles, manuelles Geschick
Max	Maximalwert
Min	Mindestwert
ML	Maximum-Likelihood-Methode
n	Stichprobengröße, Anzahl
naturwiss.	naturwissenschaftlich
natur. Verst.	naturwissenschaftliches Verständnis
Natur	naturwissenschaftliches Verständnis (hypothetisches Konstrukt)
NFI	Normed Fit Index
Nr.	Nummer
n.s.	nicht signifikant
OE	Orientierungseinheit
p	Signifikanzwert
$p_m$	Itemschwierigkeit
PA	planes Aufliegen
PAT	Perceptual Ability Test

---

PAT-2D	Subskala des PAT (zweidimensionale Objekte)
PAT-3D	Subskala des PAT (dreidimensionale Objekte)
PCLOSE	p of close fit of RMSEA
Pha	Phantomkurs
PLS	Partial Least Squares
PMAT	Perceptual Motor Ability Test
$Q^2$	Stone-Geissers $Q^2$
QB	Qualität der Biegung
QLS	Kleinste-Quadrate-Schätzung
r	Korrelationskoeffizient
R	Rater
$r_{ij}$	Kreuzkorrelation
$r_{(it)}$	Trennschärfe
$r_{i(t-i)}$	korrigierte Trennschärfe
$r_{tt}$	Splitt-Half Reliabilität
$R^2$	Bestimmtheitsmaß
$\Delta R^2$	inkrementelle Validität
Raum	räumliches Vorstellungsvermögen
RB	Rundbögen
rel ( $x_i$ )	Indikatorreliabilität
rel ( $\xi_j$ )	Faktorreliabilität
RMPFB	Revised Minnesota Paper Form Board
RMSEA	Root Mean Square Error of Approximation
RR	robuste Regressionsanalyse
RV	räumliches Vorstellungsvermögen (hypothetisches Konstrukt)
s.	siehe
S	Varianz-Kovarianz-Matrix
S.	Seite
SAN	Schulabschlussnote (hypothetisches Konstrukt)
Sch	Schiefe
Schul.	Schulabschlussnote
SD	Standardabweichung
S.E.	Standardfehler der Schätzung
SfH	Stiftung für Hochschulzulassung

---

Sig.	Signifikanz
SJT	Situational Judgment Test
SJI	Situational Judgment Inventory
SRMR	Standardized Root Mean Residual
SWT	Shapiro-Wilk-Test
t	Testprüfgröße des t-Tests
tr	Summe der Diagonalelemente (KFA)
Tab.	Tabelle
TAM	Technology Acceptance Mode
TMS	Test für medizinische Studiengänge
TN	Teilnote
TPK	Technisch-Propädeutischer-Kurs
UT	Untertest
vgl.	vergleiche
x	Variable, Indikator
$x_i^T$	transponierten Vektor der Regressoren (RR)
y	Variable, Indikator
ZÄPrO	Approbationsordnung für Zahnärzte
z.B.	zum Beispiel

## 1. Einleitung

Die Antwort auf die Frage, über welche Fähigkeiten, Eigenschaften und Kenntnisse ein Zahnarzt\* zur erfolgreichen Ausübung seines Berufes verfügen muss, sollte auch die Antwort auf die Frage sein, wie ein Auswahlverfahren für Studienplatzbewerber in der Zahnmedizin inhaltlich zu gestalten ist. Eine eingehende Betrachtung dieser Frage fördert jedoch beträchtliche Widersprüche zutage, die überwiegend auf die im Studium zu erlernende zahnärztliche Berufsausbildung zurückzuführen sind. Dabei ist die Frage der Studierendenauswahl immer ambivalent zu betrachten. Während sich für Bewerber die Frage der Studienfachwahl stellt, ist es für Universitäten die Frage der Studienfacheignung eines Bewerbers (Schuler & Hell, 2008). Im Feld der Studieneignungsdiagnostik bieten universitäre Auswahlverfahren auf beide Fragen eine Antwort. Einerseits erhalten Bewerber eine Rückmeldung über die Anforderungen eines Studiengangs, andererseits können Hochschulen in Erfahrung bringen, ob der Bewerber für das von ihm gewählte Studium geeignet ist.

Um die Eignung für einen Studiengang festzustellen, werden im Allgemeinen zunächst die benötigten Fähigkeiten, Persönlichkeitseigenschaften und Kenntnisse von Experten definiert, für deren Beurteilung unter anderem auf Schulnoten der Bewerber und fachbezogene Studierfähigkeitstests oder spezielle psychometrische Testverfahren zurückgegriffen wird (Brandstätter & Farthofer, 2002). Zur Überprüfung der an die Bewerber eines Studiengangs gerichteten Anforderungen erfolgt überwiegend der Einbezug des Studienerfolges als Kriterium. Hier gilt als Faustregel, je besser der Studienerfolg durch ein Auswahlinstrument vorhergesagt wird, desto besser ist das Instrument zur Studierendenauswahl geeignet (Brandstätter, 2008). Gleichzeitig kann beurteilt werden, ob dem Auswahlinstrument zugrundeliegende Fähigkeit(en), Eigenschaft(en) oder Kenntnisse benötigt werden, um erfolgreich die Anforderungen des Studiums zu bewältigen.

Der Studiengang der Zahnmedizin beinhaltet neben einer theoretischen auch eine praktische Ausbildung, in der die Studierenden zahnärztliche Untersuchungs- und Behandlungsmethoden erlernen. In der Literatur zur Auswahlforschung in der Zahnmedizin besteht Konsens, dass kognitive Fähigkeiten ausschlaggebend für die Leistungen in den theoretischen Lehrveranstaltungen sind (vgl. Spratley, 1990), in denen neben naturwissenschaftlichen auch spezielle zahnmedizinische Themen unterrichtet werden, welche beispielsweise die Zahnanatomie und die in der Patientenbehandlung verwendeten Werkstoffe und Materialien umfassen.

\* Für eine bessere Lesbarkeit wird nur die männliche Form geführt; die weibliche ist gleichzeitig impliziert.

Es besteht daher kein Zweifel über Einsatz und Eignung von beispielsweise Schulnoten sowie Studierfähigkeitstests als Auswahlkriterien für Bewerber um einen Studienplatz in der Zahnmedizin. Ein komplett konträres Bild bietet sich jedoch dazu bei der Frage, welche Fähigkeiten die Leistungen in den praktischen Laborkursen der Zahnmedizin beeinflussen (vgl. Ranney, Wilson & Bennett, 2005). In etlichen Studien wurde die kognitive Fähigkeit räumliches Vorstellungsvermögen als maßgebliche, die Leistung determinierende Fähigkeit identifiziert, in anderen hingegen die nichtkognitive Fähigkeit manuelles Geschick (u.a. Boyle & Santelli, 1986; Evans & Dirks, 2001; Kramer, 1986). Es finden sich in der Literatur auch Studien, in denen sogar beiden Fähigkeiten ein substantieller Einfluss auf die Aufgabenleistungen in den Laborkursen abgesprochen und resümiert wurde, dass ein Einsatz entsprechender Testverfahren zur Studierendenauswahl in der Zahnmedizin nicht vertretbar sei (z.B. Oudshoorn, 2003). Die Relevanz der Diskussion spiegelt sich schon allein in ihrer Dauer wider, deren Beginn auf die Debatte um die Validierung des zahnmedizinischen Studierfähigkeitstest „Dental Admission Test“ in den USA zurückzuführen ist, die kurz nach Einführung des Verfahrens zur Auswahl von Studienplatzbewerbern in 1951 ihren Anfang nahm.

Im Gros der Primärstudien zur Vorhersage der Laborkursleistungen wurden aber oftmals sehr unterschiedliche Definitionen und Testverfahren zur Messbarmachung der kognitiven bzw. nichtkognitiven Fähigkeit verwendet, die aber, wie auch die Überprüfung der Messung, gar nicht oder nur in sehr geringem Umfang berichtet wurden. Ebenfalls fehlte in vielen empirischen Arbeiten ein theoretisches Fundament, sodass eine wissenschaftlich fundierte Definition der Beziehungen zwischen Fähigkeiten und dem Erlernen zahnmedizinischer Untersuchungs- und Behandlungsmethoden in der Zahnmedizin ausblieb. Dadurch gestaltete sich nicht nur der Umgang mit kontroversen Studienergebnissen schwierig, sondern es führte auch zu Verschleppungen im Forschungsprozess. Außerdem erfolgte die Auswertung der Untersuchungsdaten in fast allen Studien mit statistischen Verfahren der ersten Generation in Form von Korrelations- und Regressionsanalysen. Vor allem Regressionsmodelle erlauben immer nur den gleichzeitigen Einschluss eines Kriteriums, weshalb in vergangenen Studien oftmals Mittelwerte über mehrere Laborkursaufgaben gebildet und in die Auswertung einbezogen wurden.

In der vorliegenden Arbeit werden die zuvor aufgeführten Kritikpunkte an früheren Studien zur Vorhersage der Aufgabenleistungen beim Erlernen zahnmedizinischer Untersuchungs- und Behandlungsmethoden in zwei Fragenkomplexen aufgegriffen.

Zuerst wird die Fragestellung untersucht, ob die Fähigkeiten räumliches Vorstellungsvermögen und manuelles Geschick zuverlässig operationalisierbar sind, um nachfolgend unter Anwendung von Ackermans (1988) Theorie des Fertigkeitserwerbs zu prüfen, ob und unter welchen Bedingungen für die kognitive und nichtkognitive Fähigkeit ein Einfluss auf die Aufgabenleistungen in den Laborkursen besteht. Hierfür werden die Laborkursaufgaben im vorklinischen Studienabschnitt entsprechend der Komplexität und Konsistenz klassifiziert, für die Ackerman zufolge Unterschiede in der Beziehung zwischen Fähigkeiten und dem Erlernen von Fertigkeiten bestehen. Die statistische Auswertung der Untersuchungsdaten erfolgt mit dem Verfahren der „partial least squares“ als Vertreter varianzbasierter Strukturgleichungsmodelle. Das Verfahren ist für Stichprobengrößen kleiner 50 Teilnehmer anwendbar und erlaubt die Modellierung latenter Variablen, sodass der Erwerb zahnmedizinischer Untersuchungs- und Behandlungsmethoden als Kriterium in den Prädiktionsmodellen über die einzelnen Laborkursaufgaben als Indikatoren abgebildet werden kann.

Aus den Ergebnissen der vorliegenden Arbeit sollen Erkenntnisse über die Applikation räumlicher und manueller Tests zur Bewerberselektion in der Zahnmedizin resultieren, um einen unnötigen Einsatz entsprechender Verfahren zu vermeiden. Dadurch wären nicht nur die Kosten und Ökonomie des Auswahlprozesses präziser kontrollierbar, sondern im Hinblick auf die Güte und Testethik des Verfahrens ist ebenfalls gewährleistet, dass für die Bewerberauswahl nur Leistungsergebnisse auf Basis von Fähigkeiten einbezogen werden, welche tatsächlich relevant für das Erlernen zahnmedizinischer Fertigkeiten im Studium sind. Zugleich können aus den vorliegenden Untersuchungsergebnissen auch Erkenntnisse für den diagnostischen Einsatz räumlicher und manueller Tests bei Erstsemesterstudierenden in der Zahnmedizin übertragen werden. So wäre es möglich, Studierenden vor Beginn der Laborkurse ein individuell angepasstes Training gering ausgeprägter Fähigkeiten anzubieten, die zum Erlernen der Untersuchungs- und Behandlungstechniken benötigt werden, um Schwierigkeiten bei der Aufgabenbearbeitung in den Laborkursen vorzubeugen.

## **2. Theoretische und empirische Herleitung**

Die Unterkapitel 2.1 und 2.2 geben eine kurze Übersicht zu den Studierendenauswahlverfahren, die im Unterkapitel 2.1 zur Einleitung in die Thematik allgemein gehalten und in 2.2 speziell auf die Auswahl von Zahnmedizinbewerbern fokussiert ist. Im Abschnitt 2.3 sind unterschiedliche Ansätze zur Vorhersage des praktischen Studienerfolgs im Zahnmedizinstudium aufgeführt. Im Unterkapitel 2.4 sind Ackermans Modell des Fertigkeitserwerbs zur Vorhersage individueller Leistungsunterschiede dargestellt und die empirische Prüfung der Theorie im Kontext der Studierenden- und Berufseignungsdiagnostik.

### **2.1 Studierendenauswahl**

Mit dem 7. Gesetz zur Änderung des Hochschulrahmengesetzes vom 28. August 2004 wird den Hochschulen in Deutschland die Möglichkeit eingeräumt, zusätzlich zur Abiturdurchschnittsnote weitere Kriterien zur Studierendenauswahl zu verwenden. Ziel der Gesetzesänderung ist eine Wettbewerbssteigerung zwischen den Hochschulen um die besten Studierenden, die im Studium sehr gute Prüfungsleistungen erzielen und das Studium maximal schnell absolvieren (Wedler, Troche & Rammsayer, 2008), wodurch die Auswahl von Studierenden zu einem universitärem Qualitätssicherungsinstrument hinsichtlich der Profilbildung deutscher Hochschulen wird.

#### **2.1.1 Ziele hochschuleigener Auswahlverfahren**

Ein qualitativ gutes Auswahl- und Zulassungsverfahren führt zu einem verbessertem Leistungsverhalten in Form von Noten, höherer Zufriedenheit der Studierenden, kürzeren Studienzeiten und geringeren Studienabbruchquoten, wodurch die Studienplatzkosten für Studierende und Hochschulen sinken (Schuler & Hell, 2008). In ihrem 2008 veröffentlichten Projektbericht zeigten Heublein, Schmelzer, Sommer und Wank eine bundesweite Studienabbruchquote deutscher Studierender von 21 Prozent auf, die gegenüber der vorangegangenen Erhebung von 2005 leicht zurückgegangen ist. Für die Zahnmedizin weist der Projektbericht eine Studienabbruchquote von fünf Prozent auf, die bereits in den neunziger Jahren niedrig ausgefallen war, sich aber im zeitlichen Verlauf noch weiter verringerte. Ursachen dafür sahen Heublein, Hutzsch, Schreiber, Sommer und Besuch (2010) auf Basis einer bundesweiten Befragung von Exmatrikulierten des Studienjahres 2007/08 in den Zulassungsbeschränkungen zum

Zahnmedizinstudium, einer hohen intrinsischen Studienmotivation, guten Vorstellungen von den Studieninhalten und Studienanforderungen sowie einem klaren Berufsbild.

Falls es zu einem Studienabbruch kommt, sind die Gründe in der Nichterfüllung der Studienaufgaben zu suchen, wobei es sich fast ausschließlich um das Nichtbestehen von Zwischenprüfungen, weniger von Staatsexamina handelt. Studienabbrecher sind als ehemalige Studierende definiert, die zwar durch Immatrikulation ein Erststudium an einer deutschen Hochschule aufgenommen haben, dann aber das Hochschulsystem ohne (erstes) Abschlussexamen verlassen. Fachwechsler, Hochschulwechsler wie auch erfolglose Studierende in einem Zweitstudium gehen nicht in die Berechnung der Abbruchquote ein; sie finden lediglich bei den entsprechenden Schwundquoten Berücksichtigung (Heublein et al., 2010).

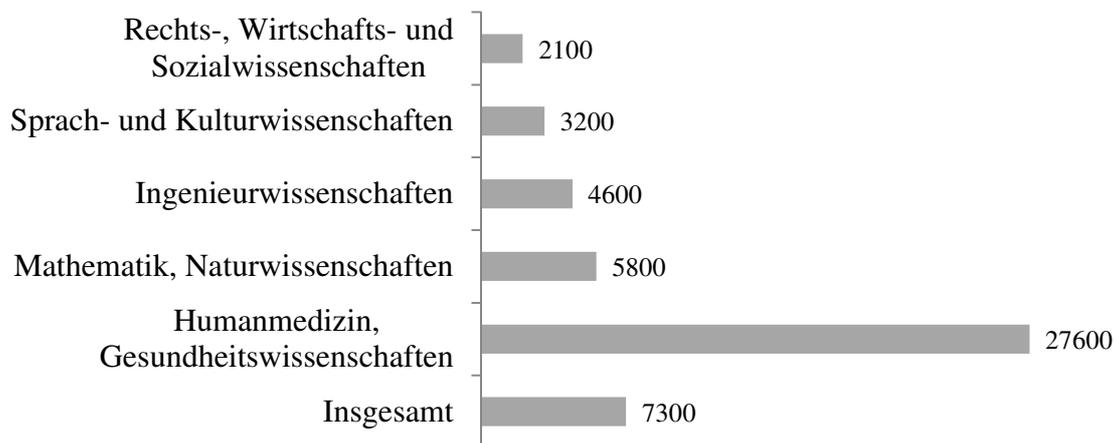


Abbildung 2.01: Jährliche Kosten in Euro pro Studierenden an deutschen Hochschulen in 2008 (Quelle: Statistisches Bundesamt, 2010)

Die Kostenverhältnisse relativieren die niedrigen Studienabbruchsquote in der Zahnmedizin und erfordern daher unter anderem hochschuleigene Studierendenauswahlverfahren als ein Qualitätssicherungsinstrument. 2008 beliefen sich die Ausgaben pro Studierenden in der Medizin, Zahnmedizin und Gesundheitswissenschaften auf jährlich 27.600 Euro (s. Abbildung 2.01). Diese fielen fast viermal so hoch aus, wie die Kosten im Durchschnitt aller Studienfächer, die 7.300 Euro betragen (Statistisches Bundesamt, 2010).

### 2.1.2 Studierfähigkeit und Studienerfolg

Die Eignung für ein Hochschulstudium wird anhand der Kriterien Studierfähigkeit und Studienerfolg festgestellt. Studierfähigkeit beruht nach Heldmann (1984, S. 184) für

jedes Studium auf grundlegenden allgemeinen Leistungskriterien und fachspezifisch notwendigen Fähigkeiten, Kenntnisse und Fertigkeiten. Leistungskriterien wie Fächerprofile sind dabei voneinander untrennbar. Konegen-Grenier (2002) spezifiziert die allgemeine Definition von Heldmann in ihrer Untersuchung mit der Einführung einer kognitiven, persönlichen, sozialen und fachlichen Dimension als Studierfähigkeitskonstrukt, welche von Hochschullehrenden hinsichtlich der Wichtigkeit für die Studierfähigkeit eingeschätzt wurde. Auf Basis der Umfrageergebnisse ihrer Studie definiert Konegen-Grenier Studierfähigkeit wie folgt:

„Analytische Fähigkeiten, Abstraktionsfähigkeit und Differenzierungsvermögen sind unabdingbare Voraussetzungen für ein Hochschulstudium. Sie müssen kombiniert sein mit guten Kenntnissen in den Fächern Englisch, Mathematik und Deutsch. Das Fachwissen ist durch Arbeitstechniken wie Präsentationsfähigkeit, Kenntnisse in der Textverarbeitung und Recherchetechniken zu ergänzen. Um seine kognitiven Fähigkeiten und sein fachliches Wissen anwenden zu können, muss der Studienanfänger eine persönliche Arbeitshaltung mitbringen, die inhaltliches Interesse mit Leistungsbereitschaft und Genauigkeit vereint.“ (Konegen-Grenier, 2002, S. 168).

Lewin und Lischka (2004) schlagen für die Definition der Studierfähigkeit einen Paradigmenwechsel hin zum Begriff Passfähigkeit vor, die zu verstehen ist „[...] als möglichst hohe Übereinstimmung individueller Kompetenzen der StudienanfängerInnen mit den grundlegenden und spezifischen Anforderungen eines Studiums, differenziert nach Inhalt und Profil.“ (Lewin & Lischka, 2004, S. 87). Besonders an ihrem Ansatz ist, dass sich Studierende und Hochschule aneinander anpassen, anstatt sich jeweils einseitig auf die Anforderungen oder Voraussetzungen des anderen anzupassen.

Im allgemeinen Sprachgebrauch ist Studienerfolg als das Erreichen des Studienziels definiert, welches das Abschließen eines einmal begonnenen Studiums mit dem Hauptexamen beinhaltet (Schuler & Hell, 2008). Die Operationalisierung des Studienerfolgs ist dabei recht vielfältig und reicht von Noten in Zwischen- oder Abschlussprüfungen, Noten in einzelnen Lehrveranstaltungen, Studiendauer, Prüfungswiederholungen, Beurteilung des Studienerfolgs durch Dozenten oder Kommilitonen bis hin zur Zufriedenheit mit dem Studium. Viele empirische Studien verwenden zur Validierung von Auswahlverfahren das Erfolgskriterium Studiennoten, da sie eine gute Quantifizierung der Studienleistung auf einer definierten Skala darstellen und zugleich ökonomisch zu erheben sind (Hell, Trapmann & Schuler, 2007).

### **2.1.3 Vorhersage des Studienerfolgs**

Schuler (2007, S. 291 ff.) übertrug den von ihm entwickelten trimodalen Ansatz der Berufseignungsdiagnostik auf die Studierendenauswahl (s. Tabelle 2.01) und bietet

damit ein Konzept an, welches unterschiedliche Verfahren aus der psychologischen Diagnostik zur Vorhersage der Studierfähigkeit und des Studienerfolgs berücksichtigt. Verfahrensspezifische Messfehler können somit besser kontrolliert werden, sodass eine gesteigerte soziale Akzeptanz sowie Fairness für die Auswahlverfahren zu erwarten sind.

Tabelle 2.01: Trimodaler Ansatz der Berufseignungsdiagnostik in der Studierendenauswahl

Diagnostische Zugänge	Generelle Studieneignung	Studienfachspezifische Eignung
Eigenschaftsorientierte Verfahren	<ul style="list-style-type: none"> <li>• Intelligenztests (Allgemeine Intelligenz)</li> <li>• Wissenstests (Allgemeinwissen)</li> <li>• Persönlichkeitstests (z.B. Gewissenhaftigkeit, Leistungsmotivation)</li> <li>• Arbeitshaltungen (generelles Interesse an wissenschaftlichen Inhalten und akademischer Arbeitsweise)</li> </ul>	<ul style="list-style-type: none"> <li>• Spezifische Fähigkeitstests</li> <li>• Fachwissenstests</li> <li>• Persönlichkeitstests (spezifische Faktoren)</li> <li>• Interessentests (fachspezifische Interessen)</li> <li>• Selbsteinschätzung Fähigkeiten</li> </ul>
Biografieorientierte Verfahren	<ul style="list-style-type: none"> <li>• Schulnoten</li> <li>• Interviews</li> <li>• Biografische Fragebogen</li> </ul>	<ul style="list-style-type: none"> <li>• Schulnoten in einschlägigen Fächern</li> <li>• Interviews</li> <li>• Arbeits- und Interessenbiografie</li> </ul>
Simulationsorientierte Verfahren	<ul style="list-style-type: none"> <li>• Simulationen allgemeiner Studiensituationen (z.B. Präsentationen, Gruppendiskussionen, Arbeitsplanung)</li> <li>• Studientypische Konflikte und Belastungssituationen</li> </ul>	<ul style="list-style-type: none"> <li>• Simulationen fachspezifischer Studiensituationen (z.B. Präsentationen, Gruppendiskussionen, Rollenspiele)</li> <li>• Situational Judgment Test</li> </ul>

Anm.: Darstellung nach Schuler und Hell (2008)

Schulers Konzept des trimodalen Ansatzes beinhaltet neben dem biografischen Ansatz, der zur Erhebung vergangenen Verhaltens und Leistungen dient, eigenschaftsorientierte Verfahren, in welchen persönliche Potentiale gemessen werden und den Simulationsansatz, der den Fokus auf eher kriterienbezogene Kompetenzen ausrichtet. Die Entscheidung über den Einsatz eines Verfahrens oder einer Kombination mehrerer Verfahren, zum Beispiel zur Vorhersage des Studienerfolgs in Form von Prüfungsergebnissen, bedingt sich durch deren Validitäten, Ökonomie, Akzeptabilität und Fairness (Schuler & Hell, 2008). Eine Auswahl der am häufigsten eingesetzten Verfahren und Kriterien als Prädiktoren des Studienerfolgs in der Zahnmedizin ist im nachfolgenden Abschnitt 2.2 dargestellt. Zuvor wird für eine bessere Verständlichkeit der Thematik der Ablauf des Zahnmedizinstudiums im Unterkapitel 2.2.1 berichtet.

## **2.2 Vorhersage des Studienerfolgs in der Zahnmedizin**

Der Ablauf des zahnmedizinischen Studiums wird zunächst in Unterkapitel 2.2.1 skizziert, um die in den nachfolgenden Unterkapiteln berichteten Auswahlinstrumente und -kriterien zur Vorhersage des Studienerfolgs in der Zahnmedizin zuverlässig beurteilen zu können. Ein Resümee findet sich im Unterkapitel 2.2.8.

### **2.2.1 Ablauf des Zahnmedizinstudiums**

Der Aufbau des zahnmedizinischen Studiums in Deutschland ist maßgeblich über die Approbationsordnung für Zahnärzte (ZÄPrO; Verlag, 2013) von 1955 geregelt und mit anderen internationalen Studiengängen vergleichbar. An der medizinischen Fakultät der Universität Hamburg kann das Zahnmedizinstudium nur zum Wintersemester eines Jahres begonnen werden. Das Studium ist in einen fünfsemestrigen vorklinischen und fünfsemestrigen klinischen Abschnitt unterteilt.

Im vorklinischen Studienabschnitt werden in der theoretischen Ausbildung bis zur Naturwissenschaftlichen Vorprüfung nach dem zweiten Fachsemester die Fächer Biologie, Chemie, Physik und bis zur Zahnärztlichen Vorprüfung nach dem fünften Semester die Fächer Anatomie, Biochemie, Physiologie und Terminologie unterrichtet. Die praktische Ausbildung des vorklinischen Studienabschnitts beinhaltet die Zahnersatzkunde, die sich aus dem Kursus der technischen Propädeutik, dem Phantomkurs der Zahnersatzkunde I und Phantomkurs der Zahnersatzkunde II zusammensetzt, die ebenfalls theoretische Ausbildungselemente wie Vorlesungen oder Seminare in den Lehrplan implementieren.

Im klinischen Abschnitt schließen sich Kurse und Vorlesungen in den Fächern Zahnerhaltungskunde, Zahnersatzkunde, Kieferorthopädie, Zahnärztliche Radiologie, Zahn-, Mund-, Kieferheilkunde, Allgemeine Pathologie, Pharmakologie, Hygiene/Medizinische Mikrobiologie, Innere Medizin, Chirurgie, Dermatologie, Hals-, Nasen-, Ohrenheilkunde an. Nach dem zehnten Semester erfolgt die Zahnärztliche Prüfung als Staatsexamen zum Abschluss des klinischen Studienabschnitts, mit der die Approbation als Zahnarzt beantragt werden kann.

### **2.2.2 Schulabschlussnoten**

Die Metaanalyse von Trapmann, Hell, Weigand und Schuler (2007) beinhaltet für den Studiengang Zahnmedizin insgesamt elf seit 1986 in Deutschland publizierte Studien zur Vorhersage des zahnmedizinischen Studienerfolgs und ergab eine mittlere,

gewichtete Validität von  $\rho = 0.375$  (reliabilitäts- und selektionskorrigiert) für das Studienerfolgskriterium Abiturdurchschnittsnote im Studiengang Zahnmedizin, weshalb die Autoren den Einsatz bei der Auswahl von Studienplatzbewerbern empfehlen. Arnold, Gonzalez und Gaengler (2011) berichteten schwach positive, signifikante Korrelationen in ihrer Primäruntersuchung zwischen Abiturdurchschnittsnote und der Naturwissenschaftlichen Vorprüfung ( $r = 0.343$ ,  $p < .001$ ), der Zahnärztlichen Vorprüfung ( $r = 0.268$ ,  $p < .001$ ) und dem Staatsexamen ( $r = 0.269$ ,  $p < .001$ ), die eine Eignung des Kriteriums zur Studierendenauswahl in der Zahnmedizin befürworteten. Mittels Regressionsanalyse konnten die Autoren zeigen, dass zwölf Prozent der Leistungsvarianz in der Naturwissenschaftlichen Vorprüfung durch die Abiturdurchschnittsnote aufgeklärt wird, jeweils sieben Prozent für die Zahnmedizinische Vorprüfung und das Staatsexamen. Die Ergebnisse der aus Deutschland stammenden Studie deuten auf einen Zusammenhang zwischen der Abiturdurchschnittsnote als Prädiktor und dem mittels Prüfungsnoten operationalisierten Studienerfolg in der Zahnmedizin hin.

Frühere Studien kamen nicht nur zu vergleichbaren Schlussfolgerungen, sondern berichteten ähnliche Ergebnisse im europäischen und internationalen Ausland. Lynch, McConnell und Hannigan (2006) konnten zwar nicht die Examensergebnisse der staatlichen Abschlussprüfung mittels der durchschnittlichen Schulabschlussnote in Irland vorhersagen, aber für die erste zahnmedizinische Prüfung, die theoretische Kenntnisse in Anatomie, Physiologie und Biochemie nach dem ersten Studienjahr prüft, wurden zwölf Prozent Varianzanteil der Prüfungsleistung durch die Ergebnisse der Schulabschlussnote erklärt. In den USA und in Kanada werden Schulabschlussnoten in Form des „grade point average“ (GPA) für die Zulassung zum Zahnmedizinestudium eingesetzt. Der GPA verfügt über die höchste Prädiktionsstärke des akademischen Studienerfolgs in der Zahnmedizin, wie in zahlreichen Primärstudien nachgewiesen wurde. Holmes, Doering und Spector (2008) berichteten für den GPA eine moderat positive, signifikante Korrelation in Beziehung zum Studiennotendurchschnitt ( $r = 0.53$ ,  $p < .05$ ), die vergleichbar zu den Untersuchungen von Curtis, Lind, Plesh und Finzen (2007), Downey, Collins und Browning (2002), Kreit und McDonald (1968), Sandow, Jones, Peek, Courts und Watson (2002) und Ward, Downey, Thompson und Collins (2010) ausfielen.

Schulnoten stellen zwar den besten Einzelprädiktor zur Vorhersage des Studienerfolgs in Form von Studiennoten dar (Brandstätter & Farthofer, 2002; Zimmerhofer & Trost, 2008), jedoch erweist sich die überwiegende Fokussierung auf kognitive Fähigkeiten als Nachteil, der zu einer Einschränkung prognostischer Validitäten führt, da auch Leistungsmotivation, Selbstwirksamkeitsüberzeugungen, biographische und persönlichkeitsbezogene Aspekte bei der Prädiktion des Studienerfolgs einen Beitrag leisten (Oswald, Schmitt, Kim, Ramsay & Gillespie, 2004; Robbins, Lauver, Le, Davis, Langley & Carlstrom, 2004).

### **2.2.3 Studierfähigkeitstests**

Studierfähigkeitstest sind nach Zimmerhofer und Trost (2008) eignungsdiagnostische Testverfahren zur Messung intellektueller Fähigkeiten, die für ein erfolgreiches Studium bedeutsam und somit zur Auswahl von Studienbewerbern heranzuziehen sind. Hierbei kann eine Unterteilung in zwei Gruppen vorgenommen werden. Allgemeine Studierfähigkeitstests erfassen kognitive Fähigkeiten, die für alle Studiengänge wichtig sind. Fachspezifische Studierfähigkeitstests messen hingegen spezifische kognitive Fähigkeiten, die nur für ein bestimmtes Studienfach erforderlich sind.

#### **2.2.3.1 Test für medizinische Studiengänge**

Der Test für medizinische Studiengänge (TMS) ist der bisher im deutschsprachigen Raum am gründlichsten erforschte fachspezifische Studierfähigkeitstest. Der TMS wurde im Rahmen des besonderen Auswahlverfahrens zum Wintersemester 1986/87 eingeführt und war bis zum Wintersemester 1997/98 fester Bestandteil aller Bewerbungen um einen Studienplatz in der Medizin, Zahnmedizin und Pharmazie. Am TMS haben bisher 300.727 Personen teilgenommen (Zimmerhofer & Trost, 2008). Für den TMS ermittelten Hell et al. (2007) in ihrer Metaanalyse eine prognostische Validität des Studienerfolgs in der Zahnmedizin von  $\rho = 0.353$ , in welche sie acht Primärstudien einschlossen. Trapmann et al. (2007) zeigten in ihrer dazu inhaltsgleichen Metaanalyse, dass die Abiturdurchschnittsnote im Vergleich zum TMS ein stärkerer Prädiktor des zahnmedizinischen Studienerfolgs ist ( $\rho = 0.375$ ). Eine Kombination von Schulnoten und Studierfähigkeitstest in einem Auswahlverfahren ist sinnvoll, wenn beide Prädiktoren gering miteinander korrelieren und gute inkrementelle Validitäten zu erwarten sind (Hell et al., 2007).

### 2.2.3.2 Besonderer Auswahltest Zahnmedizin

Der in Deutschland als Vorläufer zum TMS eingesetzte fachspezifische Studierfähigkeitstest Besonderer Auswahltest Zahnmedizin (BATZ) stellt eine Testbatterie zur Diagnose der Studierfähigkeit im Studiengang Zahnmedizin dar. Bereits die Validierung der BATZ von Hitpass (1978) unterstreicht den Einsatz mehrerer Fähigkeiten zur Auswahl von Zahnmedizinstudierenden, zu denen unter anderem das räumliche Vorstellungsvermögen, Wahrnehmungsprozesse oder manuelles Geschick zählten (s. Tabelle 2.02). Für die aus 792 Studierenden der Zahnmedizin sowie aus Studienbewerbern der Psychologie und Zahnmedizin bestehende Stichprobe ermittelte Hitpass eine Retest-Reliabilität der BATZ-Gesamttestbatterie von  $r = 0.96$  und Korrelationen zwischen den Testaufgaben, die auf gute diskriminante Validitäten der zugrundeliegenden Fähigkeiten schließen lassen. Beispielsweise finden sich zwischen der Drahtbiegeaufgabe und den räumliches Vorstellungsvermögen erfassenden Testaufgaben Flächenabwicklungen ( $r = 0.28$ ), Würfelaufgaben ( $r = 0.19$ ) und Bauprobe ( $r = 0.24$ ) positiv schwache Korrelationen.

Tabelle 1.02: Testaufgaben der BATZ und erfasste Merkmale (Quelle: Hitpass, 1978)

Testaufgabe	erfasste Merkmale
Drei Stäbchen	Fingerfein-, Arm- und Handgeschick
Pinzette und Schraubenzieher	
Spiegelprobe	Geschwindigkeit, Genauigkeit, Bewegungskoordination, Gelenkigkeit
Tremometer	
Sinusoide	
Gurnee-S-Test	
Formenlegen	
Flächenabwicklungen	Fähigkeit, sich in neuen Situationen praktischer Art zu bewähren
Würfelaufgaben	
Bauprobe	
Dynamometer	Fähigkeit, eine bestimmte Muskelkraftaufwendung vorzunehmen, zu behalten und zu reproduzieren
Reaktion	Fähigkeit, auf einen visuellen Reiz rasch zu antworten
Drahtbiegeprobe	Fähigkeit, aus Draht oder Gips bestehendes Rohmaterial mittels Finger-, Arm- und Handgeschick sowie der Bewegungskoordination gemäß einer Vorlage formgerecht zu bearbeiten
Schnitzprobe	
Pseudo-isochrom. Tafeln	Visuelle Fähigkeit, einfarbige Zahlen- oder Buchstabenkonfigurationen aus verwirrender Farbenvielfalt zu erkennen
Orthorater	Fähigkeit des Gesichtssinnes zur Wahrnehmung von Gestalten im Nah-, Fern- und Tiefenbereich

In der von Barkowski (1986) vorgenommenen Längsschnittvalidierung der BATZ wurden die Abiturdurchschnittsnote und das BATZ-Gesamtergebnis als Prädiktoren des Studienerfolgs in der Zahnmedizin miteinander verglichen. Die Staatsexamensnote wurde besser durch den BATZ ( $r = 0.51$ ) vorhergesagt als durch die Abiturdurchschnittsnote ( $r = 0.26$ ), jedoch konnten die Naturwissenschaftlichen Vorprüfungsergebnisse nach dem zweiten Semester besser durch die Abiturdurchschnittsnote ( $r = 0.33$ ) als durch den BATZ ( $r = 0.29$ ) vorhergesagt werden, was sich aber nicht für die regulär nach dem fünften Fachsemester stattfindende Zahnärztliche Vorprüfung bestätigen ließ. Hier wies der BATZ eine größere positive Korrelation in Höhe von  $r = 0.31$  im Vergleich zur Abiturdurchschnittsnote mit  $r = 0.25$  auf. Barkowski führte als Erklärung an, dass die von der BATZ gemessenen Fähigkeiten erst im Verlauf des Zahnmedizinstudiums zum Tragen kommen. Für die Längsschnittvalidierung konnte Barkowski 438 der 792 Probanden der Stichprobe aus 1978 gewinnen. Neben der Testökonomie der BATZ, die circa sechs Stunden dauert, kritisiert vor allem Pawlik (1979) die Validierung von Hitpass (1978) in den Punkten Versuchsdesign, Prädiktorenauswahl und Interventionsstrategie.

### **2.2.3.3 Dental Admission Test**

Der 1951 von der American Dental Association (ADA) eingeführte „Dental Admission Test“ (DAT) besteht gegenwärtig in den USA aus einem Naturwissenschaftstest, der Kenntnisse in Biologie, Chemie und organischer Chemie abfragt, einem Test räumlich-perzeptiver Fähigkeiten, dem „Perceptual Ability Test“ (PAT), einem Lese- und Verständnistest sowie einen Test zum schlussfolgernden Denken. Die Auswertung erfolgt in acht Kategorien, beispielsweise für die akademische Durchschnittsleistung „Academic Average“, die aus den Testergebnissen zum logischen Schlussfolgern, Leseverständnis und den naturwissenschaftlichen Kenntnissen in Biologie und Chemie gemittelt wird. Der fachspezifische Studierfähigkeitstest ist aufgrund seiner zuverlässigen Prädiktion des akademischen Studienerfolgs ein etabliertes Instrument zur Auswahl von Zahnmedizinstudierenden, die aber im Vergleich zum GPA schwächer ausfällt (Holmes et al., 2008). In Kanada wird der DAT als „Dental Aptitude Test“ bezeichnet und ist bis auf den Subtest „Carving Dexterity“ inhaltsgleich zum Pendant aus den USA. Eine ausführliche Darstellung des DAT findet sich unter anderem bei Boyd und Teteruck (1979) oder Kramer (1986).

#### 2.2.4 Interviews

Ein Auswahlgespräch bzw. Interview definiert Trost (1986) als gelenktes Gespräch zum Zwecke der Gewinnung von Informationen über objektive, z.B. biographische Angaben, Qualifikationen und subjektive Daten, wie persönliche Meinungen. In der deutschen Hochschullandschaft haben Interviews mit Ausnahmen in Auswahlverfahren medizinischer Studiengänge und privater Hochschulen in der Vergangenheit nur geringe Anwendung gefunden, hingegen sind sie Auswahlverfahren ausländischer Hochschulen etabliert (Gentsch & Gold, 2008). Kay, Bennett, Allison und Coombes (2010) fanden für strukturierte Interviews positiv schwache, nicht-signifikante Korrelationen im Bereich von  $0.007 \leq r \leq 0.036$  zur zahnmedizinischen Studienleistung in Australien. Zur Operationalisierung der Studienleistung wurden nach dem ersten Studienjahr zwei Progress-Tests, die Verständnis und die Fähigkeit zum Anwenden von Wissen erfassen und zwei vorklinische Wissenstests verwendet. In ihrer Langzeitstudie mit einer Stichprobe von 373 kanadischen Zahnmedizinistudierenden des ersten bis fünften Ausbildungsjahres zeigten Poole, Catano und Cunningham (2007), dass strukturierte Interviews besser die klinische Leistung im dritten und vierten Ausbildungsjahr vorhersagen als der kanadische Studierfähigkeitstest DAT. Ursache dafür könnte die im klinischen Studienabschnitt beginnende Interaktion zwischen Studierenden und Patienten sein, was bereits Walker, Killip und Fuller in ihrer Untersuchung 1985 vermuteten. Poole et al. (2007) empfahlen zur Auswahl von Zahnmedizinistudierenden eine Kombination von strukturierten Interviews und dem DAT (vgl. Smithers, Catano & Cunningham, 2004).

#### 2.2.5 Nichtkognitive Merkmale

Die Berücksichtigung nichtkognitiver Merkmale, zum Beispiel persönliche Interessen, Studienmotivation oder Einstellungen zum Studium, ist bei der Studierendenauswahl bisher im deutschsprachigen Raum nur selten erfolgt. Gallagher, Niven, Donaldson und Wilson (2009) zeigten in ihrer Studie aus Großbritannien, dass mittels nichtkognitiver Merkmale, wie höherer sozialer Status und ethnische Minoritätszugehörigkeit, erfolgreiche Studierende in der zahnmedizinischen Vorklinik identifiziert werden konnten. Darüber hinaus fanden sie ebenfalls Hinweise für die Variablen Alter, Geschlecht, Ethnizität und Schultyp, die mit einer erfolgreichen Studienplatzbewerbung in der Zahnmedizin assoziiert zu sein schienen. Geschlecht und Ethnizität konnten durch eine logistische Regressionsanalyse der zahnmedizinischen Bewerbungs- und

Studiendaten von 1994 bis 1998 in Großbritannien als soziodemographische Variablen zur Charakterisierung erfolgreicher Bewerber von Newton, Buck, Thorogood und Skelly (2003) identifiziert werden. Gewissenhaftigkeit, Verträglichkeit und Neurotizismus als Persönlichkeitsfaktoren der „Big Five“ erwiesen sich als signifikante Prädiktoren zur Vorhersage des akademischen und praktischen Studienerfolgs in der zahnmedizinischen Vorklinik, womit Chamberlain, Catano und Cunningham (2005) ihre Forderung nach einer Erweiterung der Auswahlkriterien um Persönlichkeitsfaktoren untermauerten. Hingegen konnten Fenlon, McCartan, Sheriff und Newton (2001) mit 302 Studierenden aus Irland und Großbritannien keine Persönlichkeitsfaktoren zur Vorhersage der Leistung in der zahnmedizinischen Vorklinik finden. 1967 analysierten Chen und Podshadley berufliche Interessen, manuelle Fähigkeiten, Persönlichkeitseigenschaften, kognitive und psychologische Variablen zur Vorhersage des Erfolgs im zahnmedizinischen Studium. Die Schulabschlussnote bestätigte sich als bester Prädiktor, die mit dem „California Psychological Inventory“ gemessenen Persönlichkeitseigenschaften erwiesen sich hingegen als wenig geeignete Prädiktoren.

### **2.2.6 Situational Judgment Tests**

Der „Situational Judgment Test“ (SJT) bzw. das „Situational Judgment Inventory“ (SJI) gehören zur Kategorie der simulationsorientierten Verfahren in der Berufseignungsdiagnostik und werden primär zur Personalauswahl eingesetzt (Pixner & Schübach, 2008). Der in Belgien von Buyse und Lievens (2011) zur Studierendenauswahl in der Zahnmedizin eingesetzte videobasierte „Situational Judgment Test“ (SJT) beinhaltete Szenen, die Interaktionen zwischen Ärzten bzw. Zahnärzten und Patienten darstellten, um interpersonale und kommunikative Fertigkeiten der Bewerber zu erfassen. Nach jeder kritischen Situation wurde das Bild eingefroren, sodass die 796 Bewerber um einen Studienplatz in den Jahren von 1997 bis 2008 aus vier von einem Erzähler vorgetragenen Antwortmöglichkeiten die Antwort auswählen konnten, welche ihrer Ansicht nach die effizienteste Reaktion in der Situation bedeutete. Die Auswertung der Untersuchungsdaten ergab nur bei der Vorhersage der Leistung im fünften Studienjahr für den SJT eine inkrementelle Validität von drei Prozent in Kombination mit der Schulabschlussnote und einem papierbasierten Multiple-Choice-Test zum Verständnis (zahn)medizinischer Texte, der ebenfalls im Auswahlverfahren eingesetzt wurde. Als Erklärung führen die Autoren den

Umgang mit Patienten und die dafür notwendigen zwischenmenschlichen sowie kommunikativen Fertigkeiten an.

Der Einsatz von SJTs als Auswahlinstrument in der Zahnmedizin zur Gewinnung weiterer, nichtkognitiver Prädiktoren ist nach Ansicht von Lievens, Buyse und Sackett (2005) sowie von Lievens und Coetsier (2002) empfohlen, da ihr inkrementeller Beitrag in Kombination mit kognitiven Tests zu einer erhöhten prädiktiven Validität des Zulassungsverfahrens führt. Eine ausführliche Darstellung zur Entwicklung, Historie und Validierung von SJTs findet sich im Review von Lievens, Peeters und Schollaert (2008).

### **2.2.7 Universitäre Auswahlverfahren**

Heintze, Radeborg, Bengtsson und Stenlås (2004) evaluierten das in den 1990er Jahren entwickelte und seit 2001 in Schweden eingesetzte Auswahlverfahren für Studienbewerber in der Zahnmedizin. Es beinhaltete neben Testverfahren zur Messung der allgemeinen Intelligenz, des räumlichen Vorstellungsvermögens, dem manuellem Geschick, Empathie und sozialer Kompetenz auch ein halbstrukturiertes Interview. Die Stichprobe bestand aus insgesamt 191 bereits zugelassenen Studierenden in den Jahren von 1992 bis 1997, von denen 108 nicht alle Testverfahren absolvierten. Die Auswertung der Studienergebnisse ergab negativ schwache, signifikante Korrelationen zwischen der Anzahl der nichtbestandenen Kursprüfungen im Studium und den Punktzahlen der Auswahltests. Zur Leistung des im dritten Semester stattfindenden Laborkurses der Zahn- und Kieferchirurgie berichteten Heintze et al. positiv schwache und signifikante Zusammenhänge für die Tests zum räumlichen Vorstellungsvermögen (Test 1:  $r = 0.24$ ,  $p < .01$ ,  $n = 131$ ; Test 2:  $r = .27$ ,  $p < .01$ ,  $n = 121$ ), in denen Objekte rotiert und Faltaufgaben richtig gelöst werden mussten. Vergleichbar stark fiel der Zusammenhang zu den praktischen Laborkursnoten für den Test zum manuellen Geschick aus ( $r = 0.23$ , *n.s.*), der sich aber als nicht-signifikant erwies, was die Autoren auf die geringe Stichprobengröße ( $n = 50$ ) zurückführten. Der manuelle Test erforderte eine in einem Modell vorgegebene Zahnvertiefung in Körper aus Plexiglas zu schnitzen, die von zwei Ratern auf einer dreistufigen Skala bewertet wurden. Die Laborkursleistung wurde über eine fünfstufige Notenskala operationalisiert, in der fünf die beste Note darstellt und eins die schlechteste.

An der Medizinischen Fakultät der Innsbrucker Universität in Österreich wird seit 2000 ein zweitägiges Zulassungsverfahren zur Auswahl von Zahnmedizinistudierenden

durchgeführt, welches aus einem theoretischen Multiple-Choice Test mit 60 Fragen und einem praktischen Teil besteht (Beier, Kapferer, Ostermann, Staudinger & Dumfahrt, 2010). Der praktische Teil setzt sich aus einem Wachstumstest, einem Drahtbiegetest, dem Polieren eines Acrylblockes, dem Ausfräsen einer vorgegebenen Linie mittels Dentalbohrer und dem Bohren von Löchern zusammen. Zwischen der Studienleistung in Form von Noten im ersten Studienjahr und dem Gesamtttestwert des Auswahlverfahrens fanden die Autoren eine negativ moderate und signifikante Korrelation ( $r = -0.46$ ,  $p < .01$ ), die für den theoretischen ( $r = -0.39$ ,  $p < .01$ ) und praktischen Testabschnitt ( $r = -0.37$ ,  $p < .01$ ) beinahe gleich stark ausfielen. Die Stichprobe umfasste insgesamt 97 Erstsemesterstudierende der Jahre 2001 bis 2005.

Das seit 1993 eingesetzte zahnmedizinische Auswahlverfahren an der Universität Witten/Herdecke in Deutschland besteht aus einem halbstündigen, nicht-strukturierten Interview, in welchem die Bewerber in drei Klassifikationen eingestuft werden und einen praktischen Test (Arnold et al., 2011). Der praktische Test beinhaltet das Erstellen einer dreidimensionalen Zeichnung eines Objekts. Die Zeichnung diente zugleich als Vorlage, um das gezeichnete Objekt aus Wachs zu schnitzen. Abschließend muss ein in der Kieferorthopädie verwendeter Draht nach vorgegebenem Muster gebogen werden. Die Stichprobe bestand aus 196 über das Auswahlverfahren in den Jahren 1994 bis 2001 zugelassenen Studierenden. Zwischen der Interviewleistung und den Prüfungsergebnissen der Naturwissenschaftlichen und der Zahnärztlichen Vorprüfung sowie dem Staatsexamen können Arnold et al. einen positiven Zusammenhang nicht eindeutig nachweisen. Für den mit Schulnoten bewerteten praktischen Test berichteten die Autoren eine schwach positive und signifikante Korrelation zur praktischen Note in der Zahnärztlichen Vorprüfung ( $r = 0.20$ ,  $p < .05$ ).

### **2.2.8 Zusammenfassung und Resümee**

Die in dem Unterkapitel vorgestellten Instrumente verdeutlichen die gelungene Transformation eignungsdiagnostischer Konzepte und Verfahren in das Feld der zahnmedizinischen Studierendenauswahl. Die Vorhersage des Studienerfolgs ist hierbei ausschlaggebend für die Evaluation der zur Studierendenauswahl eingesetzten Instrumente. Bei der Operationalisierung des Studienerfolgs besteht im Studiengang der Zahnmedizin die Besonderheit der Unterteilung in einen theoretischen und praktischen Bereich. Für die Vorhersage des theoretischen bzw. akademischen Studienerfolgs, der naturwissenschaftliches und (zahn)medizinisches Wissen widerspiegelt, bewähren sich

die auch in anderen Studiengängen etablierten Auswahlkriterien, wie die Schulabschlussnote. Die Metaanalyse von Trapmann et al. (2007) und auch die spätere Primärstudie von Arnold et al. (2011) zeigen die Eignung der Abiturdurchschnittsnote als Auswahl von Zahnmedizinistudierenden in Deutschland auf, indem moderate prognostische Beziehung zur akademischen Studienleistung in der Zahnmedizin nachgewiesen wurde, die sich auch in europäischen und nordamerikanischen Primärstudien zeigten (z.B. Curtis et al., 2007; Downey et al., 2002; Poole et al., 2007). Als weitere zuverlässige Prädiktoren der akademischen Studienleistung erwiesen sich fachspezifische Studierfähigkeitstests, wie aus den Validierungsstudien von Trost et al. (1997) zum TMS oder Kramer (1986) zum nordamerikanischen DAT hervorgeht. Die prognostischen Beziehungen der Studierfähigkeitstests zur akademischen Studienleistung in der Zahnmedizin fielen ebenfalls überwiegend moderat aus, besitzen aber im direkten Vergleich zur Schulabschlussnote eine schwächere, prädiktive Stärke. Interviews, sozio-demographische Faktoren und SJTs bzw. Persönlichkeitstests sind als Prädiktoren der akademischen Studienleistung in der Vorklinik umstritten. Einen Einsatz empfahlen Hell et al. (2007) in ihrer Synopse von Metaanalysen zur Prognostizierbarkeit der Studienleistung eher im Hinblick auf ihren inkrementellen Beitrag in Kombination mit etablierten Kriterien, wie der Abiturdurchschnittsnote oder Studierfähigkeitstests, um zu einer erhöhten prädiktiven Validität des Zulassungsverfahrens zu führen.

Vergleicht man den aktuellen TMS mit der Vorläuferversion BATZ, fällt das Fehlen von Subtests auf, die psychomotorische Fähigkeiten auf Basis manueller Tätigkeiten, wie einer Schnitz- oder Drahtbiegeprobe, erfassen. Universitäre Auswahlverfahren zahnmedizinischer Fakultäten in Schweden (Heintze et al., 2004), Österreich (Beier et al., 2010) oder Deutschland (Arnold et al., 2011) verwenden zwar entsprechende Instrumente, jedoch fehlt in den zugehörigen Publikationen eine ausführliche und theoretisch fundierte Analyse ihrer Beziehung zum praktischen Studienerfolg in der Zahnmedizin. Eine ökonomisch sinnvolle und inhaltlich zu rechtfertigende Anwendung eignungsdiagnostischer Instrumente zur Selektion von Studienbewerbern erfordert jedoch eine grundlegende Auseinandersetzung der Beziehung zwischen Prädiktor und Kriterium, weshalb im nachfolgenden Unterkapitel 2.3 frühere und aktuelle Studienergebnisse in diesem Forschungsfeld berichtet werden.

## **2.3 Vorhersage der praktischen Studienleistung in der Zahnmedizin**

Im Unterkapitel 2.3.1 werden zunächst die Begriffe Fähigkeit und Fertigkeit grundlegend definiert sowie einzelne Fähigkeiten beschrieben, deren Kenntnis zur Beurteilung der in den Unterkapiteln 2.3.2 bis 2.3.6 dargestellten Studien förderlich ist. Die ersten Studien zur Vorhersage des praktischen Studienerfolgs finden sich in Unterkapitel 2.3.2. In der Literatur findet sich eine Vielzahl von Studien zur Vorhersage des praktischen Studienerfolgs in der Zahnmedizin, deren Untersuchungsergebnisse aber in der Mehrheit nicht in ein theoretisches System eingebunden sind, welches durch die empirischen Beiträge eine Weiterentwicklung erfährt. Aufgrund dessen orientiert sich die Darstellung der Studien an den thematisch unterschiedlichen Forschungssätzen. Die im Kontext der Validierung des amerikanischen Studierfähigkeitstest DAT kontrovers diskutierten Fähigkeiten zur Prädiktion der Leistung in den praktischen Studienkursen sind im Unterkapitel 2.3.3 aufgeführt. Im darauf folgenden Unterkapitel 2.3.4 werden Untersuchungen berichtet, deren Fokus auf die Fähigkeiten als Prädiktoren der praktischen Leistungen konzentriert ist und nicht einem Verfahren wie dem DAT. Die Betrachtung der Studien in diesem Unterkapitel fällt umfassender aus, um die Vergleichbarkeit und Beurteilung der Studienergebnisse aus mehr als 60 Jahren Forschung zu gewährleisten. Kritik an den Forschungsmethoden sowie Ansätze, diese zu kompensieren, sind in Unterkapitel 2.3.5 beschrieben. Im vorletzten Unterkapitel 2.3.6 ist der Erwerb motorischer Fertigkeiten in den praktischen Kursen als Erklärungsansatz zur Prädiktion des praktischen Studienerfolgs aufgeführt. Abschließend finden sich in Unterkapitel 2.3.7 eine Kapitelzusammenfassung und ein Resümee.

### **2.3.1 Definitionen**

#### **2.3.1.1 Fähigkeit**

Fleishman (1972, S. 1018) bezieht sich in seiner Definition von Fähigkeit mehr auf eine allgemeine Eigenschaft eines Individuums und führt genauer aus: „Many of these abilities are, of course, a product of learning and develop at different rates, mainly during childhood and adolescence. They represent organismic factors that the individual brings with him when he begins to learn a new task.“. Damit betont Fleishman die festgelegte und statische Natur von Fähigkeiten, die sich mit Beendigung des Lernens und der Entwicklung nach Abschluss der Adoleszenz verfestigt haben. Van der Maas,

Molenaar, Maris, Kievit und Borsboom (2011) beschreiben Fähigkeit als Kapazität, etwas zu tun. Verfügt man über eine Fähigkeit, die allein zur Lösung einer Aufgabe notwendig ist, so kann diese Aufgabe bewältigt werden, insofern ausreichend Zeit zur Verfügung steht.

### **2.3.1.2 Fertigkeit**

Fertigkeiten werden hingegen als mehr spezifisch und eingegrenzter definiert, verglichen mit Fähigkeiten (Henry & Hulin, 1989). Fleishman (1972, S. 1018) beschreibt Fertigkeit als „The term *skill* refers to the level of proficiency on a specific task.“ und führt weiter aus: “The assumption is that the skills involved in complex activities can be described in terms of the more basic abilities.”. Hoyos (1974, S. 143) stellt Fähigkeiten in den Vergleich zu Fertigkeiten, „die stärker mit konkreten Aufgaben zusammenhängen und auf Fähigkeiten aufbauen. Sie werden als organisierte und koordinierte Aktivitäten in bezug auf ein Objekt verstanden und umfassen Ketten von sensomotorischen, zentralen und motorischen Mechanismen.“ Das Ausüben einer Fertigkeit ist somit eine organisierte Abfolge bekannter Handlungen zur Bearbeitung einer Aufgabe.

Zimmer (1990, S. 153) unterscheidet zudem zwischen einer kognitiven und motorischen Fertigkeit: „Eine Abfolge von Bewegungen wird als integrierte motorische Fertigkeit bezeichnet, wenn die sie konstituierenden Einzelbewegungen zusammen eine wohlgeordnete Struktur bilden (,sequencing‘) und jeweils zeitlich eindeutig strukturiert sind (,timing‘)“. Die Unterschiede bestehen Zimmer zufolge in der Zeitstruktur und dem Wissensformat, d.h. für die Erzeugung von Bewegungen kann auch auf analoges Wissen zurückgegriffen werden, dem keine Abstraktion der Bedeutung durch sprachliche, logische oder allgemein symbolische Aussagen zugrunde liegt, wie es bei propositionalen Wissenskomponenten der Fall ist, auf die kognitive Fertigkeiten ausschließlich basieren.

### **2.3.1.3 Manuelles Geschick**

Allgemein gesagt, ist Geschicklichkeit die Fähigkeit die Hände zu gebrauchen oder die Fähigkeit Objekte mit den Händen zu manipulieren (Farber, 1991). Nach Poirier (1987, S. 71) ist Geschicklichkeit definiert als „[...] the manual ability that requires rapid coordination of gross or fine voluntary movements, based on a certain number of capacities, which are developed through learning, training and experience.“. Manuelle

Geschicklichkeit lässt sich in die Fein- und Grobmotorik unterteilen, wobei Grobmotorik die Koordination von Hand und Arm bezeichnet, die im Vergleich zur Feinmotorik weniger präzise Bewegungen der Hand und Finger erfordert (Desrosiers, Hebert, Bravo & Dutil, 1995). Fleischman und Ellison (1962, S. 101) definieren Feinmotorik als „[...] the ability to make rapid, skillful, and controlled manipulative movements of small objects, using primarily the fingers.“. In der Literatur wird Feinmotorik auch als Fingergeschicklichkeit bezeichnet, für die im Englischen ebenfalls die Begriffe „digital dexterity“ und „fine finger dexterity“ öfter Anwendung finden, hingegen Grobmotorik auch einfach als manuelles Geschick verwendet wird (Desrosiers et al., 1995). Geschicklichkeit hängt von leistungsstarken, dezentralen neuronalen Netzwerken ab, weshalb Läsionen im Gehirn besonders beeinträchtigte Auswirkungen haben können (Wiesendanger & Serrien, 2001). McPhee (1987) unterteilt manuelles Geschick in eine statische und dynamische Phase, wobei sich jedoch die meisten Definitionen nur auf die statische Phase beziehen. Die statische Phase definiert McPhee als nicht-fassbare Phase (nonprehensile phase), die keine manuellen, manipulativen Fertigkeiten beinhaltet, beispielsweise die Verwendung der Faust als Hammer. Die dynamische Phase lässt sich unterteilen in eine fassbare Phase, die grob- und feinmotorische Aufgaben umfasst und in das exakte Hantieren, welches die Veränderung von Objekten unter Verwendung des Daumen und der Finger einschließt. Die Vielzahl an hier berichteten Definitionen zeigt die Diversität auf, mit der manuelles Geschick nicht nur als Begriff verwendet wird, sondern auch definiert ist. Einheitlich ist in den meisten Definitionen der Bezug auf die Feinmotorik, die vor allem den Einsatz der Finger umfasst.

#### **2.3.1.4 Räumliches Vorstellungsvermögen**

Das räumliche Vorstellungsvermögen bzw. das räumliche Denken ist definiert als kognitive Fähigkeit „[...] in der Vorstellung räumlich zu sehen und zu denken, d.h. im Gedächtnis gespeicherte (mehrdimensionale) Vorstellungsbilder zu reproduzieren und mit ihnen mental zu operieren.“ (Quaiser-Pohl, Lehmann & Schirra, 2001, S. 42). Bisher besteht aber kein einheitlicher Konsens in der Literatur hinsichtlich einer verbindlichen Definition des Konstrukts räumliches Vorstellungsvermögen (Linn & Peterson, 1985; Quaiser-Pohl, 2001). Nach Thurstone und Thurstone (1941) zählt das räumliche Vorstellungsvermögen zu den sieben Primärfaktoren der Intelligenz und kann in die Faktoren Veranschaulichung (visualization), räumliche Beziehungen (spatial

relations) und räumliche Orientierung (spatial orientation) unterteilt werden. Der Faktor Veranschaulichung bezeichnet die mentale Vorstellung von Bewegungen, welche die mentale Rotation, die räumliche Verschiebung oder das Falten von Objekten beinhaltet. Der Faktor räumliche Beziehungen repräsentiert die Fähigkeit, Objekte aus unterschiedlichen Perspektiven zu erkennen. Räumliche Orientierung definierten Thurstone und Thurstone als die richtige räumliche Einordnung eines Individuums in eine Situation.

Linn und Petersen (1985) unterteilen in ihrer Metaanalyse das Konstrukt räumliches Vorstellungsvermögen hingegen in die Faktoren räumliche Wahrnehmung, mentale Rotationen und räumliche Veranschaulichung. Die räumliche Wahrnehmung beschreiben Linn und Petersen als die Wahrnehmung der räumlichen Beziehungen und Verhältnisse zwischen Objekten und in Bezug zur Position der eigenen Person. Der Faktor räumliche Veranschaulichung ist als komplexe Verarbeitung und Manipulation räumlicher Informationen in mehreren Schritten zu verstehen und die mentale Rotation als gedanklich vollzogene Drehung zwei- oder dreidimensionaler Objekte.

### **2.3.1.5 Psychomotorik**

Perzeptiv-motorische Fähigkeiten oder psychomotorische Fähigkeiten bezeichnen die Fähigkeitsart, welche notwendig ist, um sensorische Informationen und muskuläre Reaktionen zur Bewältigung einer bestimmten Aufgabe zu koordinieren (Fleishman & Reilly, 1992). Zur erfolgreichen Bewältigung einer komplexen psychomotorischen Aufgabe werden Fähigkeiten, wie Diskriminationsreaktionszeit, Reaktionszeit, mechanische Erfahrung, psychomotorische Koordination, verbales Verständnis, Armbewegungsgeschwindigkeit, Wahrnehmungsgeschwindigkeit, Geschicklichkeit und räumliche Beziehungen, benötigt (Fleishman & Hempel, 1955). Johnston und Catano (2002, S. 77) zählen zu psychomotorischen Fähigkeiten auch “[...] specific abilities of control precision, multilimb coordination, response orientation, rate control, reaction time, arm-hand steadiness, manual dexterity, finger dexterity, wrist-finger speed, and speed-of-limb movement.”. Psychomotorik ist nach Clauß (1995) definiert als ein Sammelbegriff für psychisch regulierbare Motorik, zu der Bewegungen, Haltungen und Tonus gehören. Erweitert man die Betrachtung der Definitionen zum manuellen Geschick um den Kontext der Bewegungskoordination, so wird manuelles Geschick als eine der Psychomotorik zugehörige Fähigkeit beschrieben, die im Zusammenhang mit

Wahrnehmungsprozessen und -fähigkeiten steht, zu welchen auch das räumliche Vorstellungsvermögen zählt.

### **2.3.2 Frühe Studien**

Die Fragestellung, welche Fähigkeiten und Fertigkeiten Studierende benötigen, um erfolgreich die vorklinischen Laborkurse der Zahnmedizin zu absolvieren, beantworteten Experten in den 1930er und 1940er Jahren noch einheitlich. Ausschlaggebend für den praktischen Studienerfolg in den vorklinischen Laborkursen waren motorische und mechanische Fähigkeiten (Thompson, 1942). Hintergrund dieser Fragestellung war neben einem curricularen Interesse die Einbindung zusätzlicher Kriterien in die Auswahl von Zahnmedizinistudenten, welche den Anforderungen des Studiums entsprechen (Spratley, 1990).

1937 konnte bereits Harris mit der Hinzunahme von Ergebnissen mechanischer Fähigkeitstests als weitere Auswahlkriterien zum Schulabschlussnotendurchschnitt GPA die Vorhersage der Studienleistung verbessern, welche über den Mittelwert der Anatomie- und Laborkursnoten von 50 Erstsemestern gebildet wurde. Von den vier eingesetzten Testverfahren zeigte aber nur die Kombination des „O'Connor Finger Dexterity test“ und des „O'Connor Tweezer Dexterity test“ eine signifikante Beziehung zur Studienleistung ( $r = 0.465$ ). Der räumliches Vorstellungsvermögen erfassende „Wiggly Block test“, in welchen aus Holzstücken eine Figur zusammengelegt werden muss, als auch der die Handfestigkeit und -führung prüfende „Hand-Steadiness test“ besaßen hingegen schwache, positive Korrelationen zur Studienleistung.

### **2.3.3 Dental Admission Test**

#### **2.3.3.1 Subtests „Spatial Relations“ und „Carving Dexterity“**

Der praktische Studienerfolg in den Laborkursen sollte in der ersten Version des DAT von 1951 durch die grob umrissenen Fähigkeiten, Muster und Beziehungen zu visualisieren, im Untertest „Space Relations“ (bzw. „Space Visualization“) zum einen und zum anderen durch den geschickten Einsatz von Händen und Finger im Untertest „Carving Dexterity“ vorhergesagt werden. Ursache für die langanhaltende Diskussion und Vielzahl an Validierungsstudien zu beiden DAT-Untertests war die geringe Durchführungsökonomie des im Untertest „Carving Dexterity“ eingesetzten Kalkschnitttests, der neben einer Testdauer von sechs Stunden eine sehr aufwendige Materialorganisation erforderte. Der Test erfasst manuelle Fähigkeiten, indem aus

mehreren Kalkblöcken Zahnmuster geschnitzt werden müssen. Ansteigende Bewerberzahlen seit Einführung des DAT erhöhten in den USA und Kanada den Druck, ein alternatives und praktikableres Verfahren zu finden (Spratley, 1990).

Derevere schlug bereits 1961 auf Basis der Zulassungsdaten von 118 Zahnmedizinstudierenden in den USA vor, den Kalkschnittstest durch einen papierbasierten Multiple-Choice Test zur Objektvisualisierung zu ersetzen, da beide Testverfahren gut miteinander ( $r = 0.78$ ) und mit der Laborkursleistung im ersten Studienjahr (Kalkschnittstest:  $r = 0.37$ , Objektvisualisierungstest:  $r = 0.39$ ) sowie identisch mit den Noten der klinischen, praktischen Kurse im vierten Studienjahr korrelierten ( $r = 0.26$ ). Auch Hood (1963) gelangte in seiner Untersuchung mit 300 Erstsemesterstudierenden zu der Schlussfolgerung, dass in der DAT-Batterie der Kalkschnittstest des Untertests „Carving Dexterity“ ( $r = 0.38, p < .05$ ) als auch der räumliche Beziehungen erfassende Testbestandteil „Space Relations“ ( $r = 0.30, p < .05$ ) aufgrund ihrer korrelativen Beziehungen zur praktischen Leistung in den Laborkursen des ersten Studienjahres als Prädiktoren gleich geeignet seien (vgl. Ginley, 1966).

Widersprüchliche und kontroverse Studienergebnisse berichteten hingegen Manhold und Manhold (1965; 1967) für die Untertests „Space Relation“ und „Carving Dexterity“ im Rahmen der Validierung des DAT. Über die vier Untersuchungsstichproben mit Zahnmedizinstudierenden ( $28 \leq n \leq 34$ ) änderten sich die Zusammenhänge beider Untertests mit den technischen Kursnoten im vorklinischen und klinischen Studienabschnitt in ihrer Stärke, Signifikanz und Richtung, ohne dass die Autoren außer den zu geringen Stichprobengrößen eine zufriedenstellende Erklärung finden konnten. Bellanti, Mayberry und Tira (1972) hingegen sprachen sich deutlich auf Basis ihrer Studienergebnisse für einen Verbleib des Kalkschnittstests im DAT aus. Die ursprüngliche Intention der DAT-Untertests „Carving Dexterity“ und „Space Visualization“ war die Fähigkeit zu erfassen, wie geschickt jemand mit seinen Händen umgehen kann. Aufgrund dessen verwendeten Bellanti et al. als Kriterium des praktischen Studienerfolgs nur die Noten des vorklinischen Prothetikurses von 344 Erstsemesterstudierenden der Jahre 1966 bis 1968, da in diesem verschiedene komplexe Fertigkeiten manueller und kognitiver Art erforderlich waren, beispielsweise für das Erstellen und Einsetzen von Zahnkronen aus Wachs und Porzellan. Die schrittweise multiple Regressionsanalyse ergab für den Kalkschnittstest als Prädiktor 13.83 Prozent an vorhergesagter Leistung im Prothetikkurs. Die Hinzunahme des GPA als zweiten Prädiktor führte jedoch nur zu einer Erhöhung der aufgeklärten Varianz von 1.77

Prozent, die beinahe identisch hoch ausfiel wie die Aufnahme des „Space Visualization“-Test mit 1.79 Prozent als dritten und letzten Prädiktor in das Modell. Kritik an den DAT-Subtests „Carving Dexterity“ und „Spatial Relations“ übten unter anderem Brigante und Lamb (1968) und entwickelten den „Perception and Control Test“ mit dem Ziel, nicht nur einen Leistungsstand, sondern die zugrundeliegende Fähigkeit zu messen. Der „Perception and Control Test“ beinhaltet eine Batterie an Testverfahren, die taktiles Erfassen, die zielgerichtete Steuerung der Hand, Tiefenwahrnehmung, visuelle Schärfe, taktile Diskriminierung, hart-weich Sensitivität, Vergleich von Oberflächenkonturen, Koordination des Fingerdrucks, Koordination der Fingerspannung und die Sensitivität der Handfestigkeit umfassen. Die Testergebnisse korrelierten moderat mit den Noten in den Laborkursen ( $0.416 \leq r \leq 0.583$ ), die stärker im Vergleich zu den DAT-Subtests „Space Relations“ ( $0.240 \leq r \leq 0.348$ ) und „Carving Dexterity“ ( $0.209 \leq r \leq 0.277$ ) ausfielen, weshalb die Autoren ihre Zielsetzung nicht nur bestätigt sahen, sondern den „Perception and Control Test“ als geeigneter zur Vorhersage der praktischen Studienleistung hielten. Neben der Auswertungsobjektivität der Laborkursnoten kritisierten Brigante und Lamb den geringen Stichprobenumfang der in ihrer Untersuchung einbezogenen Jahrgänge ( $40 \leq n \leq 65$ ).

### **2.3.3.2 Dimensionalität der Subtests „Carving Dexterity“ und „Spatial Relations“**

Zweifel an der Dimensionalität des DAT veranlassten Dworkin (1970) zu einer Überprüfung des Studierfähigkeitstests. Die Ergebnisse der Faktoranalyse (Varimax-Rotation) mit 91 Studierenden der Zahnmedizin ergaben die Faktoren „Science“, „Manual“ und „Intelligence“. Die DAT-Subtests „Manual Average“ ( $r = 0.93$ ), „Spatial Relations“ ( $r = 0.87$ ) und „Carving Dexterity“ ( $r = 0.43$ ) luden nur auf dem zweiten Faktor „Manual“, anstatt auf den ursprünglich postulierten Faktoren „Space Conception“ und „Carving Dexterity“. Die Frage, ob es sinnvoll ist über das arithmetische Mittel aus den Punktzahlen der Untertests „Spatial Relations“ und „Carving Dexterity“ den Subtest „Manual Average“ zu bilden, verneint Dworkin, da dieses Vorgehen nur vertretbar wäre, falls der gebildete Durchschnittswert eine inhaltlich nachweisbare Beziehung zu einer Fähigkeit besitzt, die er repräsentieren soll. Young berichtete 1972 vergleichbare Ergebnisse zu Dworkin (1970) für den DAT in Kanada. Die Hauptkomponentenanalyse mit 43 Erstsemesterstudierenden ergab signifikante, hohe Faktorladungen der DAT-Subtests „Manual Average“ ( $r = 0.986$ ),

„Spatial Relation“ ( $r = 0.843$ ) und „Carving Dexterity“ ( $r = 0.830$ ) auf den Faktor „Spatial“.

Die Ergebnisse dieser Untersuchungen stellten eindeutig die Konstruktvalidität beider DAT-Subtests infrage. Dworkins Ansatz wurde in Jahr später von Zullo (1971) mit dem Ziel aufgegriffen, die durch die DAT-Subtests „Carving Dexterity“ und „Spatial Relations“ gemessenen Fähigkeiten zu identifizieren. 100 Erstsemesterstudierende der Zahnmedizin absolvierten zu Studienbeginn eine umfangreiche Testbatterie mit 14 psychomotorischen Verfahren und erklärten sich mit der Verwendung ihrer DAT-Zulassungsergebnisse einverstanden. Die Ergebnisse der Hauptkomponentenanalyse ergaben insgesamt sechs Faktoren. Der im DAT-Subtest „Carving Dexterity“ eingesetzte Kalkschnittstest zeigte die höchste Ladung auf den Faktor „Small Instruments Dexterity“ ( $r = 0.76$ ), während der DAT-Subtest „Spatial Relations“ wie erwartet am höchsten auf den Faktor „Spatial Relations I“ lud ( $r = 0.68$ ), der dreidimensionales, räumliches Vorstellungsvermögen repräsentierte. Zullo wertete die Ergebnisse als Bestätigung der Konstruktvalidität beider DAT-Subtests. Der Kalkschnittstest scheint das manuelle Geschick im Umgang mit Handinstrumenten als eine spezifische Fähigkeit zu erfassen, weshalb die Vorhersage von Laborkursnoten wenig aussichtsvoll ist, sobald andere Faktoren, wie kognitive Fähigkeiten oder didaktische Elemente, einen ausschlaggebenden Anteil an der Notenbildung haben.

### **2.3.3.3 Entnahme des Subtests „Carving Dexterity“**

Trotz widersprüchlicher Studienergebnisse ersetzte 1972 der papierbasierte „Perceptual Motor Ability Test“ (PMAT) den DAT-Untertest „Carving Dexterity“, welcher aufgrund der stark angestiegenen Bewerberzahlen in den USA um einen Studienplatz in der Zahnmedizin zu aufwendig durchzuführen war. Die zugrundeliegende Gemeinsamkeit des PMAT zum Kalkschnittstest war das visuelle Wahrnehmungsvermögen, von dem man annahm, ein valider Prädiktor zur Beurteilung des Erfolges in den Laborkursen zu sein (Graham, 1972). Den zunächst inhaltlich geäußerten Zweifeln folgten erste Untersuchungen, ob das Vorgehen auf empirischer Ebene zu vertreten ist. So widersprach beispielsweise Peterson (1974) der Entnahme des Kalkschnittstestes aus dem DAT, welcher nicht gleichwertig durch den papierbasierten PMAT als kognitiver Fähigkeitstest ersetzt werden kann. Aufgrund des Fehlens motorischer Fertigkeiten zur Aufgabenbearbeitung bestünde beim PMAT immer die Gefahr, Studierende mit fünf Daumen an einer Hand zuzulassen. Dieser

Argumentation widersprachen Deubert, Smith, Jenkins und Berry (1975) aus Großbritannien, die für papierbasierte Testverfahren zum mechanischen Verständnis und räumlichen Beziehungen die üblichen Korrelationen zu den vorklinischen Laborkursnoten fanden ( $0.33 < r < 0.58$ ,  $p < .05$ ), welche keine schriftlichen Prüfungsergebnisse beinhalteten. Ein gleichzeitig mit den 41 Erstsemesterstudierenden der Zahnmedizin durchgeführter papierbasierter Intelligenztest für Akademiker wies hingegen keine Beziehung zum praktischen Studienerfolg auf ( $0.03 < r < 0.15$ , *n.s.*).

1974 identifizierte Graham für die Multiple-Choice-Aufgaben des papierbasierten PMAT mittels Faktorenanalyse in seiner Primäruntersuchung an 1.163 Teilnehmern mehrere unabhängige Faktoren, die er als „block design“, „space relations“, „length of lines“, „sequence of ideas“, „three-dimensional drawing“ und „passing an object through an aperture“ bezeichnete. Trotz mehrjähriger Entwicklung des PMAT, war empirisch nicht zweifelsfrei nachweisbar, dass der Kalkschnittstest durch den papierbasierten Test in seiner Prädiktivität der praktischen Leistung in den Laborkursen gleichwertig ersetzt werden konnte.

Nur drei Jahre nach der Entnahme wurde aufgrund der bestehenden Zweifel in Kanada der Kalkschnittstest 1975 von der Canadian Dental Association wieder in den DAT integriert (Spratley, 1990), was durch etliche Studien gestützt wurde. Wood (1979) schlug auf Basis seiner Untersuchungsergebnisse mit 312 Zahnmedizinistudierenden des ersten und zweiten Studienjahrs aus Kanada sogar vor, die Punktzahlen des Kalkschnittstests im DAT gleichberechtigt zum GPA bei der Studierendenauswahl als Kriterium einzusetzen. Für die Ergebnisse des PMAT hingegen ergaben sich keine Korrelationen zu den praktischen Kursen, die psychomotorische Fertigkeiten erforderten, während die des DAT-Untertest „Carving Dexterity“ im Bereich zwischen 0.31 und 0.36 lagen. Im gleichen Jahr wie Wood kritisierten auch Wong, Watson und Thye (1979) die Prädiktivität der erstsemestrigen Laborkursleistung durch den PMAT, die sie als zweifelhaft beschrieben. Zwischen dem Median der vier Laborkursnoten und dem Zulassungsergebnis des PMAT zeigte sich für 100 Erstsemesterstudierende eine von den Autoren als moderat interpretierte Korrelation ( $r = 0.349$ ,  $p < .01$ ), die aber für 98 nachfolgende Erstsemesterstudierende des nächsten Jahres nicht bestätigt werden konnte ( $r = 0.106$ , *n.s.*).

#### 2.3.3.4 Die Einführung des „Perceptual Ability Tests“

Die Kritik an der Prädiktivität der praktischen Studienleistungen durch die DAT-Subtests PMAT und „Spatial Relations“ veranlasste 1981 die American Dental Association in den USA zu einer Überarbeitung der Verfahren. Die Aufgabenbestandteile beider Untertests, welche die höchsten prädiktiven Validitäten zur praktischen Studienleistung aufwiesen, wurden in den „Perceptual Ability Test“ (PAT) zusammengeführt. Der PAT erfasst auf mehrdimensionaler Ebene räumliche Fähigkeiten, die anhand von 90 Multiple-Choice-Aufgaben mit zwei- als auch dreidimensionalen Problemfragestellungen operationalisiert sind. Beispielsweise muss bei der Aufgabe Papier falten aus einer vorgegebenen Faltvorlage die richtige Figur bestimmt werden, hingegen ist in einer anderen Aufgabe die richtige Anzahl dreidimensionaler Blöcke zu ermitteln, ohne das alle Blöcke eines Stapels direkt zu sehen sind (Spratley, 1990).

Kramer, Kubiak und Smith (1989) evaluierten den PAT auf der Datenbasis von 2.000 Studierenden aus 54 zahnmedizinischen Fakultäten in den USA. Die zur Konstruktvalidierung durchgeführte Hauptkomponentenanalyse mit Varimax-Rotation ergab den Faktor räumliche Fähigkeit, auf dem die Aufgabentypen „Orthographic Projections“ und „Apertures“ luden, den Faktor dreidimensionale, räumliche Wahrnehmung, zu welchem die Aufgabentypen „Cubes“ und „Form Development“ eine Ladung aufwiesen und den Faktor zweidimensionale, räumliche Wahrnehmung, auf den ausschließlich der Aufgabentyp „Angles“ lud. Zu den Noten des ersten vorklinischen Laborkurses in 47 Fakultäten berichteten die Autoren eine prädiktive Validität in Höhe von  $R^2 = 0.14$ , die nur geringfügig kleiner ausfiel, als der Zusammenhang zu den Durchschnittsnoten mehrerer vorklinischer Laborkurse in 33 Fakultäten mit  $R^2 = 0.19$ . Der PAT gilt als zuverlässiges Instrument zur Vorhersage der praktischen Studienleistung in der Zahnmedizin (z.B. Boyle & Santelli, 1986; Gansky, Pritchard, Kahl, Mendoza, Bird, Miller & Graham, 2004; Hegarty, Keehner, Khooshabeh und Montello, 2009; Kramer, 1986; Lundergan, Soderstrom & Chambers, 2007). Es wurden aber beispielsweise von Coy, McDougall und Sneed (2003), Oudshoorn (2003) oder Walcott, Knight und Charlick (1986) Zweifel an der Prognostik des PAT geäußert, ohne dass die Autoren hierfür eine zufriedenstellende Erklärung geben konnten.

### **2.3.4 Fähigkeiten als Prädiktoren**

#### **2.3.4.1 Räumliches Vorstellungsvermögen**

##### **2.3.4.1.1 Frühere Studien**

Der Zusammenhang zwischen der kognitiven Fähigkeit räumliches Vorstellungsvermögen und der praktischen Studienleistung in der Zahnmedizin wurde unabhängig von der Validierung des Studieneignungstests DAT in zahlreichen Untersuchungen thematisiert. Eine hohe Ausprägung der räumlichen Fähigkeit scheint zwar mit guten praktischen Studienleistungen einherzugehen, darf aber Smith (1976) zufolge nicht zur Annahme verleiten, dass ein adäquates Maß an räumlichem Vorstellungsvermögen den erfolgreichen Erwerb von praktischen Fertigkeiten eher garantiert, als ein adäquates Maß an Intelligenz den akademischen Studienerfolg sicherstellt. Weitere Faktoren, wie eine ausreichende Motivation und geeignete Persönlichkeitsfaktoren, bedingen den Studienerfolg ebenfalls und sollten im Rahmen der Studierendenauswahl nicht ignoriert werden. Das Fehlen von einem gewissen Minimum an räumlichem Vorstellungsvermögen verringert jedoch enorm die Wahrscheinlichkeit, erfolgreich Zahnmedizin zu studieren, schließt Smith seine Überlegungen, denen zufolge eine nichtlineare Beziehung zwischen Fähigkeit und praktischem Studienerfolg bestehen könnte.

In Analogie zu Holliday und Koepchen (1954), die visuelles Leistungsvermögen als einen Einflussfaktor psychomotorischer Fertigkeiten identifizierten, konnten Beck, Taft und Zimmer (1978) in ihrer Untersuchung bestätigen, dass die Fähigkeit zum räumlichen Denken geeignet ist, um das Erlernen grundlegender psychomotorischer Fertigkeiten in der Zahnmedizin vorherzusagen. 70 Zahnmedizinstudierende absolvierten vor Beginn der Laborkurse im zweiten Studienjahr den DAT-Untertest „Space Relations“ zur Operationalisierung des räumlichen Vorstellungsvermögens. Als Kriterien verwendeten Beck et al. nur die Prüfungsnoten für die praktischen Arbeiten in den Laborkursen „Oral Anatomy“ und „Dental Therapy“ sowie die Noten der vorklinischen und klinischen Zahnpflegeprüfungen, die von den Autoren auf eine zwölfstufige Punkteskala transformiert wurden. Im Laborkurs „Oral Anatomy“ mussten unter anderem Zahnzeichnungen vervollständigt werden, um anschließend den Zahn aufzuwachsen, während im zweiten Laborkurs Projekte bewertet wurden, die das Einsetzen von Amalgamfüllungen oder die Zahnrestauration aus Kunstharz umfassten. In der vorklinischen Prüfung „Preclinical Dental Hygiene“ demonstrieren Studierende

an Modellen Untersuchungstechniken, während die Note der „Clinical Dental Hygiene“-Prüfung nach der Behandlung von 80 Patienten vergeben wurde. Der DAT-Untertests „Space Relations“ zeigte zwar eine gegen Null tendierende Korrelation zur „Preclinical Dental Hygiene“-Note auf ( $r = 0.036$ , *n.s.*), aber eine moderate lineare Beziehung zu den Laborkursen „Oral Anatomy“ ( $r = 0.550$ ,  $p < .01$ ) und „Dental Therapy“ ( $r = 0.350$ ,  $p < .01$ ), die stärker ausfiel als die zur klinischen Note „Clinical Dental Hygiene“ ( $r = 0.254$ , *n.s.*). Beck et al. vermuteten als Erklärung, dass Aufgaben mit hohem Imitationscharakter weniger Visualisierungsleistungen erfordern, aber mehr Muskelkoordination in Hand und Arm, sodass andere, nicht spezifizierte Variablen ausschlaggebend für die praktische Leistung in den Laborkursen waren. Als mögliche, die Ergebnisinterpretation limitierende, Faktoren führten Beck et al. den zeitlichen Abstand von fast einem Jahr beider Laborkurse an, die unterschiedliche Komplexität der Aufgaben und der Mangel an zuverlässigen Beurteilungskriterien der praktischen Arbeiten und Aufgaben in den Laborkursen.

Ein schwach ausgeprägtes räumliches Vorstellungsvermögen führt bei Zahnmedizinstudierenden nicht nur zu schlechten Leistungen in den praktischen Kursen, sondern geht auch mit Gefühlen der Unzufriedenheit und Unzulänglichkeit einher, argumentiert Smith (1989). Darin ist ein Grund zu sehen, weshalb Studierende freiwillig das Zahnmedizinstudium abbrechen bzw. aufgrund nicht ausreichender Studienleistungen abbrechen müssen oder es zu mindestens nicht in der vorgegebenen Mindestdauer abschließen. Zur Stützung seiner Behauptungen ließ Smith in seiner Primärstudie von 1979 bis 1982 insgesamt 1391 Studienplatzbewerber in Großbritannien den „Space Relations Test“ absolvieren, der als Untertest des nordamerikanischen Studierfähigkeitstest DAT räumliches Vorstellungsvermögen erfasst (vgl. Unterkapitel 2.3.3.1), aber dessen Ergebnisse nicht in die Zulassung zum Zahnmedizinstudium eingingen. Zwischen den Ergebnissen des „Space Relations Test“ und dem GPA der Bewerber ergab sich kein korrelativer Zusammenhang, worin Smith die Diskriminanz des „Space Relations Test“ bestätigt sah. Studierende, die den empfohlenen Mindestwert im „Space Relations Test“ nicht erreichten, studierten nicht nur länger, sondern scheiterten auch häufiger im Studium als Studierende, die den Grenzwert erreichten ( $\chi^2 = 19.84$ ,  $p < .001$ ). Wäre der „Space Relations Test“ als Zulassungsinstrument eingesetzt worden, hätten 21 von 178 Erstsemestern keinen Studienplatz erhalten, von denen aber 14 das Zahnmedizinstudium absolvierten,

weshalb Smith forderte, dass Tests zum räumlichen Vorstellungsvermögen mit Vorsicht und anderen verfügbaren Selektionskriterien kombiniert eingesetzt werden sollten.

#### **2.3.4.1.2 Aktuellere Studien**

Das Ergebnisbild früherer Studien, vor allem im Rahmen der Validierung des DAT, ist nicht einheitlich, ob räumliches Vorstellungsvermögen als kognitive Fähigkeit einen Einfluss auf die praktische Leistung in den Laborkursen der Zahnmedizin besitzt und demzufolge als Auswahlkriterium für Bewerber um einen Studienplatz eingesetzt werden sollte (Spratley, 1990). In aktuelleren Studien ist die Beziehung zwischen dem räumlichen Vorstellungsvermögen als Fähigkeit und der praktischen Studienleistung oftmals nur eine Teilfragestellung, die aus Gründen einer übersichtlichen Darstellung ausschließlich in diesem Abschnitt berichtet wird. Dabei fällt die inhaltliche Argumentation in aktuelleren Studien im Vergleich zu früheren Untersuchungen detaillierter aus.

Nach Evans und Dirks (2001) erfordern die Aufgaben in den praktischen Kursen der Zahnmedizin motorische Fertigkeiten im Umgang mit Handinstrumenten, wofür präzise Bewegungen der Finger, der Hand und der Handgelenke erforderlich sind. Die dreidimensionale Wahrnehmung von Objektformen, ihren Proportionen und ihrer Symmetrie ist für die erfolgreiche Koordination und Steuerung motorischer Handlungen ausschlaggebend, beispielsweise um einen Zahn zu präparieren, d.h. vorbereitend für das Einsetzen einer Füllung abzuschleifen. Demzufolge sollte räumliches Vorstellungsvermögen als kognitive Fähigkeit einen Einfluss auf die vorklinische Laborkursleistung besitzen. Um ihre Vermutung empirisch zu überprüfen, ließen Evans und Dirks 80 Zahnmedizinstudierende in der ersten Studienwoche das „Revised Minnesota Paper Form Board“ (RMPFB), den „Blocks test“ und den „Dimensions test“ absolvieren, um ihre Fähigkeit zum räumlichen Denken zu erfassen. Die Einzelnoten als auch die Durchschnittsnote von drei vorklinischen Laborkursen wurden als Kriterien der praktischen Studienleistung von Evans und Dirks herangezogen, die Arbeiten an Zahnbrücken, Kronen sowie Teil- und Totalprothesen beinhalteten, deren Bewertung auf einer Punkteskala von null bis einhundert erfolgte. Die höchste Beziehung zu den Laborkursergebnissen wiesen die Ergebnisse des „Blocks tests“ auf ( $r = 0.400$ ,  $p < .05$ ), in welchem die Anzahl von dreidimensionalen Blöcken eines Stapels ermittelt werden muss. Aufgrund der mangelnden Signifikanz berichteten die Autoren keine Ergebnisse für die zwei anderen Testverfahren zum räumlichen Vorstellungsvermögen, deren

Aufgaben den Umgang mit zweidimensionalen Objekten erforderten. Evans und Dirks sahen ihre Überlegungen für die dreidimensionalen Aspekte des räumlichen Vorstellungsvermögens bestätigt, die sie als Prädiktor der praktischen, vorklinischen Laborkursleistung identifizieren konnten.

Coy, McDougall und Sneed (2003) hingegen zweifelten eine ausreichende Vorhersage der Leistung praktischer Fertigkeiten in den vorklinischen Laborkursen durch die kognitive Fähigkeit räumliches Vorstellungsvermögen an und verwiesen auf die Ergebnisse der Studien von Boyd und Teteruck (1979), Boyd, Teteruck und Thompson (1980), Boyle und Santelli (1986), Derevere (1961), Potter, McDonald und Sagraves (1982b), Salvendy, Boydston, Ferguson, Birch und Cunningham (1977), Suddick, Yancey, Devine und Wilson (1982), Suddick, Yancey und Wilson (1983), Walcott, Knight und Charlick (1986) und Wood und Boyd (1982). Um mögliche Effekteinschränkungen durch „range restriction“ im Kriterium zu vermeiden, die aufgrund der Studienplatzkapazitäten pro Semester bestehen könnten, akkumulierten Coy et al. die Zulassungsergebnisse des DAT-Subtests PAT von 492 Studierenden in den Jahren 1993 bis 2002. Das räumliche Vorstellungsvermögen setzten die Autoren in Beziehung zu den praktischen Prüfungsnoten der Laborkurse im ersten Studienjahr. Die Aufgaben in den Laborkursen beinhalteten beispielsweise das Anfertigen von Füllungsmischungen bzw. Plomben, die in vorgefertigte Zahnaushöhlungen eingeführt werden mussten. Die Benotung erfolgte durch fünf Fakultätsmitglieder, deren Ergebnisse gemittelt wurden. Lediglich fünf bis neun Prozent der Laborkursleistung konnten durch die Zulassungsergebnisse des DAT-Subtests PAT mittels Regressionsanalyse vorhergesagt werden, deren Koeffizienten auf dem Niveau  $p < .01$  signifikant ausfielen. Solche schwach positiven Beziehungen seien zwar in der Literatur üblich, aber der praktische Nutzen ist nach Ansicht der Autoren zu eingeschränkt. Geschlechtsspezifische Unterschiede im räumlichen Vorstellungsvermögen der Studierenden wären nach Coy et al. eine mögliche Erklärung, welche die Autoren in ihrer Untersuchung nicht nur für Männer ( $\bar{x} = 16.94$ ) und Frauen ( $\bar{x} = 15.37$ ;  $F = 37.54$ ,  $p \leq .05$ ) nachweisen konnten, sondern auch zeigten, dass die PAT-Ergebnisse von Frauen nur ein bis drei Prozent der Laborkursleistung prädizierten, während die der Männer sieben bis dreizehn Prozent der Laborkursleistung vorher sagten.

Der geschlechtsspezifische Leistungsunterschied im räumlichen Vorstellungsvermögen wurde in zahlreichen Studien bestätigt, zu denen die Metaanalysen von Linn und Petersen (1985), Maeda und Yoon (2013), Peters, Lehmann, Takahira, Takeuchi und

Jordan (2006) oder Voyer, Voyer und Bryden (1995) zählen. Coy et al. (2003) hielten aber den Einfluss auf die vorklinische Ausbildung für vernachlässigbar, da männliche und weibliche Studierende über eine äquivalente Entwicklung technischer Fertigkeiten verfügen. In zwei Experimenten gingen Hegarty, Keehner, Khooshabeh und Montello (2009) der Frage nach, ob die Ausbildung in den Laborkursen auch das räumliche Vorstellungsvermögen der Zahnmedizinierenden fördert und ob das räumliche Vorstellungsvermögen eine prognostische Beziehung zur praktischen Leistung in den Laborkursen besitzt. Die Ergebnisauswertung des ersten Experiments verneinte zwar beide Fragestellungen, jedoch führte der Testungszeitpunkt zu einem Konfundierungseffekt, da die Studierenden im zweiten Studienjahr die Testbatterie erst nach Bestehen der Laborkurse in der Zahnerhaltungskunde absolvierten.

Auf Grund dessen wurden im zweiten Experiment von Hegarty et al. 79 Studierende im ersten und 65 Studierende im vierten Studienjahr der Zahnmedizin sowie 62 Psychologiestudierende im Grundstudium rekrutiert. Die Testbatterie in der zweiten Untersuchung umfasste Verfahren zum räumlichen Vorstellungsvermögen, zu denen unter anderem der „Vandenberg Mental Rotation test“ und der „Tooth-Cross-Section-test“ zählten, in dem speziell Zähne als dreidimensionale Objekte mental rotiert werden mussten. Darüber hinaus wurden in die Auswertung die Ergebnisse des Studierfähigkeitstest DAT der Zahnmedizinierenden einbezogen, die den Untertest PAT zum räumlichen Vorstellungsvermögen und den Untertest „Abstract Reasoning“ zum schlussfolgernden Denken beinhalteten. Die ursprüngliche Bewertung der vorklinischen Laborkursaufgaben erfolgte mit Buchstaben als Noten, die zuerst transformiert, sodass die sehr gute Note A einem Punktwert von zehn entsprach und anschließend über alle Laborkurse gemittelt in die Auswertung einbezogen wurden.

Nur die Ergebnisse des DAT-Subtest PAT der Zahnmedizinierenden im ersten Studienjahr wiesen einen positiv moderaten, signifikanten Zusammenhang von  $r = 0.35$  ( $p < .01$ ) zur durchschnittlichen Laborkursleistung auf, was sich auch in den berechneten Partialkorrelationen widerspiegelte ( $r = 0.33$ ,  $p < .01$ ), in denen der Einfluss des schlussfolgernden Denkens auf den Zusammenhang zwischen räumlichem Vorstellungsvermögen und der Laborkursleistung eliminiert war. Als Ursache vermuteten die Autoren, dass der PAT zum einen in einer realen Prüfungssituation durchgeführt wurde und zum anderen unterschiedliche räumliche Aufgaben beinhaltet, die eher den Anforderungen in den Laborkursen entsprechen. Demzufolge scheint das räumliche Vorstellungsvermögen als Fähigkeit ausschlaggebend für die Leistung in den

zahnmedizinischen Laborkursen des ersten Studienjahres zu sein, da anatomische Zahnstrukturen, deren Formen, Positionen und Verbindungen zueinander, vor allem der oralen Bereiche, die nicht direkt einsehbar bzw. sichtbar sind, mittels der räumlichen Vorstellungskraft als mentale Modelle operationalisiert werden.

Die Vergleiche der Testbatterieergebnisse über die drei Studierendengruppen konnten nicht die von Hegarty et al. postulierte Vermutung einer generellen Verbesserung des räumlichen Vorstellungsvermögens durch das Zahnmedizinstudium bestätigen. Es scheint eher der Fall zu sein, so schlussfolgerten die Autoren, dass Zahnmedizinstudierende spezifische mentale Modelle entwickeln, die wiederherum zu einer Leistungsverbesserung bei speziellen Aufgaben in der Zahnmedizin führen, die sich aber nicht für allgemeine Aufgaben, zu deren Bearbeitung räumliches Vorstellungsvermögen erforderlich ist, zeigte. Sowohl für die Zahn- als auch Psychologiestudierenden zeigte sich ein auf dem Niveau  $p < .01$  signifikanter, geschlechtsspezifischer Unterschied mittlerer Effektstärke ( $d \geq 0.50$ ) in der räumlichen Fähigkeit. Da die Stichproben der Zahnmedizinstudierenden des ersten und vierten Studienjahres hierfür von den Autoren kombiniert wurden, konnten keine Aussagen darüber getroffen werden, ob der Unterschied im räumlichen Vorstellungsvermögen zwischen Männern und Frauen auch nach Abschluss der Laborkurse im vierten Studienjahr besteht. Abschließend vermuteten Hegarty et al. den Einfluss weiterer Variablen auf die Laborkursleistung, beispielsweise der Motivation, sodass der Zusammenhang zwischen räumlichem Vorstellungsvermögen und der Laborkursleistung nur von maximal moderater Stärke ausfallen könnte.

### **2.3.4.2 Manuelles Geschick**

#### **2.3.4.2.1 Frühere Studien**

Vor allem der hohe Durchführungsaufwand von Testverfahren zum manuellen Geschick stand im Vordergrund der Kritik, die in den USA und Kanada im Rahmen der Validierungsdebatte des Studierfähigkeitstests DAT geäußert wurde, aber weniger die prognostische Güte der praktischen Studienleistung durch manuelle Fähigkeiten (vgl. Unterkapitel 2.3.3). Entgegengesetzt dazu berichtete Smith (1979) über die Auswahl von Zahnmedizinstudierenden in Großbritannien, dass zwar Aufgaben, die zur Bearbeitung manuelles Geschick erfordern, von einzelnen zahnmedizinischen Fakultäten eingesetzt, aber nicht als Report oder Studie in der Fachliteratur veröffentlicht wurden. Als Ursache vermutete Smith die geringe prognostische Güte der

praktischen Studienleistung jener Testverfahren, da seiner Ansicht nach Schnitz- und Drahtbiegetests Leistungstests darstellen, die manuelle Fähigkeiten nur in variierenden Ausprägungen erfassen, welche von Training beeinflusst sind. Nachfolgende Studien konnten Smiths Überlegungen nicht teilen. In den USA zeigten im gleichen Jahr Wong, Watson und Thye für die „Ivory Block Exercise“ eine zuverlässige Prädiktivität der praktischen Laborkursleistung im ersten zahnmedizinischen Semester auf. Die Übung misst manuelles Geschick, indem eingedruckte Muster in einem Wachsblock nach Maß herausgearbeitet werden müssen. Die Bewertung der drei anzufertigen Wachsböcke erfolgt auf einer dreistufigen Skala, dessen Maximum bei Erfüllen der Vorgabe drei Punkte beträgt. Die zu Studienbeginn absolvierte Übung ergab sowohl für 100 als auch für 98 im nächsten Jahr nachfolgende Erstsemesterstudierende positive Beziehungen zur durchschnittlichen Leistung im ersten Laborkurs der Zahnerhaltungskunde, die in Höhe von  $r = 0.431$  und  $r = 0.329$  signifikant auf dem Niveau  $p < .01$  ausfielen. Die vier Prüfungsarbeiten im Laborkurs umfassten das Präparieren von Zähnen zum Einsetzen von Füllungen, deren Bewertung über mehrstufige Skalen erfolgte, auf denen eine höherwertige Qualität durch ansteigende Punktwerte verzeichnet war.

Auch Walcott, Knight und Charlick konnten 1986 vergleichbare Ergebnisse mit einem Wachstumstest erzielen, in welchem ein Zahn aus Wachs modelliert werden musste. Der im Vergleich zum Wachstumstest in die Auswertung einbezogene DAT-Subtest PAT zum räumlichen Vorstellungsvermögen zeigte um 23 Prozent schwächere positive Korrelationen zu der Durchschnittsnote in den vorklinischen Laborkursen des ersten und um 32 Prozent des zweiten Studienjahres auf, die für den Wachstumstest  $r = 0.39$  sowie  $r = 0.40$  betragen und für beide Testverfahren signifikant ausfielen ( $p < .01$ ). Der GPA und der naturwissenschaftliche Kenntnisse prüfende DAT-Subtest „Academic Average“ besaßen hingegen nur sehr schwache, nichtsignifikante Korrelationen zu den Laborkursleistungen, deren Punktbewertungen z-transformiert in die Auswertung eingingen. Walcott et al. berichteten weiterführend, dass in einem Regressionsmodell die Hinzunahme des PAT zum Wachstumstest, der alleine 16.4 Prozent der durchschnittlichen Laborkursleistung im ersten Studienjahr prädizierte, zu einem inkrementellen Anstieg der erklärten Varianz von 5.8 Prozent führte. Die Autoren schränkten die Interpretation ihrer Untersuchungsdaten mit 131 Studierenden ein, da die Laborkursleistung von der Didaktik der Lehrenden abhängig sei, die Reliabilität der Aufgabenbeurteilungen in den Laborkursen aufgrund fehlender Bewertungskriterien nicht hoch ausfallen könnte und das Vorliegen von „restrictions of range“ aufgrund der

Stichprobenbeschaffenheit zu vermuten sei und somit verringerte Effektstärken nicht auszuschließen waren.

#### **2.3.4.2 Aktuellere Studien**

In aktuelleren Studien traten vermehrt Zweifel an der Vorhersage des praktischen Studienerfolgs durch manuelle Fähigkeiten und der Verwendung als Auswahlkriterium für Studienplatzbewerber in der Zahnmedizin auf. Luck, Reitemeier und Scheuch (2000) führten mit 88 Zahnmedizin- und 23 von insgesamt 29 zufällig ausgewählten Medizinstudierenden zu Beginn des zweiten und nach Ende des sechsten Semesters eine Testbatterie mit vier Verfahren durch, um das feinmotorische Geschick zu erfassen. Im Tremometertest mussten Metallstifte in vorgefertigte Löcher gesteckt werden, ohne dass die Lochränder berührt werden dürfen, während bei der Adaption des Verfahrens mit Spiegelaufsatz, die direkte Sicht auf die Lochplatte versperrt war und nur über den aufgesteckten Spiegel eingesehen werden konnte. Der Zweihandsinusoidtest besteht aus einer Apparatur, in welcher ein Metallstift über zwei Regler gesteuert wird. Aufgabe war es, mit dem Metallstift möglichst genau dem Verlauf einer sinusförmigen Linie zu folgen. Das Computerspiel Bogenschießen erforderte es, mit sechs Pfeile innerhalb einer Minute eine Zielscheibe bestmöglich zu treffen. Die Auswertung zeigte keinen Leistungsunterschied zwischen den Studierenden der Human- und Zahnmedizin für die Testung im zweiten Semester, weshalb eine Auswahl von Bewerbern auf Basis der Feinmotorik Luck et al. zufolge als nicht sinnvoll erscheint. Die Testungswiederholung nach Ende des sechsten Semesters ergab nur eine signifikante Leistungsverbesserung für die Zahnmedizinstudierenden. Ursache ist nach Ansicht von Luck et al. ein Übungseffekt, bedingt durch die Anforderungen der vorklinischen Laborkurse des Zahnmedizinstudiums, die zu einer Verbesserung der Feinmotorik führten. Daher empfehlen Luck et al. Feinmotorik erfassende Testverfahren eher zu diagnostischen Zwecken im Studium einzusetzen, um Studierende mit Defiziten im manuellen Geschick frühzeitig zu identifizieren und rechtzeitig in Förderungsprogramme aufnehmen zu können, anstatt sie auf Basis von Testergebnissen einer trainierbaren Fähigkeit vom Studium auszuschließen.

Konform zu den Schlussfolgerungen von Luck et al. (2000) hielten auch Giuliani, Lajolo, Clemente, Querqui, Viotti, Boari und Miani (2007) einen Ausschluss vom Studium auf Basis des grundlegenden manuellen Geschicks für nicht vertretbar, zumal durch gezieltes Training im Studium eine Verbesserung der manuellen Fähigkeit

erfolgen würde und somit zu Studienbeginn bestehende Defizite ausgeglichen werden könnten. Zu empirischen Stützung ihrer inhaltlichen Argumentation variierten Giuliani et al. die Testwiederholungszeitpunkte der im Auswahlverfahren durchgeführten Testbatterie im Bereich von acht bis 52 Monaten, die von 433 zahnmedizinischen Bewerbern in fünf aufeinanderfolgenden Jahren an der Universität Rom in Italien absolviert wurden. Die Testbatterie beinhaltete zehn Aufgaben, die etablierten Verfahren wie dem „Vayer psycho-motor profile“ entstammten. Alle Teilnehmer waren darüber informiert, dass ihre Testergebnisse keine Auswirkung auf das aktuelle Auswahlverfahren hatten. Die Testbatteriewiederholung der 15 Studierenden umfassenden Kohorten zeigte eine signifikante Verbesserung ( $p < .05$ ) des grundlegenden, manuellen Geschicks im Zahnmedizinstudium nach 32 Monaten, worin die Autoren ihre Behauptung bestätigt sahen.

Gansky et al. (2004) argumentierten, dass ein Einsatz von manuellen Testverfahren zur Studierendenauswahl in der Zahnmedizin nicht sinnvoll sei, da ein zuverlässiger Informationsgewinn über den Erfolg bzw. Misserfolg in den vorklinischen Laborkursen ausbleibt. Blommers kritisierte bereits 1956 die fehlende Möglichkeit, abgelehnte Bewerber in die Überprüfungen einzubeziehen, ob die getroffenen Auswahlentscheidungen als richtig oder falsch zu werten sind. Gansky et al. griffen diesen Gedanken in ihrer Studie auf, indem sie die Leistungen von Zahnmedizinstudierenden untersuchten, die keine Zulassung zum Studium erhalten hätten, falls die Auswahl über einen manuellen Test erfolgt wäre. 244 Erstsemesterstudierende der Zulassungsjahrgänge von 2000 bis 2002 absolvierten in den ersten Studienwoche einen zweistündigen „Block Carving“-Test zur Erfassung des manuellen Geschicks, in welchem mit Hilfe eines Messers ein vorgegebenes Muster in einen Kalkblock geschnitzt werden musste. Die Zulassung der Studierenden erfolgte regulär über den in den USA eingesetzten Studieneignungstest DAT und den Schulabschlussnotendurchschnitt GPA. Die Auswertung der Ergebnisse zeigte, dass von 23 Studierenden, die aufgrund ihrer Resultate im „Block Carving“-Test keine Zulassung zum Studium erhalten hätten, nur vier zu den zehn Prozent der Leistungsschwächsten in den Laborkursen der ersten zwei Studienjahre zählten. 33 Prozent der Studierenden, die den „Block Carving“-Test nicht bestanden hätten, wurden zudem durch Beurteilungen von Lehrenden anhand der Kriterien Kompetenz, Persönlichkeit und Potential als zukünftiger Zahnarzt auf Fragebögen als geeignet eingestuft. Mittels des „Block Carving test“-Tests konnten die zehn Prozent der leistungsschwächsten Studierenden

anhand der Durchschnittsnoten aus fünf Laborkursen der Zahnerhaltungskunde nicht signifikant ( $p > .05$ ) vorhergesagt werden. Interessant ist außerdem die Randnotiz der Autoren, dass Studierende, die den „Block Carving“-Test nicht bestanden hätten, eine geringere Punktzahl beim räumlichen Vorstellungsvermögen erfassenden DAT-Subtest PAT besaßen, was auf einen positiven Zusammenhang zwischen beiden Fähigkeitsarten schließen lässt.

### **2.3.4.3 Kognitive Stile**

Kognitive Stile beschreiben bipolare, intraindividuelle Eigenarten und Präferenzen der Informationswahrnehmung und -verarbeitung. Zu den bekanntesten, empirisch untersuchten zählen „Impulsivität vs. Reflexivität“ oder „Feldabhängigkeit vs. Feldunabhängigkeit“ (Hasselhorn & Gold, 2009). Der kognitive Stil „Feldabhängigkeit vs. Feldunabhängigkeit“ könnte für die Vorhersage des praktischen Studienerfolgs relevant sein, da die Vermutung bestand, individuelle Unterschiede in der Wahrnehmung und Verarbeitung von Informationen würden einen Einfluss auf die Leistungen in den Laborkursen der Zahnmedizin ausüben. Das Konstrukt „Feldabhängigkeit vs. Feldunabhängigkeit“ bezeichnet die Fähigkeit geometrische Figuren, die in zunehmend komplexere Figuren eingebettet werden, wieder zu erkennen, d.h., visuell zu isolieren. In Abhängigkeit des Erfolges bzw. Misserfolges die eingebettete Figur zu erkennen, gilt eine Person als feldunabhängig bzw. feldabhängig (Witkin, Lewis, Hertzman, Machover, Meissner & Wapner, 1954). Feldunabhängige Studierende erreichten im Vergleich zu feldabhängigen höhere Punktzahlen im DAT-Subtest PMAT und im „Mirror-Tracing-Test“, der die visio-motorische Koordination misst, weshalb die Involvierung vergleichbarer kognitiver Wahrnehmungsprozesse in beiden Testverfahren vermutet wurde. Falls für diese Prozesse auch Ähnlichkeiten in den Lernkurven bestünden, so könnte darin ein Indiz gesehen werden, dass Studierende mit Feldunabhängigkeit rascher ein asymptotisches Lernlevel bei der Aufgabenbearbeitung in den vorklinischen Laborkursen erreichen, um auf Problemstellungen schneller andere psychomotorischen Techniken und Methoden anzuwenden (Wilson, Suddick & Shay, 1981; Wilson, Suddick, Shay & Hustmyer, 1981). Empirische, die Überlegung stützende, Ergebnisse wurden beispielsweise 1982 in der Studie von Suddick, Yancey, Devine und Wilson erbracht oder ein Jahr später von Suddick, Yancey und Wilson (1983). Linder, Janus, Bauer und Dishman (1991) konnten ebenfalls in einer Untersuchung zeigen, dass feldunabhängige Studierende

bessere Noten in den vorklinischen Laborkursen erzielten und geschlechtsspezifische Unterschiede ausschließen.

Der von Evans und Dirks (2001) eingesetzte Eingebettete-Figuren-Test zur Messung des kognitiven Stils „Feldabhängigkeit/-unabhängigkeit“ wies eine signifikant schwache Beziehung zu der Durchschnittsnote im Laborkurs des ersten Semesters auf ( $r = -0.300$ ,  $p < .05$ ). Feldunabhängige Studierende hätten nach Ansicht der Autoren weniger Schwierigkeiten, individuelle Bestandteile in komplexen Strukturen wahrzunehmen, wie es beispielsweise bei Zahnprothesen der Fall ist. Ranney et al. (2005) sahen in ihrem Review den Nutzen des Eingebettete-Figuren-Tests zur Auswahl von Studierenden als begrenzt an, da Übung die Testergebnisse stark verzerren könnte.

#### **2.3.4.4 Psychomotorik**

##### **2.3.4.4.1 Frühere Studien**

Fleishman und Hempel vertraten schon 1955 den Standpunkt, dass die Vorhersage individueller Leistungen bei komplexen Aufgaben eine Kombination von Fähigkeiten erfordert. Dem folgend wurde in Studien zur Vorhersage des praktischen Studienerfolgs in der Zahnmedizin nicht hinterfragt, ob eine einzelne Fähigkeit bei der Vorhersage des praktischen Studienerfolgs gegenüber anderen hervorsteicht, sondern es wurde der Ansatz verfolgt, mehrere psychomotorische Fähigkeiten in Beziehung zur praktischen Studienleistung zu setzen. Ziel war die Identifikation der Fähigkeiten, die alle für eine erfolgreiche Bewältigung der Aufgaben in den praktischen Kursen erforderlich sind, wobei aber gewisse Grenzen des Möglichen berücksichtigt werden müssen. Die Leistung der zahnmedizinischen Abschlussprüfungen, insbesondere die Note der praktischen Abschlussprüfung, könnten daher nicht zuverlässig durch psychomotorische Fähigkeiten vorhergesagt werden, wie die Auswertung der Studienergebnisse von Tsutsui und Rich (1981) ergab. Als Ursache vermuteten die Autoren zum einen die zu komplexe Notenbildung der Abschlussprüfung, in welche auch Faktoren, wie das Patientenmanagement der Studierenden eingingen, die in keiner Beziehung zu psychomotorischen Fähigkeiten stünden und zum anderen die psychometrische Güte bestehender Testverfahren, die nicht ausreichend war, um komplexere Fähigkeitsniveaus zu diskriminieren. In der Untersuchung von Tsutsui und Rich absolvierten 193 Zahnmedizinische Studierende der Jahrgangsklassen von 1977 bis 1980 den „Minnesota Paper Form Board Test“, der mechanisches Verständnis als auch perzeptiv-

räumliche Fähigkeiten erfasst, indem anhand von zweidimensionalen Formstücken die entsprechende, komplette Figur vorgestellt werden musste.

Boyle und Santelli (1986) sahen die Vermittlung der benötigten psychomotorischen Fertigkeiten in den Laborkursen als eine der größten Herausforderungen in der zahnmedizinischen Lehre an. Der motorische Aspekt beim Erlernen der zahnmedizinischen Untersuchungs- und Behandlungstechniken bezieht sich auf das koordinierte Ausführen von Bewegungen, wofür ein hohes Maß an räumlicher Wahrnehmung benötigt wird, um Formen, Tiefen und räumliche Beziehungen von Objekten richtig zu unterscheiden, sodass mehrere Fähigkeiten involviert sein müssen, vor allem aber räumliches Vorstellungsvermögen und manuelles Geschick. Bewerber mit Defiziten in diesen Fähigkeitsarten sollten Boyle und Santelli zufolge weniger bei der Studienplatzvergabe berücksichtigt werden, da ihre Chancen das Studium zu absolvieren kleiner und der Lehraufwand zugleich höher ausfallen würde. Um ihre inhaltlichen Ausführungen empirisch zu überprüfen, ließen Boyle und Santelli in der zweiten Studienwoche 76 Zahnmedizinistudierende den „Crawford Small Parts Dexterity Test“ absolvieren, einen Leistungstest zur Augen-Hand-Koordination, in welchem unter Zeitdruck auf einer Bordseite Pins gesteckt und auf der anderen Schrauben eingedreht werden, der von 71 Studierenden nach vier Jahren wiederholt wurde. Zur durchschnittlichen Laborkursleistung der Vorklinik wiesen die Resultate des „Crawford Small Parts Dexterity Test“ schwache positive, aber signifikante Korrelationen auf (Pins:  $r = 0.29$ ,  $p < .05$ ; Schrauben:  $r = 0.25$ ,  $p < .05$ ). Die von den Studierenden erreichten Zulassungsergebnisse im PAT, einem DAT-Subtest zum räumlichen Vorstellungsvermögen, zeigten zur durchschnittlichen Laborkursleistung eine vergleichbare Beziehungen mit  $r = 0.24$  ( $p < .05$ ) auf. Die Testwiederholung ergab zudem eine signifikante Verbesserung der psychomotorischen Fähigkeiten ( $p < .01$ ) insbesondere bei Studierenden, die zu Beginn des Studiums niedrigere Testwerte erzielten, was Boyle und Santelli als Bestätigung für den Bedarf psychomotorischer Fähigkeiten im Zahnmedizinistudium interpretierten und sie in ihrer Empfehlung bestärkte, für die Auswahl von Studierenden in der Zahnmedizin Tests einzusetzen, die psychomotorische Fähigkeiten erfassen. Aus der Korrelationstabelle im Ergebnisteil ist außerdem ersichtlich, dass zwischen der Schraubenaufgabe des „Crawford Small Parts Dexterity Test“ und dem PAT eine schwache, positive Korrelation in Höhe von 0.31 bestand ( $p < .01$ ), die sich aber nicht für die Steckaufgabe des „Crawford Small Parts Dexterity Test“ ergab ( $r = 0.07$ , *n.s.*).

Kao, Ngan, Wilson und Kunovich (1990) zufolge ist ein Drahtbiegetest sehr gut geeignet, um die Leistung in den Laborkursen der zahnmedizinischen Vorklinik vorherzusagen, da neben manuellen Fähigkeiten auch Wahrnehmungsprozesse und das Erlernen von räumlichen Beziehungen zum erfolgreichen Bearbeiten der Drahtbiegeaufgabe erforderlich sind, die aber von Zulassungskriterien wie dem GPA nur in sehr geringem Ausmaß erfasst werden. In ihrer Untersuchung mit 105 Erstsemestern der Zahnmedizin verwendeten sie einen Drahtbiegetest mit einer Durchführungszeit von einer Stunde, bei dem der Gebrauch einer Spitzzange zum Biegen der Drähte zugelassen war. Die Studierenden hatten eine Stunde Zeit zum Üben und anschließend freien Zugang zum Muster, nachdem ihnen das korrekte Biegen des Drahtes vorgeführt wurde. Die gebogenen Drähte wurden von zwei Ratern nach Form, Ebenheit, planem Aufliegen des Drahtes, Genauigkeit der Schleifen und Genauigkeit des Anliegens der Schleifen auf vierstufigen Skalen bewertet.

Als Leistungskriterien wurden die Punktskoren von sieben Übungen aus den vorklinischen Laborkursen herangezogen, die das Schnitzen eines Wachszahns, das Präparieren einfacher und komplexer Zahnkronen, das Anfertigen eines Provisoriums und einer Brücke sowie das Präparieren, Herstellen und Einsetzen von Amalgam-, Kunstharz und Goldfüllungen umfassten. Die Ergebnisse des Drahtbiegetests korrelierten positiv moderat und signifikant ( $p < .01$ ) mit den Laborkursaufgaben, außer mit der Aufgabe Präparieren und Einsetzen von Amalgamfüllungen ( $r = 0.236$ ,  $p < 0.1$ ). Für die komplementär zum Drahtbiegetest in die Auswertung einbezogenen Ergebnisse des Studierfähigkeitstests DAT sowie für die GPA Punktwerte der Studierenden fanden Kao et. al. nur nichtsignifikante Korrelationen um Null zu den Laborkursaufgaben. Lediglich der räumliches Vorstellungsvermögen erfassende DAT-Subtest PAT wies eine schwache, positive Korrelation mit der Aufgabe Präparieren einer einfachen Krone und Erstellen eines Provisoriums auf ( $r = 0.248$ ), die signifikant ausfiel ( $p < .01$ ). Die Kombination aus GPA und den DAT-Untertests „Academic Average“ sowie PAT ergab in einem regressionsanalytischen Modell ein  $R^2$  von 0.142. Die Hinzunahme der Drahtbiegetestergebnisse führte zu einem Anstieg des Determinationskoeffizienten  $R^2$  auf 0.305, d.h., 30.5 Prozent der durchschnittlichen Aufgabenleistung konnten insgesamt aufgeklärt werden.

Kao et al. betonten in ihrer Schlussfolgerung nicht nur die höhere prädiktive Stärke des Drahtbiegetests gegenüber den klassischen Zulassungskriterien DAT und GPA zum Zahnmedizinstudium, sondern empfahlen aufgrund der inkrementellen Validität von

16.3 Prozent den zusätzlichen Einsatz von Drahtbiegetests zur Studierendenauswahl, damit in die Zulassungsentscheidung der zukünftige Erfolg in den vorklinischen Laborkursen stärker berücksichtigt wird. Ein weiteres Argument für den Einsatz eines Drahtbiegetests im Auswahlverfahren sei dessen deutlich höhere Testökonomie im Vergleich zu Wachsschnitt- bzw. Kalkschnitttests.

#### **2.3.4.4.2 Aktuellere Studien**

Aktuellere Studien empfehlen nicht den Einsatz psychomotorischer Testverfahren zur Studierendenauswahl in der Zahnmedizin. Kritisiert wurde nicht nur deren schwache Vorhersage der praktischen Studienleistungen, sondern auch die geringe Testsensitivität der Verfahren. Vor allem Studierende, die in den Laborkursen nur ungenügende Leistungen erzielten, konnten nicht im Vorfeld der Studierendenauswahl identifiziert werden. In Kanada setzte Oudshoorn (2003) die DAT-Subtests PAT und „Carving Dexterity“ (vgl. Unterkapitel 2.3.3) als Prädiktoren der Laborkursleistungen im ersten Studienjahr ein, in denen das Präparieren von Zähnen und das Einsetzen von Amalgamfüllungen am Phantomkopf geübt wurden. In Kombination konnten beide Untertests 19.58 Prozent der durchschnittlichen Laborkursleistung vorhersagen, als Einzelprädiktoren waren es 16.56 Prozent durch den manuelles Geschick messenden „Carving Dexterity“ und 12.02 Prozent durch den räumliches Vorstellungsvermögen erfassenden PAT. Oudshoorn schlussfolgerte auf Basis seiner Ergebnisse, dass die durch beide Subtests erfassten psychomotorischen Fähigkeiten aufgrund der schwachen Prädiktion praktischer Studienleistungen im ersten Studienjahr nur von geringem Nutzen sind.

Mit der Zielsetzung, die in den zahnmedizinischen Laborkursen benötigten psychomotorischen Fähigkeiten zu identifizieren, führten in Spanien de Andres, Sanchez, Hilgado und Diaz (2004) eine umfangreiche Testbatterie mit 306 Zahnmedizinistudierenden des ersten bis vierten Studienjahrs durch. Die teilnehmenden Studierenden wurden auf Basis ihrer Laborkursnoten in die Leistungsklassen hoch, mittel und gering eingestuft, die sich entsprechend in den Ergebnissen der Testbatterie widerspiegeln sollten, welche im letzten Monat des akademischen Jahres durchgeführt wurde. Die Untertests „tracing“ und „dotting“ des MacQuarrie-Tests, die räumliche Intelligenz und manuelle Geschwindigkeit messen, konnten signifikant ( $p < .05$ ) Studierende mit einer hohen Laborkursleistung ermitteln, was aber leistungsschwächere ausschließt, die zum regulären Lehrangebot Förderung benötigen. Mittels des

Feinmotorik erfassenden „O‘ Connor Tweezer Test“ schien lediglich identifizierbar zu sein, welche Studierende wahrscheinlich mehr Probleme in den Laborkursen haben könnten. Der räumliches Vorstellungsvermögen erfassende „test on rotation of solid shapes“ und der „indirect vision test“, welcher die Fähigkeit der Studierenden abbildet, mit indirekter Sicht umzugehen sowie der den kognitiven Stil „Feldabhängigkeit vs. Feldunabhängigkeit“ prüfende Eingebettete-Figuren-Test (vgl. Unterkapitel 2.3.4.3) konnten keine der drei Leistungsklassen replizieren. De Andres et al. betonten die Bedeutung der frühzeitigen Identifikation leistungsschwacher Studierende in den praktischen Kursen durch psychomotorische Tests, um rechtzeitig spezifische Förderkurse anbieten zu können, hielten diese aber aufgrund ihrer Studienergebnisse nicht zur Auswahl von Studienplatzbewerbern geeignet.

### **2.3.5 Methodische Kritik**

#### **2.3.5.1 Psychometrische Güte des praktischen Studienerfolgs**

##### **2.3.5.1.1 Noten als Kriterien der praktischen Studienleistung**

Die Noten in den praktischen Kursen der Zahnmedizin scheinen wenig akzeptable Kriterien des praktischen Studienerfolgs zu sein, da kognitive Fähigkeiten und Persönlichkeitseigenschaften in ihre Bildung mit einfließen, die aber nicht in Beziehung zu psychomotorischen Fähigkeiten stehen (Thompson, 1947). Natkin und Guild (1967) sowie später auch Spratley (1992) vermuteten in der inkonsistenten Bewertungen der Aufgabenleistungen in den vorklinischen Laborkursen, die oftmals ohne definierte und standardisierte Kriterien erfolgten, eine mögliche Ursache für die kontroversen Studienergebnisse zur Prädiktion der praktischen Studienleistung in der Zahnmedizin, die Phipps, Fishman und Scott (1968) zufolge auch für die Noten in den klinischen Laborkursen der Zahnmedizin bestünden.

Die Kritik an der psychometrischen Güte der Noten wurde in etlichen Studien aufgegriffen, in denen man versuchte, zuverlässigere Methoden und Instrumente zur Operationalisierung der praktischen Studienleistung zu finden. Beispielsweise Evans und Dirk (2001) verwendeten anstatt Noten eine Punkteskala von null bis einhundert, um die Arbeiten in den Laborkursen des ersten Studienjahr zu bewerten. Silvestri, Cohen und Singh (1979) stellten in ihrer Untersuchung die Frage, ob und falls ja, inwieweit feinmotorische Fähigkeiten in den Laborkursen des Zahnmedizinstudiums benötigt werden. Aufgrund der bestehenden Kritik an Noten als Maß der praktischen Studienleistung, führten Silvestri et al. mit 154 Zahnmedizinstudierenden vor Beginn

der Laborkurse die Übung „Stylized Class II Cavity Preparation“ durch, in welcher ein Muster mittels Instrument in einen Plastikblock geschnitzt wird, die nach 18 Monaten mit 122 der 154 Studierenden wiederholt wurde. Die Auswertung der Prä-Post-Messungen ergab eine signifikante Verbesserung der Übungsleistung, welche sich für alle drei Rater bestätigte ( $\chi^2_{R1} = 29$ ,  $\chi^2_{R2} = 35$ ,  $\chi^2_{R3} = 41$ ,  $p < .001$ ). Zudem benötigten die Studierenden bei der Übungswiederholung nur noch die Hälfte der ursprünglichen Bearbeitungszeit zur Fertigstellung der Schnitzarbeiten. Somit konnten Silvestri et al. eine Verbesserung der Feinmotorik in den vorklinischen Laborkursen nachweisen, ohne Limitationen durch die Güte der Kursnoten zu erfahren.

#### **2.3.5.1.2 Faktoren der zahnmedizinischen Studienleistung**

Einen anderen Ansatz zur Operationalisierung der Studienleistung verfolgten Potter, McDonald und Sagraves (1982a). Die von ihnen durchgeführte Faktorenanalyse konnte die Leistungen der ersten drei Studienjahre in 57 zahnmedizinischen Fakultäten auf die Faktoren „Basic Science“, „Dental Science“ und „Manual Skills“ reduzieren, um valide und zuverlässige Kriterien zur Beurteilung der Studienleistung zu gewinnen. In einer Folgestudie versuchten Potter, McDonald und Sagraves (1982b) Prädiktoren dieser drei Leistungsfaktoren zu identifizieren. Hierfür setzten die Autoren die Ergebnisse des Studierfähigkeitstest DAT und des Schulabschlussnotendurchschnitts GPA von Studierenden der Abschlussklassen 1980 ( $n = 123$ ) und 1981 ( $n = 133$ ) in Beziehung zu den von ihnen ermittelten Leistungsfaktoren des Zahnmedizinstudiums. Die retrospektive Analyse ergab für beide Jahrgänge eine schwache, positive und signifikante Beziehung zwischen dem DAT-Subtest PMAT und dem Faktor „Manual Skills“ (1980:  $r = 0.23$ ,  $p < .05$ ; 1981:  $r = 0.34$ ,  $p < .05$ ). Ebenfalls konnte der Faktor „Basic Science“ zuverlässig durch den DAT und GPA vorhergesagt werden, jedoch nicht der Faktor „Dental Science“, weshalb Potter et al. anregten, die Kriterien der Zulassungsverfahren zu ändern. Der vielversprechende Ansatz von Potter et al. wurde jedoch in nachfolgenden Studien nur noch von Potter und McDonald (1985) aufgegriffen, aber ansonsten nicht weiter fortgeführt.

#### **2.3.5.2 Statistische Auswertungsverfahren**

Der Einsatz multipler Regressionsmodelle in Studien zur Prädiktion der praktischen Studienleistungen erfuhr Kritik wegen der geringen Stichprobengrößen, niedriger Reliabilitäten der Kriterien und Einschränkungen in der Datenvariabilität, da nicht

zugelassene Bewerber von den Untersuchungen ausgeschlossen waren. Weitere Kritik bestand in den fehlgeschlagenen Kreuzvalidierungen prädiktiver Variablenbeziehungen, der Auswahl redundanter Prädiktoren und des akademischen Charakters der Leistungskriterien (Chambers, 1972). Auch Potter und McDonald (1985) hielten Korrelations- und Regressionsanalysen als Verfahren weniger zur Prognostik der praktischen Studienleistung in der Zahnmedizin geeignet, da die Ergebnisse in Abhängigkeit der Stichprobengröße sehr stark variieren könnten. Mehrere Prädiktoren korrelieren zudem immer unterschiedlich in einer multiplen Regressionsanalyse miteinander, wodurch der Einfluss der Prädiktoren auf das Kriterium verzerrt werden kann, sodass die Ergebnisse nur bedingt interpretierbar bzw. die Modelle miteinander vergleichbar sind. Darüber hinaus kann in einem Regressionsmodell immer nur ein Kriterium, meist über Durchschnittswerte gebildet, implementiert und betrachtet werden. Ebenfalls kann die Reihenfolge der Prädiktoren einen Einfluss auf den Anteil der erklärten Varianz haben, wobei in einem regressionsanalytischen Modell immer nur Prädiktoren und Kriterium spezifiziert, aber nie ein inhaltliches Modell formuliert werden muss. Im Vergleich dazu erlauben kovarianzbasierte Strukturgleichungsmodelle latente Variablen über beobachtbare Variablen, d.h. Indikatoren, zu operationalisieren, für die auch reziproke Wirkungszusammenhänge ermittelt und Messfehler berücksichtigt werden können. Für die Prädiktion der Studienleistung erweist sich die Betrachtung der Fehlervarianzen in Beziehung zu der durch die Modellvariablen erklärten Varianz als Vorteil, da der verbleibende Varianzanteil indiziert, ob möglicherweise andere Ursachen von Bedeutung sein könnten, die bisher nicht im Modell berücksichtigt sind.

Aufgrund ihrer Kritik an den statistischen Verfahren der ersten Generation stellten Potter und McDonald ein kovarianzbasiertes Strukturgleichungsmodell zur Vorhersage der akademischen und praktischen Studienleistungen in der Zahnmedizin auf, in dem die Fähigkeiten und Kenntnisse von 256 Studierenden über die latenten Variablen „Manual Ability Construct“, „Science Aptitude Construct“ und „Verbal-Reading Construct“ operationalisiert und in Beziehung zur Studienleistung gesetzt wurden. Das Modell verfügte insgesamt über einen akzeptablen Fit ( $\chi^2 = 82.4$ ,  $df = 67$ ,  $GoF = 0.94$ ,  $RMSR = 0.04$ ), wobei nichtsignifikante Beziehungen in der Modellevaluation entfernt wurden. Aus Gründen der Übersichtlichkeit ist nur der Auszug zur praktischen Studienleistung dargestellt. Zur Operationalisierung der latenten, exogenen Variable „Manual Ability Construct“ verwendeten Potter und McDonald die Subtests PAT-2D

und PAT-3D des Studierfähigkeitstests DAT, die räumliches Vorstellungsvermögen für zwei- und dreidimensionale Objekte messen. Aufgrund seiner Ladung von  $\lambda = 0.40$  wurde der Leseverständnis prüfende DAT-Subtest „Reading Comprehension“ auch als Indikator der latenten Variable „Manual Ability Construct“ eingesetzt, auf die der PAT-3D die höchste Ladung mit  $\lambda = 0.74$  aufwies und die des PAT-2D  $\lambda = 0.49$  betrug. Die Fehlervarianzen der Indikatoren „Reading Comprehension“, PAT-2D und PAT-3D waren mit 0.64, 0.77 und 0.47 recht hoch, was folglich ihre Reliabilität in Frage stellte. Die Pfadstärke der latenten Variable „Manual Ability Construct“ auf die vorklinische Laborkursleistung, die anhand der Kursnoten operationalisiert wurde, betrug 0.56, deren Fehlervarianz einen Wert von 0.70 annahm und darauf hinwies, dass andere Fähigkeiten als nur räumliches Vorstellungsvermögen einen Einfluss auf die Leistung in den vorklinischen Laborkursen besitzen könnten.

### **2.3.5.3 Stichprobenbeschaffenheit**

Die Beschaffenheit der Stichproben in Untersuchungen zur Vorhersage der praktischen Studienleistung in der Zahnmedizin stand oftmals im Mittelpunkt methodischer Kritik. Bedingt durch die Zulassung zum Studium, kann das Auftreten einer „range restriction“ in den oftmals sehr homogenen und weniger als 100 Studierende umfassenden Stichproben nicht ausgeschlossen werden. Untersuchungsvariablen verfügen somit über einen kleineren Wertebereich, der nicht repräsentativ die gesamte Spannweite möglicher Werte enthält, die in der zugrundeliegende Population vorkommen, sodass statistische Effekte in ihrer Stärke verzerrt werden könnten (Sackett, Lievens, Berry und Landers, 2007). Die schwache Prognostik der praktischen Studienleistung durch Fähigkeiten könnte eine mögliche Folge aufgrund von „restrictions of ranges“ sein, weshalb ein allgemein akzeptiertes, häufig angewandtes Vorgehen zum Akquirieren einer repräsentativen Stichprobe in der Akkumulierung von Bewerbern oder Studierenden aus mehreren, aufeinander folgenden und vergleichbaren Jahren besteht (vgl. Coy et al. 2003; Giuliani et al., 2007). Oudshoorn (2003) beispielsweise bezog insgesamt 213 Erstsemesterstudierende aus den akademischen Jahren 1999 bis 2002 ein, deren Zulassungsdaten er mit denen von 735 Bewerbern verglich, um verringerte Effektstärken durch „range restrictions“ bei der Ergebnisinterpretation ausschließen zu können. Ob dadurch die benötigte Diversität in der Merkmalsausprägung von Fähigkeiten zur Vorhersage der praktischen Leistungen in den Laborkursen erreicht wird, ist allerdings zu bezweifeln.

Interessant ist auch der Ansatz von Raybould, Raggard und Norton (1983), die in ihrer Untersuchung Absolventen durch Fakultätsmitglieder mittels eines Fragebogens in den Punkten technische Fertigkeiten, Potential als Praktiker, Intelligenz und Einstellung zum Lernen beurteilten ließen. Die so in eine hohe und geringe Leistungsklasse unterteilten Studierenden absolvierten anschließend eine Testbatterie, zu denen beispielsweise der „Mirror Tracing test“ und das „Purdue Pegboard“ zählten. Zwischen beiden Gruppen aus jeweils zehn Absolventen zeigte sich aber kein signifikanter Unterschied in ihren psychomotorischen Fähigkeiten. Raybould et al. vermuteten als Erklärung, dass die in der Testbatterie eingesetzten Verfahren nicht geeignet waren, um die Feinheiten zahnmedizinischer Fertigkeiten zu identifizieren und gleichzeitig höhere kognitive und psychomotorische Fähigkeiten zu erfassen.

### **2.3.5.4 Validität der Testinstrumente**

#### **2.3.5.4.1 Gütekriterien der Klassischen Testtheorie**

Die in Studien zur Vorhersage der praktischen Studienleistung in der Zahnmedizin eingesetzten psychomotorischen Testverfahren sind in ihrer Güte umstritten. Die zur Messung der Fähigkeiten eingesetzten Instrumente hinterließen manchmal sogar den Eindruck, eher intuitiv und weniger nach psychometrischen Gütekriterien entwickelt worden zu sein (Fernandez-Pabon, 1968). Abgesehen von selbstentwickelten bzw. modifizierten Testinstrumenten konnten Yancosek und Howell (2009) aber auf Basis ihres Reviews zu psychomotorischen Testverfahren uneingeschränkt den „Minnesota Rate of Manipulation Test“, das „Purdue Pegboard“ und den „O'Connor Finger Dexterity Test“ empfehlen, während der „Crawford Small Parts Dexterity Test“ und das „Grooved Pegboard“ hingegen weniger ausreichend validiert sind und hinsichtlich eines Einsatzes mit Sorgfalt im Vorfeld geprüft werden sollten. Die Interrater-Reliabilitäten der Testverfahren überschritten die geforderte Höhe von 0.90 (Bortz & Döring, 2006), die Retest-Reliabilitäten lagen im Bereich zwischen 0.60 und 0.90. Die in der Literatur geforderten Reliabilitätswerte von 0.80 bis 0.90 wurden nur geringfügig vom „Grooved Pegboard“ mit 0.72 unterschritten, sind aber ansonsten als erfüllt anzusehen, beispielsweise vom „Minnesota Rate of Manipulation Test“ mit 0.95 oder auch vom „Crawford Small Parts Dexterity Test“ mit 0.89. Die von Yancosek und Howell aufgeführten externen Validitäten zu etablierten Tests lagen im Wertebereich zwischen 0.60 und 0.90, die nicht immer der in der Literatur geforderten Güte größer 0.70 entsprachen (Bühner, 2011).

#### 2.3.5.4.2 Inhaltliche Güte

Weinstein und Kiyak (1981) vermuteten eher die fehlende Abbildung klinischer Tätigkeiten in psychomotorischen Tests als Ursache für ihre geringe Vorhersagekraft der praktischen Studienleistung im zahnmedizinischen Studium. Der Umgang mit Instrumenten auf beschränktem Raum würde gut die Verhältnisse einer zahnmedizinischen Behandlung widerspiegeln, die spezifische Anforderungen an feinmotorische Fertigkeiten stellt, welche Weinstein und Kiyak in dem von ihnen entwickelten „Dental Dexterity Test“ (DDT) abbildeten. Mittels einer Pinzette mussten Stifte unter Zeitdruck in halbkreisförmig angeordnete Vertiefungen gesteckt und wieder entfernt werden, bei der die Halterung der Vertiefungen von einer nachgebildeten Mundhöhle umgeben war.

Die Durchschnittsnoten aus neun vorklinischen Laborkursen der ersten zwei Studienjahre und sieben klinischen Kursen der letzten zwei Studienjahre von 38 Zahnmedizinistudierenden verwendeten Weinstein und Kiyak, um die praktische Studienleistung zu operationalisieren. Zur praktischen Leistung in den vorklinischen Laborkursen zeigte der DDT moderat negative, signifikante Korrelationen ( $p < .05$ ), die für den zu Vergleichszwecken herangezogenen DAT-Subtest PAT schwach negativ und nicht signifikant ausfielen. In den Ergebnissen sahen die Autoren ihre Vermutung bestätigt, da der papierbasierte Test zum räumlichen Vorstellungsvermögen keine Aufgaben mit zahnmedizinisch relevanten Inhalten besitzt. Zwischen DDT und PAT berichteten die Autoren nur für das Auswertungskriterium Extraktionszeit eine signifikante, schwach negative Korrelation ( $r = -0.31$ ,  $p < .05$ ). Jedoch konnten Weinstein und Kiyak für ihre Untersuchung aufgrund der Stichprobengröße verringerte Effektstärken durch „restrictions of range“ nicht ausschließen, wodurch die Generalisierbarkeit der Ergebnisse vor allem für den PAT eingeschränkt wurde.

Testverfahren, die zur Messung des manuellen Geschicks Pinzetten einsetzen, sind nach Ansicht von Lundergan, Soderstrom und Chambers (2007) wenig zur Operationalisierung von manuellen Fähigkeiten zur Studierendenauswahl in der Zahnmedizin geeignet. In ihrer Langzeituntersuchung wählten Lundergan et al. von 134 Erstsemesterstudierenden der Zahnmedizin 50 zufällig aus, die zu Studienbeginn die Versionen Nr. 32022 und Nr. 18 des „O'Connor Tweezer Dexterity Test“ absolvierten. Während im Test Nr. 32022 unter Zeitdruck Metallstifte mittels Pinzetten in vorgefertigte Löcher einer Stecktafel platziert werden mussten, erforderte Test Nr. 18

das Übertragen von Metallstiften einer Stecktafelseite auf die andere, in welchem zusätzlich die Genauigkeit der Leistung bewertet wird, d.h., fallengelassene Pins führen zu Strafpunkten. Mit 49 der 50 Erstsemesterstudierenden wurde nur der „O’Connor Tweezer Dexterity Test“ Nr. 18 in den letzten Studienwochen wiederholt, da die Testversion Nr. 32022 von der „Johnson O’Connor Research Foundation“ nicht mehr eingesetzt wird. Die Noten der Laborkurse bzw. Kurse mit praktischen Arbeitsanteilen des ersten Studienjahres sowie die kumulierten Noten aller Labor- und klinischen Kurse bis zur Abschlussprüfung bildeten die Leistungskriterien, die aber auch theoretische Wissensaspekte beinhalteten, wobei sehr gute Leistungen mit maximal fünf Punkten und sehr schlechte mit null Punkten bewertet wurden.

Die Korrelationen beider Testversionen miteinander ( $r = 0.318$ ,  $p < .05$ ) wie auch zu den Laborkursleistungen im ersten Studienjahr fielen schwach positiv und überwiegend signifikant aus ( $p < .05$ ), jedoch erzielten die Erstsemesterstudierenden keine höheren Testergebnisse im Vergleich zur allgemeinen Bevölkerungsnorm. Die Wiederholung der Testversion Nr. 18 zum Studienende ergab zudem keinen signifikanten Leistungsunterschied zu den Ergebnissen der Testung zu Studienbeginn. Der zu Vergleichszwecken eingesetzte, räumliches Vorstellungsvermögen erfassende, DAT-Subtest PAT besaß einen höheren positiven Zusammenhang zu der kumulierten Leistung aller praktischen Kurse ( $r = 0.423$ ,  $p < .05$ ) als die Tests Nr. 32022 ( $r = 0.279$ ,  $p < .05$ ) und Nr. 18 ( $r = 0.260$ ,  $p < .05$ ), zu denen der PAT schwach positive, signifikante Korrelationen mit 0.245 und 0.045 aufwies. Zu den Kursen des ersten Studienjahres zeigte der PAT ebenfalls überwiegend schwach positive, aber signifikante Beziehungen auf ( $p < .05$ ). Ein Einfluss des räumlichen Vorstellungsvermögens auf den Zusammenhang zwischen den „Tweezer Dexterity Tests“ und der praktischen Studienleistung konnte durch berechnete Partialkorrelationen ausgeschlossen werden. Eine Erklärung der Ergebnisse oder alternative Strategien bzw. Testverfahren zur Studierendenauswahl in der Zahnmedizin wurden von Lundergan et al. nicht aufgezeigt.

### **2.3.6 Fertigkeitserwerb in den vorklinischen Laborkursen**

#### **2.3.6.1 Vielfalt zahnmedizinischer Fertigkeiten**

Die hohe Anzahl der in den Laborkursen benötigten motorischen Fertigkeiten zur Zahnuntersuchung und Behandlung könnte eine Ursache für die inkonsistenten Studienergebnisse zur Prädiktion des praktischen Studienerfolgs sein, welche entweder eine Vielzahl an einzelnen Fähigkeiten oder eine Variation dieser erfordern würde.

Damit ginge die Schwierigkeit einher, die ausschlaggebenden Fähigkeiten zu erfassen, wodurch in Studien die geringe Prädiktionsstärke der Leistung in den vorklinischen Laborkursen bedingt sein könnte. Wilson und Husak (1988) reduzierten daher in ihrer Primärstudie das in Laborkursen erforderliche Spektrum an spezifischen motorischen Fertigkeiten, indem von 128 Studierenden 33 zufällig ausgesuchte zwei Aufgaben zu Studienbeginn absolvierten, die das Entfernen von Belägen auf der Wurzeloberfläche („scaling“) und das Glätten der obersten Zahnwurzelschicht („root planning“) erforderten. Die von geschulten Juroren nach definierten Kriterien beurteilten Aufgabenleistungen konnten weder durch den Schulabschlussnotendurchschnitt GPA, den Studierfähigkeitstest DAT oder durch soziodemographische Variablen, wie die Anzahl der Collegejahre oder das Familieneinkommen, vorhergesagt werden, sei es als Einzelprädiktor oder in Kombination. Die Aufgabenwiederholung nach 14 Wochen zeigte zwar eine signifikante Verbesserung der Aufgabenleistung, in der Wilson und Husak, bedingt durch den zwischenzeitlich absolvierten Laborkurs, einen Lernprozess motorischer Fertigkeiten bestätigt sahen, die aber weder durch die zuvor verwendeten Prädiktoren noch durch die Ergebnisse der ersten Testung zu Studienbeginn vorhergesagt werden konnten. Wilson und Husak vermuteten, dass die zugrundeliegenden Faktoren beim Erwerb motorischer Fertigkeiten entweder sehr unterschiedlich, zahlreich oder noch unbekannt sind, wobei zukünftig zu klären wäre, welche zahnmedizinischen Fertigkeiten welche Fähigkeiten erfordern und ob diese in frühen Lernphasen ähnlich geprägt sind wie in späteren.

#### **2.3.6.2 Frühe Phase des Erlernens zahnmedizinischer Fertigkeiten**

Zahnmedizinstudierende des ersten Studienjahrs werden von Chambers und Glassman (1997) als Anfänger oder Novizen beschrieben, die beim Erlernen motorischer Fertigkeiten einer zahnmedizinischen Behandlung, wie dem Präparieren eines Zahns zum Einsetzen einer Füllung, in den praktischen Kursen auf spezifisches und sofortiges Feedback der Lehrenden angewiesen sind. Aufgabeninstruktionen sollten daher für Anfänger sehr detailliert ausfallen und in kleine Zwischenschritte unterteilt werden (Hauser & Bowen, 2009). Für den Erwerb psychomotorischer Fertigkeiten ist insbesondere zu Beginn der zahnmedizinischen Ausbildung praktischer Fertigkeiten zu beachten, dass die Lernsituation den Anforderungen der Anwendungssituation entsprechen sollte, um den Lernerfolg zu gewährleisten (Knight, Guenzel & Fitzgerald, 1990). Als psychomotorische Fähigkeiten, die einen Einfluss auf das frühe Stadium des

Erwerbs von zahnmedizinischen Fertigkeiten ausüben, konnten Waldman, Macdonald und Wilson (1995) die Augen-Hand-Koordination, die Präzision der Handführung und die Zielgenauigkeit identifizieren. 45 Studierenden absolvierten zu Studienbeginn eine umfangreiche psychomotorische Testbatterie, die neben dem „Grooved Pegboard“, den „Purdue Hand Precision Test“ auch den „O'Connor Tweezer Dexterity Test“ umfasste, während die praktische Studienleistung anhand von drei Instrumentenaufgaben im ersten Trimester erfasst wurde, die eine Untersuchung der Zähne und das Entfernen der Beläge von der Wurzeloberfläche erforderten. Nur der „Purdue Hand Precision Test“ zeigte als einziger Test einen signifikanten Zusammenhang zu den praktischen Instrumentenaufgaben ( $0.17 \leq R^2 \leq 0.22$ ). Die Identifikation der in späteren Stadien des Erwerbs zahnmedizinischer Fertigkeiten ausschlaggebenden Fähigkeiten folgte einem anderen Ansatz, der im nachfolgenden Abschnitt vorgestellt wird.

### **2.3.6.3 Lernphasen zahnmedizinischer Fertigkeiten**

Chambers (1993) unterteilte das Lernen motorischer Fertigkeiten von Zahnmedizinstudierenden über das gesamte Studium in die Kompetenzphasen „novice“, „beginner“, „competent“, „proficient“ und „expert“. Für jede Phase existiert ein definierter Kriterienkatalog an benötigten Kompetenzen, zu präferierende didaktische Methoden und zur Beurteilung des Könnens einzusetzende Evaluationsinstrumente. Während beispielsweise Studierende der Phase „expert“ selbstständig mit Skripten neue Behandlungstechniken erlernen, im Umgang mit Patienten Erfahrung besitzen und nur selten spezifisches Feedback von Dozenten brauchen, sind Studierende in der Phase „beginner“ sehr auf Demonstrationen und häufiges Feedback von Lehrenden angewiesen, um Behandlungstechniken am Modell zu erlernen, zu denen unter anderem das Einsetzen einer Zahnfüllung zählt. In der Kompetenzphase „novice“ erweisen sich Prüfungen in Form von Tests bzw. Vorführungen als geeignete Instrumente zur Beurteilung des Lernstands, während in der letzten Stufe „expert“ auch das Urteil des Studierenden über die eigene Arbeit herangezogen werden kann.

Chambers und Geissberger (1997) konnten die Kompetenzstufen jedoch empirisch nicht bestätigen, da sich zwischen Studierenden der Phasen „beginner“ und „competent“ kein Unterschied in der Leistung der Untersuchungsaufgabe zeigte, die das Einsetzen einer okklusalen Amalgamfüllung an einem Typodonten, einem frei zugängigen Gebissmodell, erforderte. Die Kompetenzstufe „beginner“ war dabei von fünf Studierenden repräsentiert, die den vorklinischen Operationstechnikkurs bestanden

hatten, während sechs Studierende, die einen Monat vor Abschluss des Studiums standen, die Phase „competent“ repräsentierten. Die „beginner“-Studierenden verwendeten aber deutlich mehr Zeit für Verfeinerungen und Nachbesserungen beim Einsetzen der Füllung, die keinen direkten Bezug zur eigentlichen Aufgabe hatten, als die Studierenden in der Phase „competent“. Chambers und Geissberger führten die geringere Effizienz auf die fehlende Erfahrung der Zahnmedizinierenden in den ersten Semestern zurück, die noch Schwierigkeiten haben, ihre eigenen Tätigkeiten zu beurteilen und somit auf Feedback von Lehrenden angewiesen sind.

Die von Chambers (1993) aufgestellten Kompetenzstufen basierten auf der „Motor Skills Theory“ von Fitts und Posner (1967), die für den Erwerb von Fertigkeiten die kognitive, assoziative und autonome Phase vorschlugen. Darauf beruht auch die aktuellere Theorie des Fertigkeitserwerbs von Ackerman (1988), die bereits in ersten Studien zur Vorhersage der praktischen Studienerfolgs aufgegriffen wurde, um die für den Erwerb praktischer, zahnmedizinischer Fertigkeiten bestimmenden Fähigkeiten in der Vorklinik (Gray & Deem, 2002a) und Klinik zu ermitteln (Gray & Deem, 2002b). Aufgrund der Komplexität des Modells von Ackerman erfolgt eine ausführliche Darstellung im nachfolgenden Unterkapitel 2.4.

### **2.3.7 Zusammenfassung und Resümee**

#### **2.3.7.1 Dental Admission Test**

Die Vorhersage der praktischen Studienleistung im Zahnmedizinstudium ist Gegenstand zahlreicher Studien gewesen, deren Ergebnisse ein sehr widersprüchliches Bild zeichnen. Die Frage, welche Fähigkeiten erforderlich sind, um die Aufgaben in den praktischen Kursen erfolgreich zu bewältigen, wurde zunächst zwar einheitlich, aber wenig präzise beantwortet, indem man allgemein motorische und mechanische Fähigkeiten für ausschlaggebend hielt. Die Einführung des amerikanischen Studierfähigkeitstests DAT in 1951 führte zu einer etwas konkreteren Definition der für die zahnärztlichen Tätigkeiten benötigten Fähigkeiten, die durch drei Untertests repräsentiert waren. Einerseits sollte der geschickte Einsatz von Händen und Fingern durch einen Kalkschnittstest im Untertest „Carving Dexterity“ erfasst werden und andererseits das Visualisieren von Mustern und Beziehungen im Untertest „Space Relations“ (bzw. „Space Visualization“) beinhalten, deren Ergebnisse im Untertest „Manual Average“ zusammengefasst wurden. Die widersprüchlichen Studienergebnisse (Manhold & Manhold, 1967) und inhaltliche Kritik (Brigante & Lamb, 1968) an den

DAT-Untertests zur Prädiktivität des praktischen Studienerfolgs führten zu Zweifeln und Überprüfungen ihrer Konstruktvalidität. Die Ergebnisse der Faktoranalysen von Dworkin (1970) oder Young (1972) ergaben Ladungen der DAT-Untertests „Carving Dexterity“, „Spatial Relations“ und „Manual Average“ auf einen gemeinsamen Faktor, anstelle auf die für den DAT postulierten Faktoren „Carving Dexterity“ und „Space Conception“. Zullo (1971) konnte hingegen die Konstruktvalidität beider DAT-Untertests bestätigen. Der Kalkschnittstest scheint das Geschick im Umgang mit Handinstrumenten zu erfassen, während der Untertest „Spatial Relations“ dreidimensionales räumliches Vorstellungsvermögen misst.

Aufgrund der geringen Testökonomie wurde letztlich 1972 der Kalkschnittstest durch den papierbasierten, perzeptiv-motorische Fähigkeiten erfassenden PMAT ersetzt. Die Gemeinsamkeit beider Verfahren bestand nach Ansicht der American Dental Association im visuellen Wahrnehmungsvermögen, von dem man annahm, ein valider Prädiktor zur Beurteilung der praktischen Leistung in den Laborkursen zu sein (Graham, 1972). Dieses Vorgehen entsprach zwar Empfehlungen empirischer Studien (z.B. Derevere, 1961; Hood, 1963), aber nachfolgende Studien konnten für räumliche Wahrnehmungsfähigkeiten keine Ergebnisse berichten, die adäquat zum Kalkschnittstest die praktischen Leistungen in den vorklinischen Laborkursen prädizierten (Bellanti et al., 1972; Wong et al., 1979; Wood, 1979). Der 1981 eingeführte „Perceptual Ability Test“ (PAT) ersetzte nicht nur die vorherigen Untertests im DAT zur räumlichen Wahrnehmung, sondern erfasst räumliches Vorstellungsvermögen anhand von 90 Multiple-Choice-Aufgaben auf mehrdimensionaler Ebene (Spratley, 1990). Die Konstruktgültigkeit des PAT wurde von Kramer et al. (1989) bestätigt, die eine prädiktive Validität für den DAT-Untertest zu den Noten im ersten vorklinischen Laborkurs von  $R^2 = 0.14$  und zu den Durchschnittsnoten mehrerer vorklinischer Laborkurse von  $R^2 = 0.19$  berichteten.

### **2.3.7.2 Fähigkeiten als Prädiktoren**

Die vor allem in Nordamerika geführte Debatte um die Validität des DAT führte maßgeblich in zahlreichen Studien zu Fragestellungen, die manuelles Geschick und räumliches Vorstellungsvermögen als Fähigkeiten zueinander in Konkurrenz stellten (Spratley, 1990). Als Vergleichskriterium beider Fähigkeiten diente die prognostische Güte der praktischen Studienleistung, wobei nicht zur Frage stand, dass ein Mindestmaß an manuellem Geschick und räumlichen Vorstellungsvermögen erforderlich ist, um

erfolgreich Zahnmedizin zu studieren und eine hohe Ausprägung beider Fähigkeiten sich für das Studium als förderlich erweist (Smith, 1976; Wong et al., 1979).

Trotz einer Vielzahl an Untersuchungen kann bis heute kein Konsens in der Antwort gegeben werden. Es finden sich mindestens genauso viele Studien, die manuelles Geschick geeigneter zur Vorhersage des praktischen Studienerfolgs beurteilen, als Studien, die sich dagegen aussprechen und räumliches Vorstellungsvermögen präferieren. So fanden z.B. Beck et al. (1978) positiv moderate, signifikante Korrelationen zwischen räumlichen Vorstellungsvermögen im Untertest „Space Relations“ des DAT und der Leistung in den vorklinischen Laborkursen, die aber von Smith (1989) in Großbritannien nicht bestätigt werden konnten, obwohl das gleiche Messinstrument verwendet wurde. Aktuellere Studien zeigten ein eher einheitliches Bild für die Beziehung zwischen räumlichem Vorstellungsvermögen und der praktischem Studienleistung in der Zahnmedizin, deren überwiegend schwachen Korrelationen kleiner 0.4 nicht immer als ausreichend für die Auswahl von Studienbewerbern angesehen werden (Coy et al., 2003). Vor allem das räumliche Vorstellungsvermögen dreidimensionaler Objekte erwies sich als zuverlässig für die Aufklärung der Leistungsvarianz in den zahnmedizinischen Laborkursen der Vorklinik (Gray & Deem, 2002a; Hegarty et al., 2009). Hierbei besteht aber ein geschlechtsspezifischer Verzerrungseffekt, d.h. Männer verfügen über ein besseres räumliches Vorstellungsvermögen als Frauen, worin Coy et al. (2003) zufolge ein Grund für die unzureichende Prädiktion der praktischen Studienleistung durch das räumliche Vorstellungsvermögen zu sehen wäre, sodass eine Auswahl von Studienbewerbern anhand dieser Fähigkeit nicht zu rechtfertigen sei.

Übungen, die es erforderten in Kalk oder Wachs vorgegebene Muster zu schnitzen bzw. mit Wachs einen Zahn zu modellieren, erwiesen sich als zuverlässige, wenn aber auch als schwache Prädiktoren der praktischen Studienleistung (Walcott et al., 1986; Wong et al., 1979). Zwar würden diese Verfahren zuverlässig manuelles Geschick im Umgang mit Handinstrumenten erfassen, sind aber sehr aufwendig durchzuführen. Hier boten Kao et al. (1990) mit einem Drahtbiegetest eine effizientere Verfahrensalternative an, die zudem die Leistungen in den vorklinischen Laborkursen besser prädizierte. Aktuellere Studien empfehlen nicht den Einsatz manueller Testverfahren zur Auswahl von Zahnmedizinistudierenden, da vor allem manuelle Fähigkeiten im Studium trainierbar seien, weshalb eine Auswahl aufgrund einer geringeren Ausprägung zu Studienbeginn nicht vertretbar wäre (Giuliani et al., 2007; Luck et al., 2000). Der Kritik,

manuelles Geschick wäre in diesen Untersuchungen nicht im Umgang mit Instrumenten operationalisiert, sondern als Feinmotorik bzw. Fingergeschicklichkeit ohne manipulative Fertigkeiten erfasst, die zudem nicht von psychomotorischen Fähigkeiten abgegrenzt wurden, sind die Ergebnisse von Gansky et al. (2004) gegenüber zu stellen. Der zu Studienbeginn absolvierte zweistündige „Block Carving“-Test konnte die leistungsschwachen Studierenden in den Laborkursen nicht zuverlässig identifizieren und in Bezug auf den DAT-PAT zu keiner Erhöhung der Prädiktion der praktischen Studienleistung führen. Daher folgten auch Gansky et al. dem aktuellen Kanon, welcher weniger manuelles Geschick zur Auswahl von Studienbewerber in der Zahnmedizin empfiehlt, sondern eher zur frühzeitigen Identifikation und gezielten Förderung Studierender, die erhebliche Schwierigkeiten bei der Bearbeitung der Aufgaben in den vorklinischen Laborkursen haben könnten.

Den auf Fleishman und Hempel (1955) zurückzuführenden Ansatz, die komplexen Aufgaben in den praktischen Kursen durch eine Kombination von Fähigkeiten vorherzusagen, folgten eine Vielzahl an Untersuchungen, die psychomotorische Fähigkeiten als ausschlaggebend für die praktische Leistung hielten. Zur Psychomotorik zählen Fähigkeiten, die benötigt werden, um sensorische Informationen und muskuläre Reaktionen zur Bewältigung einer bestimmten Aufgabe zu koordinieren, wozu neben manuellem Geschick und dem räumlichen Vorstellungsvermögen auch die Feinmotorik, Orientierungsreaktion, Wahrnehmungsgeschwindigkeit, Fingergeschicklichkeit, Reaktionszeit, Handdruck oder die Armbewegungsgeschwindigkeit zählen (Fleishman & Reilly, 1992; Johnston & Catano, 2002). Zur Identifikation der in den praktischen Laborkursen erforderlichen Fähigkeiten wurden in Studien oftmals umfangreiche, psychomotorische Testbatterien durchgeführt (z.B. Boyle & Santelli, 1986; de Andres et al., 2004), in denen aber primär Testverfahren zum räumlichen Vorstellungsvermögen und manuellen Geschick im Umgang mit Instrumenten zum Einsatz kamen. Zwar konnten die Verfahren eine signifikante Beziehung zur praktischen Studienleistung aufweisen, die aber aufgrund ihrer oftmals nur schwachen Ausprägung Kritik erfuhr (Oudshoorn, 2003). Daher wurde beispielsweise der „O‘ Connor Tweezer Test“ eher für den Einsatz zur Diagnostik von Studierenden empfohlen, die eine zusätzliche Förderung benötigen, um erfolgreich die vorklinischen Laborkursen zu absolvieren (de Andres et al., 2004). Der Frage, inwieweit manuelles Geschick und räumliches Vorstellungsvermögen in Beziehung zueinander stehen, wurde in Studien nicht nachgegangen und demzufolge auch nicht als eine mögliche Ursache für die

widersprüchlichen Untersuchungsergebnisse zur Aufklärung der praktischen Studienleistung in Betracht gezogen.

Neben manuellem Geschick und räumlichem Vorstellungsvermögen konnte auch der kognitive Stil „Feldabhängigkeit vs. Feldunabhängigkeit“ als Prädiktor der praktischen Leistungen in den Laborkursen identifiziert werden (Suddick et al., 1983). Studierende, deren Informationswahrnehmung und -verarbeitung als feldunabhängig einzuordnen sind, zeigten weniger Schwierigkeiten, individuelle Bestandteile in komplexen Strukturen wahrzunehmen, beispielsweise bei Zahnprothesen (Evans & Dirk, 2001). Der Einsatz des Eingebettete-Figuren-Tests zur Studierendenauswahl ist allerdings aufgrund von Übungseffekten kritisch zu hinterfragen (Ranney et al., 2005).

### **2.3.7.3 Methodische Kritik**

Eine weitere Vermutung für die kontroversen Ergebnisse zur Prädiktion der praktischen Studienleistung bestand in der Studienmethodik. Kritisiert wurden primär die geringe psychometrische Güte der zur Operationalisierung der praktischen Leistung verwendeten Laborkursnoten (Natkin & Guild, 1967; Phipps et al., 1968), die Beschaffenheit der Stichproben, in denen oftmals das Vorliegen einer „range restriction“ nicht ausgeschlossen werden konnte (Raybould et al., 1983) und die Güte der eingesetzten Testinstrumente zur Messung der Fähigkeiten. Oftmals wurden in den Studien selbst entwickelte oder Adaptionen bestehender Tests eingesetzt, die aber nicht hinreichend validiert waren (Fernandez-Pabon, 1968). Zudem ist insbesondere die inhaltliche Validität der verwendeten Testverfahren umstritten, in die zahnmedizinische Tätigkeiten nicht repräsentativ eingebunden sind (Weinstein & Kiyak, 1981). Kritisiert wurde Einsatz von Korrelations- und Regressionsanalysen als statistische Auswertungsverfahren, vor allem im Hinblick auf die Stichprobengrößen, sodass die Studienergebnisse nur bedingt miteinander vergleichbar sind (Potter & McDonald, 1985). Aber auch Studien, in denen die genannten Kritikpunkte aufgegriffen wurden, um methodische Einflüsse möglichst weitestgehend zu kontrollieren, konnten keine Ergebnisse präsentieren, aufgrund derer die Beziehung zwischen Fähigkeiten und praktischen zahnmedizinischen Fertigkeiten umfassend erklärt wurden.

#### **2.3.7.4 Fertigkeitserwerb in den vorklinischen Laborkursen**

Im Fertigkeitserwerb der zahnmedizinischen Techniken könnte ebenfalls eine mögliche Ursache für die inkonsistenten Ergebnisse zu den die praktische Leistung im Zahnmedizinstudium bedingenden Fähigkeiten liegen. Ansätze zur Reduktion der in den Laborkursen benötigten motorischen Fertigkeiten (Wilson & Husak, 1988) oder die Unterteilung in ein frühes und spätes Stadium des Fertigkeitserwerbs (Waldman et al., 1995) konnten zwar zu keiner Klärung der Fragestellung führen, leisteten aber dennoch einen wichtigen Beitrag. In nachfolgenden Studien wurden die unterschiedlichen Phasen des Erlernens motorischer Fertigkeiten aufgegriffen, um zu klären, welche Fähigkeiten zahnmedizinische Fertigkeiten in welchen Lernphasen bedingen und ob diese Beziehungen in frühen Phasen ähnlich geprägt sind wie in späteren (Gray & Deem, 2002a; b).

## **2.4 Vorhersage erlernter Leistung: Ackermans Theorie des Fertigkeitserwerbs**

Aufgrund der Vielzahl an verwendeten Begrifflichkeiten wird zunächst im Unterkapitel 2.4.1 die allgemeine Intelligenz definiert, um die im Unterkapitel 2.4.2 dargestellte Theorie des Fertigkeitserwerbs von Ackermann (1988) beurteilen zu können. Ackermans Modell zur Vorhersage erlernter Leistung wurde bisher nur in sehr geringem Umfang in die zahnmedizinische Studieneignungsdiagnostik eingebunden, weshalb die empirische Applikation der Theorie zusätzlich im Rahmen der Prädiktion beruflicher Leistung in Unterkapitel 2.4.3 geprüft wird. Ziel ist es zugleich, Analogien im Feld der Berufseignungsdiagnostik zur Auswahl von Zahnmedizinierenden aufzudecken.

### **2.4.1 Definition allgemeine Intelligenz**

Die Vielzahl an unterschiedlichen Definitionen zur Intelligenz trugen dazu bei, dass sowohl im Alltagsgebrauch als auch in der Fachliteratur mehrere Begrifflichkeiten, wie allgemeine kognitive Fähigkeit, allgemeine mentale Fähigkeit, Intelligenz oder IQ, verwendet werden, die letztlich aber austauschbare Synonyme für die allgemeine Intelligenz darstellen (Deary, Penke & Johnson, 2010). Allgemeine Intelligenz ist nach Hunter (1986) kurz definiert als Fähigkeit zu lernen. Gottfredsons (1997, S. 13) Definition von Intelligenz ist vergleichsweise spezifischer und weitgehend von Intelligenzforschern akzeptiert: „Intelligence is a very general capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience.“. Spearman (1904b) schlug vor, dass die allgemeine Intelligenz einen allgemeinen kognitiven Faktor widerspiegelt, welchen er als „g“ bzw. g-Faktor bezeichnet, der zugleich für jede Art kognitiver Aufgabe zutreffend sei. Deary et al. (2010, S. 202) unterstreichen die Bedeutung des g-Faktors: „[...] it accounts for a substantial large amount of variance, it is the source of most of the predictive power of cognitive tests and, [...], it is the locus of most of the genetic variance“.

### **2.4.2 Ackermans Theorie des Fertigkeitserwerbs**

Die 1988 von Ackerman vorgestellte „theory of ability determinants of skilled performance“ beinhaltet ein dreistufiges, hierarchisches Modell von Fähigkeiten zur Vorhersage des Erwerbs komplexer Fertigkeiten, welches im Wesentlichen auf der Theorie von Fitts (1964; Fitts & Posner, 1967) und der ACT-Theorie von Anderson

(1983) basiert. Mit Hilfe des in Abbildung 2.02 dargestellten dreistufigen Modells können individuelle Leistungsunterschiede in einer Aufgabe erklärt werden.

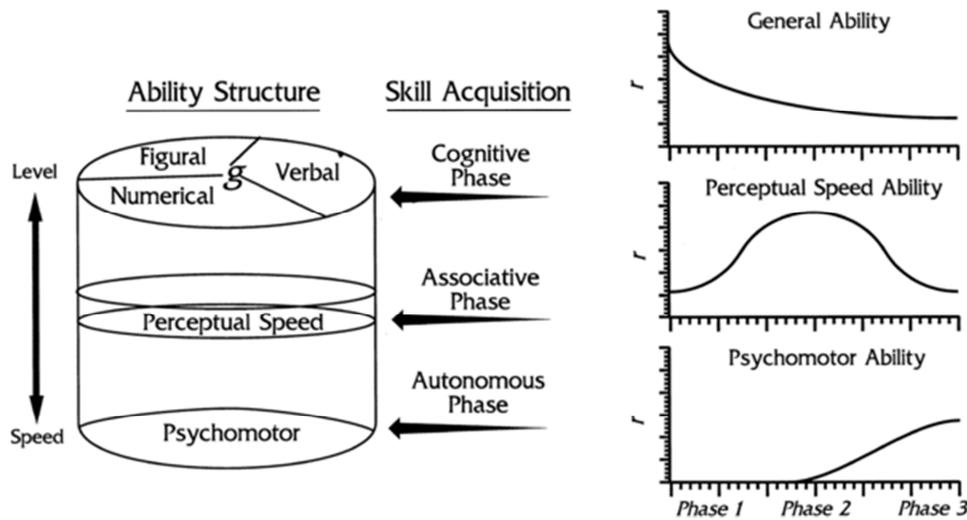


Abbildung 2.02: Ackermans Modell des Fertigkeitserwerbs (Quelle: Ackerman & Cianciolo, 2000)

Die Leistung wird in der kognitiven Phase durch eine allgemeine oder spezifische Fähigkeit bestimmt, die durch die Art des Aufgabeninhaltes erfordert wird. Die erste Phase ist charakterisiert durch einen hohen kognitiven Aufwand des Lernenden, da beispielsweise die Aufgabeninstruktionen und Aufgabenziele verstanden und Strategien zum Erreichen der Aufgabenziele entwickelt werden müssen. Sind diese Prozeduren abgeschlossen, erfolgt der Übergang in die zweite Phase, die Ackerman als die assoziative Phase bezeichnet. In dieser Phase ist die Leistung durch die Wahrnehmungsgeschwindigkeit ausschlaggebend beeinflusst, indem die Reiz-Reaktionsbeziehungen derart gefestigt werden, dass die Aufgabenleistung erhöht und die Fehlerhäufigkeit verringert wird. Die dritte und finale Phase des Fertigkeitserwerbs ist die autonome Phase, in der die Automatisierung der erlernten Fertigkeit erfolgt, sodass nur noch geringe oder gar keine Aufmerksamkeit zur Ausführung der Fertigkeit erforderlich ist. Die Aufgabenleistung ist in dieser Phase vor allem durch die psychomotorische Fähigkeit beeinflusst. Vor allem Müdigkeit, Angst, Motivation, Aufmerksamkeit und physische Voraussetzungen erwiesen sich als die motorische Leistung beeinflussende Faktoren (Schmidt & Wrisberg, 2004).

Ackermans „theory of ability determinants of skilled performance“ wurde in mehreren Studien und Experimenten bestätigt (Ackerman, 1988; Ackermann, 1990; Ackerman, 1992; Ackerman & Cianciolo, 2000; Eyring, Johnson & Francis, 1993; Gray & Deem, 2002b; Farrell & McDaniel, 2001; Kanfer & Ackerman, 1989; Keil & Cortina, 2001; Taatgen, 2002). Ackerman (1988) kritisiert selbst, dass in seinem Modell weitere

Einflüsse, vor allem Motivation, zur Erklärung individueller Leistungsunterschiede beim Fertigkeitserwerb nicht mit einbezogen sind. Diesem Kritikpunkt entgegneten Kanfer und Ackerman (1989) durch die Entwicklung des „Resource Allocation Model“, welches im Unterkapitel 2.4.2.3 dargestellt ist.

#### **2.4.2.1 Komplexität und Konsistenz von Aufgaben**

Die Aufgabenkomplexität und die Konsistenz in der Reiz-Reaktionszuordnung sind im dem Modell von Ackerman als Moderatorvariablen beim Fertigkeitserwerb repräsentiert, da sie die Beziehung zwischen Fähigkeit und Aufgabenleistung mit gestalten (Ackerman, 1988). Die Komplexität einer Aufgabe beeinflusst das Ausmaß der aufgabenbedingten Aufmerksamkeitsanforderungen und ist beispielsweise gekennzeichnet durch die Anzahl der Items, die parallel oder sequentiell bearbeitet werden müssen, die Darbietungsdauer, dem Ausmaß der Stimulus-Response Kompatibilität oder der Anzahl der zu wählenden Reaktionen.

Konsistenz in einer Aufgabe führt mit andauernder Praxis der Aufgabe zu einer Reduktion der kognitiven Kontrollprozesse, was bei inkonsistenten Aufgaben gar nicht oder nur in sehr geringem Umfang der Fall ist. Inkonsistente Aufgaben bleiben immer weitestgehend neu, da fortwährend neue Produkte der Reiz-Reaktionsbeziehung gebildet werden müssen, wodurch hohe Korrelationen zu allgemeinen Fähigkeiten entstehen (Mack, 1996). Betrachtet man die Dimension „Konsistenz-Inkonsistenz“ als einen kontinuierlichen Übergang, so kann man von kontinuierlichen Phasenübergängen des Fertigkeitserwerbs ausgehen, die in unterschiedlichen Zusammenhängen mit kognitiven Fähigkeiten, Wahrnehmungsgeschwindigkeit und psychomotorischen Fähigkeiten stehen.

#### **2.4.2.2 Prädiktion der Leistung beim Fertigkeitserwerb**

Auf Basis der Theorie des Fertigkeitserwerbs lassen sich folgende Aussagen über die Beziehung zwischen Fähigkeit und Aufgabenleistung der erlernten Fertigkeiten treffen. Für konsistente Aufgaben zeigt sich zu Beginn ein hohes Korrelat zwischen allgemeinen oder aufgabenrelevanten Fähigkeiten und der Aufgabenleistung, welches sich aber mit zunehmender Erfahrung im Umgang mit der Aufgabe verringert. Das Korrelat zwischen der Fähigkeit Wahrnehmungsgeschwindigkeit und Aufgabenleistung hingegen steigt sehr stark an, sobald die Reiz-Reaktionsbeziehungen ausgeprägter werden und fällt genauso stark wieder, wenn die Aufgabenleistung ihr Maximum erreicht hat. Das Korrelat zwischen Psychomotorik und Aufgabenleistung wird mit

zunehmender Automatisierung der erlernten Fertigkeit größer, der Anteil der kognitiven Fähigkeiten nimmt dabei immer mehr ab (Mack, 1996).

Bei inkonsistenten Aufgaben kann die Aufgabenleistung zu jedem Zeitpunkt gut durch die benötigte kognitive Fähigkeit vorhergesagt werden. Die Beziehung zwischen Wahrnehmungsgeschwindigkeit, Psychomotorik und Aufgabenleistung wird auch während der Aufgabenübung konstant bleiben. Für Aufgaben, die sich aus konsistenten und inkonsistenten Aufgabenbestandteilen zusammensetzen, gestalten sich die Leistungsvorhersagen deutlich schwieriger. Selbst bei einem geringen inkonsistenten Aufgabenanteil von ca. 30 Prozent sollte die Vorhersage der Aufgabenleistung eher für inkonsistente Aufgaben erfolgen. Auch mit ansteigender Übung in der erworbenen Fertigkeit muss das Korrelat zwischen Fähigkeiten und Aufgabenleistung nicht abnehmen. Ebenfalls sollten Zeitabstände zwischen Messung der Fähigkeit und Messung der Aufgabenleistung keinen verringernden Einfluss auf das Korrelat zeigen (Ackerman, 1992).

#### **2.4.2.3 Resource Allocation Model**

In dem 1989 von Kanfer und Ackerman entwickelten „Resource Allocation Model“ werden die Interaktion zwischen Motivation und dem allgemeinen kognitiven Fähigkeitsniveau sowie deren Auswirkungen auf die Leistung beschrieben. Diese Beziehungen werden durch kognitive Ressourcen bzw. Aufmerksamkeitsressourcen mediiert. Motivation reguliert die zur Bearbeitung einer Aufgabe aufgewendete Aufmerksamkeitsleistung über distale und proximale Prozesse. Distale Motivationsprozesse führen zu Urteilen, wieviel Aufmerksamkeit auf die Aufgabe gerichtet wird, in Abhängigkeit des daraus resultierenden, persönlichen Nutzens. Je höher der Nutzen eingeschätzt wird, desto mehr Aufmerksamkeit wird in die Erledigung der Aufgabe investiert. Proximale Motivationsprozesse bestimmen die Verteilung der Aufmerksamkeit auf aufgabenbezogene und aufgabenfremde Aktivitäten während der Aufgabenbearbeitung (Keith, Richter & Naumann, 2010). Verglichen mit distalen Motivationsprozessen erfordern proximale Motivationsprozesse selbst Aufmerksamkeit, die im Wettbewerb um die Aufmerksamkeitsressourcen mit aufgabenbezogenen und aufgabenfremden Aktivitäten stehen kann. Eine ausführliche Darstellung zu proximalen Motivationsprozessen findet sich in Kanfer und Ackerman (1989).

Für das allgemeine kognitive Fähigkeitsniveau postulierten Kanfer und Ackerman (1989) eine lineare Beziehung zu den Aufmerksamkeitsressourcen. Individuelle Unterschiede in der Ausprägung des allgemeinen Fähigkeitsniveaus bedingen

unterschiedliche kognitive Ressourcen bzw. Aufmerksamkeitsressourcen, die dem Einzelnen zur Verfügung stehen. Demzufolge kann die Leistungs-Ressourcen-Funktion in eine Leistungs-Fähigkeits-Funktion umgewandelt werden (Keith et al., 2010). Im Modell von Kanfer und Ackerman ist Selbstregulation von entscheidender Bedeutung, die fortwährend kognitive Aufmerksamkeitsressourcen steuert. Selbstregulation beschreiben die Autoren als Prozesse, die aktiviert werden, falls eine Person ein Ziel erreichen möchte und dafür benötigte Fähigkeiten oder Fertigkeiten ermittelt. Aus diesem Abgleich resultieren entsprechende Handlungen, die zur Erreichung des Ziels von der Person als erforderlich angesehen werden (Bandura, 1990). Mit Bezug auf die sozial-kognitive Theorie von Bandura (1986) kann der Prozess der Selbstregulation in die Abschnitte Selbstbeobachtung, Selbstbeurteilung und Selbstreaktion unterteilt werden, die aufeinanderfolgen und sich gegenseitig beeinflussen.

Ohne Selbstregulation würde die durch distale Motivationsprozesse zuerst initiierte Aufmerksamkeitsmenge während der gesamten Aufgabenbearbeitung beibehalten werden, die nach Einschätzung zu Aufgabenbeginn zur Erreichung der Zielsetzung benötigt wird. Mögliche Veränderungen, sei es eine Erhöhung oder Verringerung der Aufmerksamkeitsmenge, würden ohne Selbstregulationsprozesse nicht stattfinden.

#### **2.4.2.4 Prädiktion praktischer Leistung in der Zahnmedizin**

Basierend auf Ackermans (1988) Theorie des Fertigkeitserwerbs identifizierten Gray und Deem (2002a) räumliches Vorstellungsvermögen als spezifische Fähigkeit der kognitiven Phase des Fertigkeitserwerbs in den vorklinischen Laborkursen. Die Beurteilung der von Gray und Deem als inkonsistent klassifizierten Laborkursaufgaben erfolgte auf Notenskalen, auf denen für sehr gute Leistungen fünf und für sehr schlechte null Punkte vergeben wurden. Der räumliches Vorstellungsvermögen erfassende Untertest PAT des Studierfähigkeitstests DAT konnte eine höhere positive und signifikante Korrelation zur durchschnittlichen Aufgabenleistung der ersten zwei Studienjahre aufzeigen ( $r = 0.496$ ,  $p < .05$ ) als der Schulabschlussnotendurchschnitt GPA ( $r = 0.173$ , *n.s.*) und der akademisches Wissen prüfende DAT-Untertest „Academic Average“ ( $r = 0.188$ , *n.s.*). Um die Differenzierung der Zulassungsdaten zu erhöhen, exkludierten Gray und Deem die DAT-Testergebnisse im mittleren Bereich, sodass insgesamt 177 Datensätze von Zahnmedizinistudierenden der Abschlussklassen 1998 bis 2001 in die Auswertung einfließen konnten. Die Autoren sahen in ihren Untersuchungsdaten Ackermans (1988) Theorie des Fertigkeitserwerbs für das Anwendungsfeld der zahnmedizinischen Ausbildungsforschung als empirisch bestätigt.

Auch die Vorhersage der praktischen Leistung in den klinischen Kursen der Zahnmedizin durch kognitive Fähigkeiten konnten Gray und Deem (2002b) in einer methodisch identischen Folgeuntersuchung mit dem Modell von Ackerman (1988) erklären. Kognitive Fähigkeiten besitzen einen verschwindend geringen Einfluss auf die klinische Leistung, da die kognitive und assoziative Phase der in der Klinik benötigten Fertigkeiten bereits in den Laborkursen der vorherigen Semester abgeschlossen wurde und jetzt in der autonomen Phase nur noch trainiert und dadurch automatisiert werden. Die Aufgaben der in die Untersuchung einbezogenen klinischen Kurse des dritten und vierten Studienjahres wurden von Gray und Deem als konsistent klassifiziert. Die Beziehung des räumlichen Vorstellungsvermögens zum Mittelwert der klinischen Kursnoten für 169 Studierende fiel demzufolge nur schwach positiv aus ( $r = 0.051$ , *n.s.*). Beide Untersuchungen von Gray und Deem (2002a; 2002b) können aber Ackermans (1988) Theorie des Fertigkeitserwerbs nicht vollständig verifizieren, da die die Leistung beeinflussenden Fähigkeiten der assoziativen und autonomen Phase nicht eingeschlossen wurden.

Suksudaj, Townsend, Kaidonis, Lekkas und Winning (2011) gingen in diesem Kritikpunkt umfassender vor. Neben kognitiven Fähigkeiten, wie räumliches Vorstellungsvermögen und schlussfolgerndes Denken, wurden ebenfalls die Wahrnehmungsgeschwindigkeit, psychomotorische Fähigkeiten und die Motivation erfasst. Um die Phasen des Fertigkeitserwerbs abzubilden, bearbeiteten Zahnmedizinische Studierende zu Beginn der Laborkurse im dritten und vierten Semester eine Übung zum Einsetzen einer Amalgamfüllung im Phantomkopf („Cavity Preparation Exercise“), die jeweils am Ende des Semesters wiederholt wurde, während die Studierenden im fünften Semester die Übung nur zu Beginn des dritten Studienjahres absolvierten. Die Untersuchungsergebnisse ergaben zwar eine Verbesserung der Übungsleistung zwischen dem dritten und vierten Semester, jedoch variierten die Beziehungen zwischen Fähigkeiten und Aufgabenleistung in den unterschiedlichen Phasen des Fertigkeitserwerbs nicht, wie von Ackerman (1988) postuliert. Suksudaj et al. vermuteten als Ursache die geringe Anzahl der Übungswiederholungen sowie die Stichprobengröße, die über die Kohorten zwischen 50 bis 71 Studierenden schwankte. Zwischen Übungsleistung und Motivation, aber nur für den Faktor Selbsteinschätzung der investierten Anstrengung, zeigte sich eine positive, schwache und signifikante Korrelation ( $r = 0.196$ ,  $p < .05$ ), die Suksudaj et al. auf den Durchführungszeitpunkt der praktischen Übung zu Beginn des dritten Studienjahrs zurückführten, die ausschließlich

von Studierenden der zweiten Untersuchungskohorte absolviert wurde. Zu diesem Studienzeitpunkt müssen die Studierenden ausreichende Kompetenzen unter Beweis stellen, um mit der Behandlung von Patienten beginnen zu dürfen, sodass sie vermutlich mehr Anstrengungen in die praktische Übung investierten. Eine detailliertere Einschätzung der Studienergebnisse gestaltet sich jedoch insgesamt schwierig, da die Autoren nur eine Korrelationstabelle berichteten, in der die Daten der einzelnen Messzeitpunkte gemittelt waren.

### **2.4.3 Vorhersage beruflicher Leistung**

#### **2.4.3.1 Drei-Phasen-Modell von Ackerman**

Ackerman (1992) betont mit Bezug auf seine Theorie des Fertigkeitserwerbs die Wichtigkeit der Fähigkeiten Wahrnehmungsgeschwindigkeit und der Psychomotorik zur Vorhersage individueller Leistungsunterschiede, die neben der allgemeinen Intelligenz einen guten Beitrag aufweisen und somit Leistung in jeder Phase des Fertigkeitserwerbs vorhergesagt werden kann. Die Leistung wurde mittels des „Terminal Radar Approach Controller“-Programms ermittelt, einer zum Training in der Fluglotsenausbildung eingesetzten Computersimulation. Bei der Datenauswertung zeigte sich, dass neben der allgemeinen Intelligenz auch spezifische Fähigkeiten zur Vorhersage der Leistung sehr gut geeignet sind, insbesondere das räumliche Vorstellungsvermögen, das anhand der Tests „Paper Folding“, „Surface Development“ und des „Form Board“ aus dem „Kit of factor-referenced cognitive tests“ (Ekstrom, French, Harman & Dermen, 1976) bei den 102 Untersuchungsteilnehmern erfasst wurde. In 2007 spezifizierten Ackerman und Beier die Vorhersagekraft der Wahrnehmungsgeschwindigkeit für Aufgaben, die nicht nur zeit- und geschwindigkeitsbezogene Handlungen bzw. Reaktionen erforderten, sondern gleichzeitig auch Genauigkeit verlangten.

Die Forschungsergebnisse von Ackerman und Cianciolo (2000) ergaben, dass eine zuverlässige Vorhersage individueller Leistungsunterschiede bei Aufgaben, die den Erwerb von Fertigkeiten implizieren, trotz Übungseffekten und unterschiedlich entwickelten Fähigkeitsniveaus mittels des drei Phasen Modells der Theorie des Fertigkeitserwerbs möglich ist. In den drei Experimentalbedingungen wurden als kognitive Fähigkeiten verbale, numerische und räumliche Intelligenz mittels papierbasierter Intelligenztestverfahren operationalisiert. Unterschiede in der Lernfähigkeit, im Lernstil und in den verwendeten Strategien beeinflussten das psychomotorische Lernen, wodurch es zu Unterschieden in der Aufgabenleistung

kommen kann, die den Erwerb neuer Fertigkeiten erfordern. Gleichzeitig betonten die Autoren die Wichtigkeit einer validen und reliablen Operationalisierung der Wahrnehmungsgeschwindigkeit und der Psychomotorik, um prädiktive Aussagen über das Korrelat zwischen Fähigkeiten und der Leistung beim Fertigkeitserwerb zuverlässig treffen zu können.

Keil und Cortina (2001) überprüften die in Ackermans Modell postulierten prädiktiven Beziehungen zwischen Fähigkeiten und Aufgabenleistung hinsichtlich ihrer temporären Stabilität, die, wie die Ergebnisse der Studie zeigten, weniger durch die Charakteristik der Kombination aus Fähigkeit und Aufgabe bedingt ist als bisher angenommen wurde. Zudem zeigte sich, dass die Verringerung bzw. Abnahme der Leistungsprädiktion nicht linear verläuft, sondern dem Muster einer Cusp-Katastrophe folgt, die unstete, sprunghafte Veränderungen zwischen mehreren Extremen darstellt.

Farrel und McDaniel (2001) verwendeten Ackermans Theorie zur Vorhersage der beruflichen Leistung, die sie anhand von Vorgesetztenbeurteilungen operationalisierten. Die Messung der allgemeinen Intelligenz, der Wahrnehmungsgeschwindigkeit und der Psychomotorik erfolgte in ihrer Studie durch die „General Aptitude Test Battery“ (Hartigan & Wigdor, 1989). Die berufliche Leistung für Aufgaben mit überwiegend konsistentem Inhalt wurde am besten durch die allgemeine Intelligenz vorhergesagt, jedoch ab einer Berufserfahrung von sieben bis neun Jahren zeigte sich, dass psychomotorische Fähigkeiten als Prädiktor besser geeignet waren. Die Autoren erklärten dieses Ergebnis mit dem Erlernen von Berufsregeln, Aufgaben und Arbeitsroutinen, die mit der Zeit immer mehr automatisiert werden und nur zu Beginn höhere kognitive Fähigkeiten erfordern. Für Aufgaben mit überwiegend inkonsistenten Inhalten ist aber die kognitive Fähigkeit der beste Prädiktor für die Berufsleistung, auch mit zunehmender beruflicher Erfahrung, da es zu keiner Automatisierung der Aufgaben kommt und höhere kognitive Fähigkeiten kontinuierlich benötigt werden. Schmidt und Hunter (2004) hingegen behaupteten in ihrer Metaanalyse zur allgemeinen Intelligenz als Prädiktor beruflicher Leistung, dass auch mit zunehmender Berufserfahrung die allgemeine Intelligenz zuverlässig die berufliche Leistung vorhersagt, wobei aber zu beachten ist, dass ihre Aussage auf einer Korrelationsmatrix von McDaniels (1985) beruht, dessen Stichprobenumfang mit einem Anstieg der Variable Berufserfahrung (in Jahren) sehr stark abnimmt.

#### 2.4.3.2 Der „whole-Person“ Ansatz als Prädiktor von Leistung

Die Vorhersage akademischer und beruflicher Leistung durch Fähigkeiten und Persönlichkeitseigenschaften unter dem „whole-person“ Ansatz wurde von Kanfer, Wolf, Kantowitz und Ackerman (2010) untersucht. Der „whole-person“ Ansatz von Snow und Farr (1987) umgeht das „Silo“-Problem bei der Vorhersage von Leistung, indem nicht nur einzelne Eigenschaften wie Motivation oder kognitive Fähigkeiten als Prädiktoren verwendet werden, sondern ein individuelles Gesamtbild der Person abgebildet wird, welches sich aus kognitiven Fähigkeiten, Fertigkeiten, Persönlichkeits- und Motivationseigenschaften sowie Interessen zusammensetzt. Zur Erfassung des individuellen Gesamtbilds setzten Kanfer et al. elf Testverfahren zur Messung der numerischen, verbalen und räumlichen Fähigkeiten ein, vier Tests zur Bestimmung des Domänen- bzw. Fachwissens und 27 Tests zur Erhebung der Persönlichkeits- und Motivationseigenschaften sowie beruflicher Interessen und Selbstkonzepte. Wie erwartet konnten kognitive Fähigkeiten und Persönlichkeitsdispositionen die akademische Leistung vorhersagen, jedoch zeigte sich für die berufliche Leistung, dass nur Persönlichkeitsdispositionen signifikante Prädiktoren darstellten. Als Erklärung führen die Autoren die Operationalisierung der beruflichen Leistung an, die über Beurteilungsbögen in Firmen erfasst wurde, welche ein „CO-OP“-Programm anbieten. Das „CO-OP“-Programm in den USA entspricht einem dualen Studium in Deutschland, d.h., die Teilnehmer studieren und arbeiten abwechselnd bis zum Studienabschluss. Für die Arbeit erhalten die Studierenden eine Vergütung. Die Ergebnisse der Studie bestätigten aber den Autoren zufolge den „whole-person“ Ansatz der Leistungsprädiktion, weshalb sich zukünftig nicht mehr die Frage stellt, ob Persönlichkeitsdispositionen zur Vorhersage der Leistung mit eingezogen werden, sondern wie und in welchem Umfang. Der alleinige Einsatz von kognitiven Fähigkeiten ist die Ursache für die Prädiktion von nur 30 bis 40 Prozent der Leistungsvarianz einer komplexen Aufgabe (Ackerman, Kanfer & Goff, 1995).

Die Ansicht, dass allgemeine Intelligenz, Aufmerksamkeit und perzeptiv-motorische bzw. psychomotorische Fähigkeiten unterschiedliche Kategorien repräsentieren, widerlegen Nijenhuis und van der Flier (2002) in ihrer Untersuchung mit 584 Bewerbern. Psychomotorische Fähigkeiten sind bewegungsbezogene Fähigkeiten, zu denen zum Beispiel die Augen-Hand-Koordination oder die Körper-Auge-Koordination gehören. Die Korrelationen zwischen den drei Konstrukten variierten in einer Höhe von  $r = 0.14$  bis  $r = 0.56$  und lagen im Durchschnitt bei  $r = 0.39$ . Nijenhuis und van der Flier

interpretierten die Ergebnisse als Beleg dafür, dass sich allgemeine Intelligenz, Aufmerksamkeit und psychomotorische Fähigkeiten zu einem beträchtlichen Teil überschneiden und nicht als vollständig voneinander trennbar angesehen werden könnten.

#### **2.4.4 Zusammenfassung und Resümee**

Das von Ackerman (1988) entwickelte Drei-Phasen-Modell des Fertigkeitserwerbs trifft zuverlässige Vorhersagen der Leistung durch kognitive, perzeptive und psychomotorische Fähigkeiten, die sich empirisch bestätigen ließen (Taatgen, 2002). Mittels der Moderatorvariablen Aufgabenkomplexität und -konsistenz lässt sich die prädiktive Beziehung von Fähigkeiten auf individuelle Leistungsunterschiede beim Fertigkeitserwerb weiterführend charakterisieren, indem die Aufgabenleistung als Kriterium detaillierter spezifiziert wird. Die Erweiterung der Theorie des Fertigkeitserwerbs durch das „Resource Allocation Model“ (Kanfer & Ackerman, 1989) ermöglicht ergänzende Betrachtungen der Leistungsprädiktion unter Einbezug der Motivation. Die empirische Anwendung Ackermans Theorie des Fertigkeitserwerbs erstreckt sich primär auf das Forschungsfeld der Berufseignungsdiagnostik. Für die Vorhersage der praktischen Leistung in den technischen Kursen der Zahnmedizin ist eher der Standpunkt von Keil und Cortina (2001) oder Ackerman und Cianciolo (2000) in der Frage mehrerer, geeigneter Prädiktoren beruflicher Leistung zu übernehmen, auch um die noch unbeantwortete Frage nach der Stärke der Vorhersageleistung des praktischen Studienerfolgs in der Zahnmedizin für das manuelle Geschick und das räumliche Vorstellungsvermögen zu klären.

Kognitive Fähigkeiten, Wahrnehmungsgeschwindigkeit und psychomotorische Fähigkeiten können individuelle Leistungsunterschiede in der kognitiven, assoziativen und autonomen Phase des Fertigkeitserwerbs zuverlässiger vorhersagen als nur durch den Prädiktor allgemeine Intelligenz. Somit könnten die inkonsistenten Ergebnisse von Studien zur Vorhersage der praktischen Studienleistung in der Zahnmedizin (Spratley, 1990) in ein theoretisches Modell eingebunden werden, um zukünftig zielgerichteter die Beziehungen zwischen Fähigkeiten und praktischen Fertigkeiten in den Laborkursen zu bestimmen. Der „whole person“ Ansatz spiegelt eine Methodik der Leistungsprognostik wider, welche mehrere Fähigkeiten und Persönlichkeitseigenschaften als Prädiktoren einbezieht, um ein möglichst genaues Abbild des Individuums zu erschaffen und so die größtmögliche Vorhersage von Leistung zu erzielen (Kanfer et al., 2010). Einzelne Fähigkeiten konnten maximal moderate Effektstärken in der Prognostik der praktischen

Studienleistung erzielen. Diesen Überlegungen stellen sich aber einige Hürden entgegen. Einerseits gibt es bisher nur wenige Studien die Ackermans (1988) Theorie des Fertigkeitserwerbs in Studien zur Prädiktion des praktischen Studiererfolgs in der Zahnmedizin einsetzen (Gray & Deem, 2002a; Gray & Deem, 2002b; Suksudaj et al., 2011) und andererseits widersprechen Nijenhuis und van der Flier (2002) einer strikten Trennung von separaten Fähigkeiten, wie sie unter anderem in Ackermans Drei-Phasen-Modell postuliert sind. Am Beispiel der Kovarianzen von allgemeiner Intelligenz, Aufmerksamkeit und Psychomotorik wiesen Nijenhuis und van der Flier gegenseitige, partielle Überschneidungen ihrer Taxonomie auf.

### **3. Fragestellung und Hypothesen**

Im Unterkapitel 3.1 sind einleitend die Fragestellungen und folgend im Unterkapitel 3.2 die Ziele der vorliegenden Arbeit aufgeführt. Im Abschnitt 3.3 werden die Untersuchungshypothesen hergeleitet, die in aggregierter Form in Tabelle 3.01 sowie in einem Konzeptmodell graphisch in Unterkapitel 3.3.3 dargestellt sind.

#### **3.1 Vorhersage der praktischen Studienleistung**

Die Änderung des Hochschulrahmengesetzes in 2004 erlaubt Hochschulen in Deutschland zusätzlich zur Abiturdurchschnittsnote weitere Kriterien zur Auswahl von Studierenden einzubeziehen, die sehr gute Studienleistung erzielen (Wedler et al., 2008). Daraus sollen eine höhere Zufriedenheit der Studierenden, kürzere Studienzeiten und geringere Studienabbruchquoten resultieren, die wiederherum günstige Studienplatzkosten für Studierende und Hochschulen bedeuten. Im Vergleich zum Durchschnitt aller Studienfächer fielen 2008 in der Zahnmedizin die Studienplatzkosten fast viermal so hoch aus (vgl. Unterkapitel 2.1.1), weshalb ein qualitativ gutes Auswahl- und Zulassungsverfahren erforderlich ist, zumal die Bewerberzahlen die vorhandenen Ausbildungskapazitäten bei weitem überschreiten. Die Entscheidung über den Einsatz von Verfahren zur Studierendenauswahl ist dabei nicht nur durch deren prognostische Güte der Studienleistung bedingt, sondern auch durch ihre Ökonomie, Akzeptabilität und Fairness (Schuler & Hell, 2008). In der Zahnmedizin umfasst der vorklinische Studienabschnitt fünf Semester, die neben naturwissenschaftlichen und (zahn)medizinischen Fächern auch technische Laborkurse der Zahnersatzkunde beinhalten, in denen zahnmedizinischen Behandlungstechniken erlernt werden (vgl. Unterkapitel 2.2.1). Daher ist es erforderlich, dass bei der Auswahl von Zahnmedizinistudenten nicht nur Instrumente verwendet werden, die sich durch eine zuverlässige Vorhersage theoretischer Studienleistungen auszeichnen, sondern auch Verfahren zum Einsatz kommen, um die in den praktischen Kursen benötigten Fähigkeiten der Bewerber zu erfassen.

Sehr kontrovers wird die Frage diskutiert, welche Testverfahren bzw. Fähigkeiten zuverlässig die Leistungen in den praktischen Kursen des Zahnmedizinstudiums vorhersagen. Die zugelassenen Studenten sollen nicht nur die akademischen, sondern auch die Anforderungen in den praktischen Kursen erfüllen. Spratley gab 1990 in seinem Review zur Studieneignung in der Zahnmedizin einen umfassenden Überblick zu den zahlreichen und inkonsistenten Studienergebnissen in dieser Frage. Im Rahmen

der Validierung des amerikanischen DAT wurde ab Mitte des 20. Jahrhunderts versucht, manuelle Geschicklichkeitstests durch ökonomischere Multiple-Choice Tests zum räumlichen Vorstellungsvermögen zu ersetzen, mit der Begründung vergleichbarer prognostischer Beziehungen zu den praktischen Studienleistungen. Diese Annahme erfuhr in einer Vielzahl von Studien Zustimmung, aber auch Kritik und wird bis heute in der Literatur diskutiert, wie das aktuellere Review von Ranney et al. (2005) aufzeigt. Die Ursache für die anhaltende Debatte liegt in den diffusen Untersuchungsergebnissen, da unterschiedliche, oftmals sogar gegensätzliche Zusammenhänge für die Fähigkeiten räumliches Vorstellungsvermögen und manuelles Geschick in Bezug zur praktischen Studienleistung in den Punkten Stärke, Richtung und Signifikanz veröffentlicht wurden. Offen blieb auch im Kontext mehrerer Konstruktvalidierungen, vor allem der des amerikanischen Studierfähigkeitstest DAT, die Frage, in welcher Beziehung manuelles Geschick und räumliches Vorstellungsvermögen zueinander stehen. Dabei besteht aber im Allgemeinen kein Zweifel an der Erforderlichkeit beider Fähigkeiten, um die auf den Beruf des Zahnarztes vorbereitenden praktischen Kurse im Zahnmedizinstudium erfolgreich bewältigen zu können (Smith, 1989; Wong et al., 1979).

### 3.2 Ziele der vorliegenden Arbeit

Ranney et al. (2005) empfahlen in ihrem Literaturreview zur Auswahl von Zahnmedizinierenden, manuelle und räumliche Tests nicht als prädiktive, sondern prüfende Selektionsverfahren einzubeziehen, d.h., in den Tests würde ein Mindestwert definiert werden, der von den Bewerbern erreicht werden muss, um im Zulassungsverfahren weiterhin berücksichtigt zu werden. Dieser Vorschlag kann eher als temporärer Kompromiss, aber nicht als zufriedenstellende Antwort auf die offenen Fragen bei der Vorhersage der praktischen Studienleistung in der Zahnmedizin angesehen werden. Zwar würde das Festlegen einer „cut-off“-Grenze in den Auswahltests das benötigte Mindestmaß an manuellem Geschick und räumlichen Vorstellungsvermögen bei zukünftigen Studierenden gewährleisten, aber anhand der Fähigkeiten könnte nicht mehr zwischen den Bewerbern in ihrer Merkmalsausprägung diskriminiert werden, was sich vor allem bei einem Bewerberüberhang als gravierender Nachteil erweisen würde.

Eine Schwierigkeit, die Diversität der Studienergebnisse zur Prädiktion der praktischen Studienleistung systematisch nachzuvollziehen, könnte weitestgehend im Fehlen eines theoretischen Fundaments liegen, in welchem die Beziehungen zwischen den Fähigkeiten als Prädiktoren und den in den praktischen Kursen benötigten Fertigkeiten als Kriterien eingebettet sind. Diese Problematik aufgreifend, bezogen Gray und Deem (2002a; b) das Modell zum Fertigkeitserwerb von Ackerman (1988) in ihre Untersuchungen ein, in denen die Autoren räumliches Vorstellungsvermögen als Fähigkeit der kognitiven Phase des Fertigkeitserwerbs identifizierten konnten (vgl. Unterkapitel 2.4.2.4). Ackerman stellte 1988 in seiner Theorie des Fertigkeitserwerbs ein Modell zur Vorhersage individueller Unterschiede erlernter Leistung durch Fähigkeiten in drei Phasen auf (vgl. Unterkapitel 2.4.2). Die Theorie gilt als empirisch bewährt und besitzt auch für den Erwerb komplexer Fertigkeiten Gültigkeit, sodass eine Anwendung des Modells im Feld der Zahnmedizin möglich ist.

Der Einfluss von manuellem Geschick auf die praktische Leistung, der gemäß Ackermans Theorie primär in der autonomen Phase des Fertigkeitserwerbs auftritt, wurde neben kognitiven Fähigkeiten der vorhergehenden Phasen in der Studie von Suksudaj et al. (2011) berücksichtigt. Die Autoren fanden keinen Einfluss des manuellen Geschicks auf den Erwerb zahnmedizinischer Fertigkeiten in den Laborkursen des Zahnmedizinstudium, jedoch lässt die von den Autoren gewählte Auswertung und Ergebnisdarstellung keine Interpretationen für die unterschiedlichen

Ergebniszeitpunkte in der Untersuchung zu, die im Kontext der Phasen des Fertigkeitserwerbs besonders aufschlussreich gewesen wären. Neben der geringen Stichprobengröße sahen die Autoren zudem die geringe Anzahl der Übungswiederholungen als weitere Limitation ihrer Studienergebnisse an.

Die Ziele der vorliegenden Arbeit lassen sich in zwei Fragenkomplexe unterteilen. Zunächst wird die Güte der Operationalisierung der Fähigkeitsarten räumliches Vorstellungsvermögen und manuelles Geschick untersucht, ob diese als autonome Konstrukte messbar sind. Der zweite Fragenkomplex beinhaltet den Einfluss des räumlichen Vorstellungsvermögen und des manuellen Geschicks auf die praktische Leistung in den vorklinischen Laborkursen, der unter Anwendung von Ackermans (1988) Theorie des Fertigkeitserwerbs analysiert wird. Als Konsequenz aus den Untersuchungsergebnissen sollen Empfehlungen für den Einsatz von psychometrischen Testverfahren zur Auswahl von Studienbewerbern in der Zahnmedizin resultieren, um möglichst zuverlässig deren praktische Studienleistungen zu präzisieren. Hierfür sind insbesondere die inkrementellen Beiträge der Prädiktoren räumliches Vorstellungsvermögen und manuelles Geschick entscheidend, die Auskunft geben, ob die Hinzunahme eines Tests in ein Auswahlverfahren ökonomisch im Hinblick auf den Durchführungsaufwand zu vertreten ist. Aus zeitökonomischen Gründen wird die praktische Leistung im klinischen Studienabschnitt nicht in die vorliegende Arbeit einbezogen.

### 3.3 Hypothesen

#### 3.3.1 Diskrimination der Fähigkeitsarten

In einer empirischen Untersuchung wiesen Nijenhuis und van der Flier (2002) zwischen den Fähigkeiten allgemeine Intelligenz, Aufmerksamkeit und Psychomotorik substanzielle Überschneidungen in Höhe von durchschnittlich  $r = 0.39$  nach, weshalb nach Meinung der Autoren die Fähigkeiten nicht wie bisher postuliert als drei unterschiedliche Taxonomien angesehen werden könnten (s. Unterkapitel 2.4.3.2). Die Beziehung zwischen räumlichem Vorstellungsvermögen als Vertreter kognitiver Fähigkeiten und manuellem Geschick als nichtkognitive bzw. manuelle Fähigkeit wurde in Studien bisher nur in sehr geringem Umfang untersucht.

Im Rahmen der Validierungsdebatte des Studierfähigkeitstest DAT in den USA und Kanada wurde die Beziehung zwischen beiden Fähigkeitsarten oftmals nur in der Ergebnisdarstellung aufgeführt, aber kaum thematisiert. Eine Ausnahme bildete hiervon die Forderung in einigen Primärstudien, den in der Durchführung aufwendigen Kalkschnittstest im DAT durch einen papierbasierten Test zum räumlichen Vorstellungsvermögen zu ersetzen (s. Unterkapitel 2.3.3). Zwischen beiden Testverfahren berichtete z.B. Derevere (1961) in seiner Untersuchung eine Korrelation von  $r = 0.78$ , ohne aber auf die Beziehung zwischen beiden Fähigkeiten inhaltlich näher einzugehen. Interpretiert man die Korrelation, zeigt sich, dass Zahnmedizinierende mit einem höher ausgeprägten manuellem Geschick auch über ein besseres räumliches Vorstellungsvermögen verfügen. Zwar konnte in voneinander unabhängigen Studien zur Konstruktvalidierung des DAT bestätigt werden, dass der Kalkschnittstest als DAT-Untertest „Carving Dexterity“ manuelles Geschick im Umgang mit Handinstrumenten erfasst (Zullo, 1971) und der DAT-Untertest „Space Relations“ bzw. auch der ihn 1981 ersetzende Untertest PAT, die kognitive Fähigkeit räumliches Vorstellungsvermögen auf mehrdimensionaler Ebene misst (Kramer et al., 1989), aber die früheren Resultate von Dworkin (1970) oder Young (1972) zur Konstruktvalidierung des DAT wurden nicht klärend aufgegriffen. Beide Autoren berichteten als Ergebnisse ihrer Faktoranalyse zum DAT in den USA und Kanada Ladungen der Subtests „Carving Dexterity“ und „Space Relations“ auf einem Faktor, anstatt auf den ursprünglich bei der Testentwicklung konzeptionierten Faktoren „Space Conception“ und „Carving Dexterity“ (vgl. Unterkapitel 2.3.3).

Im Kontext der Validierung des als Vorläufer des TMS in Deutschland eingesetzten Studierfähigkeitstests BATZ (vgl. Unterkapitel 2.2.3.2) berichtete Hitpass (1978)

positiv schwache Korrelationen zwischen der Drahtbiegeaufgabe und den räumliches Vorstellungsvermögen erfassenden Aufgaben Flächenabwicklungen ( $r = 0.28$ ), Würfelaufgaben ( $r = 0.19$ ) und Bauprobe ( $r = 0.24$ ), die aber auf inhaltlicher Ebene nicht analysiert wurden. In Studien zur Vorhersage der praktischen Studienleistung finden sich für Testverfahren, in denen das manuelle Geschick im Umgang mit Handinstrumenten erfasst wird, ebenfalls schwach positive, aber signifikante Korrelationen zum räumlichen Vorstellungsvermögen, welches über den DAT-Untertest PAT erhoben wurde. So berichteten Boyle und Santelli (1986) zwischen dem „Crawford Small Parts Dexterity Test“ und dem PAT eine auf dem Niveau  $p < .01$  signifikante Korrelation von 0.38 (vgl. Unterkapitel 2.3.4.4.1), die aber von den Autoren nicht diskutiert bzw. näher betrachtet wird. Auch den Ergebnissen von Lundergan et al. (2007) kann man ebenfalls eine schwach positive, signifikante Beziehung von  $r = 0.245$  ( $p < .05$ ) zwischen dem „O’Connor Tweezer Dexterity Test“ und dem DAT-PAT entnehmen, auf welche die Autoren aber nicht erklärend eingingen.

Dem Ansatz von Potter und McDonald (1985) folgend, wären räumliches Vorstellungsvermögen und manuelles Geschick besser in ein Strukturgleichungsmodell zur Vorhersage der praktischen Studienleistung einzuschließen, anstatt Korrelations- und Regressionsanalysen zu verwenden, um einerseits den synchronen Einfluss mehrerer Prädiktoren sowie andererseits deren Beziehungen untereinander abzubilden. Zuvor stellen sich aber die Fragen, ob manuelles Geschick, das im Umgang mit Handinstrumenten benötigt wird, als autonomes Konstrukt in einem Modell operationalisiert werden kann und von räumlichem Vorstellungsvermögen als kognitive Fähigkeit zu diskriminieren ist. Sollten beide Fähigkeitsarten nicht als eigenständige Konstrukte messbar sein, könnte darin vielleicht eine Ursache für die unterschiedlichen Forschungsergebnisse zur prognostischen Güte der Laborkursleistungen in der zahnmedizinischen Vorklinik begründet sein. Es stellt sich demzufolge die Frage, ob bei der Operationalisierung von manuellem Geschick im Zahnmedizinstudium, das primär im Umgang mit Handinstrumenten benötigt wird, nicht eher Definitionen herangezogen werden sollten, die manuelles Geschick und das räumliche Vorstellungsvermögen als der Psychomotorik zugehörige Fähigkeiten beschreiben, welche im Zusammenspiel benötigt werden, um die Aufgaben in den vorklinischen Laborkursen erfolgreich zu bewältigen (vgl. Unterkapitel 2.3.1.5). Nachfolgend soll von instrumentellem, manuellem Geschick gesprochen werden, das manuelles Geschick im manipulativen Umgang mit Handinstrumenten bezeichnet.

Die berichteten Studienergebnisse lassen aber auf einen insgesamt schwachen Zusammenhang zwischen beiden Fähigkeitsarten schließen, da die Stärken der Koeffizienten nicht mehrheitlich über 0.4 liegen. Demzufolge sollte bei der Operationalisierung zwischen beiden Konstrukten in einem Modell diskriminiert werden können. Aus diesen Überlegungen leiten sich die folgenden Hypothesen ab:

#### Hypothese 1:

Es besteht eine hohe Konstruktreliabilität für das instrumentelle, manuelle Geschick und räumliche Vorstellungsvermögen.

#### Hypothese 2:

Es besteht eine hohe diskriminante Validität für die Konstrukte instrumentelles, manuelles Geschick und räumliches Vorstellungsvermögen.

### **3.3.2 Vorhersage der Leistung in den vorklinischen Laborkursen**

Frühere und aktuellere Reviews geben einen umfassenden Überblick zu den inkonsistenten Ergebnissen von Primärstudien in der Frage, welche Prädiktoren zur Vorhersage der praktischen Studienleistung in der Zahnmedizin geeignet sind, um in Bewerberauswahlverfahren eingesetzt zu werden (z.B. Ranney et al., 2005; Spratley, 1990). Insbesondere für manuelle und kognitive Fähigkeiten bestehen differente Ansichten über deren prognostischen Validitäten der praktischen Studienleistung. Eine Ursache könnte weitestgehend im Fehlen einer theoretisch fundierten Charakterisierung der Beziehungen zwischen Fähigkeiten und in den Laborkursen zu erlernenden Fertigkeiten bestehen, sodass die widersprüchlichen und teilweise gegensätzlichen Studienergebnisse von beispielsweise Manhold und Manhold (1965; 1967) nur schwer nachzuvollziehen sind. Setzt man Ackermans (1988) dreistufiges Modell des Fertigkeitserwerbs in Bezug zu den im Theorieteil dargestellten Studienergebnissen, so lassen sich für die Fähigkeiten zur Vorhersage des Erwerbs zahnmedizinischer Fertigkeiten in den vorklinischen Laborkursen die nachfolgenden Korrelate aufstellen.

Die praktische Studienleistung bzw. die Aufgabenleistungen in den Laborkursen werden in der kognitiven Phase durch das räumliche Vorstellungsvermögen bestimmt (Gray & Deem, 2002a), da die dreidimensionale Wahrnehmung von Objekten bzw. Zähnen, deren Formen, anatomische Strukturen, Positionen und Verbindungen zueinander, vor allem der Bereiche, die nicht direkt sichtbar sind, als mentale Modelle imaginiert werden müssen. Diese mentalen Modelle werden zum Erlernen der

zahnmedizinischen Fertigkeiten bzw. Dentaltechniken benötigt, die beispielsweise die zur Präparation eines Zahnes (Abschleifen) und zum Einsetzen einer Füllung benötigten Arbeitsschritte und -methoden darstellen. Es ist zu erwarten, dass die Prognostik der Aufgabenleistung durch die Fähigkeit räumliches Vorstellungsvermögen in der kognitiven Phase zum Prüfungszeitpunkt in den Laborkursen nur noch auf einem schwachen Niveau existent ist, da der Übergang in die nachfolgende assoziative Phase des Fertigkeitserwerbs zugleich den Abschluss der kognitiven Phase darstellt (Ackerman, 1989).

Ist das Erlernen der Dentaltechniken einer Aufgabe abgeschlossen, erfolgt der Übergang in die zweite Phase, die Ackerman als die assoziative Phase bezeichnet. In dieser Phase ist die Leistung durch die Fähigkeit Wahrnehmungsgeschwindigkeit ausschlaggebend beeinflusst, indem die Reiz-Reaktionsbeziehungen derart gefestigt werden, dass die Aufgabenleistung erhöht und Fehlerhäufigkeit verringert wird. In der vorliegenden Untersuchung wird die Wahrnehmungsgeschwindigkeit als Prädiktor nicht weiter betrachtet, da Ackerman und Beier (2007) eine zulässige Vorhersagekraft nur für Aufgaben nachwiesen, die Handlungen bzw. Reaktionen mit einem hohen Genauigkeitsgrad unter hohem Zeitdruck beinhalten.

Die dritte und finale Phase des Fertigkeitserwerbs ist als autonome Phase bezeichnet, in der die Automatisierung der erlernten Dentaltechniken erfolgt. In dieser Phase sind für individuelle Leistungsunterschiede in den Laborkursaufgaben primär manuelle Fähigkeiten ausschlaggebend. Demzufolge sollte instrumentelles, manuelles Geschick, d.h., manuelles Geschick im manipulativen Umgang mit Handinstrumenten, die Laborkursleistungen maßgeblich beeinflussen, da zum Prüfungszeitpunkt der praktischen Arbeiten die Automatisierung der erlernten Dentaltechniken noch vorhält (Mack, 1996). Kontrollierend sollte ebenfalls die Schulabschlussdurchschnittsnote sowie das naturwissenschaftliche Verständnis als Prädiktoren der vorklinischen Laborkursleistungen einbezogen werden, obwohl keine Zusammenhänge zu erwarten sind. Beide Kriterien basieren auf kognitiven Fähigkeiten wie allgemeine Intelligenz (vgl. Wittmann & Suess, 1999), die nicht ausschlaggebend für den Erwerb der Dentaltechniken in den praktischen Kursen benötigt werden (Gray & Deem, 2002a; Kao et al., 1990).

### 3.3.2.1 Konsistente, einfache Aufgaben

Die zuvor angestellten Überlegungen stehen aber im Widerspruch zu aktuellen Studien, die für räumliches Vorstellungsvermögen moderate prognostische Effekte in Bezug auf die Laborkursleistungen fanden (z.B. Giuliani et al., 2007). Integriert man in Ackermans (1988) Drei-Phasen-Modell des Fertigkeitserwerbs jedoch zusätzlich die Bedingungen Aufgabenkomplexität und Aufgabenkonsistenz (vgl. Unterkapitel 2.4.2.1), ergibt sich ein erweiterter Blickwinkel auf die Beziehungen zwischen Fähigkeiten und den Leistungen in den Laborkursen beim Erlernen der Dentaltechniken. Die im Technisch-Propädeutischen Kurs des ersten Semesters zu erstellenden Prüfstücke sind als konsistente, einfache Aufgaben zu charakterisieren, da deren Bearbeitung grundlegende Dentaltechniken erfordert, für deren Erlernen bei der Aufgabenbearbeitung nicht fortwährend neue Reiz-Reaktionsverbindungen ausgebildet werden müssen (vgl. Unterkapitel 4.1.5.2).

Mit zunehmender Übung erfahren Zahnmedizinstudierende bei der Bearbeitung konsistenter Aufgaben in der assoziativen Phase eine erhöhte Arbeitsgeschwindigkeit und -genauigkeit, sodass gleichzeitig die Anforderungen an das räumliche Vorstellungsvermögen als kognitive Fähigkeit verringert werden. Darüber hinaus können die Prüfstücke jederzeit von allen Seiten betrachtet werden, da sie nicht in einen Typodonten (Gebissmodell) oder Phantomkopf eingespannt sind, weshalb keine Notwendigkeit besteht, mentale Modelle von nicht direkt beobachtbaren Zahn- und Gebissstrukturen zu entwickeln bzw. abzurufen.

In der abschließenden autonomen Phase des Fertigkeitserwerbs sind die grundlegenden Dentaltechniken weitestgehend automatisiert, sodass die individuellen Leistungsunterschiede zur Anfertigung des jeweiligen Prüfstücks in dieser Phase am höchsten mit dem instrumentellen, manuellen Geschick korrelieren sollten. Daher ist vergleichbar zu der Untersuchung von Kao et al. (1990) zu erwarten, dass die Fähigkeit instrumentelles, manuelles Geschick einen moderaten Einfluss auf die Leistung konsistenter, einfacher Aufgaben beim Fertigkeitserwerb in den vorklinischen Laborkursen ausübt, hingegen räumliches Vorstellungsvermögen mit Bezug zu den Untersuchungen von Beck et al. (1978), Evans und Dirks (2001), Hegarty et al. (2009) und Kao et al. (1990) hierbei als Prädiktor einen eher schwachen Einfluss aufweist. Aus diesen Überlegungen lassen sich die folgenden Hypothesen ableiten:

### Hypothese 3:

Es besteht ein signifikanter, positiv schwacher Zusammenhang zwischen räumlichem Vorstellungsvermögen und der Leistung konsistenter, einfacher praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten.

### Hypothese 4:

Es besteht ein signifikanter, positiv moderater Zusammenhang zwischen instrumentellem, manuellen Geschick und der Leistung konsistenter, einfacher praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten.

Die Fragestellung, ob Schulabschlussnoten zur Vorhersage der praktischen Leistung in den vorklinischen Laborkursen geeignet sind, wird einheitlich in Primärstudien verneint (vgl. Bellanti, et al., 1972; Lundergan et. al, 2007; Walcott et al., 1986; Weinstein & Kiyak, 1981; Wong et al., 1979). Schulabschlussnoten, wie auch fachspezifische Verständnistests basieren nicht auf räumlichen oder manuellen Fähigkeiten, weshalb beide Kriterien nicht zur Prädiktion der vorklinischen Laborkursleistungen geeignet sind (Deubert et al., 1975; Kao et al., 1990). Vergleichbar zur Untersuchung von Gray und Deem (2002a) werden in der vorliegenden Arbeit die Schulabschlussnote und das naturwissenschaftliche Verständnis aus komplementären Gründen eingebunden, um räumliches Vorstellungsvermögen als kognitive Fähigkeit des Erwerbs von Dentaltechniken in den vorklinischen Laborkursen zu bestätigen. Für die Schulabschlussnote als auch für das naturwissenschaftliche Verständnis wird konform zu den in der Literatur berichteten Studienergebnissen kein Zusammenhang zur Leistung konsistenter, einfacher Aufgaben beim Erlernen der Dentaltechniken in den vorklinischen Laborkursen erwartet (vgl. Ranney et al., 2005). Aus den Überlegungen leiten sich die Hypothesen ab:

### Hypothese 5:

Es besteht kein Zusammenhang zwischen der Schulabschlussnote und der Leistung konsistenter, einfacher praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten.

### Hypothese 6:

Es besteht kein Zusammenhang zwischen naturwissenschaftlichem Verständnis und der Leistung konsistenter, einfacher praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten.

Folgt man den angestellten Überlegungen zu den vorangegangenen Hypothesen, ist für die Vorhersage konsistenter, einfacher Aufgabenleistungen in den vorklinischen Laborkursen zu erwarten, dass die kognitive Fähigkeit räumliches Vorstellungsvermögen über instrumentelles, manuelles Geschick, Schulabschlussnoten und naturwissenschaftliches Verständnis eine schwache inkrementelle Validität besitzt (vgl. Bellanti et al., 1972; Oudshoorn, 2003; Walcott et al., 1986). Ackermans (1988) Theorie des Fertigkeitserwerbs zufolge, verringert sich mit zunehmender Erfahrung im Umgang mit konsistenten, einfachen Aufgaben das Korrelat von räumlichem Vorstellungsvermögen und der Aufgabenleistung, während mit zunehmender Automatisierung der erlernten Fertigkeiten das Korrelat zwischen manueller Fähigkeit und Aufgabenleistung ansteigt. Fügt man somit instrumentelles, manuelles Geschick als Prädiktor in ein Modell zur Vorhersage der Leistung konsistenter, einfacher Aufgaben beim Erwerb von zahnmedizinischen Fertigkeiten hinzu, in das bereits räumliches Vorstellungsvermögen, die Schulabschlussnote und naturwissenschaftliches Verständnis eingeschlossen sind, so ist ein moderater Anstieg an aufgeklärter Leistungsvarianz durch die manuelle Fähigkeit zu erwarten (Kao et al., 1990). Aus den Überlegungen leiten sich die Hypothesen ab:

#### Hypothese 7:

Die kognitive Fähigkeit räumliches Vorstellungsvermögen besitzt eine schwache inkrementelle Validität über die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick, die Schulabschlussnote und das naturwissenschaftliche Verständnis für die Vorhersage der Leistung konsistenter, einfacher praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten.

#### Hypothese 8:

Die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick besitzt eine moderate inkrementelle Validität über die kognitive Fähigkeit räumliches Vorstellungsvermögen, die Schulabschlussnote und das naturwissenschaftliche Verständnis für die Vorhersage der Leistung konsistenter, einfacher praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten.

Für die Schulabschlussnote und das naturwissenschaftliche Verständnis als tradierte Prädiktoren der akademischen Studienleistung in der Zahnmedizin werden keine Zusammenhänge zur Leistung konsistenter, einfacher Aufgaben beim Erlernen

zahnmedizinischer Fertigkeiten postuliert (vgl. Unterkapitel 4.3.5), sodass Hypothesen zu ihrem Inkrement redundant sind.

### **3.3.2.2 Inkonsistente, komplexe Aufgaben**

Die Aufgaben in den vorklinischen Laborkursen stehen in didaktischen Bezug zueinander, d.h., einige der in einer Aufgabe benötigten Dentaltechniken werden bereits in der vorherigen erlernt. So muss beispielsweise im ersten Phantomkurs ein tief zerstörter Zahn im Phantomkopf aufgebaut werden, was unter anderem in den Aufgaben Aufwachsübung und Interimsprothese des vorherigen Laborkurses geübt wurde. Über die Laborkursaufgaben ist demzufolge ein Anstieg der Komplexität zu verzeichnen, der aber maßgeblich durch die Einführung des Phantomkopfes, einer Nachbildung des oralen Kopfsegmentes eines Patienten, erhöht wird. Grundlegende Dentaltechniken der Zahnbehandlung, wie das Präparieren (Abschleifen, Bohren) eines Zahnes zum Einsetzen einer Füllung, das Anfertigen und abschließende Einsetzen der Füllung, werden jetzt am Phantomkopf mit praxisgleichen Materialien und Instrumenten durchgeführt. Die größere Anzahl an Aufgabenschritten und erforderlichen Dentaltechniken sowie die veränderte Darbietungsform durch den Phantomkopf führen zu einer höheren Aufgabenkomplexität, die eine Automatisierung der Dentaltechniken erschwert, sodass Ackermans (1988) Theorie des Fertigkeitserwerbs zufolge, keine Reduktion der kognitiven Fähigkeit räumliches Vorstellungsvermögen zu erwarten ist.

Die Arbeiten am Phantomkopf bedeuten zudem auch eine Veränderung der Aufgabenkonsistenz, in denen das Korrelat zwischen kognitiver Fähigkeit und Aufgabenleistung nicht abnimmt (Ackerman, 1988). Der Phantomkopf verhindert nicht nur die direkte Sicht auf alle Zahnbereiche, weshalb mentale, zahnatomische Modelle für die erfolgreiche Bearbeitung der Laborkursaufgaben erforderlich sind, sondern erfordert auch eine fortwährende Adaption der mentalen Modelle. Die Fertigung, Kontrolle und Korrektur von Passung und Sitz zahntechnischer Arbeiten im Phantomkopf stellen somit inkonsistente Aufgaben dar, welche die fortwährende Bildung neuer Reiz-Reaktionsbeziehungen erfordern, verstärkt durch den Umstand, dass Studierende zu Beginn des Erlernens neuer Dentaltechniken viele Verfeinerungen und Nachbesserungen vornehmen, die keinen direkten Bezug zur eigentlichen Aufgabe haben (Chambers & Geissberger, 1997).

Nach Ackerman (1988) stoppt bei inkonsistenten Aufgaben die Prozeduralisierung in der kognitiven Phase, sodass auch mit zunehmender Übung kein verringernder Einfluss

des räumlichen Vorstellungsvermögens auf die Aufgabenleistung zu erwarten ist. Selbst bei einem inkonsistenten Aufgabenanteil von circa 30 Prozent sollte die Vorhersage der Aufgabenleistung eher für inkonsistente anstatt für konsistente Aufgaben erfolgen (Ackerman, 1992). Somit wären die Studienergebnisse von Gansky et al. (2004), Hegarty et al. (2009), Kramer et al. (1989) und Lundergan et al. (2007) zu erklären, die räumliches Vorstellungsvermögen als Prädiktor der praktischen Laborkursleistung nachwies bzw. Durchschnittsnoten über mehrere (vor)klinische Laborkurse zur Operationalisierung der praktischen Studienleistung verwendeten, in denen die Bearbeitung der Aufgaben primär am Phantomkopf erfolgten. Weiterführend könnte darin auch begründet sein, weshalb in Studien keine oder nur schwache Zusammenhänge zwischen räumlichem Vorstellungsvermögen als kognitive Fähigkeit und der Laborkursleistung im ersten Semester bzw. Studienjahr als Kriterium der praktischen Studienleistung zu finden waren, da hier konsistente, einfache Aufgaben von Studierenden bearbeitet wurden (z.B. Bellanti et al., 1972; Coy et al., 2003).

Es ist mit Bezug auf Ackermans (1988) Theorie des Fertigkeitserwerbs daher zu erwarten, dass die kognitive Fähigkeit räumliches Vorstellungsvermögen und die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick einen moderaten Einfluss auf die Leistung inkonsistenter, komplexer Aufgaben beim Erlernen der Dentaltechniken in den vorklinischen Laborkursen besitzen. Trotz Übungseffekten durch zusätzliche, freiwillige Laborzeiten oder unterschiedlich entwickelten Fähigkeitsniveaus aufgrund von Hobbies oder einer beruflichen Ausbildung sowie der Abstände zwischen den Messungszeitpunkten von Fähigkeiten und Aufgabenleistungen, sollte eine zuverlässige Vorhersage der individuellen Leistungsunterschiede von Zahnmedizinstudierenden möglich sein und keinen verringenden Einfluss auf ihr Korrelat ausüben (Ackerman, 1992; Ackerman & Cianciolo, 2000). Die inhaltlichen Überlegungen führen zu den nachstehenden Hypothesen:

#### Hypothese 9:

Es besteht ein signifikanter, positiv moderater Zusammenhang zwischen räumlichem Vorstellungsvermögen und der Leistung inkonsistenter, komplexer praktischer Aufgaben beim Erwerb von zahnmedizinischen Fertigkeiten.

Hypothese 10:

Es besteht ein signifikanter, positiv moderater Zusammenhang zwischen instrumentellem, manuellen Geschick und der Leistung inkonsistenter, komplexer praktischer Aufgaben beim Erwerb von zahnmedizinischen Fertigkeiten.

Das Erlernen von zahnmedizinischen Untersuchungs- und Behandlungsmethoden am Phantomkopf erfordert keine kognitiven Fähigkeiten wie allgemeine Intelligenz oder logisches Schlussfolgern (Deubert et al., 1975; Kao et al., 1990). Demzufolge sollten Schulabschlussnoten und das naturwissenschaftliche Verständnis auch keinen Einfluss auf die Leistung inkonsistenter, komplexer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten ausüben (Deem, 2002b). Primär sind hierfür räumliches Vorstellungsvermögen als Vertreter kognitiver Fähigkeiten (vgl. Hegarty et al., 2009) und instrumentelles, manuelles Geschick als nichtkognitive Fähigkeit ausschlaggebend (vgl. Lundergan et al, 2007). Aus diesen Annahmen leiten sich die Hypothesen ab:

Hypothese 11:

Es besteht kein Zusammenhang zwischen der Schulabschlussnote und der Leistung inkonsistenter, komplexer praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten.

Hypothese 12:

Es besteht kein Zusammenhang zwischen naturwissenschaftlichem Verständnis und der Leistung inkonsistenter, komplexer praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten.

Die Leistung inkonsistenter, komplexer Aufgaben kann in jeder Phase des Fertigkeitserwerbs gut durch die benötigten kognitiven Fähigkeiten vorhergesagt werden, wie Ackerman (1988) in seiner Theorie zum Fertigkeitserwerb postuliert. Das räumliche Vorstellungsvermögen sollte den Überlegungen der vorangegangenen Hypothesen zufolge eine moderate inkrementelle Validität über das instrumentelle, manuelle Geschick, die Schulabschlussnote und das naturwissenschaftliche Verständnis besitzen, da bei inkonsistenten Aufgaben ständig neue Reiz-Reaktionsverbindungen produziert werden müssen, wie am Beispiel einer Zahnpräparation im Phantomkopf deutlich wird. Das Präparieren eines Zahns erfordert nicht nur eine fortwährende Kontrolle des Arbeitsfortschritts, ob das Bearbeiten des Zahnes im exakten Maß erfolgt

und mit den schrittweisen Vorgaben übereinstimmt, sondern für nicht direkt einsehbare Zahnbereiche auch eine Adaption des mentalen Zahnmodells aufgrund der strukturellen anatomischen Veränderungen. Ebenfalls sollte instrumentelles, manuelles Geschick eine moderate inkrementelle Validität über das räumliche Vorstellungsvermögen, die Schulabschlussnote und das naturwissenschaftliche Verständnis aufweisen, da der Einfluss der nichtkognitiven Fähigkeit auf die Leistung inkonsistenter, komplexer Aufgaben auch während der Aufgabenübung konstant ausfällt. Aus den postulierten Annahmen leiten sich die Hypothesen ab:

Hypothese 13:

Die kognitive Fähigkeit räumliches Vorstellungsvermögen besitzt eine moderate inkrementelle Validität über die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick, die Schulabschlussnote und naturwissenschaftliches Verständnis für die Vorhersage der Leistung inkonsistenter, komplexer praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten.

Hypothese 14:

Die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick besitzt eine moderate inkrementelle Validität über die kognitive Fähigkeit räumliches Vorstellungsvermögen, die Schulabschlussnote und das naturwissenschaftliche Verständnis für die Vorhersage der Leistung inkonsistenter, komplexer praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten.

Da für die Schulabschlussnote und das naturwissenschaftliche Verständnis keine Zusammenhänge zur Leistung inkonsistenter, komplexer Aufgaben beim Erlernen der Dentaltechniken am Phantomkopf erwartet werden, entfallen Aussagen zu ihrem Inkrement in Form von Hypothesen.

### 3.3.3 Graphische und tabellarische Darstellung der Hypothesen

Das konzeptionelle Untersuchungsmodell zur Vorhersage der praktischen Aufgabenleistungen beim Erwerb zahnmedizinischer Fertigkeiten in den vorklinischen Laborkursen ist in Abbildung 3.01 dargestellt. Die praktische Studienleistung in den vorklinischen Laborkursen wird in konsistente, einfache Aufgaben und inkonsistente, komplexe Aufgaben unterteilt und entsprechend als Aufgabenleistung I und II in Abbildung 3.01 bezeichnet. Die vorklinischen Laborkursleistungen werden von den zum Erwerb der Dentaltechniken benötigten Fähigkeiten räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick in Abhängigkeit des Aufgabentypus unterschiedlich bedingt, was mit Bezug auf Ackermans (1988) Theorie des Fertigkeitserwerbs in den Untersuchungshypothesen postuliert ist.

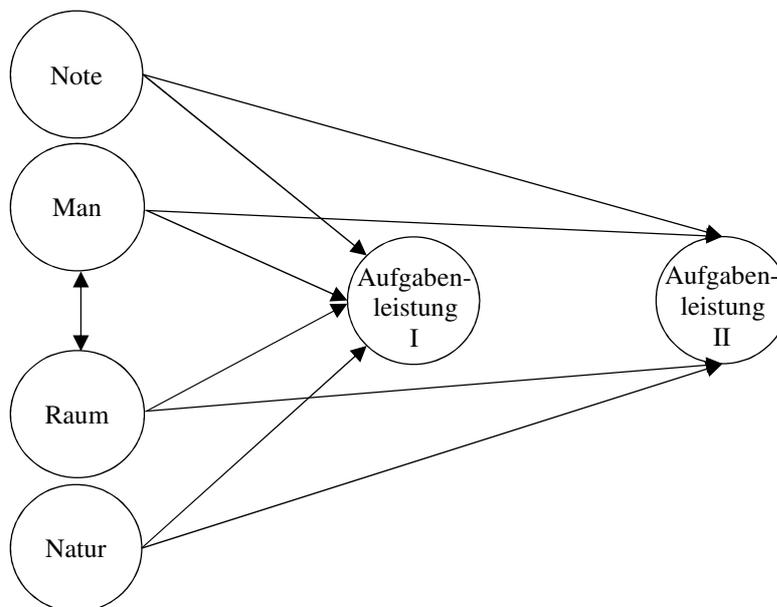


Abbildung 3.01: Untersuchungsmodell zur Vorhersage der praktischen Aufgabenleistungen

Anm.: Note = Schulabschlussnote; Man = instrumentelles, manuelles Geschick; Raum = räumliches Vorstellungsvermögen; Natur = naturwissenschaftliches Verständnis; Aufgabe I = konsistente, einfache praktische Aufgaben; Aufgabe II = inkonsistente, komplexe praktische Aufgaben

Die Hypothesen sind in aggregierter Form in Tabelle 3.01 zusammengefasst. Die Entscheidung über die Annahme der Hypothesen, in denen Zusammenhänge zwischen den Leistungen in den vorklinischen Laborkursen und dem räumlichen Vorstellungsvermögen sowie dem instrumentellen, manuellen Geschick postuliert sind, erfolgen mit einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ . Keine Zusammenhänge zu den Aufgabenleistungstypen werden hingegen für die Schulabschlussnote und dem naturwissenschaftlichen Verständnis vermutet, weshalb es beim Hypothesentesten

inhaltlich schwerwiegender wäre, einen Fehler 2. Art zu begehen. Dem vorbeugend wird das Signifikanzniveau in diesen vier Hypothesen auf  $\alpha = 0.2$  erhöht werden, um die Wahrscheinlichkeit eines  $\beta$ -Fehlers zu verringern.

Tabelle 3.01: Übersicht zu den Untersuchungshypothesen

Hypothese – Inhalt	
1	Hohe Konstruktreliabilität für das instrumentelle, manuelle Geschick und räumliche Vorstellungsvermögen.
2	Hohe diskriminante Validität für die Konstrukte instrumentelles, manuelles Geschick und räumliches Vorstellungsvermögen.
3	Signifikanter, schwach positiver Zusammenhang zwischen räumlichem Vorstellungsvermögen und der Leistung konsistenter, einfacher praktischer Aufgaben.
4	Signifikanter, moderat positiver Zusammenhang zwischen instrumentellem, manuellen Geschick und der Leistung konsistenter, einfacher praktischer Aufgaben.
5	Kein Zusammenhang zwischen der Schulabschlussnote und der Leistung konsistenter, einfacher praktischer Aufgaben.
6	Kein Zusammenhang zwischen naturwissenschaftlichem Verständnis und der Leistung konsistenter, einfacher praktischer Aufgaben.
7	Schwache inkrementelle Validität des räumlichen Vorstellungsvermögens über die anderen Prädiktoren für die Vorhersage der Leistung konsistenter, einfacher praktischer Aufgaben.
8	Moderate inkrementelle Validität des instrumentellen, manuellen Geschicks über die anderen Prädiktoren für die Vorhersage der Leistung konsistenter, einfacher praktischer Aufgaben.
9	Signifikanter, moderat positiver Zusammenhang zwischen räumlichem Vorstellungsvermögen und der Leistung inkonsistenter, komplexer praktischer Aufgaben.
10	Signifikanter, moderat positiver Zusammenhang zwischen instrumentellem, manuellen Geschick und der Leistung inkonsistenter, komplexer praktischer Aufgaben.
11	Kein Zusammenhang zwischen der Schulabschlussnote und der Leistung inkonsistenter, komplexer praktischer Aufgaben.
12	Kein Zusammenhang zwischen naturwissenschaftlichem Verständnis und der Leistung inkonsistenter, komplexer praktischer Aufgaben.
13	Moderate inkrementelle Validität des räumlichen Vorstellungsvermögens über die anderen Prädiktoren für die Leistung inkonsistenter, komplexer praktischer Aufgaben.
14	Moderate inkrementelle Validität des instrumentellen, manuellen Geschicks über die anderen Prädiktoren für die Leistung inkonsistenter, komplexer praktischer Aufgaben.

## **4. Methodik**

Das folgende Kapitel ist in drei Abschnitte zur Darstellung des quasiexperimentellen Untersuchungsdesigns unterteilt. In Unterkapitel 4.1 ist die Operationalisierung der Untersuchungsvariablen beschrieben. Die Untersuchungsdurchführung findet im Rahmen des zahnmedizinischen Studierendenauswahlverfahrens an der Medizinischen Fakultät der Universität Hamburg statt und ist inklusive der Kontrolle möglicher Störvariablen in Unterkapitel 4.2 dargestellt. Abschließend sind in Unterkapitel 4.3 die statistischen Verfahren zur Überprüfung der Untersuchungshypothesen aufgeführt.

### **4.1 Operationalisierung der Untersuchungsvariablen**

In den nachfolgenden fünf Unterkapiteln 4.1.1 bis 4.1.5 ist die Operationalisierung der Untersuchungsvariablen detailliert beschrieben. Das Alter (in Jahren) und das Geschlecht der Bewerber (1 = Mann, 2 = Frau) werden als demographische Daten erhoben. Hypothetische Konstrukte, wie instrumentelles, manuelles Geschick oder räumliches Vorstellungsvermögen, lassen sich im Vergleich zum Geschlecht oder der Körpergröße nicht direkt beobachten, weshalb sie als latente Konstrukte bezeichnet werden. Für latente Konstrukte ist es nicht möglich, direkte empirische Werte zu messen. Die Messbarkeit theoretischer Konstrukte in Strukturgleichungsmodellen erfolgt mittels Operationalisierung über messfehlerbehaftete Indikatoren, die zugleich das Messmodell bilden (Herrmann, Huber & Kressmann, 2006).

Die Beziehung zwischen latentem Konstrukt und Indikatoren kann reflektiver oder formativer Art sein. Ein reflektiver Zusammenhang bedeutet, dass die Indikatoren durch die latente Variable in ihrer Ausprägung beeinflusst werden. Das Gegenteil ist beim formativen Zusammenhang der Fall. Hier üben die Indikatoren einen Einfluss auf eine latente Variable aus (Fornell & Bookstein, 1982). Die Änderung eines formativen Indikators beeinflusst daher den Wert der latenten Variable(n), ohne jedoch dabei zu einer Veränderung der Ausprägung aller übrigen Messvariablen führen zu müssen. Eine gute Darstellung der Entscheidungskriterien, ob der Zusammenhang zwischen Indikatoren und latentem Konstrukt reflektiver oder formativer Natur ist, findet sich bei Christophersen und Grape (2006), Diamantopoulos (1999) und Jarvis, MacKenzie und Podsakoff (2003). In der Literatur werden in Strukturgleichungsmodellen unabhängige als exogene und abhängige als endogene Variablen bezeichnet.

#### 4.1.1 Instrumentelles, manuelles Geschick

Das instrumentelle, manuelle Geschick wird mittels eines Drahtbiegetests erfasst. Dafür wird der als Hamburger Auswahlverfahren für Medizinische Studiengänge - Manueller Test (HAM-Man) bezeichnete Drahtbiegetest eingesetzt.

##### 4.1.1.1 Beschreibung des HAM-Man

Der HAM-Man ist ein von der Arbeitsgruppe Auswahlverfahren an der Medizinischen Fakultät der Universität Hamburg nach testtheoretischen Kriterien entwickelter Drahtbiegetest, bei dem innerhalb von 45 Minuten mit Hilfe einer handelsüblichen Spitzzange Drähte nach Vorlagen gebogen werden müssen. Der HAM-Man erfasst manuelles Geschick im manipulativen Umgang mit Handinstrumenten, in der vorliegenden Arbeit als instrumentelles, manuelles Geschick bezeichnet. Bei der Testentwicklung des HAM-Man wurde auf die erfahrungsgeleitete, intuitive Methode zurückgegriffen, um das Erlernen von Dentaltechniken in den vorklinischen Laborkursen abzubilden. Die Figuren des HAM-Man (s. Abbildung 4.01) sind im Größenverhältnis 1:1 als Vorlage zusammen mit den Testinstruktionen auf einer DIN A4 Seite aufgeführt.

Bei der Figur Gleichseitiges Dreieck beginnt der Draht in der rechten Ecke des Dreieckes, verläuft dann zur oberen Spitze, nach unten links und dann nach unten rechts. Der Draht muss unten rechts nicht mit dem überhängenden Ende abschließen.

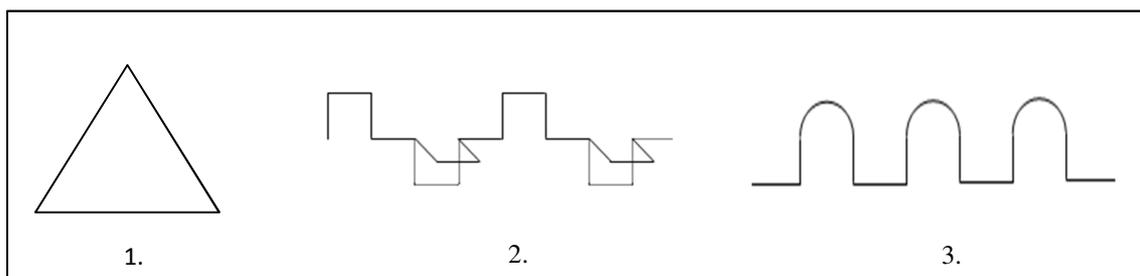


Abbildung 4.01: Figurenvorlagen des HAM-Man (keine Originalgröße), 1. Gleichseitiges Dreieck, 2. Dritte Dimension, 3. Rundbögen

In der Figur Dritte Dimension müssen das zweite und vierte Quadrat senkrecht aus der Ebene nach oben stehen. Alle Winkel haben 90 Grad, alle Strecken sind gleich lang. Der Draht muss nicht mit den Enden der Figur abschließen. Die zweidimensionale Figur Rundbögen soll aus drei identischen Rundbögen bestehen, wobei der Draht nicht mit den Enden der Figur abschließen muss. Insgesamt umfasst das Testmaterial sechs harte

Chromium-Drähte mit einem Durchmesser von 0.8 und einer Länge von 150 Millimetern. Drei Drähte können zu Probiebungen verwendet werden.

#### **4.1.1.2 Auswertung des HAM-Man**

Jede gebogene Drahtfigur wird nach drei Kriterien von zwei Ratern unabhängig voneinander beurteilt. Maximal können sechs Punkte für die Deckungsgleichheit erreicht werden, d.h., der gebogene Draht muss exakt auf der Zeichnung bzw. Vorlage verlaufen. Das plane Aufliegen des Drahtes bewertet, inwieweit der gebogene Draht plan auf dem Tisch liegt und nicht von der Tischebene absteht, wofür maximal sechs Punkte vergeben werden. Bei der Figur Dritte Dimension wird das plane Aufliegen für beide Ebenen geprüft. Für die Qualität der Biegung bzw. die Sauberkeit der Biegung dürfen keine mehrfachen Knicke und somit Sollbruchstellen vorliegen, um maximal drei Punkte zu erzielen, wobei auch die Vergabe von halben Punkten möglich ist. Dieses Kriterium ist nur mit 50 Prozent der üblichen Punktzahl gewichtet, um die technischen Beschränkungen der verwendeten Materialien, vor allem der handelsüblichen Spitzzange, für ein optimales Ergebnis zu berücksichtigen. Die Punkte werden über zwei Rater gemittelt; insgesamt können maximal 45 Punkte erreicht werden. Aufgrund der fehlenden Testnormierung nach beispielsweise Alter und Geschlecht wird als normorientiertes Maß eine Rangreihenfolge der Ergebnisse aller Teilnehmer eines Durchgangs gebildet.

#### **4.1.1.3 Güte des HAM-Man**

Bisher ist eine Überprüfung der Güte des Drahtbiegetests HAM-Man nicht erfolgt, weshalb in der vorliegenden Arbeit die Validität und Reliabilität des Testverfahrens nach Kriterien der Klassischen Testtheorie ermittelt wird (s. Unterkapitel 4.3.1). Zur Ermittlung der Kriteriumsvalidität wird Lienerts Drahtbiegeprobe (DBP) in der 2. Auflage aus dem Jahr 1967 eingesetzt, welche ein bewährtes Instrument zur Erfassung des manuellen Geschicks ist.

In der DBP muss ein weicher Draht mit einer Länge von 25 Zentimetern und einem Durchmesser von einem Millimeter gemäß der in Abbildung A01 im Anhang aufgeführten Vorlage mit der Hand gebogen werden. Die Vorlagenfigur ist gegenüber der zu biegenden Drahtfigur im Verhältnis sieben zu fünf vergrößert abgebildet. Der von Lienert (1961; 1967) entwickelte Test besitzt kein Zeitlimit, es wird im Handbuch eine durchschnittliche Bearbeitungsdauer von 15 bis 20 Minuten berichtet. Die

Auswertung der Strecken, Winkel und Kreisbögen der gebogenen Drahtfigur erfolgt mittels Messschablone nach 29 Bewertungskriterien, die in die Bereiche Form und Sauberkeit unterteilt sind.

Basierend auf einer Stichprobe von 1.700 Teilnehmern liegen für die DBP alters-, berufs- und schulspezifische Prozentrang- und Standardwerte vor. Lienert (1967) berichtete für die DBP Splitt-half-Reliabilitäten von  $r_{tt} = 0.91$  und  $r_{tt} = 0.88$  sowie eine Retest-Reliabilität von  $r = 0.78$ , deren Messwiederholungszeitraum 14 Wochen betrug. Zwischen der DBP und der Leistung Auszubildender in Metallberufen ermittelte Lienert einen Zusammenhang von  $r = 0.47$  ( $n = 40$ ), wofür die Beurteilungen von Vorgesetzten eineinhalb Jahre nach Absolvieren der Drahtbiegeprobe herangezogen wurden.

#### **4.1.1.4 Operationalisierung als latentes Konstrukt**

Das instrumentelle, manuelle Geschick wird als Bi-Faktor-Modell operationalisiert, welches konzeptionell in Abbildung 4.02 dargestellt ist. In einem Bi-Faktor-Modell besteht für jedes Item bzw. Indikator eine Beziehung zu einem allgemeinen Faktor, der sowohl die Gemeinsamkeit, als auch die Unterschiedlichkeit aller Indikatoren auf eine Zieldimension reflektiert (Reise, Moore & Haviland, 2010). In der vorliegenden Untersuchung repräsentiert das instrumentelle, manuelle Geschick als latente Variable die Zieldimension, d.h., den allgemeinen Faktor. Die Bewertungskriterien Deckungsgleichheit, planes Aufliegen und Qualität der Biegung bilden für die Drahtfiguren Dreieck, Dritte Dimension und Rundbögen die insgesamt neun manifesten Variablen, die auf dem allgemeinen Faktor des Bi-Faktor-Modells reflektiv laden. Der Grad des instrumentellen, manuellen Geschicks beeinflusst die Güte der Drahtfiguren, sodass die Beziehung zwischen allgemeinem Faktor und allen neun Indikatoren reflektiver Art ist (s. Abb. 4.02).

Darüberhinaus erlaubt ein Bi-Faktor-Modell gegenüber einem Ein-Faktor-Modell weitere Gruppenfaktoren zu spezifizieren, welche die Kovariation der Indikatoren erfassen, die unabhängig von der durch den allgemeinen Faktor erfasste Kovariation ist (Reise, Morizot & Hays, 2007). Es daher empfohlen, eher eine bi-faktorielle als eine mehrfaktorielle Modellstruktur zu postulieren, falls Überlegungen auf Abhängigkeiten der Items zu anderen Faktoren hinweisen, die einen Teil der Itemvarianz bedingen, der nicht durch den allgemeinen Faktor erklärt wird (Gibbons & Hedeker, 1992).

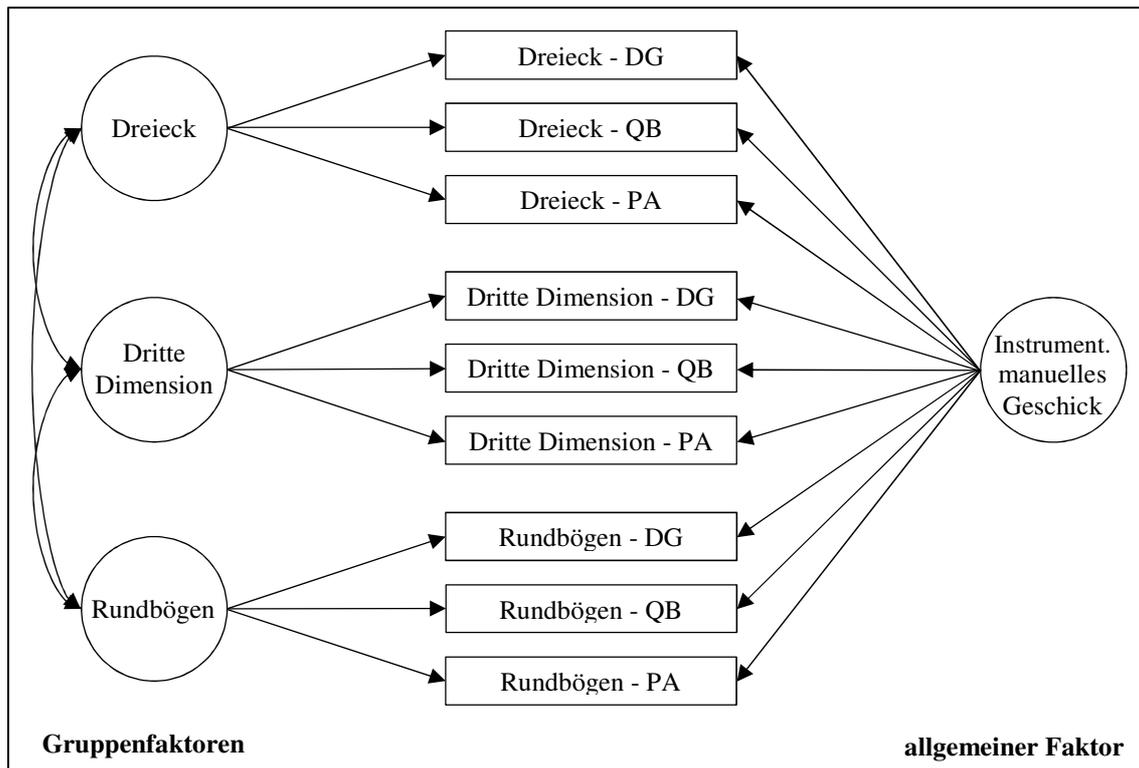


Abbildung 4.02: Bi-Faktor-Modell des instrumentellen, manuellen Geschicks (konzeptionelle Darstellung)

Anm.: DG = Deckungsgleichheit, QB = Qualität der Biegung, PA = planes Aufliegen, instrument. = instrumentelles

Die Leistung im HAM-Man sollte primär durch das instrumentelle, manuelle Geschick bedingt sein. Es besteht jedoch auch ein Einfluss auf die Varianz der Indikatorenwerte durch die Formen der zu biegenden Figuren (vgl. Abb. 4.01), für die unterschiedliche Schwierigkeiten bestehen. Im Bi-Faktor-Modell ist der allgemeine Faktor unabhängig von den Gruppenfaktoren, die miteinander in korrelativer Beziehung stehen. Im Vergleich zu hierarchischen Messmodellen mit Konstrukten zweiter Ordnung stellt der allgemeine Faktor des Bi-Faktor-Modells gegenüber den Gruppenfaktoren keine inhaltlich übergeordnete Ebene dar, sondern ist ihnen gleichgestellt (Reise et al., 2007). Die Multidimensionalität des postulierten Messmodells ist charakterisiert durch die drei Gruppenfaktoren Dreieck, Dritte Dimension und Rundbögen sowie durch den allgemeinen Faktor instrumentelles, manuelles Geschick.

Die Überprüfung der postulierten Bi-Faktor-Struktur zur Erfassung des manuellen Geschicks erfolgt mittels konfirmatorischer Faktorenanalyse, die als Sonderfall kovarianzbasierter Strukturgleichungsmodelle gilt und im Unterkapitel 4.3.2 beschrieben ist. Dem Bi-Faktor-Modell werden zusätzlich ein Ein-Faktor-Modell und ein Drei-Faktoren-Modell als konkurrierende Modelle gegenüber gestellt. Das Ein-Faktor-Modell beinhaltet instrumentelles, manuelles Geschick als Faktor, auf welchem

die neun Indikatoren reflektiv laden (vgl. Abbildung A01 im Anhang). Im Drei-Faktoren-Modell stellt jede Drahtfigur des HAM-Man jeweils einen Faktor dar, dem die zugehörigen Bewertungskriterien Deckungsgleichheit, Qualität der Biegung und planes Aufliegen als reflektive Indikatoren zugeordnet sind (vgl. Abbildung A02 im Anhang).

#### 4.1.2 Räumliches Vorstellungsvermögen

Das räumliche Vorstellungsvermögen wird mittels der Skala technische Begabung des Leistungsprüfsystems (LPS) von Horn (1983) erfasst.

##### 4.1.2.1 Testbeschreibung des LPS

Die zweite Auflage des Leistungsprüfsystems von Horn (1983) basiert auf dem von Thurstone 1938 entwickelten Primärfaktorenmodell und erfasst über eine Testbatterie die kognitive Leistungsfähigkeit, zu der räumliches Vorstellungsvermögen zählt. In Tabelle 4.01 sind die Untertests (UT) der Skala technische Begabung beschrieben. Der LPS kann als Einzel- oder Gruppentest bei Kindern ab neun Jahren, bei Erwachsenen bis zu 50 Jahren eingesetzt werden. Die Untertests der Skala technische Begabung beinhalten je 40 Aufgaben, die mit einem Zeitlimit versehen sind. Für UT 7 stehen zwei Minuten, für UT 8 vier und für die UT 9 sowie 10 je drei Minuten für die Aufgabebearbeitung zur Verfügung.

Tabelle 4.01: Skalen des LPS (2. Auflage, 1983)

Aufgabenbezeichnungen nach Horn	Faktorenzugehörigkeit nach Horn	Skala ( $r_{tt}^*$ )
UT 7: spiegelbildlich gedruckte Zeilen erkennen	Räumliche Beziehungen	technische Begabung (.99)
UT 8: räumliches Vorstellen	Räumliche Vorstellungsfähigkeit	
UT 9: Flächen bei geometrischen Körpern zählen	Räumliche Orientierung	
UT 10: versteckte Muster erkennen	Flexibilität der Gestaltbildung	

Anm.: \* Splitt-Half Reliabilität

##### 4.1.2.2 Auswertung des LPS

Die Auswertung des LPS erfolgt mittels vorgefertigter Schablobe. Die erhaltenen Rohwertsummen werden anhand von Normtabellen in altersrelativierte C-Werte transformiert und in ein Profil direkt auf die Testbögen eingetragen.

##### 4.1.2.3 Gütekriterien des LPS

Als Gütekriterien für den LPS finden sich eine Retest-Reliabilität  $r_{tt} = 0.95$  für das Gesamtergebnis sowie Splitt-Half Reliabilitäten der einzelnen Untertests, welche zwischen  $r_{tt} = 0.90$  und  $r_{tt} = 0.99$  ( $n = 200$ ) liegen. Der LPS weist eine Korrelation in

Höhe von  $r = 0.74$  zum Intelligenz-Struktur-Test von Amthauer (1970) auf. Faktorenanalytische Verfahren identifizierten für den LPS die Skalen „reasoning“, „verbal factor“, „space“ und „perceptual speed“. Die Normierung erfolgte mit 10000 Personen, sodass altersgestaffelte C-Werte für alle Untertests vorliegen.

#### **4.1.2.4 Operationalisierung als latentes Konstrukt**

Das räumliche Vorstellungsvermögen wird als Multi-Item-Messmodell operationalisiert, um die unterschiedlichen Facetten des latenten Konstrukts besser zu erfassen und die Indikatorenreliabilitäten für die Gütebeurteilung des Messmodells ermitteln zu können. Mit diesen Vorteilen geht aber auch der Nachteil des erhöhten Messaufwandes einher (Bergkvist & Rossiter, 2007). Die Skala technische Begabung des LPS zur Messung des räumlichen Vorstellungsvermögens bildet mit den in Tabelle 4.01 aufgeführten Untertests spiegelbildlich gedruckte Zeiten erkennen, räumliches Vorstellen, Flächen bei geometrischen Körpern zählen und versteckte Muster erkennen die manifesten Variablen zur Operationalisierung des latenten Konstrukts. Der Grad des räumlichen Vorstellungsvermögens beeinflusst die Ergebnisse der vier Untertests, weshalb ein reflektiver Zusammenhang zwischen den manifesten Variablen und der latenten Variable hergeleitet werden kann. Horns Testvalidierung von 1983 zeigte mittels Faktorenanalyse die Zugehörigkeit der Untertests sieben bis zehn zu einem Faktor, weshalb Eindimensionalität für das latente Konstrukt räumliches Vorstellungsvermögen angenommen werden kann.

Die im Unterkapitel 2.3.1.4 aufgeführten Definitionen sowie die Studien zur Konstruktvalidierungen des PAT, der als Subtest des Studierfähigkeitstests DAT in den USA und Kanada räumliches Vorstellungsvermögen erfasst, weisen auf ein mehrdimensional zu operationalisierendes Konstrukt hin (vgl. Unterkapitel 2.3.3.4). Colom, Contreras, Botella und Santacreu (2002) zufolge ist aber eine eindeutige Separierung einzelner räumlicher Fähigkeiten, wie „visualization“, „spatial relations“ oder „spatial orientation“, nicht möglich, da substantielle Überschneidungen zwischen den Fähigkeiten bestehen. Die unterschiedlichen von den Autoren getesteten Modelle bestätigten eine Operationalisierung räumlichen Vorstellungsvermögens als Ein-Faktor-Konstrukt gegenüber einer zwei- bzw. dreifaktoriellen Lösung.

#### 4.1.2.5 Neukonzeption des LPS

Die von Kreuzpointner, Lukesch und Horn (2013) vorgenommene Überarbeitung des LPS führte zu einer Neukonzeption der Faktorenstruktur sowie der Bezeichnung als Leistungsprüfsystem 2 (LPS-2) und greift die von Kreuzpointner (2010) in seiner Dissertationsschrift aufgezeigten Strukturschwächen des LPS in der zweiten Auflage von 1983 auf. Das latente Konstrukt räumliches Vorstellungsvermögen ist im LPS-2 als Faktor visuelle Wahrnehmung der allgemeinen Intelligenz (g-Faktor) operationalisiert, der die Fähigkeiten zur mentalen Bildverarbeitung, zur Visualisierung und zum Erkennen von Mustern umfasst und über die Untertests mentale Rotation (UT 6), Flächenzahl (UT 7) und Linienmuster (UT 8) erfasst wird.

Der Untertest 8 zum räumlichen Vorstellen der Skala technische Begabung im LPS von 1983 wurde aufgrund der fehlenden Diskriminanz zu den Untertests 9 und 10 nicht in den LPS-2 aufgenommen. Die Untertests 7 und 9, die im LPS räumliche Beziehungen und Orientierung erfassten, wurden im LPS-2 als Untertest 6 und 7 eingefügt. Der Untertest 8 erfasst im LPS-2 die Fähigkeit Wahrnehmungsgeschwindigkeit, indem unter Zeitdruck in Linienmustern eine vorgegebene Form wiedererkannt werden muss. Die Antwortformate der Untertest 7 und 8 wurden im LPS-2 leicht überarbeitet, ansonsten blieben die Aufgaben unverändert. Die von Kreuzpointner et al. (2013) mittels konfirmatorischer Faktorenanalyse vorgenommene Überprüfung der postulierten Messstruktur des LPS-2 ( $GFI = 0.96$ ,  $RMSEA = 0.43$ ,  $\chi^2 = 60.39$ ,  $\chi^2/df = 1.51$ ,  $df = 40$ ) ergab für die Untertests 6 bis 8 als Indikatoren Faktorladungen  $\lambda_{ij}^2$  in Höhe von 0.55, 0.63 und 0.59 auf die latente Variable visuelle Wahrnehmung. Der LPS-2 konnte in der vorliegenden Arbeit nicht als Messinstrument eingesetzt werden, da dessen Veröffentlichung im Frühjahr 2013 nach der Untersuchungsdurchführung erfolgte.

#### 4.1.3 Schulabschlussnote

Die Operationalisierung der Schulabschlussdurchschnittsnote erfolgt anhand der Abiturdurchschnittsnote der Bewerber um einen Studienplatz in der Zahnmedizin am Universitätsklinikum Hamburg-Eppendorf.

##### 4.1.3.1 Beschreibung der Abiturdurchschnittsnote

Die Schulabschlussnote ist repräsentiert durch die Abiturdurchschnittsnote als numerische Zahl mit einer Nachkommastelle, die von der Stiftung für

Hochschulzulassung im Rahmen der Durchführung des Studienplatzauswahlverfahrens an die Arbeitsgruppe Auswahlverfahren an die Medizinischen Fakultät der Universität Hamburg übermittelt wird. Das Zeugnis der allgemeinen Hochschulreife gilt als Nachweis für die uneingeschränkte Studienbefähigung an einer Universität oder sonstigen Hochschule in der Bundesrepublik Deutschland. Die Abiturprüfungen erfolgen schriftlich und mündlich, jedoch kann das Verfahren sich von Bundesland zu Bundesland unterscheiden. Bundesweit werden die in der Qualifikationsphase der gymnasialen Oberstufe und in den Abiturprüfungen erreichten Noten zur Abiturdurchschnittsnote akkumuliert (Cortina, Baumert, Leschinsky, Mayer & Trommer, 2008).

#### **4.1.3.2 Güte der Schulabschlussnote**

Zwischen Schulabschlussnote bzw. Abiturdurchschnittsnote und Intelligenz werden in der Literatur Korrelationen zwischen 0.30 bis 0.50 berichtet. Aufgrund der mangelnden Reliabilität von Schulnoten können keine stärkeren Korrelationen als 0.60 gefunden werden, weshalb in der Literatur die Beziehung zwischen allgemeiner Intelligenz und Abiturdurchschnittsnote als gut interpretiert wird (Stemmler, Hagemann, Amelang & Bartussek, 2011, S. 196; Asendorpf & Neyer, 2012, S. 150 f.; Wittmann & Suess, 1999). Krumm, Ziegler und Bühner (2008) konnten in ihrer Untersuchung die kognitive Fähigkeit Schlussfolgern in einem Regressionsmodell als besten Prädiktor für Schulnoten nachweisen. Die Arbeitsgedächtniskomponente Koordination beinhaltet die Fähigkeit, neue Beziehungen zwischen Elementen herzustellen und diese Beziehungen in Strukturen zu implementieren. Die gemeinsamen Varianzanteile der kognitiven Fähigkeit Schlussfolgern und der Arbeitsgedächtniskomponente Koordination konnten insbesondere die Schulleistungen in Noten der naturwissenschaftlichen Fächer, wie Mathematik, Physik, Chemie und Biologie, vorhersagen.

#### **4.1.3.3 Operationalisierung als latentes Konstrukt**

Die Schulabschlussnote wird als Single-Item-Messmodell operationalisiert. In einem Single-Item-Messmodell wird die latente Variable nur über einen Indikator abgebildet (Bergkvist & Rossiter, 2007). Das hypothetische Konstrukt Schulabschlussnote wird anhand der Abiturdurchschnittsnote als Indikator formativ operationalisiert, da eine Veränderung in der Abiturdurchschnittsnote auch zu einer Veränderung in der latenten Variable führt. Gravierende Nachteile von Single-Item-Messmodellen sind die fehlende

Abbildung der unterschiedlichen Facetten eines latenten Konstrukts sowie die fehlende Reliabilität zur Gütebeurteilung. Es ist jedoch in der Praxis nicht immer möglich, latente Konstrukte optimal, wie in der Literatur gefordert, über drei oder mehr Indikatoren zu operationalisieren (Rossiter, 2002). Die über die Bundesländer unterschiedlichen Abiturprüfungsbestimmungen lassen nur eine Operationalisierung der Schulabschlussnote über die Abiturdurchschnittsnote als Einzelindikator zu. Zur Einbindung in das Prädiktionsmodell soll die Schulabschlussnote als hypothetisches Konstrukt modelliert werden, obwohl Noten direkt beobachtbar sind und somit keine latente Variable darstellen.

Die Schulabschlussnote kann als Single-Item-Messmodell automatisch nur über eine Dimension verfügen, da der Indikator zu ein und demselben theoretischen Konzept bzw. Dimension gehört (Albers & Hildebrandt, 2006). Dabei können Single-Item-Konstrukte über gleiche prädiktive Validitäten wie Messmodelle verfügen, welche dieselbe latente Variable über mehrere manifeste Variablen operationalisiert (Bergkvist & Rossiter, 2007), insbesondere falls kleine Stichproben vorliegen und schwache Effektgrößen zu erwarten sind (Diamantopoulos, Sarstedt, Fuchs, Wilczynski & Kaiser, 2012).

#### **4.1.4 Naturwissenschaftliches Verständnis**

Der naturwissenschaftliche Studierfähigkeitstest „Hamburger Auswahlverfahren für Medizinische Studiengänge - Naturwissenschaftstest“ (HAM-Nat) wird zur Auswahl von Zahnmedizinierenden an der Medizinischen Fakultät der Universität Hamburg eingesetzt. In der vorliegenden Arbeit soll das Gesamtergebnis zur Messung des naturwissenschaftlichen Verständnisses herangezogen werden.

##### **4.1.4.1 Beschreibung des HAM-Nat**

Der HAM-Nat ist ein papierbasierter Multiple-Choice-Test mit 80 Fragen zu (zahn)medizinisch relevanten Aspekten der Fächer Mathematik, Physik, Chemie und Biologie, deren Schwierigkeitsgrad auf gymnasialem Oberstufenniveau angesiedelt ist. Jedes Jahr wird ein nicht festgelegter Anteil neuer Fragen entwickelt, der in einen Fragenpool eingeht, aus welchem die Testfragen entnommen werden. Als Bearbeitungszeit stehen 120 Minuten zur Verfügung. Das Testverfahren wurde nach zweijähriger Entwicklungszeit zum ersten Mal in 2008 zur Auswahl von Studienbewerbern in der Humanmedizin an der Medizinischen Fakultät der Universität

Hamburg verwendet. Seit 2009 wird der HAM-Nat in identischer Ausführung als Auswahlinstrument für Bewerber um einen Studienplatz in der Zahnmedizin eingesetzt.

#### **4.1.4.2 Auswertung des HAM-Nat**

Jede Multiple-Choice Frage des HAM-Nat besitzt eine richtige Lösungsantwort von fünf Antwortmöglichkeiten, die mit einem Punkt bewertet ist. Falsch oder gar nicht beantwortete Fragen werden nicht mit Strafpunkten versehen. Die Anzahl der richtigen beantworteten Fragen wird zu einem Gesamtwert summiert. Eine normierte Auswertung nach Alter und Geschlecht oder eine Differenzierung in Subskalen existieren zum Anfertigungszeitpunkt der vorliegenden Arbeit nicht. Daher wird für die Auswertung als normorientiertes Maß eine Rangreihenfolge der Ergebnisse aller Teilnehmer eines Durchgangs gebildet.

#### **4.1.4.3 Güte des HAM-Nat**

Hampe, Klusmann, Buhk, Münch-Harrach und Harendza (2008) fanden zwar nur eine geringe prognostische Validität des HAM-Nat ( $R^2 = 0.095$ ) für den Studienerfolg in der Medizin, die aber höher ausfiel als die der Abiturdurchschnittsnote ( $R^2 = 0.066$ ). Der Studienerfolg wurde anhand der bestandenen Leistungsnachweise der ersten zwei Fachsemester, dem Wintersemester 2006/07 und dem Sommersemester 2007 der Humanmedizin an der Medizinischen Fakultät der Universität Hamburg operationalisiert. Der Zuwachs an erklärter Leistungsvarianz betrug für den HAM-Nat über die Abiturdurchschnittsnote 6.4 Prozent. Paralleltest- und Retest-Reliabilitäten im Bereich von  $0.53 \leq r_{tt} \leq 0.67$  fanden Hissbach, Klusmann und Hampe (2011) für den HAM-Nat anhand einer aus 361 Studienanfängern bestehenden Stichprobe. Das Testmodul naturwissenschaftliches Denken, welches inhaltlich und strukturell dem Modul medizinisch-naturwissenschaftliches Grundverständnis des TMS entspricht, korrelierte zwischen  $0.21 \leq r \leq 0.34$  mit dem HAM-Nat; zur Abiturdurchschnittsnote berichteten die Autoren eine Korrelation von  $r = -0.24$ .

#### **4.1.4.4 Operationalisierung als latentes Konstrukt**

Das naturwissenschaftliche Verständnis wird als Single-Item-Messmodell operationalisiert (vgl. Unterkapitel 4.1.3.3). Das hypothetische Konstrukt wird anhand des Gesamtergebnisses im Naturwissenschaftstest HAM-Nat als Indikator reflektiv umgesetzt, da die Ausprägung des naturwissenschaftlichen Verständnisses die Höhe des

Testergebnisses bedingt. Das latente Konstrukt naturwissenschaftliches Verständnis verfügt als Single-Item-Messmodell automatisch nur über eine Dimension (Albers & Hildebrandt, 2006).

#### **4.1.5 Praktische Studienleistung**

Die Aufgaben der vorklinischen Laborkurse sind in Tabelle 4.02 aufgeführt, welche an der Medizinischen Fakultät der Universität Hamburg den Technisch-Propädeutischen-Kurs (TPK), den Phantomkurs I (PhaI) und den Phantomkurs II (PhaII) der Zahnersatzkunde umfassen. Die Benotung erfolgt anhand von Kriterien (s. Tabelle A22 im Anhang) als numerische Zahl mit einer Nachkommastelle, deren Notenintervalle von 1.0 bis 1.5 mit sehr gut, 1.6 bis 2.5 mit gut, 2.6 bis 3.5 mit befriedigend, 3.6 bis 4.5 mit ausreichend, 4.6 bis 5.5 mit mangelhaft und ab 5.6 mit ungenügend codiert sind.

##### **4.1.5.1 Beschreibung der Laborkurse**

Die im Kursverlauf anzufertigenden Arbeiten (s. Tabelle 4.02) sind aus dem zu Kursbeginn ausgehändigten Aufgabenheften ersichtlich, in denen die einzelnen Arbeitsschritte ausführlich beschrieben sind. Alle Arbeiten und ebenfalls die Aufgabenhefte dürfen nicht aus den Laborräumen entfernt werden. Im TPK erfolgt die Bewertung der Arbeiten zu zwei Kurszeitpunkten, während in den beiden Phantomkursen jede Arbeit bzw. jedes Prüfstück an festgelegten Kurszeitpunkten separat benotet wird. Eine ausführliche Beschreibung der Arbeiten in den Laborkursen der zahnmedizinischen Vorklinik findet sich unter anderem bei Eichner und Kappert (2005), Lehmann, Hellwig und Wenz (2008) und Strub (1994).

An der Medizinischen Fakultät der Universität Hamburg müssen der regulär im ersten Semester stattfindende TPK erfolgreich abgeschlossen werden, um zum Phantomkurs I im folgenden Sommersemester zugelassen zu werden. Das Bestehen des ersten Phantomkurses ist Voraussetzung für die Teilnahme am zweiten Phantomkurs, der ab dem fünften Studiensemester absolviert werden kann. Der TPK wird in jedem Semester angeboten, während die Phantomkurse nur einmal im Jahr stattfinden. Neben dem Anfertigen der praktischen Arbeiten muss eine Vorlesung besucht und eine schriftliche Klausur bestanden werden, um den Kursschein zu erhalten. Die Klausurergebnisse werden nicht zur Operationalisierung der praktischen Laborkursleistung herangezogen, da diese überwiegend durch kognitive Fähigkeiten und Persönlichkeitseigenschaften beeinflusst sind (vgl. Bellanti et al., 1972; Thompson, 1947).

Tabelle 4.02: Aufgaben in den vorklinischen Laborkursen

Laborkurs	Aufgabe	Beschreibung	Aufgabentyp*
TPK	Schnitzübung (Note1)	Maßstabsgetreues Nacharbeiten eines ein- und dreiwurzeligen Zahnmodells aus einem Wachsbloch.	konsistente, einfache Aufgabe
	Aufwachsübung (Note2)	Die aus dem Gebissmodell entfernten Zähne (26, 35, 36) werden durch eine spezielle Aufwachstechnik ersetzt.	konsistente, einfache Aufgabe
	Vollgusskrone (Note3)	Herstellen einer Vollgusskrone für einen Zahn (36).	konsistente, einfache Aufgabe
	Interimsprothese (Note4)	Herstellen einer provisorischen Prothese mit drei Klammern, einem Schalt- und einem Freundsattel für den Unterkiefer.	konsistente, einfache Aufgabe
PhaI	Totalprothese (Note1)	Herstellen einer Totalprothese im Ober- und Unterkiefer nach dem Aufstellungskonzept von Gysi.	inkonsistente, komplexe Aufgabe
	Stiftaufbau und Schalenprovisorium (Note2)	Aufbau eines tief zerstörten Zahns (21) mit einem Titanstift und provisorischer Versorgung mittels eines Schalenprovisoriums.	inkonsistente, komplexe Aufgabe
	Präparation und Tangentialbrücke (Note3)	Präparation zweier Zähne (44, 46) für eine Brücke und definitive Versorgung des Lückengebisses mit einer Vollguss-Tangentialbrücke.	inkonsistente, komplexe Aufgabe
PhaII	Totalprothese (Note1)	Herstellen einer Totalprothese im Ober- und Unterkiefer nach dem Aufstellungskonzept von Gerber.	inkonsistente, komplexe Aufgabe
	Präparation und Schalenprovisorium (Note2)	Präparation zweier Zähne (23, 25) und provisorische Versorgung mittels eines Schalenprovisoriums.	inkonsistente, komplexe Aufgabe
	Michiganschiene (Note3)	Herstellen einer Stabilisierungsschiene bei schmerzhafter, verspannter Muskulatur, Bruxismus und Kiefergelenkarthrose.	inkonsistente, komplexe Aufgabe

\* Klassifizierung der Aufgabenkomplexität und -konsistenz (vgl. Unterkapitel 2.4.2.1)

#### 4.1.5.2 Operationalisierung als latentes Konstrukt

Die Operationalisierung der praktischen Studienleistung erfolgt über den Erwerb zahnmedizinischer Fertigkeiten bzw. den Fertigkeitserwerb als latente Variable, da die Aufgabenleistungen in den vorklinischen Laborkursen abhängig von den erlernten Dentaltechniken sind. Das Erlernen der erforderlichen zahnmedizinischen Fertigkeiten bestimmt die Qualität der Arbeiten bzw. die Aufgabenleistungen in den Laborkursen und somit auch die Benotung der Prüfstücke. Zwischen dem latenten Konstrukt Fertigkeitserwerb und den vorklinischen Laborkursaufgaben als Indikatoren besteht demzufolge eine reflektive Beziehung, für die Eindimensionalität angenommen werden kann. In den vorklinischen Laborkursen ist zudem zwischen konsistenten, einfachen und inkonsistenten, komplexen Aufgaben zu unterscheiden (vgl. Unterkapitel 2.4.2.1). Die für die Anfertigung der Arbeiten im TPK (vgl. Tabelle 4.02) benötigten Dentaltechniken erfordern beispielsweise das Schneiden, Abtragen, Ausritzen, Fräsen, Polieren und Säubern von Materialien bzw. Werkstoffen sowie die Maßkontrolle, das

Abformungen von Modellen, Ausgießen von Gipsformen, das Aufwachsen oder das Biegen von Halteklammern, um die Werkstoffe nach den Vorgaben im Aufgabenheft zu bearbeiten. Des Weiteren müssen die Zahnmedizinierenden im ersten Laborkurs zwischen Akzeptablem und Nichtakzeptablem unterscheiden, Fehler erkennen und korrigieren sowie Hinweisinformationen richtig interpretieren. Mit Bezug auf die Theorie des Fertigkeitserwerbs von Ackerman (1988) sind die vier Aufgaben im TPK als konsistent und von der Schwierigkeit her als einfach einzustufen.

Die für die Anfertigung der Arbeiten in den Phantomkursen benötigten Dentaltechniken müssen am Phantomkopf angewendet werden. Der Phantomkopf ist eine Modellnachbildung des menschlichen Kopfes zu Übungszwecken für Studierende (Eichner & Kappert, 2005). In der Regel bestehen für den Phantomkopf als Modell für die Studierenden keine Vorerfahrungen, sodass nicht nur die Prüfstücke und Werkstoffe nach den Vorgaben im Übungsheft zu bearbeiten sind, sondern es auch gilt, sie in den Phantomkopf einzupassen. Der Phantomkopf verhindert die direkte Sicht auf bestimmte Zahnbereiche, weshalb nicht nur mentale Zahnmodelle gebildet, sondern auch fortwährend adaptiert werden müssen, da die Arbeiten in den Phantomkursen die ständige Produktion neuer Reiz-Reaktionsverbindungen erfordern. Die in den Phantomkursen zu erlernenden Fertigkeiten beinhalten Tätigkeiten am Phantomkopf, wie beispielsweise das Präparieren bzw. Separieren (Bohren) am Zahn, das Anschleifen von Zähnen zum Einsetzen von Füllungen, die Maßkontrolle mit Spiegel, das Anpassen von Schalenprovisorien, Kieferrelationsbestimmungen sowie das Probieren und Anpassen von Zahnbrücken. Aufgrund der größeren Anzahl an benötigten Aufgabenschritten und Dentaltechniken gegenüber den Arbeiten im TPK sowie der Arbeiten am Phantomkopf sind mit Bezug auf Ackermans (1988) Theorie des Fertigkeitserwerbs, die Aufgaben in den Phantomkursen als inkonsistent und komplex zu klassifizieren.

## **4.2 Untersuchungsdurchführung**

Die Erhebung der Untersuchungsdaten ist im Unterkapitel 4.2.1 dargestellt, die im Studierendenauswahlverfahren, in der Orientierungseinheit und in den vorklinischen Laborkursen der Zahnmedizin an der Medizinischen Fakultät der Universität Hamburg erfolgt. Die Untersuchungsstichprobe und der Umgang mit möglichen Störvariablen sind in den Unterkapitel 4.2.2 und 4.2.3 beschrieben.

### **4.2.1 Datenerhebung**

Die mit der Software PGP 8.1 verschlüsselten Bewerbungsdaten werden von der Stiftung für Hochschulzulassung (SfH) über das Referat 33 der Universität Hamburg eine Woche vor Stattfinden des Studierendenauswahlverfahrens an die Medizinische Fakultät per E-Mail übermittelt. Die zur Datenverarbeitung in Microsoft Excel 2003 importierten Textdateien beinhalten unter anderem Name, Geschlecht und die Abiturdurchschnittsnote aller Bewerber. Im Auswahlverfahren der Zahnmedizin absolvieren alle Bewerber am Vormittag des Testungstages, am 13. August 2010, den Naturwissenschaftstest HAM-Nat und am Nachmittag zuerst die Skala technische Begabung des LPS, gefolgt von dem Drahtbiegetest HAM-Man und der DBP (vgl. Unterkapitel 4.1.1.3). In der ersten Studienwoche wird der HAM-Man wiederholt, um anhand der über das Auswahlverfahren zugelassenen Studierenden die Retest-Reliabilität zu ermitteln. Die vorklinischen Laborkurse werden von der Poliklinik für Zahnärztliche Prothetik des Zentrums für Zahn-, Mund- und Kieferheilkunde am Universitätsklinikum Hamburg-Eppendorf veranstaltet. Die mittels Microsoft Excel 2003 erfassten Leistungsdaten können vom Institutsstudiensekretariat nach Beendigung der Laborkurse bezogen werden.

### **4.2.2 Stichprobe**

Die Untersuchungsstichprobe umfasst 200 zum Auswahlverfahren eingeladene Bewerber um einen Studienplatz in der Zahnmedizin an der Medizinischen Fakultät der Universität Hamburg. 39 von ihnen erhalten eine Zulassung zum Zahnmedizinstudium und können damit an den Laborkursen der Vorklinik teilnehmen. Falls die Bewerberanzahl die Kapazitäten des Auswahlverfahrens übersteigt, erfolgt eine Einladung zum Auswahlverfahren anhand der Abiturdurchschnittsnote und der 1. Ortspräferenz, für die Hamburg als Studienort in der Bewerbung bei der SfH gesetzt

sein muss. Die Zulassung erfolgt über zwei Listen. Die Rangreihenfolge der ersten Liste wird über die linear transformierten Ergebnisse des Naturwissenschaftstests HAM-Nat (maximal 59 Punkte) und der Abiturdurchschnittsnote (maximal 60 Punkte) berechnet. Falls in der ersten Liste ein Rang unter den 70 besten Teilnehmern erreicht wird, besteht die Möglichkeit einer Zulassung über die zweite Liste, die aus den linear transformierten Ergebnissen des Drahtbiegetests HAM-Man (maximal 59 Punkte) und der Abiturdurchschnittsnote (maximal 60 Punkte) gebildet wird. Bewerber, die eine Zulassung über die erste Liste erhalten, werden der zweiten entnommen, in die nachfolgende Bewerber entsprechend nachrücken.

#### **4.2.3 Kontrolle von Störvariablen**

Das quasiexperimentelle Untersuchungsdesign erlaubt nur eine begrenzte Kontrolle möglicher Störvariablen. Aufgrund des zulassungsbeschränkten und gesetzlich regulierten Bewerbungsverfahrens um einen Studienplatz in der Zahnmedizin ist eine randomisierte Aufteilung der Studienplatzbewerber in eine Experimental- und Kontrollgruppe nicht möglich. Vor allem geschlechtsspezifische Leistungsunterschiede im räumlichen Vorstellungsvermögen (vgl. Unterkapitel 2.3.4.1.2) können nicht über eine zufallsbedingte Gleichverteilung kontrolliert werden, sondern sind als natürliche Gegebenheit bzw. Charakteristik der Bewerberpopulation zu betrachten.

Kurzfristige Übungeffekte, die eine deutliche Verbesserung des manuellen Geschicks bewirken, sind nicht anzunehmen (Waldron & Anton, 1995). Unterschiedliche Ausprägungen der manuellen Fähigkeit könnten beispielsweise durch Hobbies oder eine Berufsausbildung zum Zahntechniker bedingt sein, weshalb über den Internetauftritt der Medizinischen Fakultät detaillierte Anweisungen veröffentlicht sind, um das Biegen von Drahtfiguren über einen mehrmonatigen Zeitraum trainieren zu können. Darüber hinaus stehen im HAM-Man drei Drähte für Probeversuche zur Verfügung.

Unterschiede in der Körperhaltung können beim Drahtbiegen auf das manuelle Geschick bis zu zehn Prozent Leistungsdifferenz verursachen (Buffington, MacMurdo & Ryan, 2006) in ihrer Studie aufzeigten, weshalb alle Teilnehmer instruiert werden, nur in aufrecht sitzender Position den Drahtbiegetest durchzuführen. Bhattacharya, Roy und Chatterjee (1978) berichteten einen Anstieg der manuellen Leistung in lauten Umgebungen, weshalb in den Untersuchungsräumen nicht nur eine ruhige Arbeitsatmosphäre, sondern auch weitere Umgebungseinflüsse, wie das Raumklima und

die Lichtverhältnisse, kontrolliert werden, um optimale Durchführungsbedingungen zu gewährleisten.

Es ist außerdem nicht anzunehmen, dass die Bewerber im Auswahlverfahren als auch die Zahnmedizinierenden in den vorklinischen Laborkursen nicht motiviert sind, jeweils bestmögliche Leistungen zu erzielen (vgl. Kanfer & Ackerman, 1989). Gründe dafür sind die freiwillige Studienplatzbewerbung und Teilnahme am Auswahlverfahren sowie das Ziel, die vorklinischen Laborkurse erfolgreich zu bestehen, um zur Zahnmedizinischen Vorprüfung nach dem fünften Semester zugelassen zu werden (vgl. Unterkapitel 2.2.1).

In der Einführungswoche bzw. Orientierungswoche (OE) wird der HAM-Man mit den Erstsemesterstudierenden der Zahnmedizin durchgeführt, um die Retest-Reliabilität anhand der über das Auswahlverfahren zugelassenen Studierenden zu ermitteln. Zur Etablierung einer zum Auswahlverfahren vergleichbaren Prüfungsatmosphäre und Testmotivation werden die Studierenden entsprechend verbal instruiert und die zehn besten Testergebnisse mit je einem Buchgutschein im Höhe von je 50 Euro prämiert.

### 4.3 Auswertung der Untersuchungsfragestellung

In Übereinstimmung mit Potter und McDonald (1985) werden zur Überprüfung der postulierten Hypothesen lineare Strukturgleichungsmodelle als statistische Verfahren eingesetzt, um den gleichzeitigen Einfluss der Prädiktoren räumliches Vorstellungsvermögen, instrumentelles, manuelles Geschick, Schulabschlussnote und naturwissenschaftliches Verständnisses auf die praktische Leistung in den vorklinischen Laborkursen zu analysieren. Insgesamt kann die Laborkursleistung von maximal 39 über das Auswahlverfahren zugelassenen Studierenden erhoben werden. Das ermöglicht in der vorliegenden Untersuchung jedoch nicht den Einsatz kovarianzbasierter Strukturgleichungsmodelle, die in der Regel Stichproben von mindestens 150 Teilnehmern erfordern (Muthén & Muthén, 2002).

Alternativ wird das 1975 von Wold entwickelte Verfahren der „partial least squares“ (PLS) zur Auswertung der Untersuchungsfragestellungen verwendet, welches als varianzbasiertes Strukturgleichungsmodell weniger restriktiv gegenüber geringen Stichprobengrößen und in Unterkapitel 4.3.3 beschrieben ist. Das Vorgehen zur Berechnung der inkrementellen Validitäten für die Prädiktoren räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick ist in Unterkapitel 4.3.3.4 beschrieben. Bevor die in den Hypothesen postulierte Vorhersage der vorklinischen Laborkursleistung aufgegriffen wird, sind zunächst für das räumliche Vorstellungsvermögen und das instrumentelle, manuelle Geschick die Frage zur Konstruktvalidierung zu analysieren, die in den ersten zwei Hypothesen formuliert wurden. Zur Prüfung der Operationalisierung empfiehlt sich das Verfahren der konfirmatorischen Faktorenanalyse, das in Unterkapitel 4.3.2 beschrieben ist. Aufgrund der geringen Stichprobengrößen in den vorklinischen Laborkursen, in denen überzufällig häufig Extremwerte auftreten können, werden komplementär zur PLS-Methode robuste Regressionsanalysen berechnet. Das Verfahren ist im Unterkapitel 4.3.4 erläutert. Abschließend sind in den Unterkapiteln 4.3.5 die Prädiktionsmodelle zur Vorhersage der vorklinischen Laborkursleistung beschrieben und in 4.3.6 die zur Auswertung eingesetzte Software aufgezählt.

Zu Beginn der Auswertung werden die statistischen Grundvoraussetzungen zur Anwendbarkeit der hypothesentestenden Verfahren analysiert, d.h., ob die empirischen Daten vollständig erhoben wurden, methodenbedingte Verzerrungen der Ergebnisse vorliegen, extreme Ausreißer auftreten und ob Normalverteilung für die

Untersuchungsvariablen besteht. Zuvor sind als Exkurs in Unterkapitel 4.3.1 die Gütekriterien nach Klassischer Testtheorie dargestellt, die zur Beurteilung des Drahtbiegetests HAM-Man ermittelt werden.

#### **4.3.1 Exkurs: Güte des HAM-Man nach Klassischer Testtheorie**

Zur Beurteilung der Güte des Drahtbiegetests HAM-Man nach Klassischer Testtheorie (Lienert & Raatz, 1998) erfolgt die Bestimmung der Kriteriumsvalidität, wofür der HAM-Man in Beziehung zur bewährten Drahtbiegeprobe von Lienert (1967) gesetzt wird (vgl. Unterkapitel 4.1.1.3). Ein hoher Zusammenhang ( $r > 0.6$ ) zwischen den Ergebnissen beider Drahtbiegetests deutet darauf hin, dass beide Verfahren dasselbe Konstrukt erfassen.

Die Übereinstimmung der Urteile beider Rater wird durch die Interrater-Reliabilität wiedergegeben, indem die Varianz zwischen verschiedenen Urteilen in Bezug auf ein Messgegenstand mit der über alle Urteile und Messgegenstände entstandenen Varianz verglichen wird (McGraw & Wong, 1996). Als Koeffizient wird hierfür die Intra-Klassen-Korrelation (*ICC*) gegenüber der Korrelation nach Pearson vorgezogen, da die Urteile als Variablen eine gemeinsame Metrik und Varianz aufweisen (Bartko, 1966) und *ICCs* sich außerdem als sehr robust gegenüber Verletzungen der Normalverteilung erweisen (Shoukri, Asyali & Donner, 2004). Im HAM-Man ist verlangt, dass beide Juroren gleiche, absolute Werte vergeben, weshalb als *ICC*-Maß für die Interrater-Reliabilität ein unjustiertes Modell mit zweifaktoriellem Design auf Basis von Einzelwerten als Ratings gewählt wird. In der Literatur gelten *ICC*-Werte ab 0.9 als sehr reliable Beurteilungsübereinstimmung, Werte ab 0.7 als gut, ein *ICC* unter 0.5 gilt als nicht akzeptabel (Wirtz & Caspar, 2002).

Für die Retest-Reliabilität wird als *ICC*-Maß ein unjustiertes Modell mit zweifaktoriellem Design auf Basis von Mittelwerten der Ergebnisse gewählt. Die Retest-Reliabilität zeigt an, ob eine Wiederholung der Messung bei Konstanz der zu messenden Eigenschaft die gleichen Messwerte liefert, wodurch sich der Genauigkeitsgrad an Übereinstimmung zeigt (Cicchetti, 1994). Wirtz (2004) weist für die Interpretation von Reliabilitätskoeffizienten auf den Umstand hin, dass die Reliabilität einer Messung sich unter sonst gleichen Bedingungen verringert, wenn die Merkmalsvarianz in der untersuchten Stichprobe abnimmt. In der Literatur gelten Retest-Reliabilitäten ab 0.9 als sehr reliabel, Werte ab 0.7 als gut (Bühner, 2011).

### 4.3.2 Konfirmatorische Faktorenanalyse

Der Vorteil der konfirmatorischen Faktorenanalyse (*KFA*) mittels kovarianzbasierten gegenüber varianzbasierten Verfahren besteht in den konsistenteren Schätzern, die eher als PLS-Schätzer die tatsächlichen Faktorladungen abbilden, der Möglichkeit konkurrierende Modelle zu testen und der ganzheitlichen Beurteilung des Messmodells mittels globaler Maße. Zu beachten ist die Voraussetzung, nur reflektive Messmodelle durch eine konfirmatorische Faktorenanalyse prüfen zu können. Reflektive unterscheiden sich von formativen Messmodellen dadurch, dass Veränderungen in den Messwerten der Indikatorvariablen durch die latente Variable kausal verursacht werden (Moosbrugger & Kelava, 2008). Anhand des Pfaddiagramms werden in der *KFA* die Beziehungen zwischen den latenten Konstrukten und zu den Indikatoren charakterisiert, die theoretisch erklärbar sein sollten. Die Beziehungen zwischen latenter Variable und den Indikatorvariablen lassen sich nach Bühner (2011) wie folgt beschreiben:

$$x_i = \lambda_i \xi + \delta_i \quad (i=1, \dots, n).$$

Die Wertigkeit eines Indikators ( $x_i$ ) ergibt sich aus der Summe des gewichteten ( $\lambda_i$ ) Faktorwerts ( $\xi$ ) und einem Fehlerterm ( $\delta_i$ ) unter der Annahme  $z$ -standardisierter Werte. Die Stärke des Zusammenhangs zwischen der latenten Variable und dem einzelnen Indikator wird durch die Gewichtunggröße bestimmt und als Faktorladung bezeichnet, die im Fall standardisierter Werte als Korrelation interpretiert werden kann. Der durch die Indikatoren der latenten Variablen nicht aufgeklärte Varianzanteil stellt die Fehlervarianz des Fehlerterms dar (Moosbrugger & Kelava, 2008).

#### 4.3.2.1 Modellidentifikation

Die Modellidentifikation erfordert die Prüfung, ob ausreichend empirische Informationen aus den Indikatoren zur Bestimmung der Modellparameter vorliegen. Darüber hinaus erfolgt die Festlegung einer Metrik für die latenten Variablen und die Fehlervariablen, da deren Varianzen nicht wie bei manifesten Variablen direkt berechnet werden können (Moosbrugger & Schermelleh-Engel, 2008). Ein Modell gilt als identifizierbar, wenn für alle unbekanntes Modellparameter Schätzwerte berechnet werden können. Aus  $p$  empirischen Indikatoren lassen sich insgesamt  $p/2(p+1)$  empirische Varianzen und Kovarianzen berechnen, sodass für die zu schätzenden Modellparameter  $t$  mehr empirische Informationen als zu schätzenden Modellparameter vorhanden sein müssen, was in überidentifizierte Modelle ( $df > 0$ ) gegeben ist.

#### 4.3.2.2 Parameterrestriktionen

Parameterrestriktionen sind eine Vorgehensweise, um die Anzahl der zu schätzenden Parameter für eine erfolgreiche Modellidentifikation zu verringern. Hierfür bieten sich zwei Möglichkeiten zur Festlegung der Metrik für latente Variablen an. Einerseits kann ein Indikator als Referenzvariable einer latenten Variable zugewiesen werden, indem seine Faktorladung auf eins gesetzt wird. Bei der Auswahl sollte auf sachlogische Überlegungen oder auf die Reliabilität der in Frage kommenden Indikatoren zurückgegriffen werden. Dieses Vorgehen ist bei der Bestimmung der Metrik auch für die Fehlervariablen anwendbar, deren Pfade auf die Indikatoren mit der Festlegung auf eins fixiert werden können. Andererseits können die Varianzen der latenten Variablen auf eins fixiert werden, um für alle Indikatorvariablen die freien Faktorladungen zu schätzen. In diesem Fall entspricht die Kovarianz der latenten Variablen ihrer Korrelation (Bühner, 2011).

Eine weitere Vorgehensweise zur Verringerung der Parameterschätzungen besteht in der Bildung eines übergeordneten Indikators durch Aggregation verschiedener Indikatoren bzw. Rohdaten zu Indikatoren. Diese Methode wird in der Literatur als „item parceling“ bezeichnet, für die unterschiedliche Techniken zur Verfügung stehen (Meade & Kroustalis, 2006). Dabei sollte zunächst immer die Dimensionalität der Indikatoren bzw. Rohdaten berücksichtigt werden, um sie beispielsweise entsprechend dieser Gemeinsamkeit zufällig zu Paketen zu aggregieren. Kritisiert wird am „item parceling“, dass Fehlspezifikationen in Modellen nicht aufgedeckt werden können, wodurch die wahre Faktorenstruktur nicht identifiziert wird. Ein mehrdimensionales Modell kann beispielsweise irrtümlicherweise als eindimensional betrachtet werden und Mehrfachladungen als Indizes für eine mehrfaktorielle Struktur unbemerkt bleiben, wodurch die Zusammenhänge und Beziehungen der latenten Variablen im Strukturmodell verzerrt und theoretisch nur bedingt interpretierbar sind. Eine ausführliche Beschreibung zu den Techniken des „item parceling“ finden sich bei Little, Cunningham, Shahar und Widaman (2002).

Eine andere Option zur Verringerung der Parameterschätzungen für eine erfolgreiche Modellidentifikation stellt die Eliminierung von manifesten oder latenten Variablen im Messmodell dar (Bollen & Lennox, 1991). Der Ausschluss bestimmter Indikatoren oder latenten Variablen erfolgt hierbei nur zur Überprüfung der Konstruktvalidierung der verbleibenden Variablen im Rahmen der KFA. Bedingungen für die Vorgehensweise sind geringe bivariate Korrelationen der auszuschließenden Variablen zu den im

Messmodell verbleibenden, sodass keine hohen Zusammenhänge im Messmodell erwartet werden können (MacCallum & Browne, 1993). Gleichzeitig muss die Eliminierung der Indikatoren oder latenten Variablen auf theoretischer Ebene vertretbar sein (Bollen, 2002).

#### 4.3.2.3 Parameterschätzungen

Nur für erfolgreich identifizierte Modelle können Parameter geschätzt werden. Einfach dargestellt, wird in der KFA versucht, die mittels Indikatorenmesswerte berechnete empirische Varianz-Kovarianz-Matrix ( $S$ ) möglichst genau durch geeignete Parameterschätzungen des Messmodells zu replizieren. Die daraus resultierende, d.h., die auf den Parameterschätzungen basierende Varianz-Kovarianz-Matrix ( $\Sigma$ ) bezeichnet man als modelltheoretisch, welche auf dem Fundamentaltheorem der Faktorenanalyse nach Thurstone (1931) basiert und folgendermaßen dargestellt werden kann:

$$\Sigma = \Lambda\Phi\Lambda^T + \Theta.$$

Der Term  $\Lambda$  bezeichnet dabei die Matrix der Faktorladungen,  $\Phi$  die Kovarianzmatrix der Faktorwerte und  $\Theta$  die Kovarianzmatrix der Fehlerterme. Die Güte des Messmodells sowie der Parameterschätzungen zeigt sich darin, inwieweit die empirische Varianz-Kovarianz-Matrix durch die modelltheoretische Varianz-Kovarianz-Matrix replizierbar ist, wofür die Maximum-Likelihood-Methode (ML) am häufigsten als Schätzmethode eingesetzt wird (Browne, 1984). Die ML-Funktionsgleichung ( $F_{ML}$ ) ist formal zu beschreiben als:

$$F_{ML} = \log|\Sigma| + \text{tr} \left( S\Sigma^{-1} \right) - \log|S| - (p + q).$$

Die Terme  $\Sigma$  und  $S$  stellen die modelltheoretische und empirische Kovarianzmatrix dar,  $\text{tr}$  repräsentiert die Summe der Diagonalelemente („trace“) einer quadratischen Matrix und  $p$  sowie  $q$  stehen für die Anzahl der Indikatoren bzw. der zu schätzenden Parameter. Der Ausdruck  $|F_{ML}(N - 1)|$  folgt dabei der Chi-Quadrat-Verteilung, weshalb die ML-Schätzmethode einem statistischen Test unterzogen werden kann. Einfach gesagt, strebt die ML-Methode die Minimierung des Funktionswertes und somit eine Minimierung der Differenz zwischen den modelltheoretischen und empirischen Matrizen an, die zugleich die effizienteste Schätzmethode darstellt (Backhaus, Erichson & Weiber, 2010). Ein weiterer Vorteil ist die Skaleninvarianz der ML-Methode, die Parameterschätzungen werden in ihrer Größe nicht durch Skalentransformationen

beeinflusst. Dafür sind eine Stichprobengröße  $n > 100$  und Multinormalverteilung sowie intervallskalierte Daten der manifesten Variablen Voraussetzungen zur Anwendung des ML-Schätzverfahrens (Kline, 2010). Bühner (2011) zufolge ist die ML-Methode relativ robust gegenüber Verletzungen der Multinormalverteilungsannahme, weshalb eine Verletzung nicht unbedingt zum Ausschluss führen muss. Grewal, Cote und Baumgartner (2004) zeigten anhand ihrer Ergebnisse aus zwei Monte Carlo Simulationen auf, dass die Auswirkungen von Multikollinearität ( $r > 0.90$ ) in Strukturgleichungsmodellen im Vergleich zu Regressionsmodellen geringer ausfallen. Die Autoren betonen die Wichtigkeit der Reliabilität für genaue und zuverlässige Parameterschätzungen in Strukturgleichungsmodellen, sodass selbst das Auftreten von moderater Multikollinearität toleriert werden kann, solange die Indikatoren über hohe Messgenauigkeiten verfügen, eine große Stichprobe zur Verfügung steht und das Modell eine hohe Aufklärungsstärke besitzt.

Führen Parameterschätzungen zu negativen Varianzen oder Korrelationen größer eins, spricht man von unzulässigen Schätzungen, die auch als „Heywood Case“ oder als „out of range parameter estimates“ bezeichnet werden. Der häufigste Grund hierfür sind Modellmisspezifikationen, indem inhomogene Indikatoren im Messmodell oder zu viele bzw. zu wenige Faktoren einbezogen werden. „Heywood Cases“ können auch durch die Verwendung einer zu kleinen Stichprobe, einer zu geringen Indikatoranzahl pro Faktor, Stichprobenfehler oder durch extreme Ausreißerwerte verursacht werden (Dillon, Kumar & Mulani, 1987). Treten unzulässige Parameterschätzungen auf, empfiehlt Kaplan (1988) weniger auf kosmetische Maßnahmen, wie die Fixierung der entsprechenden negativen Varianzen auf Null, zurückzugreifen, sondern auf inhaltlicher Ebene die hergeleiteten Modellstrukturen und Modellspezifikationen auf Plausibilität zu überprüfen.

#### **4.3.2.4 Modellevaluation**

Die zur Beurteilung des Modells verwendeten Kriterien in der konfirmatorischen Faktoranalyse sind in Tabelle 4.03 inklusive der Akzeptanzniveaus aufgeführt. Weiterführend sind die verwendeten Maße zur Bestimmung der Indikator- und Konstruktgüte sowie für die des Gesamtmodells erläutert.

Die zur Beurteilung des Messmodells berechnete Itemschwierigkeit  $p_m$  für Indikatoren mit mehrstufigem Antwortformat entspricht der Empfehlung von Fissini (2004), welche Unterschiede in den Item- bzw. Indikatorenvarianzen einbezieht, indem anstelle

einfacher Punktwerte  $X_i$  ihre Quadrate verwendet werden und sich formal wie folgt darstellen lässt:

$$p_m = \frac{\sum X^2}{\sum X_{\max}^2}$$

Als Trennschärfe  $r_{(it)}$  wird auf die korrigierte Trennschärfe  $r_{i(t-i)}$  bzw. Eigentrennschärfe zurückgegriffen, welche durch die Korrelation eines Indikators mit der Summe aller übrigen Indikatoren seiner Skala gebildet wird, um als weiteres Maß die Indikatorgüte zu ermitteln (Brosius, 2011).

Tabelle 4.03: Gütekriterien der konfirmatorischen Faktorenanalyse

Gütekriterium	Akzeptanzniveau
<b>Messmodell (Indikatoren)</b>	
Kurtosis / Exzess (Exz)	Exz = 0
Schiefe (Sch)	Sch = 0
Trennschärfe ( $r_{i(t-i)}$ )	$r_{i(t-i)} \geq 0.3$
Itemschwierigkeit ( $p_m$ )	$0.2 \leq p_m \leq 0.8$
Korrelation nach Pearson ( $r_{ij}$ )	$.6 \leq r_{ij} \leq .8$
Standardfehler der Schätzung (S.E.)	geringe Ausprägung
Critical Ratio (C.R.)	C.R. > 2.58 (zweiseitig)
Faktorladung: Signifikanz (einseitig, t-Test)	$t > 1.645$ ( $p < 0.05$ , $df_n = 130$ )
Faktorladung: Vorzeichen	entsprechend Theorie
Faktorladung: Stärke ( $\lambda_i^2$ )	$\lambda_i^2 > 0.5$
Kommunalität ( $h^2$ )	$0.5 < h^2 < 0.8$
Indikatorreliabilität ( $rel(x_i)$ )	$rel(x_i) > 0.6$
<b>Strukturmodell</b>	
Faktorreliabilität ( $rel(\xi_j)$ )	$rel(\xi_j) > 0.6$
Durchschnittlich Extrahierte Varianz (DEV)	DEV > 0.5
Interne Konsistenz (Cronbachs Alpha)	$\alpha \geq 0.9$
Fornell/Larcker-Kriterium	$DEV(\xi) > r_{\xi_j, \xi_k}^2$
<b>Gesamtmodell</b>	
Chi-Quadrat / Freiheitsgrade ( $\chi^2/df$ )	$\chi^2/df \leq 2.5$
Root Mean Square Error of Approximation (RMSEA)	RMSEA $\leq 0.05$
Standardized Root Mean Residual (SRMR)	SRMR $\leq 0.11$
Comparative Fit Index (CFI)	CFI $\geq 0.97$
Normed Fit Index (NFI)	NFI $\geq 0.95$

Anm.: i, u = Indikator; j, k = Konstrukt/Faktor

Teilt man den geschätzten unstandardisierten Parameterwert durch den Standardfehler der Schätzung, erhält man die Critical Ratio (C.R.), mittels derer der Parameterbeitrag zur Modellstrukturbildung beurteilt wird. Voraussetzung ist dafür Multinormalverteilung der Daten, sodass durch einen t-Test die Nullhypothese, ob sich die geschätzten Parameter nicht signifikant von Null unterscheiden, geprüft werden kann. Ein C.R. Wert über 1.96 erlaubt die Ablehnung der Nullhypothese mit einer Irrtumswahrscheinlichkeit von fünf Prozent (Backhaus et al., 2011).

Neben der Signifikanz der Faktorladungen ermöglichen das Vorzeichen und die Größe der standardisierten Faktorladungen Beurteilungen der Indikatorengüte im Messmodell. Theoretisch und sachlogisch formulierte Zusammenhänge können anhand der Richtung des Vorzeichens überprüft werden. Die Faktorladung eines Indikators spiegelt seine Beziehung zum Faktor wider, die ab einer Stärke größer 0.5 als akzeptabel zu bezeichnen ist (Homburg, Klarmann & Pflesser, 2008). Aus dem Quadrat der standardisierten Faktorladung berechnet sich die Kommunalität  $h^2$  eines Indikators, die angibt, wie gut dieser durch den zugehörigen Faktor erklärt wird. Die Reliabilität  $rel(x_i)$  eines Indikators zeigt im Vergleich zur Kommunalität nicht nur den durch den zugehörigen Faktor aufgeklärten Varianzanteil an, sondern berücksichtigt auch mögliche auftretende Messfehler (Christophersen, 2007). Formal lässt sich die Reliabilität  $rel(x_i)$  wie folgt beschreiben:

$$rel(x_i) = \frac{\lambda_{ij}^2 \phi_{jj}}{\lambda_{ij}^2 \phi_{jj} + \theta_{ii}}.$$

Der Term  $\lambda_{ij}^2$  bezeichnet die geschätzte, standardisierte Faktorladung zwischen dem Indikator  $x_i$  und dem zugehörigen Faktor  $\xi_j$ . Die geschätzte Varianz des Faktors  $\xi_j$  ist mit  $\phi_{jj}$  dargestellt, die geschätzte Varianz des Messfehlers mit  $\theta_{ii}$ . Reliabilitätswerte von  $rel(x_i) > 1$  deuten auf eine Fehlspezifikation im Modell hin (Backhaus et al., 2011). In der Literatur finden sich Grenzwerte zwischen  $rel(x_i) \geq 0.4$  und  $rel(x_i) \geq 0.5$ , um von einer akzeptablen Reliabilität zu sprechen (Homburg & Giering, 1996).

Albers und Hildebrandt (2006) weisen deutlich darauf hin, dass eine Verletzung des Akzeptanzniveaus eines oder mehrerer Gütekriterien nicht zu einem sofortigen Ausschluss des Indikators aus dem Messmodell zur Verbesserung des Modellfits führen darf. Jede Eliminierung eines Indikators geht nicht nur mit einem Informationsverlust sondern auch mit einer Beeinträchtigung der inhaltlichen Validität einher, die sich entsprechend auf die Forschungsfragestellungen auswirken können (Rossiter, 2002). Die Entscheidung über die Entnahme bzw. das Verbleiben eines Indikators im Messmodell muss daher immer durch sachlogische und theoretische Überlegungen gestützt werden (Weiber & Mühlhaus, 2009).

Die Faktorreliabilität ( $rel$ ) und die Durchschnittlich Extrahierte Varianz ( $DEV$ ) werden zur Prüfung der Messgenauigkeit auf Konstruktebene eingesetzt und wie folgt berechnet:

$$\text{rel}(\xi_j) = \frac{(\sum \lambda_{ij})^2 \phi_{jj}}{(\sum \lambda_{ij})^2 + \sum \theta_{ii}} \quad \text{und} \quad \text{DEV}(\xi_j) = \frac{\sum \lambda_{ij}^2 \phi_{jj}}{\sum \lambda_{ij}^2 \phi_{jj} + \sum \theta_{ii}}.$$

Der Term  $\lambda_{ij}^2$  bezeichnet die geschätzte, standardisierte Faktorladung zwischen dem Indikator  $x_i$  und der zugehörigen latenten Variable  $\xi_j$ . Die geschätzte Varianz der latenten Variable  $\xi_j$  ist mit  $\phi_{jj}$  dargestellt, die geschätzte Varianz der Fehlervariable  $\delta_i$  mit  $\theta_{ii}$  (Backhaus et al., 2010). Als akzeptables Maß gelten für die Faktorreliabilität  $\text{rel}(\xi_j)$  Werte  $> 0.6$ , hingegen spricht man bei der  $\text{DEV}(\xi_j)$  von zufriedenstellenden Werten, die größer als 0.5 sind (Bühner, 2011).

Als Maß der diskriminanten Validität hat sich das Fornell-Larcker-Kriterium etabliert (Kline, 2010). Beim Fornell-Larcker-Kriterium wird die durch einen Faktor erfasste  $\text{DEV}(\xi_j)$  mit jeder quadrierten Korrelation, die der Faktor mit anderen Faktoren im Messmodell aufweist, verglichen. Nach Fornell und Larcker (1981) stellt die quadrierte Korrelation zwischen zwei Faktoren ihre gemeinsame Varianz dar, sodass von Diskriminanzvalidität gesprochen werden kann, sobald die  $\text{DEV}(\xi_j)$  des betrachteten Faktors größer ist als die gemeinsame Varianz mit dem anderen Faktor. Der Chi-Quadrat-Wert  $\chi^2$  wird als deskriptives Anpassungsmaß im Verhältnis zu den Modellfreiheitsgraden interpretiert, wobei ein guter Gesamtmodellfit angenommen werden kann, wenn das Verhältnis  $\chi^2/\text{df} \leq 2,5$  beträgt (Homburg et al., 2008). Die durch das postulierte Modell erklärten Varianz- und Kovarianzanteile werden durch den „goodness of fit index“ (*GFI*) angegeben, wobei gute Modelle einen Fit von  $\geq 0.95$  aufweisen, akzeptable einen  $\geq 0.90$  (Marsh, Balla & McDonald, 1988). Der „adjusted goodness of fit index“ (*AGFI*) ist der um die zur Modelltestung verfügbaren Freiheitsgrade angepasste *GFI*, dessen Wertebereiche gleich dem *GFI* interpretiert werden (Joreskog & Sorbom, 1981).

Weiterführend wird der „Root Mean Square Error of Approximation“ (*RMSEA*) als inferenzstatistisches und absolutes Maß herangezogen, welches die ungefähre Passung des Modells prüft. Ein *RMSEA*-Wert  $\leq 0.05$  lässt auf einen guten, ein Wert  $\leq 0.08$  auf einen akzeptablen und ein *RMSEA*-Wert  $\geq 0.10$  auf einen inakzeptablen Modellfit schließen. Mittels des „p of close fit“ (*PCLOSE*) wird zudem die Irrtumswahrscheinlichkeit für die Nullhypothese angegeben, dass der *RMSEA*-Wert gleich 0.05 ist. Liegt der *PCLOSE* Wert über  $\geq .05$ , so kann auf eine gute Modellpassung geschlossen werden, bei Werten unter  $\leq .05$  hingegen auf eine schlechte

(Browne & Cudeck, 1993, S. 146 ff.). Der *RMSEA* reagiert jedoch als Prüfmaß sensitiv beim Modellvergleich auf Veränderungen in der Anzahl der Indikatoren und Konstrukte, deren Erhöhung zu einer Verkleinerung und damit zu einer Verbesserung des *RMSEA*-Wertes führt (Breivik & Olsson, 2001). Daher wird der „Standardized Root Mean Residual“ (*SRMR*) als weiteres globales Gütemaß zur Modellbeurteilung herangezogen, der nicht die Modellkomplexität berücksichtigt und gegenüber dem Stichprobenumfang eine geringe Sensitivität aufweist (Bühner, 2011).

Der „Comparative Fit Index“ (*CFI*) und der „Normed Fit Index“ (*NFI*) beurteilen, inwieweit das zu beurteilende Modell besser zu den Daten passt als im Vergleich zu einem Unabhängigkeitsmodell bzw. Basis- oder Nullmodell. In dem Unabhängigkeitsmodell werden die Indikatoren als nicht korreliert angenommen, um eine möglichst schlechte Datenpassung zu erreichen. Ein guter Modellfit sollte sich für den *CFI* in Werten  $\geq .97$  und für den *NFI*  $\geq .95$  widerspiegeln. Nach Erfahrung von Bentler und Bonett (1980) sollten Modelle mit einem *CFI* und *NFI* Wert  $\leq 0.90$  grundlegende Verbesserungen erfahren. Im Vergleich zum *NFI* vermeidet der *CFI* eine Unterschätzung der Modellpassung, die sich vor allem bei kleinen Stichproben oft zeigt (Bentler, 1990).

#### 4.3.2.5 Graphische Darstellung des Messmodells

Das in Abbildung 4.03 dargestellte Messmodell beinhaltet die Prädiktoren räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick. Der formativ operationalisierte Prädiktor Schulabschlussnote kann nicht in das zu prüfende Messmodell eingebunden werden, da die KFA nur eine Beurteilung reflektiv operationalisierter Variablen zulässt. Der Prädiktor naturwissenschaftliches Verständnis verfügt zwar über ein reflektives Messmodell, das aber nur einen Indikator besitzt, weshalb Parameterschätzungen, die zu negativen Kovarianzen oder Korrelationen größer eins führen, nicht auszuschließen sind. Somit könnten „Heywood Case“ aufgrund unzulässiger Schätzungen auftreten, da die zu erwartende Untersuchungsstichprobe eher die Mindestanforderung der KFA erfüllt. Zudem kann eine Inhomogenität des Indikators als weitere Ursache für das Auftreten negativer Varianzen oder Korrelationen größer eins nicht ausgeschlossen werden, da bislang keine Studien über die Teststruktur bzw. die Dimensionalität des als Indikator verwendeten Naturwissenschaftstest HAM-Nat vorliegen

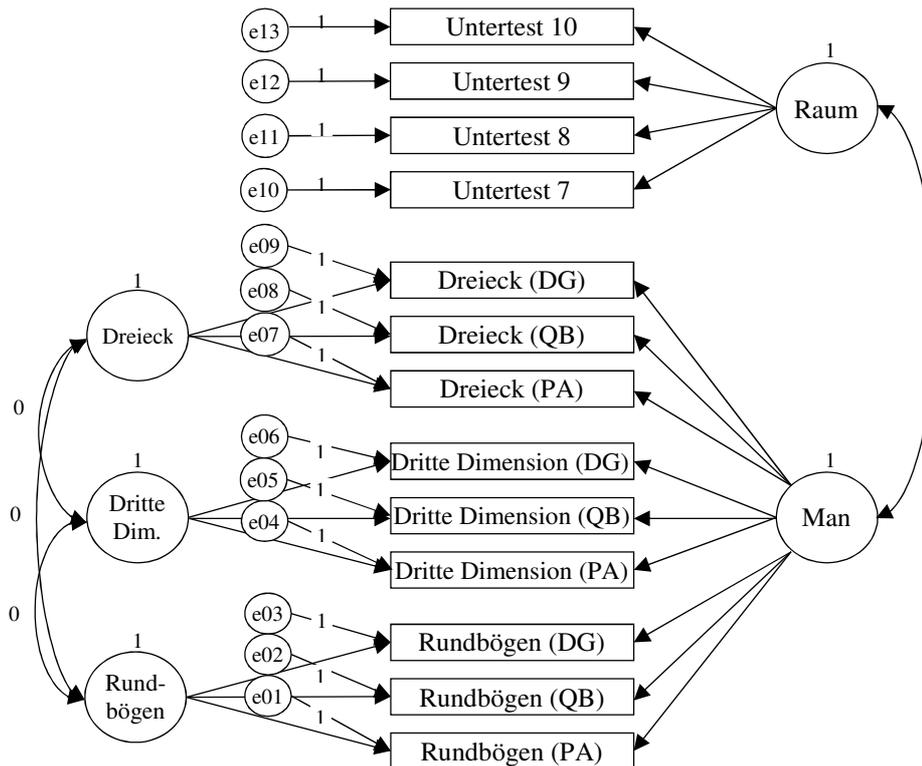


Abbildung 4.03: Messmodell im Prädiktionsmodell der praktischen Studienleistung

Anm.: DG = Deckungsgleichheit; QB = Qualität der Biegung; PA = planes Aufliegen; Dritte Dim. = Dritte Dimension; e = Fehlervariable, Raum = räumliches Vorstellungsvermögen; Man = instrumentelles, manuelles Geschick

Die Bi-Faktor-Modellstruktur des Prädiktors instrumentelles, manuelles Geschick wird zur Prüfung konkurrierenden Modellen gegenübergestellt. Das in Abbildung A02 im Anhang dargestellte einfaktorische Modell folgt der Annahme, instrumentelles, manuelles Geschick wird nur über einen Faktor erfasst, ohne dass die einzelnen Aufgaben bzw. Drahtfiguren des HAM-Man einen individuellen Beitrag auf das instrumentelle, manuelle Geschick ausüben.

Im Drei-Faktoren-Modell (s. Abbildung A03 im Anhang) wird hingegen genau diese Aussage postuliert. Die eigenständigen Faktoren Dreieck, Dritte Dimension und Rundbögen stehen in korrelativer Beziehung zueinander und erfassen instrumentelles, manuelles Geschick. Die latenten Variablen e01 bis e13 stellen im Messmodell die Fehlervariablen da, welche jeweils den Messfehler der entsprechenden Indikatorvariablen repräsentieren. Die Faktorladungen der Fehlervariablen e01 bis e13 auf die manifesten Variablen sind auf eins fixiert, um die Varianzen der Fehlervariablen zu schätzen (vgl. Unterkapitel 4.3.3.2). Die Varianzen der latenten Variablen räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick sowie der Faktoren Dreieck, Dritte Dimension und Rundbögen im Bi-Faktor-Modell sind ebenfalls auf eins fixiert, um einerseits die Modellidentifikation und andererseits das Schätzen der freien

Faktorladungen zu ermöglichen. In diesem Fall sind die Kovarianzen identisch zu den Korrelationen (Backhaus et al., 2011). Die Kovarianzen zwischen den Gruppenfaktoren sind auf Null fixiert, um negative Werte zu vermeiden, die zu einer nicht positiv definiten Kovarianz-Matrix führen würden, da aufgrund des Zeitlimits im HAM-Nat, in Abhängigkeit der gewählten Bearbeitungsstrategie und des ansteigenden Schwierigkeitsgrades im Drahtbiegetest, sich hohe Werte in einer Drahtfigur und niedrige in anderen ergeben könnten.

### 4.3.3 Partial Least Squares-Pfadanalyse

Vergleichbar mit kovarianzbasierten Strukturgleichungsmodellen besteht ein PLS-Modell aus einem Strukturmodell, welches die Beziehungen zwischen den latenten Variablen reflektiert und einem Messmodell, das anzeigt, wie latente Variablen und ihre Indikatoren miteinander über ihre Ladungen in Beziehung stehen, die zur Schätzung der Parameter der latenten Variablen verwendet werden (Chin & Newsted, 1999). Prinzipiell kann man die Schätzung von Kausalmodellen mittels PLS-Algorithmus als zweistufigen Schätzungsprozess beschreiben (Ringle, 2004). In der ersten Stufe gibt es vier Schritte, wobei mit Schritt (4) begonnen wird und die Schritte (1) bis (4) bis zur Konvergenz wiederholt werden, die sich formal in Anlehnung an Panten und Boßow-Thies (2006) wie folgt darstellen lassen:

#### 1. Berechnung der inneren Gewichte

$$(1) \zeta_{ji} = \begin{cases} \text{sign cov}(X_i; X_j) & \text{falls } X_i \text{ und } X_j \text{ angrenzend sind} \\ 0 & \text{ansonsten} \end{cases}$$

#### 2. Approximation des inneren Modells

$$(2) \tilde{X}_j = \sum_i v_{ij} X_i$$

#### 3. Bestimmung der äußeren Gewichte

$$(3) x_{hjn} = w_{hj} X_{jn} + e_{hjn} \quad \text{Mode A (reflektive Indikatoren)}$$

$$(4) X_{jn} = \sum_{hj} (w_{hj} x_{hjn}) + d_{jn} \quad \text{Mode B (formative Indikatoren)}$$

#### 4. Approximation des äußeren Modells

$$(5) X_{jn} = \hat{\eta}_{jn} = \int_j \sum_{hj} w_{hj} x_{hjn} \cdot$$

Der Term  $x_{hjn}$  bezeichnet den Indikatormesswert  $h$  des Konstrukts  $j$  bei Objekt  $n$ ,  $X_{jn}$  die geschätzten Werte der latenten Variablen  $\eta_j$ . Die Fallwerte für  $X_{jn}$  werden dabei als gewichtete Aggregate ihrer jeweiligen Indikatoren generiert, während die Koeffizienten  $\omega_{hj}$  Schätzungen für  $\pi_{hj}$  und  $w_{hj}$  darstellen. Das  $\int_j$  repräsentiert ein Skalar, welches aus schätztechnischen Gründen eine einheitliche Varianz für  $X_{jn}$  von 1 sicherstellt. In der zweiten Stufe erfolgt die Schätzung der Pfadkoeffizienten (6) des inneren sowie der Indikatorgewichte (7) und Indikatorladungen (8) des äußeren Modells mit Kleinstquadraten, die sich formal nach Panten und Boßow-Thies (2006) wie folgend beschreiben lassen:

$$(6) \eta_j = \sum_i \beta_{ji} \eta_i + \zeta_j \quad \text{für alle } j = 1, \dots, J$$

Das Symbol  $\beta_{ji}$  bezeichnet den Pfadkoeffizienten des Struktur- bzw. inneren Modells,  $\eta_j$  das Konstrukt der determinierenden, latenten Variable  $\eta_i$  und  $\zeta_j$  repräsentiert die innere Residualvariable.

$$(7) x_{hj} = \omega_{hj} \eta_j + \varepsilon_{hj} \quad \text{für } h = 1, \dots, H$$

$$(8) \eta_j = \sum_h \pi_{hj} x_{hj} + \delta_{hj}$$

Die Ladungskoeffizienten sind mit  $\omega_{hj}$  abgekürzt;  $\pi_{hj}$  bezeichnet die Gewichte,  $\varepsilon_{hj}$  und  $\delta_{hj}$  stellen die Residualvariablen des äußeren bzw. des Messmodells dar.

#### 4.3.3.1 Vor- und Nachteile

Ein Vorteil der PLS-Pfadanalyse aufgrund des iterativen Schätzverfahrens ist, dass es keine Annahmen über die Population oder Messskalen gibt, weshalb das Verfahren ohne Verteilungsannahmen mit nominal-, ordinal- und intervallskalierten Daten angewendet werden kann (Fornell & Bookstein, 1982). Zur Anwendung der PLS muss die Prädiktorspezifikation gewährleistet sein, d.h. der systematische Teil der linearen Regression muss gleich mit den konditionalen Erwartungen der abhängigen Variable sein, was aber für die meisten Fälle als erfüllt angesehen werden kann (Heanlein & Kaplan, 2004). Cassel, Hackl und Westlund (1999) zeigten in einer Monte Carlo Simulation, dass die PLS-Pfadanalyse zugleich sehr robust gegenüber weiteren Verletzungen, wie Multikollinearität oder Verteilungsschiefe, ist. Sobald für eine latente

Variable eine Vielzahl von Indikatoren vorliegen, geraten kovarianzbasierte Modelle an ihre Grenzen, hingegen kann die PLS-Pfadanalyse aufgrund ihres Konzepts sehr gut mit einer großen Anzahl von Indikatoren umgehen (Bliemel, Eggert, Fassott & Henseler, 2005). Ein weiterer Grund für die Anwendung der PLS sind Messmodelle mit formativen oder reflektiven Indikatoren bzw. mit beiden Indikatortypen. Kovarianzbasierte Strukturgleichungsmodelle haben Schwierigkeiten, Kovarianzen von Indikatoren gleich Null zu analysieren, zu welchen es bei formativen Indikatoren kommen kann (Heanlein & Kaplan, 2004). Als ein weiterer Vorteil der PLS-Pfadanalyse sind die geringen Stichprobengrößen zu nennen, die im Messmodell für eine Konvergenz des Algorithmus benötigt werden, wobei gleichzeitig keine Beschränkungen hinsichtlich der Anzahl von Indikatoren und Konstrukten bestehen. Chin und Newsted (1999) konnten daher in einer Monte Carlo Simulation zeigen, dass PLS mit einer Stichprobengröße  $n < 50$  durchführbar sei. Wold analysierte beispielsweise in einem PLS-Modell 27 Variablen, indem er zwei latente Konstrukte mit einem Datensatz aus zehn Fällen verwendete (Chin, Marcolin & Newsted, 2003). Für die Ermittlung einer Mindeststichprobengröße hat sich die Faustregel bewährt, die höchste Anzahl der reflektiven, exogenen latenten Variablen, die eine endogene latente Variablen bestimmen, mit fünf bzw. zehn zu multiplizieren, um die zur Auswertung geringste Anzahl benötigter empirischer Fälle zu erhalten (Chin, 1998b).

Problematisch ist in der PLS hingegen die Schätzerstreue. Die Schätzer in der PLS-Pfadanalyse sind nicht konsistent, jedoch bei steigender Anzahl der Indikatoren pro Konstrukt und Stichprobengröße nähern sich die Werte den geschätzten wahren Parametern an, was in der Literatur als „consistency at large“ bezeichnet wird (Lohmüller, 1989; Hui & Wold, 1982; McDonald, 1996). Weiterführend ist in der PLS problematisch, dass Messfehler im Modell nicht quantifiziert werden können, wodurch zur Einschätzung der Güte inferenzstatistische Größen verwendet werden müssen (Chin, 1998b). In der PLS-Pfadanalyse können außerdem keine globalen Gütemaße zur Beurteilung des Gesamtmodells berechnet werden, da dieses in einzelne Regressionsgleichungen aufgebrochen wird, die blockweise iterativ berechnet werden (Herrmann, Huber & Kressmann, 2006). Somit entfällt in varianzbasierten Strukturgleichungsmodellen die Möglichkeit konkurrierende Modelle zu testen, was insbesondere für die Prüfung der Messmodelle von Bedeutung ist (Chin, 1998b).

### 4.3.3.2 Modellevaluation

In Tabelle 4.04 findet sich eine Zusammenstellung aller verwendeten Kriterien zur Beurteilung der Güte eines varianzbasierten Strukturgleichungsmodells. Hierbei erfolgt eine Unterteilung der Kriterien zur Prüfung des Mess- und des Strukturmodells, die nachfolgend erläutert sind. Die Analyse des Messmodells weist für die PLS-Pfadanalyse eine Besonderheit auf. Die Faktorladungen der Indikatoren auf ihre Faktoren können wie die einer KFA bei kovarianzbasierten Strukturgleichungsmodellen interpretiert werden (Gefen, Straub & Boudreau, 2000).

Tabelle 4.04: Gütekriterien varianzbasierter Strukturgleichungsmodelle

Gütekriterium	Akzeptanzniveau
<b>Gütekriterien zur Prüfung des Messmodell</b>	
Faktorladung: Signifikanz (einseitig, t-Test)	$t \geq 2.326$ ( $p < .01$ )
Faktorladung: Stärke ( $\lambda_i^2$ )	$\lambda_i^2 > 0.707$
Standardfehler der Schätzung (S.E.)	S.E. = geringe Ausprägung
Kommunalität ( $H_j^2$ )	$H_j^2$ = hohe Ausprägung
Faktorreliabilität ( $\rho_c$ )	$\rho_c > 0.7$
Interne Konsistenz (Cronbachs Alpha)	$\alpha \geq 0.7$
durchschnittlich erfasste Varianz (DEV)	DEV > 0.5
Kreuzkorrelationen	$r_{ij} > r_{ik}, r_{ik} < 0.4$
Fornell/Larcker-Kriterium	$DEV(\xi_j) > r_{\xi_j, \xi_k}^2$
<b>Gütekriterien zur Prüfung des Strukturmodells</b>	
Bestimmtheitsmaß ( $R^2$ )	$R^2 > 0.3$
Pfadkoeffizient ( $\beta$ ), Signifikanz (einseitig, t-Test)	$\beta > 0.1; t \geq 2.326$ ( $p < .01$ )
Effektstärke ( $f^2$ )	$f^2 > 0.02$
Stone-Geissers $Q^2$	$Q^2 > 0$

Anm.: i = Indikator, j = Konstrukt/Faktor, k = Konstrukt/Faktor

Die Reliabilität reflektiver Indikatoren kann mit einer Ladung ( $\lambda$ ) größer als 0.707 auf ihre Faktoren beurteilt werden (Hulland, 1999), da in diesem Fall mehr als 50 Prozent der Varianz des Items durch das zu messende Konstrukt erklärt werden. Die Standardfehler der Schätzung (S.E.) geben Auskunft über die Zuverlässigkeit der Parameterschätzungen, indem sie deren Streuung anzeigen. Ein hoher S.E. indiziert wenig zuverlässige Parameterschätzungen (Sosik, Kahai & Piovosio, 2009).

Die Faktorreliabilität  $\rho_c$  erlaubt die Beurteilung der Faktoreignung zur Erklärung aller zu ihm in Beziehung stehenden Indikatoren (Ringle, 2004). Die Berechnung basiert auf der Formel:

$$\rho_c = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum_i \text{var}(\varepsilon_i)}$$

Die Ladung eines Indikators auf die latente Variable bzw. seinen Faktor ist mit  $\lambda_i$  angegeben. Die Varianz ( $\varepsilon_i$ ) ist über  $1 - \lambda_i^2$  definiert. Eine Faktorreliabilität  $\rho_c > 0.7$  gilt als akzeptabel (Ringle, 2004). Neben der Faktorreliabilität kann auch die interne Konsistenz als Reliabilität der Konstrukte bzw. latenten Variablen mittels Cronbachs Alpha als Gütekriterium zur Beurteilung des Messmodells herangezogen werden (Werts, Linn & Jöreskog, 1974). Die durchschnittlich erfasste Varianz (*DEV*) zeigt im Vergleich zur Faktorreliabilität  $\rho_c$  an, wie hoch der durch einen Faktor erklärte Varianzanteil der Indikatoren im Verhältnis zum nicht erklärten Varianzanteil ist (Fornell & Cha, 1994). Formal lässt sich die  $DEV(\zeta_j)$  wie folgt beschreiben:

$$DEV = \frac{\sum \lambda_i^2}{\sum \lambda_i^2 + \sum \text{var}(\varepsilon_i)}.$$

Die Ladung eines Indikators auf die latente Variable ist mit  $\lambda_i$  angegeben. Die Varianz ( $\varepsilon_i$ ) ist über  $1 - \lambda_i^2$  definiert. Werte unter 0.5 werden in der Literatur für die  $DEV(\zeta_j)$  als nicht zulässig beschrieben, da weniger als 50 Prozent der Varianz aufgeklärt werden (Homburg & Giering, 1996). Fornell und Larcker schlugen 1981 vor, als Maß der Diskriminanzvalidität die  $DEV(\zeta_j)$  zu verwenden, wobei die  $DEV(\zeta_j)$  der latenten Variablen größer sein muss als die quadrierten Korrelationen zwischen den latenten Variablen. Zusätzlich können die Kreuzkorrelationen zwischen den Indikatoren und Konstrukten bzw. Faktoren als Gütemaß für die diskriminante Validität herangezogen werden (Evermann & Tate, 2010), deren Stärke kleiner als 0.4 ausfallen sollte (Straub, Boudreau & Gefen, 2004). Eine gute diskriminante Validität des Messmodells liegt vor, wenn die Indikatoren stärker mit dem von ihnen zu messenden Konstrukt als mit den übrigen latenten Größen des Modells korrelieren und gleichzeitig alle Indikatoren des in Frage stehenden Konstrukts höher auf dieses laden als Indikatoren anderer Konstrukte (Chin, 1998b).

Die Kommunalitäten als Kreuz-Validierungs-Maß geben globale Auskunft über die Qualität des Messmodells, inwieweit Passung zwischen den manifesten Variablen und ihrer zugehörigen latenten Variable besteht, die ähnlich wie das Bestimmtheitsmaß  $R^2$  einer bei einer Regressionsanalyse interpretiert werden können (Tenenhaus, Vinzi, Chatelin & Lauro, 2005). Die durch das Bootstrapping-Verfahren erhaltenen *t*-Werte ermöglichen, die Signifikanzen der Indikatoren und Pfadkoeffizienten zu ermitteln (Panten & Boßow-Thies, 2006). Bei der Durchführung eines Bootstrappings-Verfahren empfiehlt sich die Ziehung von mindestens 100, besser jedoch um die 500 oder mehr

Zufallsstichproben aus dem vorliegenden Datensatz mit der Option Zurücklegen (Yung & Chan, 1999). Ein  $t$ -Wert größer 2.327 ( $df_n = 5000$ ) indiziert bei einem einseitigen Test Signifikanz auf dem Niveau  $p < .01$ , ein  $t$ -Wert größer 1.645 ( $df_n = 5000$ ) Signifikanz auf dem Niveau  $p < .05$ . Um für einen zweiseitigen Test Signifikanz auf dem Niveau  $\alpha = 0.2$  anzuzeigen, muss der  $t$ -Wert größer 1.281 ( $df_n = 5000$ ) ausfallen.

Zur Prüfung der Hypothesen dient das Strukturmodell, dessen Güte mittels der nachfolgend dargestellten Kriterien evaluiert wird. Hierfür können das aus dem statistischen Verfahren der Regressionsanalyse bekannte Bestimmtheitsmaß  $R^2$  und der Pfadkoeffizient  $\beta$  herangezogen werden. Das  $R^2$  ist wie bei einer klassischen Regressionsanalyse zu interpretieren, um den Streuungsanteil der endogenen latenten Variablen und somit den „goodness of fit“ der Indikatoren zu erklären. Die Pfadkoeffizienten des Strukturmodells sind wie standardisierte Beta-Koeffizienten der klassischen Regressionsanalyse hinsichtlich ihrer Richtung, Signifikanz und Einflussstärke zu betrachten (Panten & Boßow-Thies, 2006). Die Signifikanzen der Pfadkoeffizienten  $\beta$  werden mit Hilfe des Bootstrapping-Verfahrens wie bei der Ermittlung der Signifikanzen für die Faktorladungen der Indikatoren auf ihre Faktoren im Messmodell bestimmt. In der Literatur liegen unterschiedliche Empfehlungen zur Interpretation des Bestimmtheitsmaßes  $R^2$  und des Pfadkoeffizienten  $\beta$  vor. Chin (1998a) zufolge ist ein Bestimmtheitsmaß von  $R^2 > 0.67$  als substantiell zu bezeichnen, während Ergebnisse in Höhe von größer 0.33 und 0.19 als durchschnittlich bzw. schwach einzustufen sind. Ringle (2004) argumentiert auf Basis von unterschiedlichen Studiendaten, dass Ergebnisse ab einem Bestimmtheitsmaß von 0.4 bereits in PLS-Modellen akzeptiert werden sollten. Als konservatives Prüfkriterium wird von Herrmann et al. (2006) ein Bestimmtheitsmaß von  $R^2 > 0.3$  angesehen. Lohmüller (1989) empfiehlt Pfadkoeffizienten  $\beta$  in einer Höhe ab 0.1 in PLS-Modelle aufzunehmen. Die Einflussstärke eines Prädiktors, genauer, der exogenen latenten Variable auf die endogene latente Variable, lässt sich mit der Effektstärke  $f^2$  erfassen und wird wie folgt berechnet (Chin, 1998b):

$$f_j^2 = \frac{R_{\text{included}}^2 - R_{\text{excluded}}^2}{1 - R_{\text{included}}^2}.$$

Die Kennziffern  $R_{\text{included}}^2$  und  $R_{\text{excluded}}^2$  stellen Bestimmtheitsmaße der endogenen latenten Variablen unter Einschluss ( $R_{\text{included}}^2$ ) und Ausschluss ( $R_{\text{excluded}}^2$ ) einer bestimmten, über das Strukturmodell mit ihr in Beziehung stehenden exogenen latenten

Variable dar (Chin, 1998b). Dieses Verhältnis von zusätzlich und vorher nicht erklärter Varianz der Effektstärke  $f^2$  für jede exogene latente Variable lässt somit Rückschlüsse auf die Eignung zur Vorhersage einer abhängigen, d.h. endogenen latenten Variable im Untersuchungsmodell zu, deren Betrachtung für jeden Prädiktor individuell und im Vergleich zueinander erfolgen kann (Gefen et al., 2000). Dabei gilt, je größer  $f^2$ , desto gewichtiger ist der Einfluss. Werte größer gleich 0.02, 0.15 und 0.35 sind als schwach, moderat und stark anzusehen (Chin, 1998b).

Zur Beurteilung der Vorhersagevalidität des Mess- und Strukturmodells kann für reflektiv operationalisierte latente Variablen das Gütemaß Stone-Geissers  $Q^2$  berücksichtigt werden (Herrman et al., 2006). Der nichtparametrische Stone-Geisser Test folgt dabei dem Prinzip einer Blindfolding-Prozedur (Geisser, 1974; Stone, 1974) und indiziert, wie gut die Datenmatrix mit Hilfe des Strukturmodells rekonstruiert wird. Die formale Schreibweise lautet:

$$Q_j^2 = 1 - \frac{\sum_k E_{jk}}{\sum_k O_{jk}}.$$

Der Koeffizient  $E_{jk}$  stellt die quadrierten Fehler der geschätzten Werte dar, während  $O_{jk}$  die quadrierten Fehler der Originalwerte repräsentiert. Der Auslassungsabstand ist mit  $k$  bezeichnet, der die Distanz zwischen zwei nacheinander auszulassenden und daraufhin zu schätzenden Datenfällen angibt. Wold (1982) empfiehlt eine Distanzgröße von sieben. Gilt  $Q^2 > 0$ , so besitzt das Modell eine Prognoserelevanz, von der nicht gesprochen werden kann, falls  $Q^2 < 0$  ist (Panten & Boßow-Thies, 2006).

#### 4.3.3.3 Evaluation von Single-Item-Konstrukten

Die Evaluation von reflektiv operationalisierten Single-Item-Messinstrumenten kann im Fall einer Standardisierung der manifesten Variablen nicht mit den bisher vorgestellten Gütekriterien erfolgen. Die Standardisierung der manifesten Variablen, welche den Mittelwert auf null und die Varianz auf eins festlegt, führt zwangsläufig in einem Single-Item-Konstrukt zu einem Cronbachs Alpha und zu einer  $DEV(\xi_j)$  gleich eins. Bergkvist und Rossiter (2007) empfehlen in diesem Falle die prädiktive Validität auf Konstruktebene als Gütekriterium für das Messmodell heranzuziehen, da eine Messung nur valide sein kann, wenn sie auch reliabel ist. In der Literatur finden sich darüber hinaus mehrere Strategien zur Evaluation von Single-Item-Messmodellen. Wanous, Reichers und Hudy (1997) adaptierten die „correction for attenuation“ Formel von

Spearman (1904a), um die Reliabilität eines Single-Item-Konstrukts zur Messung von Arbeitszufriedenheit zu ermitteln. Dafür wurde das Einzelitem ( $y$ ) in Korrelation ( $r_{xy}$ ) mit dem Gesamtwert einer „Multi-Item-Skala“ ( $y$ ) der Arbeitszufriedenheit gesetzt, deren Reliabilität ( $r_{yy}$ ) berechnet werden konnte, um beide Kennwerte in die Formel

$$\hat{r}_{xy} = \frac{r_{xy}}{\sqrt{r_{xx}} \sqrt{r_{yy}}}$$

einzutragen. Als Unbekannte verbleiben die vermutete wahre Korrelation ( $\hat{r}_{xy}$ ) zwischen den beiden Skalen unter der Annahme der Messfehlerfreiheit und die Reliabilität des Single-Item-Messinstruments ( $r_{xx}$ ) in der Gleichung. Unter Annahme eines Wertes für  $\hat{r}_{xy}$ , von bspw. 0.95 oder 1.0, kann die Formel entsprechend umgestellt und aufgelöst werden, um die Mindestreliabilität des Single-Item-Konstrukts zu erhalten.

Wanous und Hudy (2001) schlugen die Faktoranalyse als eine weitere, alternative Möglichkeit zur Berechnung der Single-Item-Reliabilität vor, basierend auf der Formel, dass die Gesamtvarianz aus der Summe der Kommunalität, der Spezifität und dem Messfehler besteht. Die Kommunalität eines Single-Item-Konstrukts ist als ein konservatives Reliabilitätsmaß zu interpretieren, da bei einem Item das Auftreten von spezifischer Varianz als nicht gegeben betrachtet werden kann. Eine Testwiederholung des Single-Item-Messinstruments ermöglicht zudem die Berechnung der Retest-Reliabilität, um ein weiteres Gütekriterium zur Beurteilung des reflektiv operationalisierten Single-Item-Konstrukts zu gewinnen (Zimmerman, Ruggero, Chelminski, Young, Posternak, Friedman, Boerescu & Attiullah, 2006).

Formative Indikatoren sind anhand ihrer Vorzeichen und Größen der jeweiligen Gewichte sowie deren Signifikanz zu evaluieren (Panten & Boßow-Thies, 2006), die wie bei reflektiven Indikatoren mit dem Bootstrapping-Verfahren ermittelt werden (Bollen & Lennox, 1991). Die von Lohmüller (1989) geforderte Mindesthöhe der Gewichte liegt bei 0.1, während Chin (1998b) eine Stärke von 0.2 empfiehlt, wobei die Vorzeichen der Gewichte sich entsprechend den postulierten Inhalten verhalten müssen. Für die Evaluation formativ operationalisierter Single-Item-Konstrukte können jedoch diese Kriterien nicht verwendet werden, falls die manifeste Variable aus Gründen der Vergleichbarkeit standardisiert wird. Bergkvist und Rossiter (2007) empfahlen in dieser Situation die prädiktive Validität als Kriterium zur Beurteilung der manifesten Variablen zu verwenden. Den Autoren zufolge kann eine Messung nicht prädiktiv valide sein, ohne über ein reliables Messmodell zu verfügen. Im Fall geringer

prädiktiver Validitäten, die sich konform zu den postulierten Hypothesen verhalten, kann jedoch die Empfehlung von Bergkvist und Rossiter keine Anwendung finden. Eine statistische Analyse der Güte von Single-Item-Messinstrumenten kann daher in letzter Konsequenz nicht umgesetzt werden, weshalb die inhaltliche Validität auf theoretischer Ebene geprüft werden muss (Henseler, Ringle & Sinkovics, 2009; Rossiter, 2002). Die Annahme der Messfehlerfreiheit formativer Indikatoren verringert zudem die Bedeutung der Reliabilität für eine Beurteilung des Single-Item-Messmodells, wodurch die Indikatorvalidität an Bedeutung gewinnt. Die Güte der Indikatorvalidität ist damit entscheidend für die Beurteilung des formativ operationalisierten Single-Item Messinstruments (Diamantopoulos, 2012).

#### **4.3.3.4 Inkrementelle Validität**

Ein klassisches Vorgehen zur Ermittlung der inkrementellen Validität findet sich beispielsweise im Rahmen der multiplen Regressionsanalyse, in der man einen Prädiktor in ein bestehendes Modell hinzufügt, um anhand der Veränderung des Determinationskoeffizienten  $R^2$  den zusätzlich Anteil an aufgeklärter Varianz am Kriterium zu bestimmen, der die inkrementelle Validität des in das Regressionsmodell aufgenommenen unabhängigen Variable darstellt (Bühner, 2011, S.63). Ein vergleichbares Vorgehen, d.h. die Aufnahme eines weiteren Prädiktors in ein bestehendes PLS-Modell würde immer zu einem veränderten Strukturgleichungsmodell führen, in welchem der gleichzeitige Einfluss der beinhalteten exogenen auf die endogene(n) Variable(n) abgebildet wird.

Zur Berechnung der inkrementellen Validitäten der Prädiktoren räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick wird in der vorliegenden Arbeit auf das methodische Vorgehen von Mathieson, Peacock und Chin (2001) zurückgegriffen, die in ihrer Studie das in PLS modellierte „Technology Acceptance Model“ (TAM) um die exogene Variable „Perceived User Resources“ erweiterten. Um den Einfluss der neu aufgenommenen Variable zu identifizieren, wurde im ersten Schritt von den Autoren das erweiterte TAM als PLS-Modell aufgestellt und berechnet. Im zweiten Schritt wurde das TAM-Modell ohne die inhaltliche Erweiterung in Form der Variablen „Perceived User Resources“ in PLS modelliert und geschätzt. Den Autoren zufolge ist eine Veränderung im  $R^2$  der abhängigen bzw. exogenen Variable zwischen beiden Modellen als quadrierte Semipartialkorrelation zu interpretieren, die einen Anstieg in der quadrierten Korrelation reflektiert, welcher auf

die Hinzunahme der Variable „Perceived User Resources“ in das erweiterte TAM-Modell zurückzuführen ist. Nach Ansicht von Mathieson et al. ist hierbei das Maß  $f^2$  ein guter Indikator, um die Effektstärke des substanziellen Einflusses einer exogenen auf die endogene Variable zu beurteilen, da für die Berechnung Modelle verwendet werden, in welchen der Prädiktor ein- und ausgeschlossen wird (vgl. Unterkapitel 4.3.3.2). Multipliziert man zudem den  $f^2$ -Koeffizienten mit  $(n - k - 1)$ , wobei  $n$  die Stichprobengröße und  $k$  die Anzahl der Prädiktoren im Modell symbolisiert, so erhält man einen pseudo F-Test, um die Signifikanz der  $f^2$  Statistik mit 1 und  $n - k$  Freiheitsgraden zu bestimmen. Zur Bestimmung der inkrementellen Validität werden in der vorliegenden Arbeit mit Bezug auf das Vorgehen in der Studie von Mathieson et al. (2001) jeweils separiert das räumliche Vorstellungsvermögen und das instrumentelle, manuelle Geschick aus den Prädiktionsmodellen exkludiert. Anschließend werden die modifizierten Modelle erneut berechnet. Die Differenz im  $R^2$  zeigt die inkrementelle Validität der jeweiligen Fähigkeitsart an, der Koeffizient  $f^2$  wird herangezogen, um die substanzielle Stärke ihrer inkrementellen Validität zu beurteilen.

#### 4.3.4 Robuste Regressionsanalyse

Die vorliegende Untersuchungsstichprobe ist durch die Studienplatzkapazitäten von maximal 39 Zahnmedizinierenden im Auswahlverfahren der Hochschule limitiert, die anhand der Abiturdurchschnittsnote und den Ergebnissen des HAM-Nat sowie des HAM-Man zugelassen werden. Demzufolge ist das überproportional häufige Auftreten von Extremwerten nicht auszuschließen, zumal die Stichprobengröße im Verlauf der zahnmedizinischen Vorklinik durch Studienabbrüche weiter verringert werden könnte. Extremwerte in diesem Fall als Ausreißer zu definieren, die der Stichprobe entnommen werden, könnte zu systematischen Ausschlüssen führen, die inhaltlich nicht zu vertreten wären. Extremwertkonstellationen können aber insbesondere in kleineren Stichproben die Ergebnisse in die eine oder andere Richtung verzerren, weshalb komplementär zur PLS-Pfadanalyse in der vorliegenden Arbeit der Einsatz von robusten Regressionsmethoden erfolgt, die sich gegenüber Extremwerten als resistent erweisen und auch im Fall des Unterschreitens der für die PLS-Methode erforderliche Mindeststichprobengröße (vgl. Unterkapitel 4.3.31) als Äquivalent herangezogen werden können. Das statistische Verfahren der robusten Regressionsanalyse (RR) besitzt gegenüber der PLS-Pfadanalyse Nachteile, die in der vorliegenden Untersuchung insbesondere in der fehlenden Operationalisierung von latenten Variablen über

Indikatoren besteht, welche aber im Kontext der komplementären Anwendung akzeptabel sind.

Im folgenden Unterkapitel 4.3.4.1 wird nur der in der vorliegenden Arbeit verwendete MM-Schätzer von Yohai (1987) als robustes Maß berichtet. Eine gute Darstellung unterschiedlicher Schätzverfahren robuster Regressionsanalysen geben Anderson (2008), Jann (2010) und Maronna, Martin und Yohai (2006). Im Kontext robuster Verfahren bezeichnet Kontaminierung einen Datensatz, dessen beobachtete Werte aus zwei oder mehreren unterschiedlichen Prozessen resultieren, beispielsweise dem inhaltlich relevanten Zusammenhang zwischen den Untersuchungsvariablen als eigentlicher Prozess und einem Kodierungsfehler in den Variablen als zweiten Prozess. Der Anteil an Kontaminierung, den ein Schätzer maximal verkraftet, nennt man Bruchpunkt. Die Gaußsche Effizienz spiegelt die Effizienz eines robusten Schätzers wider, der sich im Fall normalverteilter Fehler als vergleichbar effizient zur Methode der Kleinste-Quadrate-Schätzung (QLS) erweist (Jann, 2010).

#### 4.3.4.1 MM-Schätzer

Der MM-Schätzer ist ein robuster Regressionsschätzer zweiter Generation, dessen Einsatz aufgrund des hohen Bruchpunktes und der guten Gaußschen Effizienz zu empfehlen ist (Anderson, 2008). Für die Berechnung eines MM-Schätzers (Darstellung in Anlehnung an Jann, 2010) wird im ersten Schritt ein Schätzer  $\hat{\beta}^{(0)}$  mit hohem Bruchpunkt aber geringer Effizienz gewählt, wofür in der Praxis ein S-Schätzer mit Bruchpunkt von 50 Prozent verwendet wird (Rousseeuw & Yohai, 1984). Es erfolgt im zweiten Schritt die Schätzung des Skalenparameters  $\hat{\sigma}$  aufgrund der Residuen

$$r_i(\hat{\beta}^{(0)}) = Y_i - x_i^T \hat{\beta}^{(0)}.$$

Der Term  $Y_i$  symbolisiert hierbei die abhängige Variable und  $x_i^T$  einen transponierten Vektor der Regressoren, welcher die Konstante einschließt. Im zweiten Schritt wird ein Bisquare-M-Schätzer berechnet (Beaton & Tukey, 1974), der zu den „redescending“ bzw. absinkenden M-Schätzern gehört (Huber, 1964), bei denen Extremwerte kaum oder gar kein Gewicht erfahren. Die Berechnung des MM-Regressionsschätzers  $\hat{\beta}^{MM}$  erfolgt als Lösung eines Bisquare-M-Schätzers mit dem Startwert  $\hat{\beta}^{(0)}$  aus dem ersten Berechnungsschritt und dem fixen Skalenparameter  $\hat{\sigma}$  des zweiten.

Ein MM-Schätzer vereint daher die Vorteile beider Schätzer in einem, d.h. den hohen Bruchpunkt des S-Schätzers und die Gewichtung weit vom Modell entfernter Punkte

mit Null oder einem geringen Gewicht durch den Bisquare-M-Schätzer. Sind hingegen keine Extremwerte in dem Untersuchungsdatensatz vorhanden, nähert sich der MM-Schätzer den Ergebnissen der OLS-Methode an (Anderson, 2008). Deutliche Unterschiede in den Resultaten zwischen der robusten und der OLS-Regression können auf Fehlspezifikationen im bzw. des gesamten Modells hinweisen. Die Interpretation der MM-Schätzer ist identisch zu den OLS-Schätzern (Jann, 2010).

#### 4.3.5 Prädiktionsmodelle der praktischen Studienleistung

Die Vorhersage der praktischen Aufgabenleistung beim Erwerb zahnmedizinischer Fertigkeiten in den vorklinischen Laborkursen der technischen Propädeutik (TPK) und der beiden Phantomkurse (Pha) erfolgt in PLS-Modellen auf Basis des in Abbildung 4.04 dargestellten Schemas.

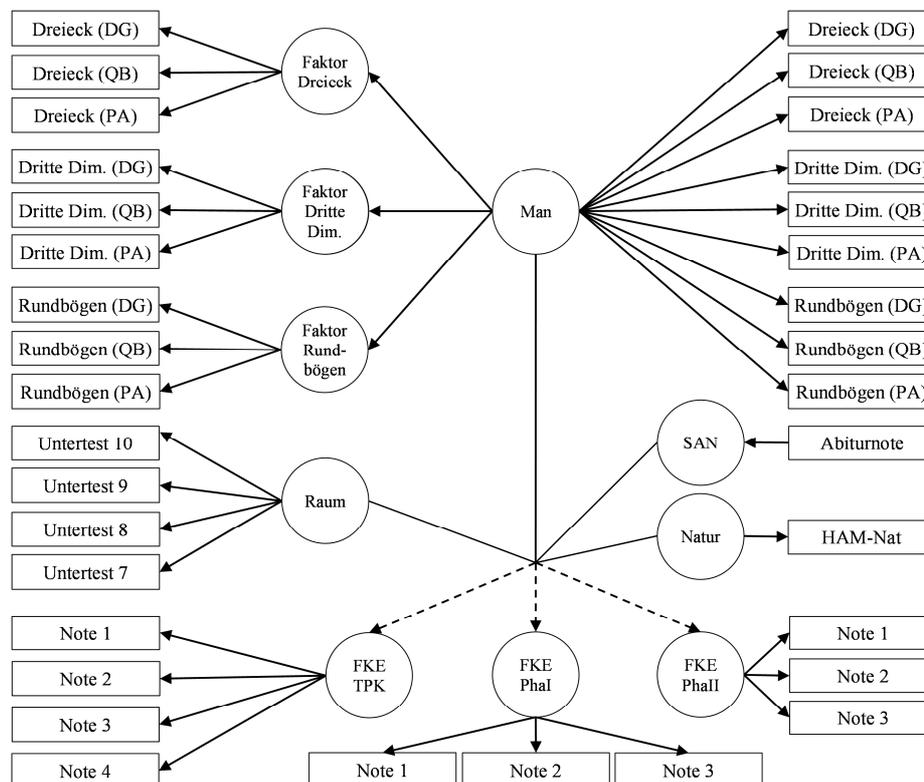


Abbildung 4.04: Konzeptionelles PLS-Prädiktionsmodell

Anm.: Man = instrumentelles, manuelles Geschick; Raum = räumliches Vorstellungsvermögen; SAN = Schulabschlussnote;

Natur = naturwissenschaftliches Verständnis; FKE = Fertigkeitserwerb, TPK = Technisch-Propädeutischer-Kurs;

Pha = Phantomkurs;

Dritte Dim. = Dritte Dimension; DG = Deckungsgleichheit; QB = Qualität der Biegung; PA = planes Aufliegen

Der Prädiktor instrumentelles, manuelles Geschick wird über den Drahtbiegetest HAM-Man erfasst, für dessen Teststruktur ein Bi-Faktor-Modell postuliert ist (vgl. Unterkapitel 4.1.1.4). Entsprechend Horns (1983) Validierung des LPS ist die räumliches Vorstellungsvermögen erfassende technische Skala als Ein-Faktor-Modell

über die Untertests 7 bis 10 operationalisiert (vgl. Unterkapitel 4.1.2.4). Die Prädiktoren Schulabschlussnote und naturwissenschaftliches Verständnis werden als Single-Item-Konstrukte in das Prädiktionsmodell aufgenommen (vgl. Unterkapitel 4.1.3.3 und 4.1.4.4). Wie aus der Abbildung 4.04 hervorgeht, besitzt nur die exogene latente Variable Schulabschlussnote ein formatives Messmodell, hingegen alle anderen ein reflektives.

Die in den vorklinischen Laborkursen zu bearbeitenden Aufgaben (s. Tabelle 4.02 in Unterkapitel 4.1.5.1) sind mit Bezug auf Ackermans Theorie des Fertigkeitserwerbs in die Aufgabentypen konsistent-einfach und inkonsistent-komplex klassifiziert worden. Die Beziehung zwischen Aufgabenleistung und den zum Erwerb der Dentaltechniken benötigten Fähigkeitsarten ist für die Aufgabentypen unterschiedlich gestaltet, sodass separate Prädiktionsmodelle aufgestellt werden. Beide Phantomkurse beinhalten per Definition zwar inkonsistente-komplexe Aufgaben, die aber auch in separate Prädiktionsmodelle eingebunden werden. Es ist beispielsweise für die Herstellung eines Schalenprovisoriums im Phantomkurs II nicht auszuschließen, dass die dafür benötigten Dentaltechniken im vorherigen Kurs bereits vollständig erworben wurden, sodass inkonsistente Aufgabenbestandteile in konsistente transformiert wurden (Mack, 1996), da jetzt nicht mehr fortwährend neue Reiz-Reaktionsverbindungen ausgeprägt werden müssen. Ebenfalls ist der zeitliche Einfluss auf den Erwerb der in den Laborkursen benötigten Techniken und Methoden nicht einzuschätzen, da beide Kurse ein Jahr versetzt angeboten werden. Zwar konnte Ackerman (1992) keinen Einfluss durch den zeitlichen Abstand zwischen Messung der Fähigkeiten und den Aufgabenleistungen nachweisen, aber die Ergebnisse sind nicht direkt auf die Situation in der vorliegenden Untersuchung zu übertragen, da im ersten Kurs erlernte Dentaltechniken möglicherweise im zweiten gar nicht oder vereinzelt wieder erlernt werden müssen.

Zur Vermeidung von „reverse coding“-Effekten in den PLS-Prädiktionsmodellen werden alle Indikatoren, die Noten enthalten, einheitlich kodiert. Folglich spiegelt eine ungenügende Leistung wider und sechs bzw. vier im Fall der Abiturdurchschnittsnote eine sehr gute. Unterschiedliche gerichtete, reflektive Indikatoren können in PLS-Modellen nicht nur die Reliabilitätsmaße der latenten Variablen beeinflussen, sondern auch ihre Beziehungen in den Strukturmodellen (Temme, Kreis & Hildebrandt, 2010).

#### **4.3.6 Software**

Die statistischen Analysen erfolgen mit der Statistiksoftware IBM SPSS Version 18.0.3. Die konfirmatorischen Faktorenanalysen werden mit SPSS Amos 18.0.0 (Arbuckle, 2009), die varianzbasierten Strukturgleichungsmodelle mit Smart PLS 2.0 Version M3 (Ringle, Wende & Will, 2005) und die robusten Regressionsanalysen mit IBM SPSS 21.0.0 berechnet, wofür zusätzlich das Erweiterungspaket „IBM SPSS Statistics-Essentials for R“ sowie die Statistiksoftware R in der Version 2.14.2 erforderlich sind.

## 5. Ergebnisse

Zur Prüfung der Hypothesen werden die in Unterkapitel 5.2 berichteten Ergebnisse der konfirmatorischen Faktoranalyse und die der varianzbasierten Prädiktionsmodelle in Unterkapitel 5.3 herangezogen. In Unterkapitel 5.1 sind die vorbereitenden Auswertungsanalysen dargestellt, die neben der deskriptiven Statistik auch die Prüfung der Testgüte des Drahtbiegetests HAM-Man nach Kriterien der Klassischen Testtheorie beinhaltet.

### 5.1 Vorbereitende Auswertungsanalysen

Die deskriptive Statistik in Unterkapitel 5.1.1 umfasst die Lage- und Verteilungsmaße, die Stichprobenbeschreibung, Prüfung auf Normalverteilung und die Interkorrelationen der Untersuchungsvariablen. In Unterkapitel 5.1.2 sind Validitäts- und Reliabilitätsmaße der Klassischen Testtheorie für den Drahtbiegetest HAM-Man berichtet.

#### 5.1.1 Deskriptive Statistik

In 2010 gingen 467 Bewerbungen um einen Studienplatz in der Zahnmedizin an der Universität Hamburg ein. 200 Bewerber wurden zum Auswahlverfahren der Medizinischen Fakultät eingeladen, von denen insgesamt 129 teilnahmen. Die Untersuchungsstichprobe umfasst somit 129 Personen, deren Altersdurchschnitt (in Jahren)  $M = 21.3$  ( $SD = 2.03$ ) beträgt und aus 81 Frauen (Alter:  $M = 21.2$ ,  $SD = 2.16$ ) und 48 Männern (Alter:  $M = 21.0$ ,  $SD = 1.79$ ) besteht.

Tabelle 5.01: Deskriptive Statistik

Variable	n	M	SD	Min	Max	Sch	Exz	p-SWT
instrumentelles, manuelles Geschick	129	17.0	9.28	0	39	0.25	-0.45	.045
räumliches Vorstellungsvermögen	129	207.4	30.15	128	262	-0.45	-0.01	.011
Schulabschlussnote	129	2.0	0.22	1.5	2.3	-0.65	-0.43	.000
naturwissenschaftliches Verständnis	129	32.7	9.79	14	61	0.51	-0.32	.007
Aufgabenleistung TPK	38	2.9	0.27	2.5	4.0	1.40	-4.04	.672
Aufgabenleistung PhaI	31	3.0	0.21	2.6	3.3	-0.26	-0.93	.449
Aufgabenleistung PhaII	28	2.5	0.32	2.0	3.3	0.52	0.06	.473
Geschlecht	129	1.6	0.49	1	2	-0.54	-1.74	.000
Alter	129	21.1	2.03	18	29	1.75	3.72	.000

Anm.: n = Stichprobengröße; M = Mittelwert; SD = Standardabweichung;  $\sigma^2$  = Varianz; Min / Max = Mindest- und Maximalwert; Sch = Schiefe der Verteilung; Exz = Exzess der Verteilung; p-SWT = Signifikanz des Shapiro-Wilk-Test; TPK = Technisch-Propädeutischer-Kurs; Pha = Phantomkurs

39 Teilnehmer am Auswahlverfahren erhielten eine Zulassung zum Studium, von denen 38 (Geschlecht: 19 Männer, 19 Frauen) den Laborkurs TPK absolvierten. Den Phantomkurs I im folgenden Semester absolvierten 31 (Geschlecht: 15 Männer, 16

Frauen) über Auswahlverfahren zugelassene Studierende. Der im fünften Semester stattfindende Phantomkurs II wurde von 28 (Geschlecht: 14 Männer, 14 Frauen) Studierenden besucht.

Weder die Teilnahme am Auswahlverfahren noch in den vorklinischen Laborkursen wurde abgebrochen, sodass in den vorliegenden Datensätzen keine fehlenden Werte auftreten. Methodenbedingte Verzerrungen der Ergebnisse sind nicht anzunehmen, da weder die Bewerber und die zukünftigen Studierenden noch die Testleiter und Juroren Kenntnis über die Untersuchungshypothesen besaßen. Die freiwillige Teilnahme am Auswahlverfahren als auch die Motivation, die vorklinischen Laborkurse erfolgreich zu absolvieren, lassen auch nicht auf die Untersuchungsergebnisse verzerrende Umstände schließen. Fähigkeitsausprägungen können aufgrund ihrer Trainierbarkeit durch Hobbies oder berufliche Tätigkeiten in sehr stark unterschiedlicher Ausprägung vorliegen (Waldron & Anton, 1995), sodass zur Kontrolle von Extremwerten komplementär robuste Regressionsanalysen berechnet werden (vgl. Unterkapitel 4.3.4).

In Tabelle 5.01 ist die deskriptive Statistik der Untersuchungsvariablen aufgeführt. Die auf dem Niveau  $p < .05$  signifikante  $W$ -Teststatistik des Shapiro-Wilk Tests führt zu einer Ablehnung der Nullhypothese und damit der Normalverteilung für die Variablen. Nur die Aufgabenleistungen in den vorklinischen Laborkursen erweisen sich von allen Untersuchungsvariablen als normalverteilt ( $p > .05$ ). Der Shapiro-Wilk-Test liefert keine Informationen über die Form bzw. die Symmetrie der Verteilung, weshalb zusätzlich die in Tabelle 5.01 aufgeführten Kriterien Schiefe ( $Sch$ ) und Exzess ( $Exz$ ) herangezogen werden, mit denen kontrolliert werden kann, ob der Shapiro-Wilk Test anfällig auf Werte mit auffallend größerer Ausprägung im Verhältnis zu den übrigen Messwerten reagiert (Shapiro & Wilk, 1965). Die Schiefe und der Exzess für die unabhängigen Variablen instrumentelles, manuelles Geschick ( $Sch = 0.25$ ,  $Exz = -0.45$ ) und räumliches Vorstellungsvermögen ( $Sch = -0.45$ ,  $Exz = -0.01$ ) nehmen insgesamt eher Werteausprägungen um null herum an als die Verteilung der Aufgabenleistungen in den vorklinischen Laborkursen, für die dem Shapiro-Wilk-Test aber zufolge Normalverteilung besteht (TPK:  $Sch = 1.40$ ,  $Exz = -4.04$ ; PhaI:  $Sch = -0.26$ ,  $Exz = -0.93$ ; PhaII:  $Sch = 0.52$ ,  $Exz = 0.06$ ).

Der Signifikanztest des Korrelationskoeffizienten  $r$  nach Pearson erweist sich äußerst robust gegenüber Verletzungen der Verteilungsannahme, sodass für die in Tabelle 5.02 abgebildete Korrelationsmatrix keine Rangkorrelationskoeffizienten nach Spearman

berechnet wurden (Bortz & Göring, 2006). Eine mögliche Unterschätzung der Beziehungen der Prädiktorvariablen aufgrund ihrer asymmetrischen Verteilung ist zu tolerieren, da die Höhe der Korrelationen inferenzstatistisch nicht zur Hypothesenprüfung verwendet werden, wodurch eine Fisher-z-Transformation der Korrelationen nicht notwendig ist (Gayen, 1951; Silver & Dunlap, 1987).

Tabelle 5.02: Interkorrelationen der Untersuchungsvariablen

Variable	<i>M (SD)</i>	1	2	3	4	5	6	7	8	9
1. instr., man. Geschick	17.0 (9.28)	1								
2. räuml. Vorstellung.	207.4 (30.5)	.29***	1							
3. Schulabschlussnote	2.0 (0.22)	.09	-.04	1						
4. naturwiss. Verständnis	32.7 (9.79)	.31***	.16†	-.21*	1					
5. Aufgabenleist. TPK	2.9 (0.27)	-.47**	.12	-.25	-.04	1				
6. Aufgabenleist. PhaI	3.0 (0.21)	-.47**	-.41*	-.08	-.05	.43***	1			
7. Aufgabenleist. PhaII	2.5 (0.32)	-.44*	.13	-.20	-.20	.62***	.57***	1		
8. Geschlecht <sup>a</sup>	1.6 (0.49)	-.22*	-.12	.01	-.34***	-.24†	-.11	.14	1	
9. Alter	21.1 (2.03)	.13	.11	.07	.08	-.17	-.11	.36*	-.06	1

Anm.: M = Mittelwert, SD = Standardabweichung, 1. instrumentelles, manuelles Geschick (HAM-Man, 0 bis 45 Pkt.), 2. räumliches Vorstellungsvermögen (LPS, C-Werte von 51 bis 272); 3. Schulabschlussnote (Abiturdurchschnittsnote, 1.0 = sehr gut bis 4.0 = ausreichend); 4. naturwissenschaftliches Verständnis (HAM-Nat, 0 bis 76 Pkt.); 5. - 7. Aufgabenleistungen in den Laborkursen (Noten, 1.0 = sehr gut bis 4.0 = ausreichend); TPK = Technisch-Propädeutischer-Kurs; Pha = Phantomkurs  
a: dichotom codierte Variable (1 = männlich, 2 = weiblich)

Stichprobe: AdH: n = 129, TPK: n = 38, PhaI: n = 31, PhaII: n = 28

Signifikanzniveau (zweiseitig): †p < .10, \*p < .05, \*\*p < .01, \*\*\*p < .001

Vergleichbar zu den berichteten Studien von beispielsweise Boyle und Santelli (1986) und Ludergan et al. (2007) besteht zwischen dem instrumentellen, manuellen Geschick und dem räumlichem Vorstellungsvermögen eine schwach positive, signifikante Korrelation ( $r = 0.29$ ,  $p < .001$ ). Zu allen Aufgabenleistungen in den Laborkursen zeigt der Prädiktor instrumentelles, manuelles Geschick moderate, signifikante Beziehungen, deren Vorzeichen negativ ausfallen, da bessere Leistungen in den vorklinischen Laborkursen kleinere Notenwerte bedeuten. Hingegen weist das räumliche Vorstellungsvermögen nur zu den Leistungen im Phantomkurs I eine moderat negative, signifikante Korrelation auf ( $r = -0.41$ ,  $p < .05$ ). Die Korrelationen der Prädiktoren Schulabschlussnote und naturwissenschaftliches Verständnis zu den vorklinischen Laborkursleistungen fallen maximal schwach negativ und nicht signifikant aus. Zwischen der kognitiven Fähigkeit räumliches Vorstellungsvermögen und der Schulabschlussnote als auch dem naturwissenschaftlichen Verständnis besteht kein korrelativer Zusammenhang. Hingegen weisen die Korrelationen in Tabelle 5.02 für die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick eine schwach positive und signifikante Beziehung zum naturwissenschaftlichen Verständnis von  $r = 0.31$  ( $p < .001$ ) auf, aber ebenfalls keine zur Schulabschlussnote ( $r = 0.09$ , *n.s.*).

Die schwach negativen, signifikanten Korrelationen der soziodemographischen Kontrollvariable Geschlecht zeigen an, dass die 48 Männer gegenüber den 81 Frauen in der vorliegenden Untersuchung ein leicht höheres instrumentelles, manuelles Geschick ( $r = -0.22, p < .05$ ) und naturwissenschaftliches Verständnis ( $r = -0.34, p < .001$ ) besitzen. Entgegen den Ergebnissen von Coy et al. (2003) und Hegarty et al. (2009) findet sich zwischen dem Geschlecht der Bewerber und dem räumlichen Vorstellungsvermögen kein korrelativer Zusammenhang ( $r = -0.12, n.s.$ ). Der  $t$ -Tests indiziert keinen signifikanten Unterschied in der räumlichen Fähigkeit zwischen Männern und Frauen in der Bewerbungskohorte ( $t = 1.515, df = 126, n.s.$ ). Auch für die zugelassenen Studierenden ( $n = 38$ ) zeigt sich kein korrelativer Zusammenhang zwischen dem Geschlecht und dem räumlichen Vorstellungsvermögen ( $r = 0.07, n.s.$ ) und ebenfalls kein signifikanter Unterschied zwischen den 19 Männern und 19 Frauen im ersten Laborkurs ( $t = -0.402, df = 36, n.s.$ ). Ebenfalls besteht keine signifikante Beziehung zwischen dem Geschlecht und den Aufgabenleistungen in allen drei vorklinischen Laborkursen, was konform zu den Resultaten von Coy et al. (2003) ausfällt. Das zum Zeitpunkt der Bewerbung ermittelte Alter der Teilnehmer korreliert weder signifikant noch in einer statistisch relevanten Stärke mit den unabhängigen Untersuchungsvariablen.

### 5.1.2 Exkurs: Testgüte des HAM-Man nach Klassischer Testtheorie

Die Beurteilung der Güte des an der Medizinischen Fakultät der Universität Hamburg entwickelten Drahtbiegetests HAM-Man nach Klassischer Testtheorie erfolgt als Exkurs zur Untersuchungsfragestellung. Zwischen dem HAM-Man und der etablierten DBP von Lienert (1967) besteht ein positiv schwacher, signifikanter Zusammenhang in Höhe von  $r = 0.33 (p < .001)$ , der auf eine geringe Kriteriumsvalidität des HAM-Man schließen lässt (s. Tabelle A02 im Anhang). Die Konstruktvalidität des HAM-Man aufgrund der geringen Übereinstimmung beider Drahtbiegetests grundlegend in Frage zu stellen und von einer weiteren Ergebnisauswertung abzusehen, wäre nicht folgerichtig. Es ist eher anzunehmen, dass der HAM-Man durch den Einsatz einer Spitzzange gegenüber der DBP eine spezifischere Form des manuellen Geschicks erfasst.

Von den 39 über das Auswahlverfahren zugelassenen Zahnmedizinierenden nahmen 37 an der Testwiederholung des HAM-Man teil, von denen 20 Männer und 17 Frauen waren. Gegenüber einem durchschnittlichen Testergebnis im Auswahlverfahren

von 21.9 Punkten ( $SD = 9.8$ ), erzielten die Studierenden in der ersten Semesterwoche einen höheren Durchschnitt von 24.3 Punkten ( $SD = 7.8$ ). Die Verbesserung bzw. der Unterschied in Höhe von 2.4 Ergebnispunkten in der Testwiederholung des HAM-Man nach circa acht Wochen fällt gegenüber der Durchführung im Auswahlverfahren signifikant aus ( $t = -2.176$ ,  $df = 36$ ,  $p < .05$ ). Die 26 nicht über das Auswahlverfahren zugelassenen Studierenden setzten sich aus sechs Männern und 14 Frauen zusammen, erzielten in der ersten Studienwoche einen Punktwert im Drahtbiegetest HAM-Man von 17.1 ( $SD = 9.0$ ), der signifikant schlechter ist als der der Testwiederholenden ( $t = -3.402$ ,  $df = 61$ ,  $p < .001$ ). Vergleicht man das Ergebnis der 26 Erstteilnehmenden am HAM-Man mit den 129 Bewerbern im Auswahlverfahren ( $M = 17.0$ , s. Tabelle 5.01), so zeigt sich kein signifikanter Unterschied ( $t = 0.055$ ,  $df = 151$ ,  $n.s.$ ) unter Auflage der Varianzhomogenität beider Verteilungen, die nach Prüfung mittels des *Brown-Forsythe Tests* als erfüllt angesehen werden kann ( $F = 0.003$ ,  $df1 = 1$ ,  $df2 = 36.606$ ,  $n.s.$ ).

Die Reliabilitäten des Drahtbiegetest HAM-Man sind überwiegend als exzellent zu bezeichnen. Die interne Konsistenz beträgt  $\alpha = 0.98$  (s. Tabelle 5.02), die beinahe identisch hoch bei der Testwiederholung in der ersten Studienwoche mit 63 Erstsemesterstudierenden ausfällt und  $\alpha = 0.97$  beträgt, womit die Homogenität des Tests zum Wiederholungsmesspunkt bestätigt ist (s. Tabelle A06 im Anhang). Vergleichbar reliabel ist die Interrater-Reliabilität des HAM-Man einzustufen, für die als Maß unjustierte, nicht mittelwertskorrigierte Intra-Klassen-Korrelationen mit zweifaktoriellem Design auf Basis von Einzelwerten als Ratings berechnet wurden, die auch zur Ermittlung der Retest-Reliabilität verwendet werden. In beiden Testdurchführungen zeigte sich eine ähnlich hohe Übereinstimmung in den Urteilen beider Juroren, die für den HAM-Man zum Zeitpunkt des Auswahlverfahrens 96 Prozent (s. Tabelle A03 im Anhang) und in der Orientierungswoche 95 Prozent betrug (s. Tabelle A08 im Anhang). Einzig die Retest-Reliabilität weicht von dem exzellenten Reliabilitätsniveau ab, da die Übereinstimmung in der Messgenauigkeit mit  $ICC = 0.70$  zu den Zeitpunkten Auswahlverfahren und Orientierungswoche als akzeptable Reliabilität einzustufen ist (Wirtz & Caspar, 2002). Eine detaillierte Ergebnisdarstellung der Testwiederholung des HAM-Man findet sich im Anhang (s. Tabellen A06 und A07).

## 5.2. Diskrimination der Fähigkeitsarten

Die Fähigkeitsarten instrumentelles, manuelles Geschick und räumliches Vorstellungsvermögen sollten als eigenständige Konstrukte operationalisierbar sein, zwischen denen diskriminiert werden kann. Entsprechend sind in Hypothese 1 für beide Fähigkeitsarten als latente Konstrukte hohe Reliabilitäten postuliert. In Beziehung zueinander sollten instrumentelles, manuelles Geschick und räumliches Vorstellungsvermögen über eine hohe diskriminante Validität verfügen, um von verschiedenen Konstrukten sprechen zu können, wie als Erwartung in Hypothese 2 formuliert wurde. Die in den PLS-Modellen zur Vorhersage der Aufgabenleistungen in den vorklinischen Laborkursen einbezogenen Prädiktoren verfügen über ein identisches Messmodell, zu dessen Beurteilung das Verfahren der konfirmatorischen Faktorenanalyse eingesetzt wird. Bevor die Prüfung der ersten und zweiten Hypothese in Unterkapitel 5.2.3 erfolgt, wird zunächst im Unterkapitel 5.2.1 analysiert, ob die Voraussetzungen zum Einsatz der KFA erfüllt sind. Der Vergleich mit konkurrierenden Modellen ist in Unterkapitel 5.2.3 dargestellt, die Modellevaluation in Unterkapitel 5.2.3 aufgeführt.

### 5.2.1 Voraussetzungen der konfirmatorischen Faktorenanalyse

Die als Schätzverfahren eingesetzte Maximum-Likelihood-Methode (ML) der konfirmatorischen Faktorenanalyse erfordert eine Stichprobengröße von mindestens 100 Probanden, welche für die vorliegende Untersuchung mit einem  $n = 129$  erreicht wird. Alle manifesten Variablen der Prädiktoren instrumentelles, manuelles Geschick und räumliches Vorstellungsvermögen verfügen über Intervallskalenniveau. Das in Unterkapitel 4.3.2.5 begründete Entfernen der Prädiktoren Schulabschlussnote und naturwissenschaftliches Verständnis aus dem KFA-Modell wird durch die Korrelate zwischen den vier Prädiktoren in Tabelle 5.02 gestützt. Diese weisen nur vereinzelt einen schwachen Zusammenhang auf, sodass die Entnahme der zwei Prädiktoren zu keiner Verzerrung der Schätzungen in der KFA führt.

Die in Tabelle A02 im Anhang aufgeführten Korrelationen zwischen den Untertests des LPS bzw. den Indikatoren des räumlichen Vorstellungsvermögen unterschreiten deutlich das geforderte Niveau von  $r < 0.85$ , weshalb das Vorhandensein von Multikollinearität auszuschließen ist. Ein anderes Bild ergibt sich für die Beziehungen zwischen den Indikatoren des Prädiktors instrumentelles, manuelles Geschick. Vier

positive, signifikante Korrelationen größer  $r > 0.85$  finden sich zwischen den Indikatoren Deckungsgleichheit und Qualität der Biegung aller drei Drahtfiguren sowie zwischen den Indikatoren Deckungsgleichheit und planes Aufliegen der Drahtfigur Dritte Dimension, die auf das Vorhandensein von Multikollinearität schließen lassen (Bühner, 2011). Nach Grewal et al. (2004) ist ein Abbruch der KFA in diesem Fall nicht zwangsläufig notwendig, da die Indikatorreliabilität eine der wichtigsten Determinanten für die Genauigkeit der Parameterschätzungen im KFA-Modell ist, welche für alle betroffenen Indikatoren über  $rel(x_i) \geq 0.6$  und damit über dem von Homburg und Giering (1996) geforderten Niveau von 0.5 liegen.

Die Prüfung auf Multinormalverteilung der Indikatoren anhand der Kennziffern Schiefe und Kurtosis sowie den zugehörigen *C.R.*-Werten ist nur für den Indikator LPS UT 7 erfolgreich (*Sch* = -0.06, *C.R.* = -0.29; *Kur* = -0.56, *C.R.* = -1.30), wie aus Tabelle 5.04 in Unterkapitel 5.2.3 zu erkennen ist. Jedoch übersteigen die Gütewerte der Indikatoren niemals eine Schiefe größer zwei und eine Kurtosis größer sieben, sodass es zwar zu einer Verletzung der Multinormalverteilungsannahme kommt, aber gegen deren Ausprägungen sich die ML-Methode als relativ robust erweist (Bühner, 2011).

### 5.2.2 Konkurrenzmodelle

Vor dem Hintergrund der vielfältigen Definitionen des manuellen Geschicks (vgl. Unterkapitel 2.3.1.3) und der bisher nicht erfolgten Konstruktvalidierung des Drahtbiegetests HAM-Man, wird das postulierte Bi-Faktor-Modell konkurrierenden Modellen zur Prüfung der Operationalisierungsgüte gegenübergestellt. In den Konkurrenzmodellen ist die latente Variable instrumentelles, manuelles Geschick als ein- und dreifaktorielles Modell erfasst, während das räumliche Vorstellungsvermögen immer über die von Horn (1983) postulierte Ein-Faktor-Lösung verfügt (vgl. Unterkapitel 4.1.1.4). Im Anhang findet sich eine umfassende Darstellung der Indikatoren- und Faktorenkennwerte beider Konkurrenzmodelle (s. Tabellen A09 bis A14).

Die globalen Gütemaße in Tabelle 5.03 weisen insgesamt eine akzeptable Qualität für das Bi-Faktor-Modell auf (*GFI* = 0.907, *RMSEA* = 0.068, *SRMR* = 0.046,  $\chi^2/df$  = 1.597, *CFI* = 0.976), das sich im Vergleich zu den konkurrierenden Modellen bewährt. Die geringe Eignung einer einfaktoriellen Messstruktur zur Operationalisierung des instrumentellen, manuelles Geschicks wird zweifelsfrei durch die globalen Kennziffern für das Ein-Faktor-Modell angezeigt (*GFI* = 0.594, *RMSEA* = 0.266, *SRMR* = 0.140,

$\chi^2/df = 10.023$ ,  $CFI = 0.574$ ), deren Interpretation nicht nur eine eindeutige Verletzung des definierten Akzeptanzniveaus eines jeden Kriteriums anzeigt, sondern auch die Fehlspezifikation des Modells zur Messung des instrumentellen, manuellen Geschicks.

Tabelle 5.03: Globale Gütemaße der KFA-Modelle

KFA Modell	$\chi^2$	df	$\chi^2/df$	p	GFI	AGFI	RMSEA (PCLOSE)	SRMR	CFI	NFI
Bi-Faktor	87.8	55	1.597	.003	0.907	0.846	0.068 (.131)	0.046	0.976	0.939
Ein-Faktor	641.5	64	10.023	.000	0.594	0.423	0.266 (.000)	0.140	0.574	0.553
Drei-Faktoren	91.4	59	1.549	.004	0.904	0.852	0.066 (.163)	0.044	0.976	0.936

Anm.:  $\chi^2$  = Chi-Quadrat-Koeffizient; df = Anzahl der Freiheitsgrade; p = Wahrscheinlichkeit des Modellfits; GFI = Goodness of Fit Index ;AGFI = Adjusted Goodness of Fit Index; RMSEA = Root Mean Square Error of Approximation; PCLOSE = p of close fit of RMSEA; SRMR = Standardized Root Mean Residual; CFI = Comparative Fit Index; NFI = Normed Fit Index

Das Drei-Faktoren-Modell besitzt im Vergleich zum Bi-Faktor-Modell auf eine beinahe identische Modellgüte ( $GFI = 0.904$ ,  $RMSEA = 0.066$ ,  $SRMR = 0.044$ ,  $\chi^2/df = 1.549$ ,  $CFI = 0.976$ ). Das geforderte Akzeptanzniveau des „Normed Fit Index“ von  $NFI \geq 0.95$  ist für alle Modelle verletzt, fällt aber für das Bi-Faktor-Modell und das Drei-Faktoren-Modell mit  $NFI$ -Werten von 0.939 bzw. 0.936 gering aus (s. Tabelle 5.03). Aufgrund der vorliegenden Stichprobengröße, die im Rahmen einer KFA mit 129 Probanden als gering einzustufen ist (Gagne & Hancock, 2006), wäre die Unterschreitung des  $NFI$ -Akzeptanzniveaus zu tolerieren und um dessen mögliche Unterschätzung bei kleinen Stichproben vorzubeugen, dem  $CFI$ -Fit mehr Gewichtung einzuräumen (Bentler, 1990). Diese Argumentation ist nicht für das Ein-Faktor-Modell zulässig, da hier mit einem Wert von 0.553 eine deutliche Unterschreitung einer akzeptablen Güte vorliegt. Die berichteten Gütekriterien weisen keine eindeutige statistische Präferenz auf, ob die Modellstruktur des instrumentellen, manuellen Geschicks über ein Bi-Faktor-Modell oder ein Drei-Faktoren-Modell umzusetzen ist, führen aber zu einer indiskutablen Ablehnung des Ein-Faktor-Modells.

Eine Entscheidung ist somit unter Einbezug inhaltlicher Argumente zu treffen, die zugunsten des Bi-Faktor-Modells ausfällt. Das Bi-Faktor-Modell bietet die Vorteile, das instrumentelle, manuelle Geschick über einen allgemeinen Faktor zu erfassen und zusätzlich den Einfluss der einzelnen Drahtfiguren als eigenständige Faktoren auf die ihnen zugehörige Gruppe von Indikatoren abzubilden, der beispielsweise durch die unterschiedlichen Figurenanforderungen, aber auch durch die vom Probanden gewählte Bearbeitungsstrategie bedingt sein kann. Die Gruppenfaktoren identifizieren spezifische, einer Gruppe von Indikatoren zugehörige Varianz, welche über die vom allgemeinen Faktor erfasste hinausgeht, da dieser über die Gesamtheit aller Indikatoren gemessen wird.

### 5.2.3 Modellevaluation

Die in Unterkapitel 4.3.2.4 berichteten lokalen Gütemaße werden zur Beurteilung der Indikatoren und Faktoren des postulierten Modells herangezogen, in dem die nichtkognitive Fähigkeit instrumentelles, manuelles Geschicks als Bi-Faktor-Modell und die kognitive Fähigkeit räumliches Vorstellungsvermögen als Ein-Faktor-Modell operationalisiert sind. In Tabelle 5.04 ist die deskriptive Statistik der Indikatoren aufgeführt.

Tabelle 5.04: Deskriptive Statistik der Indikatoren

Indikator	M (SD)	Min	Max	Schiefe (C.R.)	Kurtosis (C.R.)	$r_{i(t-i)}$	$p_m$
<b>LPS</b>							
- UT 7	25.4 (8.43)	5	40	-0.06 (-0.29)	-0.56 (-1.30)	0.421	0.403
- UT 8	34.8 (6.61)	10	40	-1.60 (-7.39)	1.81 (4.19)	0.369	0.757
- UT 9	27.6 (4.84)	14	37	-0.73 (-3.39)	0.31 (0.72)	0.288	0.476
- UT 10	31.7 (5.48)	13	40	-0.75 (-3.47)	0.20 (0.47)	0.309	0.628
<b>HAM-Man</b>							
- Dreieck - DG	3.8 (1.57)	0	6	-0.77 (-3.57)	0.04 (0.09)	0.857	0.401
- Dreieck - QB	1.9 (0.75)	0	3	-1.02 (-4.74)	0.37 (0.85)	0.908	0.401
- Dreieck - PA	3.9 (1.35)	0	6	-1.24 (-5.77)	1.61 (3.74)	0.879	0.423
- Dritte Dim. - DG	1.9 (1.87)	0	6	0.47 (2.17)	-1.13 (-2.63)	0.923	0.100
- Dritte Dim. - QB	1.0 (0.94)	0	3	0.18 (0.82)	-1.43 (-3.31)	0.927	0.111
- Dritte Dim. - PA	1.9 (1.76)	0	5.5	0.31 (1.42)	-1.26 (-2.90)	0.905	0.100
- Rundbögen - DG	1.1 (1.50)	0	6	1.32 (6.10)	0.76 (1.76)	0.837	0.033
- Rundbögen - QB	0.5 (0.64)	0	2.5	1.41 (6.55)	1.12 (2.60)	0.873	0.027
- Rundbögen - PA	1.0 (1.37)	0	5	1.24 (5.77)	0.33 (0.77)	0.895	0.028

Anm.: M = Mittelwert; SD = Standardabweichung; Min = Minimalwert; Max = Maximalwert, C.R. = Critical Ratio;  $r_{i(t-i)}$  = korrigierte Trennschärfe;  $p_m$  = Schwierigkeit; DG = Bewertungskriterium Deckungsgleichheit; QB = Qualität der Biegung; PA = planes Aufliegen der Drähte; Dim. = Dimension

Der vermutete ansteigende Schwierigkeitsgrad über die einzelnen Drahtfiguren des HAM-Nat bestätigt sich in den Itemschwierigkeiten  $p_m$  der Indikatoren, die über die einzelnen Drahtfiguren abnehmen. Die Indikatoren des latenten Konstrukts räumliches Vorstellungsvermögen verfügen über eher schwache, positive Trennschärfen, was ihre Eignung, zwischen hohen und niedrigen Ausprägungen der Fähigkeit bei Bewerbern zu trennen, mindert. Die Trennschärfen der Indikatoren des instrumentellen, manuellen Geschicks besitzen hingegen sehr hohe, positive Trennschärfen, die aufzeigen, dass zwischen Bewerbern mit hohen und niedrigen Fähigkeitsmerkmalen sehr gut diskriminiert werden kann.

Die Parameterschätzungen der Indikatoren in Tabelle 5.05 fallen für die latenten Konstrukte sehr unterschiedlich aus. Das räumliche Vorstellungsvermögen verfügt als hypothetisches Konstrukt über vier Indikatoren, die ein positives Ladungsvorzeichen besitzen, wie es in der von Horn (1983) postulierten einfaktoriellen Modellstruktur gefordert ist. Das zur Beurteilung der standardisierten Faktorladungen definierte

Akzeptanzniveau von  $\lambda_{ij} > 0.5$  ist nur für den Indikator LPS UT 9 ( $\lambda_{ij} = 0.775$ ) erfüllt; der Indikator LPS UT 10 unterschreitet das geforderte Niveau aber nur geringfügig mit  $\lambda_{ij} = 0.458$ . Entsprechend verhalten sich die Reliabilitäten  $rel(x_i)$ . Nur der Indikator LPS UT 9 verfügt über eine akzeptable Messgenauigkeit mit  $rel(x_i) = 0.603$ . Die fast identisch hohen Messfehlervarianzen  $\theta_{ii}$  der Indikatoren des räumlichen Vorstellungsvermögens deuten weniger auf zufällige, sondern eher auf systematische Fehler hin.

Tabelle 5.05: Parameterschätzungen der Indikatoren

Indikator <-- Faktor	$\lambda_{ijn}$	S.E.	C.R.	$\lambda_{ij}$	$\theta_{ii}$	$h^2$	$rel(x_i)$
LPS UT 7 <-- Raum	2.769*	0.90	3.06	0.330*	0.884	0.109	0.110
LPS UT 8 <-- Raum	2.705	0.68	3.97	0.411	0.825	0.172	0.170
LPS UT 9 <-- Raum	3.738	0.57	6.58	0.775	0.396	0.601	0.603
LPS UT 10 <-- Raum	2.501	0.60	4.16	0.458	0.784**	0.206	0.211
Dreieck - DG <-- Geschick	1.009	0.14	7.14	0.645	0.139	0.860	0.750
Dreieck - QB <-- Geschick	0.499	0.07	7.39	0.668	0.090	0.909	0.788
Dreieck - PA <-- Geschick	0.778	0.13	6.23	0.578	0.198	0.800	0.693
Dritte Dim - DG <-- Geschick	1.539	0.16	9.43	0.826	0.079	0.920	0.896
Dritte Dim - QB <-- Geschick	0.752	0.08	8.96	0.800	0.088	0.912	0.879
Dritte Dim - PA <-- Geschick	1.381	0.16	8.67	0.786	0.139	0.860	0.816
Rundbögen - DG <-- Geschick	0.954	0.13	7.10	0.637	0.043**	0.957	0.904
Rundbögen - QB <-- Geschick	0.400	0.06	6.92	0.624	0.172	0.827	0.694
Rundbögen - PA <-- Geschick	0.880	0.12	7.10	0.643	0.258	0.740	0.616
Dreieck - DG <-- Dreieck	-1.044	0.11	-9.54	-0.667	0.139	--	0.762
Dreieck - QB <-- Dreieck	-0.508	0.05	-9.86	-0.680	0.090	--	0.837
Dreieck - PA <-- Dreieck	-0.919	0.10	-9.43	-0.683	0.198	--	0.702
Dritte Dim. - DG <-- Dritte Dim.	-0.910	0.18	-5.20	-0.488	0.079	--	0.751
Dritte Dim. - QB <-- Dritte Dim.	-0.490	0.09	-5.67	-0.521	0.088	--	0.755
Dritte Dim. - PA <-- Dritte Dim.	-0.866	0.17	-5.06	-0.493	0.139	--	0.636
Rundbögen - DG <-- Rundbögen	-1.113	0.10	-11.33	-0.743	0.043**	--	0.928
Rundbögen - QB <-- Rundbögen	-0.424	0.05	-9.24	-0.661	0.172	--	0.718
Rundbögen - PA <-- Rundbögen	-0.783	0.10	-7.59	-0.572	0.258	--	0.559

Anm.:  $\lambda_{ijn}$  = nichtstandardisierte Faktorladungen; S.E. = Standardfehler der Schätzung; C.R. = Critical Ratio;  $\lambda_{ij}$  = standardisierte Faktorladungen;  $\theta_{ii}$  = Messfehlervarianzen;  $h^2$  = Kommunalitäten;  $rel(x_i)$  = Indikatorreliabilität; Raum = räumliches

Vorstellungsvermögen; Geschick = instrumentelles, manuelles Geschick; Dritte Dim. = Dritte Dimension

\* Schätzungen der Faktorladungen auf Niveau  $p < .001$  signifikant, außer für Indikator LPS UT 7 ( $p < .002$ )

\*\* Schätzungen der Messfehlervarianzen auf Niveau  $p < .001$  signifikant, außer für Indikatoren Rundbögen - DG ( $p < .177$ ) und LPS UT 10 ( $p < .008$ )

Die signifikanten Faktorladungen mit Werten  $\lambda_{ij} > 0.5$  ( $p < .001$ ) bestätigen die postulierte Beziehungen zwischen Faktoren und Indikatoren im Bi-Faktor-Modell des instrumentellen, manuellen Geschicks und sind als zuverlässig einzustufen (s. Tabelle 5.05). Leichte Verletzungen des Akzeptanzniveaus von  $\lambda_{ij} > 0.5$  bestehen nur für die Ladungen des Gruppenfaktors Dritte Dimension auf die Indikatoren Deckungsgleichheit ( $\lambda_{ij} = -0.488$ ) und planes Aufliegen ( $\lambda_{ij} = -0.493$ ), die als akzeptabel einzustufen sind. Die Vorzeichen aller Faktorladungen entsprechen den inhaltlichen Erwartungen (vgl. Unterkapitel 4.1.1.4). Die Indikatoren im Bi-Faktor-Modell des instrumentellen, manuellen Geschicks verfügen insgesamt über eine gute bis sehr gute Messgenauigkeit,

sei es in Beziehung zum allgemeinen Faktor oder zu den Gruppenfaktoren. Lediglich die Faktorladung von 0.559 des Indikators Rundbögen - PA zum Gruppenfaktor Rundbögen unterschreitet geringfügig die Vorgabe von  $rel(x_i) > 0.6$ , die aber zu tolerieren ist, da immerhin noch die einfache Mehrheit von 55.9 Prozent der Indikatorvarianz durch den Gruppenfaktor aufgeklärt wird.

Entsprechend den Empfehlungen von Albers und Hildebrandt (2006) und Rossiter (2002) werden keine Indikatoren aufgrund von Verletzungen des Akzeptanzniveaus eines oder mehrerer Gütekriterien aus dem Modell entfernt. Eine Entnahme geht nicht nur mit Informationsverlust, sondern auch mit einer Beeinträchtigung der inhaltlichen Validität einher. Der Ausschluss des Indikators LPS UT 7 würde beispielsweise zur Folge haben, dass jegliche Informationen über die Fähigkeit zur Wahrnehmung von räumlichen Beziehungen bei der Operationalisierung der latenten Variable räumliches Vorstellungsvermögen verloren gingen. Darüber hinaus nimmt keine Indikatorreliabilität eine Wertausprägung  $rel(x_i) > 1.0$  an, weshalb Fehlspezifikationen in den Beziehungen zwischen Indikatoren und Faktoren ausgeschlossen werden können.

### 5.2.3.1 Reliabilitäten der manuellen und räumlichen Fähigkeit

In der ersten Hypothese ist für die Fähigkeitsarten eine hohe Konstruktreliabilität postuliert. Die Prüfung der Messgenauigkeit auf Konstruktebene in Tabelle 5.06 ergibt für die latente Variable räumliches Vorstellungsvermögen ein wenig zufriedenstellendes Resultat. Die interne Konsistenz ist mit einem Cronbachs  $\alpha = 0.58$  als schlecht zu bezeichnen, die Faktorreliabilität  $rel(\xi_j)$  beträgt 0.572, welche die in der Literatur als Grenze für ein akzeptables Maß festgelegte Höhe größer 0.6 nicht überschreitet und die mittlere korrigierte Trennschärfe erfüllt nur knapp die Forderung von  $r_{i(t-i)} > 0.3$ . Die  $DEV(\xi_j)$  der kognitiven Fähigkeit räumliches Vorstellungsvermögen in Höhe von 0.274 unterschreitet ebenfalls die geforderte Güteschwelle von 0.5.

Tabelle 5.06: Faktorenkennwerte

Faktoren	M (SD)	$\alpha$	$r_{i(t-i)}$	$DEV(\xi_j)$	$rel(\xi_j)$
Räumliches Vorstellungsvermögen	29.9 (6.33)	0.58	0.347	0.274	0.572
Instrumentelles, manuelles Geschick	1.9 (1.26)	0.92	0.717	0.783	0.970
Faktor Dreieck	3.2 (1.22)	0.95	0.879	0.763	0.906
Faktor Dritte Dimension	1.6 (1.52)	0.96	0.918	0.711	0.880
Faktor Rundbögen	0.9 (1.05)	0.94	0.869	0.736	0.892

Anm.: M = Mittelwert; SD = Standardabweichung;  $\alpha$  = Cronbachs Alpha für standardisierte Items;  $r_{i(t-i)}$  = mittlere korrigierte Trennschärfe;  $DEV(\xi_j)$  = Durchschnittlich extrahierte Varianz;  $rel(\xi_j)$  = Faktorreliabilität

Aufgrund der verletzten Akzeptanzniveaus in den Gütemaßen interne Konsistenz, Faktorreliabilität und Durchschnittlich Extrahierte Varianz ist Hypothese 1 für das latente Konstrukt der kognitiven Fähigkeit räumliches Vorstellungsvermögen zurückzuweisen.

Die Reliabilitäten im Bi-Faktor-Modell des instrumentellen, manuellen Geschicks sind überwiegend als exzellent einzustufen (s. Tabelle 5.06). Alle Faktoren verfügen über ein Cronbachs  $\alpha$  größer 0.90, was auf eine sehr hohe Skalenskonsistenz schließen lässt. Der allgemeine Faktor instrumentelles, manuelles Geschick besitzt eine exzellente Faktorreliabilität mit  $rel(\xi_j) = 0.97$ ; die der Gruppenfaktoren Dreieck, Dritte Dimension und Rundbögen fallen zwar etwas geringer aus, sind aber als sehr gut einzustufen. Die mittleren korrigierten Trennschärfen  $r_{i(t-i)}$  der Gruppenfaktoren sind mit Ausprägungen größer 0.8 höher als die des allgemeinen Faktors ( $r_{i(t-i)} = 0.717$ ), jedoch ist dessen Differenzierung der Fähigkeit immer noch als gut einzustufen. Das Gütemaß  $DEV(\xi_j)$  ist für die latente, nichtkognitive Fähigkeit instrumentelles, manuelles Geschick mit 0.783 ebenfalls als erfüllt zu betrachten. Hypothese 1 ist für das Konstrukt instrumentelles, manuelles Geschick damit erfüllt.

Das Unterschreiten der Akzeptanzniveaus mehrerer Reliabilitätsmaße der räumlichen Fähigkeit führt dazu, dass Hypothese 1 nur teilweise angenommen werden kann. Für die weitere Ergebnisinterpretation stellt die berichtete Verletzung der Gütekriterien zugleich eine Einschränkung dar, die zwar als bedenklich einzuordnen und entsprechend zu berücksichtigen ist, bedeutet aber nicht einen Abbruch der Auswertung.

### 5.2.3.2 Diskrimination der manuellen und räumlichen Fähigkeit

Zur abschließenden Beurteilung der Faktorengüte und zugleich zur Prüfung der zweiten Hypothese werden die in Tabelle 5.07 aufgeführten diskriminanten Validitäten der Fähigkeiten räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick herangezogen.

Tabelle 5.07: Diskriminante Validitäten

LV	$r_{(\xi_j, \xi_k)}^*$		DEV( $\xi$ )	$DEV(\xi) > r_{\xi_j, \xi_k}^2$
	1	2		Fornell-Larcker-Kriterium
1. instrumentelles, manuelles Geschick $\xi_j$	1	.525	0.783	$0.783 > 0.276$
2. räumliches Vorstellungsvermögen $\xi_k$	.525	1	0.274	$0.274 < 0.276$

Anm.: LV = latente Variable;  $r_{\xi_j, \xi_k}^2$  = Korrelation zwischen den LV; DEV( $\xi_j$ ) = Durchschnittlich extrahierte Varianz;  $DEV(\xi) > r_{\xi_j, \xi_k}^2$  = Fornell-Larcker-Kriterium;  $r(\xi_j, \xi_k)$  signifikant auf dem Niveau  $p < .000$

\* Aufgrund der auf eins fixierten Varianzen der latenten Variablen entspricht die Korrelation der LV ihrer Kovarianz.

In Hypothese 2 wird eine hohe Diskriminanzvalidität zwischen den Konstrukten instrumentelles, manuelles Geschick und räumlichem Vorstellungsvermögen angenommen. Das zur Beurteilung der diskriminanten Validität auf Konstruktebene herangezogene Fornell-Larcker-Kriterium ist für die latente Variable instrumentelles, manuelles Geschick erfüllt, da dessen  $DEV(\xi_j) = 0.783$  größer ist als die quadrierte Korrelation von 0.276 zum Konstrukt räumlichem Vorstellungsvermögen. Die latente Variable räumliches Vorstellungsvermögen kann hingegen diese Forderung für das Auftreten von Diskriminanzvalidität nicht erfüllen. Die  $DEV(\xi_j) = 0.274$  ist kleiner als die quadrierte Korrelation in Höhe von 0.276, welche sich aus der Korrelation  $r = 0.525$  zur latenten Variable instrumentelles, manuelles Geschick ergibt. Aufgrund dieser Resultate ist Hypothese 2 teilweise zurückzuweisen.

Die diskriminante Validität der kognitiven Fähigkeit räumliches Vorstellungsvermögen unterschreitet die in der Literatur geforderten Auflagen einer validen Messung. Deren Qualität kann aber immer noch als ausreichend angesehen werden, um eine Überprüfung der Hypothesen zur Prädiktion der vorklinischen Laborkursleistungen vorzunehmen. Zusätzlich ist zu berücksichtigen, dass in den varianzbasierten Strukturgleichungsmodellen die Güte der Messung für jede latente Variable beurteilt wird.

### 5.3 Prädiktion der vorklinischen Laborkursleistung

Während in den Hypothesen 3 bis 8 die prognostischen Beziehungen der Fähigkeitsarten sowie der Schulabschlussnote und des naturwissenschaftlichen Verständnis zur Leistung konsistenter, einfacher Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten in den vorklinischen Laborkursen thematisiert wurde, ist die prognostische Beziehung der Prädiktoren zur Leistung beim Fertigkeitserwerb für inkonsistente, komplexe Aufgaben in den Hypothesen 9 bis 14 formuliert. Bevor eine Prüfung der Hypothesen in Unterkapitel 5.3.2 erfolgt, werden zuvor die PLS-Prädiktionsmodelle hinsichtlich ihrer Güte in Unterkapitel 5.3.1 evaluiert.

#### 5.3.1 Evaluation der Prädiktionsmodelle

Die in Unterkapitel 4.3.5 beschriebenen Prädiktionsmodelle variieren nur in der endogenen Variablen, da die Leistung beim Erwerb zahnmedizinischer Techniken und in den vorklinischen Laborkursen in Abhängigkeit des Aufgabentyps operationalisiert wird. Die praktischen Arbeiten im Laborkurs der technischen Propädeutik entsprechen konsistenten, einfachen Aufgaben, die der Phantomkurse inkonsistenter und komplexer Aufgaben, weshalb die Prädiktionsmodelle entsprechend der vorklinischen Laborkurse TPK und PhaI bezeichnet werden. Die in Unterkapitel 5.1.1 berichtete deskriptive Statistik weist für den zweiten Phantomkurs eine Stichprobengröße von 28 Teilnehmern auf, sodass die von Chin und Newsted (1999) empfohlene Mindeststichprobengröße für PLS-Modelle von 30 bis 100 Fällen unterschritten ist. Aufgrund dieser Verletzung kann nicht von zuverlässigen Parameterschätzungen im PLS-Prädiktionsmodell PhaII ausgegangen werden, weshalb die Ergebnisse im Anhang der vorliegenden Arbeit dargestellt (s. Tabellen A15 bis A20) und in die Diskussion aufgegriffen, aber nicht zur Überprüfung der Hypothesen herangezogen werden.

##### 5.3.1.1 Analyse der Messmodelle

Die konfirmatorische Prüfung der Faktorladungen (vgl. Gefen et al., 2000) ergab in den Prädiktionsmodellen TPK und PhaI eine mangelnde Passung für den Indikator LPS UT 10 der latenten exogenen Variable räumliches Vorstellungsvermögen (TPK:  $\lambda_{ij} = 0.188$ , *n.s.*; PhaI:  $\lambda_{ij} = 0.421$ ,  $p < .01$ ). Die daraus resultierende Entnahme des Indikators aus beiden Messmodellen gewährleistet deren Kommensurabilität und verbessert jeweils die Güte der latenten Variable. Die *DEV* im Messmodell TPK erhöht sich von 0.372 auf

0.485, Cronbachs  $\alpha$  steigt geringfügig von 0.53 auf 0.57 und die Faktorreliabilität  $\rho_c$  besitzt nach der Indikatorentnahme eine Höhe von 0.72 anstatt von 0.66. Im Messmodell PhaI führt der Ausschluss des Indikators LPS UT 10 zu einer Verbesserung der *DEV* von 0.382 auf 0.491, die interne Konsistenz beträgt  $\alpha = 0.57$  anstatt zuvor 0.49 und die Faktorreliabilität  $\rho_c$  erhöht sich von 0.70 auf 0.73. Der Ausschluss des Indikators stellt keine gravierende Beeinträchtigung der Inhaltsvalidität dar (vgl. Albers & Hildebrandt, 2006; Rossiter, 2002), da der Untertest 10 des LPS Horn (1983) zufolge eher Thurstones Intelligenzfaktor „closure“ entspricht, also das Erkennen des Wesentlichen trotz ablenkender Einzelheiten erfasst.

Tabelle 5.08: Faktorladungen im TPK-Messmodell

latente Variable	Indikator	$\lambda_{ij}$	S.E.	t-Wert
räumliches Vorstellungsvermögen	LPS UT 7	0.6453	0.295	2.18
	LPS UT 8	0.9269	0.356	2.60
	LPS UT 9	0.4240	0.344	1.23
Faktor Dreieck	Dreieck - DG	0.9435	0.015	63.63
	Dreieck - QB	0.9439	0.012	79.92
	Dreieck - PA	0.9498	0.011	86.21
Faktor Dritte Dimension	Dritte Dim - DG	0.9641	0.010	93.05
	Dritte Dim - QB	0.9462	0.015	61.03
	Dritte Dim - PA	0.9629	0.007	128.51
Faktor Rundbögen	Rundbögen - DG	0.9526	0.006	154.55
	Rundbögen - QB	0.9198	0.006	150.45
	Rundbögen - PA	0.9504	0.010	94.89
allgemeiner Faktor instrumentelles, manuelles Geschick	Dreieck - DG	0.7240	0.054	13.53
	Dreieck - QB	0.7075	0.056	12.64
	Dreieck - PA	0.7760	0.041	18.99
	Dritte Dimension - DG	0.8614	0.019	45.58
	Dritte Dimension - QB	0.7599	0.030	26.29
	Dritte Dimension - PA	0.7816	0.032	24.10
	Rundbögen - DG	0.6621	0.051	13.09
	Rundbögen - QB	0.6825	0.043	16.02
	Rundbögen - PA	0.7125	0.045	16.01
Schulabschlussnote	Abiturdurchschnittsnote	1	-	-
naturwissenschaftliches Verständnis	HAM-Nat	1	-	-
Fertigkeitserwerb im TPK	TPK Note1	0.6094	0.137	4.46
	TPK Note2	0.4768	0.164	2.91
	TPK Note3	0.9223	0.088	10.48
	TPK Note4	0.7075	0.263	2.69

Anm.:  $\lambda_{ij}$  = geschätzte Faktorladungen; S.E. = Standardfehler der Schätzung; t-Wert = Testprüfgröße des t-Tests, UT = Untertest; LPS = Leistungsprüfsystem; DG = Deckungsgleichheit; QB = Qualität der Biegung; PA = planes Aufliegen; HAM-Nat = Naturwissenschaftstest; TPK = Technisch-Propädeutischer-Kurs;

1. Codierung Noten: 1 = sehr schlecht, 6 = sehr gut
2. „Bootstrapping“-Verfahren: 5000 Zufallsstichproben
3. Stichprobe: n = 38

Die Prüfung der Messmodelle TPK und PhaI deutet auf eine für beide Modelle akzeptable Qualität hin, da mehrere Gütekriterien von einigen manifesten Variablen verletzt wurden. Die für Faktorladungen geforderte Güte  $\lambda > 0.707$  wurde am stärksten im Messmodell TPK (s. Tabelle 5.08) durch den Indikator LPS UT 9 ( $\lambda = 0.424$ , *n.s.*) unterschritten, hingegen im Messmodell PhaI (s. Tabelle 5.09) durch die Ladung des Indikators Dreieck - QB auf den allgemeinen Faktor ( $\lambda = 0.463$ ,  $p < .01$ ). Mehrheitlich entsprechen die Faktorladungen in beiden Messmodellen der Güteforderung, womit gewährleistet wäre, dass mehr als 50 Prozent der jeweiligen Indikatorvarianz durch das zu messende Konstrukt erklärt wird und somit die Zugehörigkeit der Indikatoren auf die Konstrukte bestätigt ist.

Tabelle 5.09: Faktorladungen im PhaI-Messmodell

latente Variable	Indikator	$\lambda_{ij}$	S.E.	t-Wert
räumliches Vorstellungsvermögen	LPS UT 7	0.5683	0.163	3.50
	LPS UT 8	0.5801	0.197	2.94
	LPS UT 9	0.9018	0.104	8.65
Faktor Dreieck	Dreieck - DG	0.8720	0.030	29.17
	Dreieck - QB	0.8884	0.022	41.20
	Dreieck - PA	0.9167	0.011	85.17
Faktor Dritte Dimension	Dritte Dim - DG	0.9510	0.008	114.15
	Dritte Dim - QB	0.9285	0.014	65.21
	Dritte Dim - PA	0.9557	0.009	105.94
Faktor Rundbögen	Rundbögen - DG	0.9575	0.009	107.80
	Rundbögen - QB	0.9197	0.015	59.89
	Rundbögen - PA	0.9407	0.008	116.26
allgemeiner Faktor instrumentelles, manuelles Geschick	Dreieck - DG	0.5338	0.085	6.29
	Dreieck - QB	0.4631	0.069	9.63
	Dreieck - PA	0.6684	0.105	4.42
	Dritte Dim. - DG	0.8382	0.025	33.10
	Dritte Dim. - QB	0.7103	0.045	15.78
	Dritte Dim. - PA	0.7791	0.032	24.04
	Rundbögen - DG	0.6807	0.042	17.31
	Rundbögen - QB	0.7279	0.054	12.61
Rundbögen - PA	0.7596	0.040	19.08	
Schulabschlussnote	Abiturdurchschnittsnote	1	--	--
naturwissenschaftliches Verständnis	HAM-Nat	1	--	--
Fertigkeitserwerb im PhaI	PhaI Note1	0.7874	0.142	5.55
	PhaI Note2	0.5517	0.152	3.64
	PhaI Note3	0.8300	0.109	7.63

Anm.:  $\lambda_{ij, um}$  = geschätzte Faktorladungen; S.E. = Standardfehler der Schätzung; t-Wert = Testprüfgröße des t-Tests, UT = Untertest; LPS = Leistungsprüfsystem; DG = Deckungsgleichheit; QB = Qualität der Biegung; PA = planes Aufliegen; HAM-Nat = Naturwissenschaftstest; PhaI = Phantomkurs I

1. Stichprobe:  $n = 31$
2. Codierung Noten: 1 = ungenügend, 6 = sehr gut
3. „Bootstrapping“-Verfahren: 5000 Zufallsstichproben

Die Standardfehler der Schätzung (*S.E.*) liefern in beiden Modellen keinen Hinweis auf wenig zuverlässige Parameterschätzungen. Die Auflage der Faktorreliabilität  $\rho_c > 0.7$  ist für alle latenten Variablen in beiden Messmodellen erfüllt, sodass von einer akzeptablen Faktoreignung zur Erklärung aller zu ihm in Beziehung stehenden Indikatoren gesprochen werden kann, wie aus den Tabellen 5.10 und 5.11 hervorgeht. Die Prüfung auf interne Konsistenz der Messskalen ergab Abweichungen vom geforderten Güteniveau  $\alpha \geq 0.70$  für die latenten Variablen Fertigkeitserwerb (TPK:  $\alpha = 0.64$ ; PhaI:  $\alpha = 0.59$ ) und räumliches Vorstellungsvermögen (TPK & PhaI:  $\alpha = 0.57$ ).

Tabelle 5.10: Gütekriterien des TPK-Messmodells

latente Variable ( $\xi_j$ )	DEV ( $\xi_j$ )	$\alpha$	$H_j^2$	$\rho_c$	Fornell-Larcker-Kriterium
Fertigkeitserwerb im TPK ( $\xi_{\text{endo}}$ ) für konsistente, einfache Aufgaben	0.487	0.64	0.488	0.78	$\xi_{\text{endo}} \geq \xi_{j1}$ 0.487 > 0.0196 $\xi_{\text{endo}} \geq \xi_{j2}$ 0.487 > 0.2916 $\xi_{\text{endo}} \geq \xi_{j3}$ 0.487 > 0.0441 $\xi_{\text{endo}} \geq \xi_{j4}$ 0.487 > 0.0025
räumliches Vorstellungsvermögen ( $\xi_{j1}$ )	0.485	0.57	0.481	0.72	$\xi_{j1} \geq \xi_{j2}$ 0.485 > 0.0081 $\xi_{j1} \geq \xi_{j3}$ 0.485 > 0.0036 $\xi_{j1} \geq \xi_{j4}$ 0.485 > 0.0001 $\xi_{j1} \geq \xi_{\text{endo}}$ 0.485 > 0.0196
Schulabschlussnote ( $\xi_{j3}$ )	--	--	--	--	--
naturwiss. Verständnis ( $\xi_{j4}$ )	1	1	--	1	--
<b>Bi-Faktor-Modell</b>					
Gruppenfaktor Dreieck	0.894	0.94	0.894	0.96	$\xi_{\text{Dreieck}} \geq \xi_{\text{Dritte Dim.}}$ 0.894 > 0.2304 $\xi_{\text{Dreieck}} \geq \xi_{\text{Rundbögen}}$ 0.894 > 0.1025
Gruppenfaktor Dritte Dimension	0.917	0.96	0.917	0.97	$\xi_{\text{Dritte Dim.}} \geq \xi_{\text{Dreieck}}$ 0.917 > 0.2304 $\xi_{\text{Dritte Dim.}} \geq \xi_{\text{Rundbögen}}$ 0.917 > 0.2025
Gruppenfaktor Rundbögen	0.886	0.94	0.886	0.96	$\xi_{\text{Rundbögen}} \geq \xi_{\text{Dreieck}}$ 0.883 > 0.2304 $\xi_{\text{Rundbögen}} \geq \xi_{\text{Dritte Dim.}}$ 0.883 > 0.2025
allgemeiner Faktor instrumentelles, manuelles Geschick ( $\xi_{j2}$ )	0.552	0.90	0.552	0.92	$\xi_{j2} \geq \xi_{j1}$ 0.552 > 0.0081 $\xi_{j2} \geq \xi_{j3}$ 0.552 > 0.0064 $\xi_{j2} \geq \xi_{j4}$ 0.552 > 0.0081 $\xi_{j2} \geq \xi_{\text{endo}}$ 0.552 > 0.2916

Anm.: DEV ( $\xi_j$ ) = Durchschnittlich extrahierte Varianz;  $\alpha$  = Cronbachs Alpha;  $H_j^2$  = Kreuzvalidierungsmaß Kommunalität;  $\rho_c$  = Faktorreliabilität; Dim. = Dimension; TPK = Technisch-Propädeutischer-Kurs; naturwiss. = naturwissenschaftlich

1. Stichprobe: n = 38

Die Konstruktreliabilität  $DEV(\xi_j)$  kann als Gütekriterium hingegen nur für die exogene latente Variable instrumentelles, manuelles Geschick mit 0.552 im Messmodell TPK und im Messmodell PhaI nur für die endogene latente Variable Fertigkeitserwerb mit einem Wert von 0.538 als erfüllt angesehen werden, die in den Tabellen 5.10 und 5.11 aufgeführt sind. Die übrigen latenten Variablen unterschreiten in beiden Messmodellen aber nur sehr geringfügig die Auflage des Gütekriteriums von  $DEV(\xi_j) > 0.5$  in einer Stärke von maximal 0.2, d.h. von also höchstens zwei Prozent, wie eine weiterführende Betrachtung des Maßes zeigt.

Tabelle 5.11: Gütekriterien des PhaI-Messmodells

latente Variable	DEV ( $\xi_j$ )	$\alpha$	$H_j^2$	$\rho_c$	Fornell-Larcker-Kriterium
Fertigkeitserwerb im PhaI ( $\xi_{endo}$ ) für inkonsistente, komplexe Aufgaben	0.538	0.59	0.530	0.77	$\xi_{endo} \geq \xi_{j1}$ 0.538 > 0.0196 $\xi_{endo} \geq \xi_{j2}$ 0.538 > 0.2916 $\xi_{endo} \geq \xi_{j3}$ 0.538 > 0.0441 $\xi_{endo} \geq \xi_{j4}$ 0.538 > 0.0025
räumliches Vorstellungsvermögen ( $\xi_{j1}$ )	0.491	0.57	0.535	0.73	$\xi_{j1} \geq \xi_{j2}$ 0.491 > 0.1444 $\xi_{j1} \geq \xi_{j3}$ 0.491 > 0.0001 $\xi_{j1} \geq \xi_{j4}$ 0.491 > 0.0529 $\xi_{j1} \geq \xi_{endo}$ 0.491 > 0.0196
Schulabschlussnote ( $\xi_{j3}$ )	--	--	--	--	--
naturwiss. Verständnis ( $\xi_{j4}$ )	1	1	--	1	--
Gruppenfaktor Dreieck	0.797	0.87	0.797	0.92	$\xi_{Dreieck} \geq \xi_{Dritte Dim.}$ 0.797 > 0.1225 $\xi_{Dreieck} \geq \xi_{Rundbögen}$ 0.797 > 0.0485
Gruppenfaktor Dritte Dimension	0.893	0.94	0.893	0.96	$\xi_{Dritte Dim.} \geq \xi_{Dreieck}$ 0.893 > 0.1225 $\xi_{Dritte Dim.} \geq \xi_{Rundbögen}$ 0.893 > 0.1936
Gruppenfaktor Rundbögen	0.883	0.93	0.883	0.96	$\xi_{Rundbögen} \geq \xi_{Dreieck}$ 0.883 > 0.0485 $\xi_{Rundbögen} \geq \xi_{Dritte Dim.}$ 0.883 > 0.1936
allgemeiner Faktor instrumentelles, manuelles Geschick ( $\xi_{j2}$ )	0.481	0.86	0.482	0.89	$\xi_{j2} \geq \xi_{j1}$ 0.552 > 0.1444 $\xi_{j2} \geq \xi_{j3}$ 0.552 > 0.0361 $\xi_{j2} \geq \xi_{j4}$ 0.552 > 0.0169 $\xi_{j2} \geq \xi_{endo}$ 0.552 > 0.2916

Anm.: DEV ( $\xi_j$ ) = Durchschnittlich extrahierte Varianz;  $\alpha$  = Cronbachs Alpha;  $H_j^2$  = Kreuzvalidierungsmaß Kommunalität;  $\rho_c$  = Faktorreliabilität; Dim. = Dimension; PhaI = Phantomkurs I; naturwiss. = naturwissenschaftlich  
Stichprobe: n = 31

Die in Tabelle 5.10 dargestellten Kreuzvalidierungs-Kommunalitäten  $H_j^2$  lassen auf eine mittlere Passung zwischen den manifesten und latenten Variablen im Messmodell TPK schließen, die sich ebenfalls im Messmodell PhaI mit Kommunalitäten  $H_j^2$  um 0.5 in Tabelle 5.11 zeigt. Einzig die Gruppenfaktoren Dreieck, Dritte Dimension und Rundbögen im Bi-Faktor-Modell des instrumentellen, manuellen Geschicks weisen in beiden Modellen Kreuzvalidierungs-Kommunalitäten  $H_j^2 \geq 0.8$  aus, was auf eine hohe Passung von Indikatoren und Faktoren hindeutet.

Tabelle 5.12: Kreuzkorrelationen im TPK-Prädiktionsmodell

latente Variable Indikator	Raum	Man	Faktor Dreieck	Faktor Dritte Dim.	Faktor Rund- bögen	Schul.	natur. Verst.	FKE im TPK
LPS UT 7	.65	-.09	-.03	-.09	-.10	-.06	.15	-.06
LPS UT 8	.93	-.11	-.05	-.07	-.14	.11	-.10	-.15
LPS UT 9	.42	.11	-.12	.14	.26	-.02	.27	-.03
Dreieck - DG	-.09	.72	.94	.40	.34	-.04	-.16	.45
Dreieck - QB	-.06	.71	.94	.45	.24	-.01	-.10	.58
Dreieck - PA	-.05	.78	.95	.52	.33	-.10	-.14	.56
Dritte Dim. - DG	-.05	.86	.53	.96	.51	-.05	.26	.32
Dritte Dim. - QB	.02	.76	.42	.95	.38	-.27	.31	.45
Dritte Dim. - PA	-.14	.78	.43	.96	.40	-.08	.21	.30
Rundbögen - DG	-.04	.66	.32	.34	.95	.05	-.03	.27
Rundbögen - QB	-.11	.68	.25	.48	.92	-.15	.17	.35
Rundbögen - PA	-.11	.71	.33	.45	.95	.16	.06	.27
Abiturnote	.06	-.08	-.05	-.13	.02	1	-.12	-.21
HAM-Nat	.01	.09	-.14	.27	.07	-.12	1	.05
TPK Note1	-.16	.29	.26	.21	.21	-.32	-.10	.61
TPK Note2	-.22	.13	.07	-.02	.27	-.08	.06	.48
TPK Note3	-.10	.53	.58	.38	.26	-.19	.19	.92
TPK Note4	0	.43	.48	.31	.20	.04	-.06	.71

Anm.: UT = Untertest; DG = Deckungsgleichheit; QB = Qualität der Biegung; PA = planes Aufliegen Drähte; Dim. = Dimension; Abiturnote = Abiturdurchschnittsnote; HAM-Nat = Naturwissenschaftstest; TPK = Technisch-Propädeutischen-Kurs; Man = instrumentelles, manuelles Geschick; Schul. = Schulabschlussnote; natur. Verst. = naturwissenschaftliches Verständnis; FKE = Fertigkeitserwerb bei konsistenten, einfachen Aufgaben; TPK = Technisch-Propädeutischer Kurs

1. Stichprobe: n = 38
2. Codierung Noten: 1 = ungenügend, 6 = sehr gut

Wie aus den Tabellen 5.10 und 5.11 zu entnehmen ist, kann in beiden Messmodellen von diskriminanter Validität auf Konstruktebene ausgegangen werden, da das Fornell-Larcker-Kriterium für alle latenten Variablen erfüllt ist. Die Kreuzkorrelationen aus den Tabellen 5.12 und 5.13 stützen als weiteres Gütemaß der Diskriminanzvalidität die Interpretationen des Fornell-Larcker-Kriteriums. Es zeigt sich keine größere Korrelation zwischen einer Indikatorvariablen und einer anderen als der ihr zugeordneten latenten Variable, sodass eine Adaption der Messmodelle weder sinnvoll noch notwendig ist und somit eine gute diskriminante Validität der latenten Variablen in beiden Messmodellen angenommen werden kann.

Tabelle 5.13: Kreuzkorrelationen im PhaI-Prädiktionsmodell

latente Variable Indikator	Raum	Man	Faktor Dreieck	Faktor Dritte Dim.	Faktor Rund- bögen	Schul.	natur. Verst.	FKE im PhaI
LPS UT 7	.57	.08	.21	.05	-.04	-.13	.26	.06
LPS UT 8	.58	.17	.22	.12	.07	.14	.11	.25
LPS UT 9	.90	.38	.26	.21	.39	-.05	.21	.49
Dreieck - DG	.20	.53	.87	.22	.24	.18	-.11	.18
Dreieck - QB	.26	.46	.89	.22	.06	.29	-.15	.35
Dreieck - PA	.36	.67	.92	.44	.26	.15	.01	.48
Dritte Dim. - DG	.27	.84	.41	.95	.47	.23	.21	.24
Dritte Dim. - QB	.19	.71	.21	.93	.38	-.06	.36	.28
Dritte Dim. - PA	.15	.78	.35	.96	.39	.19	.16	.26
Rundbögen - DG	.38	.73	.23	.39	.96	.08	0	.44
Rundbögen - QB	.22	.68	.13	.40	.92	-.11	.13	.44
Rundbögen - PA	.33	.76	.25	.45	.94	.27	.05	.41
Abiturnote	.01	.19	.22	.14	.09	1	-.11	.04
HAM-Nat	.23	.13	-.08	.25	.06	-.11	1	.09
PhaI Note1	.46	.43	.39	.22	.38	.07	.16	.79
PhaI Note2	.15	.25	.12	.15	.27	.18	.04	.55
PhaI Note3	.39	.38	.28	.23	.35	-.08	-.01	.83

Anm.: UT = Untertest; DG = Deckungsgleichheit; QB = Qualität der Biegung; PA = planes Aufliegen Drähte; Dim. = Dimension; Abiturnote = Abiturdurchschnittsnote; HAM-Nat = Naturwissenschaftstest; Raum = räumliches Vorstellungsvermögen; Man = instrumentelles, manuelles Geschick; Schul. = Schulabschlussnote; natur. Verst. = naturwissenschaftliches Verständnis; FKE = Fertigkeitserwerb bei inkonsistenten, komplexen Aufgaben; Pha = Phantomkurs

1. Stichprobe: n = 31

2. Codierung Noten: 1 = ungenügend, 6 = sehr gut

Die Güte des reflektiven Single-Item-Konstrukts naturwissenschaftliches Verständnis kann anhand der von Hissbach et al. (2011) berichteten Testgütekriterien des Naturwissenschaftstests HAM-Nat als zufriedenstellend beurteilt werden (vgl. Unterkapitel 4.1.4.3). Die Kreuzkorrelationen in den Tabellen 5.12 und 5.13 deuten zudem auf keine Überschneidung mit einer anderen latenten Variable in den Messmodellen hin. Eine statistische Analyse ist für das formative Single-Item-Messmodell der Schulabschlussnote nicht möglich, weshalb die inhaltliche Validität auf theoretischer Ebene zu prüfen ist (Henseler et al., 2009; Rossiter, 2002). Die Abiturdurchschnittsnote spiegelt die durchschnittliche in den Abiturprüfungen erreichte Leistung wieder, in welche auch Leistungen gymnasialer Oberstufenkurse eingebracht werden können (vgl. Unterkapitel 4.1.3.1), sodass keine Zweifel an der inhaltlichen Validität für das formative Single-Item-Messmodell der Schulabschlussnote bestehen.

### 5.3.1.2 Analyse der Strukturmodelle

Die Gütebeurteilung der Strukturmodelle erfolgt in diesem Unterkapitel nur unter Einbezug des Bestimmtheitsmaßes  $R^2$ , des Stone-Geisser-Kriterium  $Q^2$  und der Korrelationen zwischen den latenten Variablen. Die ebenfalls in den Tabellen 5.14 und

5.15 abgebildeten Pfadkoeffizienten  $\beta$  der Prädiktoren sowie deren Einflussstärke  $f^2$  und Signifikanz werden zur Hypothesenprüfung im Unterkapitel 5.3.2 herangezogen.

Tabelle 5.14: Gütekriterien des TPK-Strukturmodells

latente Variablen	$\beta$ (t-Wert)	$R^2$	$f_j^2$	$Q^2$
Fertigkeitserwerb im TPK	--	0.323	--	0.162
räumliches Vorstellungsvermögen	-0.080 (0.723)	--	0.004	--
instrumentelles, manuelles Geschick	0.519 (5.082)		0.252	
Schulabschlussnote	-0.163 (2.128)		0.034	
naturwissenschaftliches Verständnis	-0.017 (0.184)		0.000	

Anm.:  $\beta$  = Pfadkoeffizient; t-Wert = Testprüfgröße des t-Tests (einseitig).  $R^2$  = Bestimmtheitsmaß;  $f^2$  = Effektstärke;  $Q^2$  = Stone-Geisser-Kriterium; TPK = Technisch-Propädeutischer-Kurs (konsistente, einfache Aufgaben)

1. Stichprobe:  $n = 38$
2. Bootstrapping-Verfahren: 5000 Zufallsstichproben

Die Strukturmodelle sind insgesamt von einer akzeptablen Güte, wie die in den Tabellen 5.14 und 5.15 aufgeführten Gütemaße indizieren (vgl. Unterkapitel 4.3.3.2). Die aufgestellten Prädiktionsmodelle zur Vorhersage der Leistung beim Fertigkeitserwerb für konsistente, einfache Aufgaben im TPK und für inkonsistente, komplexe Aufgaben im ersten Phantomkurs besitzen Prognoserelevanz, da das Kriterium  $Q^2$  größer als null ist (TPK:  $Q^2 = 0.162$ ; PhaI:  $Q^2 = 0.208$ ). Das von Herrmann et al. (2006) aufgestellte konservative Prüfkriterium von  $R^2$  größer 0.3 wird durch die endogene Variable Fertigkeitserwerb in beiden aufgestellten Strukturmodellen erfüllt (TPK:  $R^2 = 0.323$ ; PhaI:  $R^2 = 0.356$ ). Folgt man der Klassifikation von Chin (1998a), so ist das Bestimmtheitsmaß im TPK-Modell noch als schwach ( $R^2 > 0.19$ ), im PhaI-Modell hingegen als moderat zu klassifizieren ( $R^2 > 0.33$ ).

Tabelle 5.15: Gütekriterien des PhaI-Strukturmodells

latente Variablen	$\beta$ (t-Wert)	$R^2$	$f_j^2$	$Q^2$
Fertigkeitserwerb im PhaI	--	0.356	--	0.208
räumliches Vorstellungsvermögen	0.362 (6.29)		0.171	
instrumentelles, manuelles Geschick	0.370 (3.92)		0.168	
Schulabschlussnote	-0.036 (0.44)		0.003	
naturwissenschaftliches Verständnis	-0.041 (0.62)		0.001	

Anm.:  $\beta$  = Pfadkoeffizient; t-Wert = Testprüfgröße des t-Tests (einseitig);  $R^2$  = Bestimmtheitsmaß;  $f^2$  = Effektstärke;  $Q^2$  = Stone-Geisser-Kriterium; PhaI = Phantomkurs I (inkonsistente, komplexe Aufgaben)

1. Stichprobe:  $n = 31$
2. Bootstrapping-Verfahren: 5000 Zufallsstichproben

Die Korrelationen zwischen den exogenen latenten Variablen in Tabelle 5.16 und 5.17 lassen nicht auf das Vorhandensein von Multikollinearität in den Strukturmodellen schließen, da sich auf Konstruktebene zwischen den vier exogenen latenten Variablen keine linearen Zusammenhänge  $r > 0.85$  zeigen. Die Richtungen der Korrelationen fallen überwiegend wie erwartet in beiden Strukturmodellen aus. Eine Ausnahme bildet hier die negative, schwache Beziehung zwischen der Schulabschlussnote und dem

Fertigkeitserwerb im TPK mit  $r = -0.21$  (s. Tabelle 5.16), für die kein Zusammenhang erwartet wurde (vgl. Unterkapitel 3.3.2.1). Gleichzeitig liefern die Korrelationen keinen Hinweis auf Homogenität der latenten Variablen im Strukturmodell.

Tabelle 5.16: Korrelationen zwischen den latenten Variablen im TPK-Strukturmodell

latente Variable	1	2	3	4	5	6	7	8
1. instrumentelles, manuelles Geschick	1							
2. räumliches Vorstellungsvermögen	-.09	1						
3. Schulabschlussnote	-.08	.06	1					
4. naturwissenschaftliches Verständnis	.09	.01	-.12	1				
5. Faktor Dreieck	.78	-.07	-.05	-.14	1			
6. Faktor Dritte Dimension	.84	-.06	-.13	.27	.48	1		
7. Faktor Rundbögen	.32	-.09	.02	.07	.32	.45	1	
8. Fertigkeitserwerb im TPK	.54	-.14	-.21	.05	.56	.37	.32	1

Anm.: TPK = Technisch-Propädeutischer-Kurs (konsistente, einfache Aufgaben); n = 38

1. Codierung Noten: 1 = ungenügend, 6 = sehr gut

Im PhaI-Strukturmodell fallen die Beziehungen zwischen den latenten Variablen im Vergleich zum TPK-Strukturmodell teilweise unterschiedlich in Richtung und Stärke aus. So zeigt sich beispielsweise zwischen den Prädiktoren instrumentelles, manuelles Geschick und räumliches Vorstellungsvermögen in Tabelle 5.17 eine positiv moderate Korrelation ( $r = 0.38$ ), während im TPK-Strukturmodell kein Zusammenhang besteht (s. Tabelle 5.16). Im PhaI-Strukturmodell weist die kognitive Fähigkeit räumliches Vorstellungsvermögen eine schwache, positive Korrelation zum naturwissenschaftlichen Verständnis auf ( $r = 0.23$ ), aber im TPK-Strukturmodell besteht kein Zusammenhang zwischen beiden Variablen. Die korrelative Beziehung der manuellen Fähigkeit zur endogenen Variable ist mit  $r = 0.50$  im PhaI-Strukturmodell fast identisch zum TPK-Modell. Für das räumliche Vorstellungsvermögen findet sich einerseits eine positiv moderate Korrelation zum Fertigkeitserwerb im PhaI ( $r = 0.49$ ), aber andererseits eine negativ schwache Korrelation zum Fertigkeitserwerb im TPK-Modell ( $r = -0.14$ ).

Tabelle 5.17: Korrelationen zwischen den latenten Variablen im PhaI-Strukturmodell

latente Variable	1	2	3	4	5	6	7	8
1. instrumentelles, manuelles Geschick	1							
2. räumliches Vorstellungsvermögen	.38	1						
3. Schulabschlussnote	.19	.01	1					
4. naturwissenschaftliches Verständnis	.13	.23	-.11	1				
5. Faktor Dreieck	.64	.31	.22	-.08	1			
6. Faktor Dritte Dimension	.82	.22	.14	.25	.35	1		
7. Faktor Rundbögen	.77	.33	.09	.06	.22	.44	1	
8. Fertigkeitserwerb im PhaI	.50	.49	.04	.09	.39	.27	.46	1

Anm.: PhaI = Phantomkurs I (inkonsistente, komplexe Aufgaben); n = 31

1. Codierung Noten: 1 = ungenügend, 6 = sehr gut

Die in Tabelle 5.18 komplementär berechneten Koeffizienten der robusten Regressionsanalyse erlauben eine zusätzliche Überprüfung der Regressionsschätzungen in den PLS-Modellen (vgl. Unterkapitel 4.3.4). Die Schätzungen der Pfadkoeffizienten im TPK-Strukturmodell fallen vergleichbar in ihrer Stärke zu den Koeffizienten der robusten Regressionsanalyse aus, abgesehen vom Prädiktor instrumentelles, manuelles Geschick, dessen Pfadstärke  $\beta_{PLS} = 0.519$  deutlich höher als die des Koeffizienten der robusten Regressionsanalyse mit  $\beta^{MM} = 0.290$  ausfällt. Sowohl die Vorzeichen als auch die zur Signifikanzbeurteilungen herangezogenen  $t$ -Werte fallen für den Prädiktor räumliches Vorstellungsvermögen in beiden Methoden identisch in der Interpretation aus. Die Regressionskoeffizienten der Prädiktoren Schulabschlussnote und naturwissenschaftliches Verständnis im PLS-Strukturmodell TPK nehmen hingegen negative Vorzeichen an, im Modell der robusten Regressionsanalyse jedoch positive. Die Diskrepanz ist für den Prädiktor naturwissenschaftliches Verständnis als akzeptabel einzustufen, da der Koeffizient keinen Zusammenhang zur Leistung in konsistenten, einfachen Aufgaben beim Erwerb von zahnmedizinischen Fertigkeiten anzeigt.

Tabelle 5.18: Regressionskoeffizienten der robusten Regressionsanalyse

Prädiktor	$b^{MM}$ (S.E.)	t-Wert	$\beta^{MM}$	PLS: $\beta$ (t-Wert)
<b>Kriterium konsistente, einfache Aufgaben (TPK)</b>				
instrumentelles, manuelles Geschick	0.009 (0.004)	2.058	0.290	0.519 (5.08)
räumliches Vorstellungsvermögen	-0.001 (0.003)	-0.514	-0.055	-0.080 (0.72)
Schulabschlussnote	0.022 (0.018)	1.217	0.172	-0.163 (1.41)
naturwissenschaftliches Verständnis	0.001 (0.005)	0.259	0.031	-0.017 (0.18)
<b>Kriterium inkonsistente, komplexe Aufgaben (PhaI)</b>				
instrumentelles, manuelles Geschick	0.012 (0.005)	2.394	0.427	0.370 (3.92)
räumliches Vorstellungsvermögen	0.006 (0.003)	2.370	0.420	0.362 (6.29)
Schulabschlussnote	-0.012 (0.017)	-0.712	-0.124	-0.036 (0.44)
naturwissenschaftliches Verständnis	-0.004 (0.005)	-0.812	-0.153	-0.041 (0.62)

Anm.: Methode = MM-Schätzer;  $b$  = Regressionskoeffizient;  $\beta$  = standardisierter Regressionskoeffizient; S.E. = Standardfehler

Modell (TPK) für  $b$ : Residuum-Standardfehler: 0.2327,  $df = 33$

Modell (PhaI) für  $b$ : Residuum-Standardfehler: 0.1975,  $df = 26$

Codierung Noten: Note 1 = ungenügend, Note 6 = sehr gut

Das gegenüber der PLS-Methode im robusten Regressionsmodell invers ausfallende Vorzeichen des Prädiktors Schulabschlussnote stellt aber keine Beeinträchtigung der Hypothesenprüfung dar, sodass diese Diskrepanz zwischen beiden Methoden zu tolerieren ist. Die Vorzeichen und Signifikanzen der PLS-Pfadkoeffizienten fallen im PhaI-Modell für alle vier Prädiktoren vergleichbar zu den Schätzungen der robusten Regressionsanalyse in Tabelle 5.18 aus. Unterschiede finden sich in der Stärke der Beta-Koeffizienten, die im Modell der robusten Regressionsanalyse höhere Werte annehmen, aber in ihrer Klassifizierung keine Veränderungen zeigen, d.h., moderat bzw. schwach bleiben. Zusammenfassend betrachtet deuten die Ergebnisse der robusten

Regressionsanalyse auf keine relevanten, die Überprüfung der Hypothesen beeinträchtigenden Verzerrungen in den PLS-Parameterschätzungen durch extreme Wertekonstellationen in den kleinen Modellstichproben hin.

### 5.3.2 Hypothesenprüfung

Zur Überprüfungen der postulierten Zusammenhänge zwischen Prädiktoren und Kriterien werden die in den Tabellen 5.14 und 5.15 berichteten Pfadkoeffizienten  $\beta$  aus den PLS-Prädiktionsmodellen herangezogen, um die Richtung, Signifikanz und Einflussstärke der exogenen latenten Variablen auf die endogene latente Variable zu interpretieren. Die mittels Bootstrapping-Verfahren berechneten  $t$ -Werte ermöglichen die Beurteilung der Signifikanz der Pfadkoeffizienten in den PLS-Modellen ( $t > 1.645$ ,  $df_n = 5000$ , einseitiger Vertrauensbereich,  $p < .05$ ).

Entsprechend der in Unterkapitel 4.3.3.4 beschriebenen Vorgehensweise finden sich in den Tabellen 5.19 und 5.20 die inkrementellen Validitäten der Prädiktoren räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick. Hierfür werden außerdem die in den Tabellen 5.14 und 5.15 berichteten Effektstärken  $f^2$  herangezogen, um die Stärke des substanziellen Einflusses der beiden Fähigkeitsarten auf die endogenen Variablen zu beurteilen, die nach Mathieson et al. (2001) auch zur Interpretation der Stärke inkrementeller Validitäten in PLS-Modellen benutzt werden können.

Die für die Prädiktoren Schulabschlussnote und naturwissenschaftliches Verständnis postulierten Annahmen über die Beziehungen zu den Leistungen der Aufgabentypen konsistent, einfach und inkonsistent, komplex gehen von keinen Zusammenhängen aus, sodass die Signifikanztests nicht einseitig auf dem Niveau  $p < .05$  erfolgen, sondern zweiseitig mit einem Signifikanzniveau von  $\alpha = 0.2$  ( $t > 1.281$ ,  $df_n = 5000$ ). Die Erhöhung der Irrtumswahrscheinlichkeit  $\alpha$  soll die Wahrscheinlichkeit eines  $\beta$ -Fehlers verringern, da es schwerwiegender wäre keinen Zusammenhang anzunehmen, obwohl ein Zusammenhang besteht. Die Effektstärken  $f^2$  werden kontrollierend zur Hypothesenprüfung einbezogen.

### 5.3.2.1 Konsistente, einfache Aufgaben

Gemäß der in Hypothese 3 formulierten Erwartung soll die kognitive Fähigkeit räumliches Vorstellungsvermögen zur Leistung konsistenter, einfacher Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten eine schwach positive, signifikante Beziehung besitzen. Der Pfadkoeffizient  $\beta$  des Prädiktors räumliches Vorstellungsvermögen nimmt zur endogenen Variable Fertigkeitserwerb im TPK-Prädiktionsmodell eine Ausprägung von -0.080 an, sodass die Voraussetzung  $\beta \geq \pm 0.2$  für eine schwache Beziehung nicht erfüllt ist und kein Zusammenhang besteht. Das Vorzeichen des Pfadkoeffizienten  $\beta$  zeigt zudem eine negative Richtung der Beziehung zwischen Prädiktor und Kriterium an, die entgegen der Erwartung eines positiv gerichteten Zusammenhangs ausfällt. Der  $t$ -Wert des Pfadkoeffizienten  $\beta$  zeigt mit 0.723 keine Signifikanz an. Aufgrund der Ergebnisse muss die Hypothese 3 abgelehnt werden. Daher ist zu vermuten, dass auch der in Hypothese 7 postulierte inkrementelle Beitrag des Prädiktors zur Varianzaufklärung des Erwerbs von Fertigkeiten im TPK-Laborkurs geringer als angenommen ausfallen könnte.

In Hypothese 4 wird ein signifikanter, positiv moderater Zusammenhang zwischen dem instrumentellen, manuellen Geschick und der Leistung konsistenter, einfacher praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten erwartet. Die exogene Variable instrumentelles, manuelles Geschick besitzt im PLS-Prädiktionsmodell eine positiv moderate Beziehung zur endogenen Variable Fertigkeitserwerb im Laborkurs TPK, da der Pfadkoeffizient  $\beta$  mit 0.519 über der Schwelle von 0.4 und unter 0.6 liegt. Der  $t$ -Wert von 5.082 zeigt eindeutig Signifikanz für die Beziehung auf dem einseitigen Niveau  $p < .001$  an ( $df_n = 5000$ ,  $t > 3.091$ ). Somit ist Hypothese 4 bestätigt. Daraus folgend ist es sehr wahrscheinlich, dass die in Hypothese 8 postulierte Erwartung über die inkrementelle Validität des Prädiktors instrumentelles, manuelles Geschick sich als zutreffend erweisen wird.

In Hypothese 5 wird kein Zusammenhang zwischen dem Prädiktor Schulabschlussnote und der Leistung konsistenter, einfacher Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten erwartet. Der die Beziehung zwischen Prädiktor und dem Fertigkeitserwerb im TPK-Prädiktionsmodell reflektierende Pfadkoeffizient  $\beta$  erfüllt mit einer Ausprägung von -0.163 knapp die Anforderung von  $\beta \geq \pm 0.2$ , um als schwach klassifiziert zu werden und somit einen schwachen, negativen Zusammenhang zwischen

beiden Variablen anzeigt, was die Effektstärke  $f^2$  in Höhe von 0.034 bestätigt. Um die Wahrscheinlichkeit eines  $\beta$ -Fehlers zu verringern, wurde das zweiseitige Signifikanzniveau auf  $\alpha = 0.2$  festgelegt. Der  $t$ -Wert des Pfadkoeffizienten  $\beta$  in Höhe von 2.128 indiziert einen signifikanten Zusammenhang zwischen Prädiktor und Kriterium. Die Ergebnisse führen insgesamt zu einer Ablehnung der fünften Hypothese. Für den Prädiktor naturwissenschaftliches Verständnis wird in Hypothese 6 kein Zusammenhang zur Leistung konsistenter, einfacher Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten angenommen. Der die Beziehung zwischen Prädiktor und Kriterium widerspiegelnde Pfadkoeffizient  $\beta$  nimmt einen Wert von -0.017 an ( $f^2 = 0.000$ ), der keinen Zusammenhang anzeigt und mit einem  $t$ -Wert von 0.184 zudem nicht signifikant ausfällt. Hypothese 6 kann aufgrund der Ergebnisse bestätigt werden.

Tabelle 5.19: Inkrementelle Validitäten im TPK-Prädiktionsmodell

Prädiktor (exogene Variable)	- R <sup>2</sup> - Prädiktor im PLS- Modell eingeschlossen	- R <sup>2</sup> - Prädiktor im PLS- Modell ausgeschlossen	$\Delta R^2$
instrumentelles, manuelles Geschick	0.322	0.151	0.171
räumliches Vorstellungsvermögen		0.319	0.030

Anm.:  $\Delta R^2$  = inkrementelle Validität

Die in Hypothese 7 postulierte Verbesserung der Vorhersage der Leistung konsistenter, einfacher praktischer Aufgaben beim Fertigkeitserwerb soll für das räumliche Vorstellungsvermögen über die Prädiktoren instrumentelles, manuelles Geschick, Schulabschlussnote und naturwissenschaftliches Verständnis hinaus nur schwach ausfallen. Die Entnahme des Prädiktors räumliches Vorstellungsvermögen aus dem TPK-Prädiktionsmodell führt zu einer sehr geringen Differenz im Bestimmtheitsmaß  $R^2$ , sodass dessen inkrementelle Validität drei Prozent beträgt (s. Tabelle 5.19). Die Effektstärke  $f^2$  des Prädiktors räumliches Vorstellungsvermögen auf die endogene Variable Fertigkeitserwerb im TPK-Prädiktionsmodell beträgt 0.004, deren Ausprägung nicht der Forderung von  $f^2 \geq 0.02$  entspricht (Chin, 1998b), um eine schwache inkrementelle Validität anzuzeigen. Hypothese 7 ist somit zurückzuweisen.

Es wird in Hypothese 8 für die Vorhersage der Leistung konsistenter, einfacher praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten erwartet, da das instrumentelle, manuelle Geschick eine moderate inkrementelle Validität über die Prädiktoren räumliches Vorstellungsvermögen, Schulabschlussnote und naturwissenschaftliches Verständnis besitzt. Die in Tabelle 5.19 aufgeführte inkrementelle Validität des instrumentellen, manuellen Geschicks beträgt 17.1 Prozent,

deren Effektstärke  $f^2$  mit 0.252 als moderat einzustufen ist. Hypothese 8 kann demzufolge als bestätigt angesehen werden.

### 5.3.2.2 Inkonsistente, komplexe Aufgaben

Zur Leistung inkonsistenter, komplexer praktischer Aufgaben beim Erwerb von zahnmedizinischen Fertigkeiten ist in Hypothese 9 sowohl für die kognitive Fähigkeit räumliches Vorstellungsvermögen als auch in Hypothese 10 für die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick ein signifikanter, positiv moderater Zusammenhang postuliert. Beide Fähigkeitsarten verfügen über annähernd identisch starke Pfadkoeffizienten  $\beta$  mit 0.362 ( $t = 6.29$ ) für das räumliche Vorstellungsvermögen und  $\beta = 0.370$  ( $t = 3.92$ ) für das instrumentelle, manuelle Geschick (s. Tabelle 5.15), die einseitig auf dem Niveau  $p < .001$  signifikant ausfallen, aber nur knapp die Auflage von  $\beta \geq \pm 0.4$  erfüllen, um als positiv moderat eingestuft zu werden. Beide Hypothesen können aufgrund der Ergebnisse bestätigt werden. Weiterführend ist aufgrund der Resultate zu erwarten, dass auch die in den Hypothesen 13 und 14 postulierten inkrementellen Validitäten für beide Fähigkeiten wie erwartet ausfallen.

Zwischen der Schulabschlussnote und der Leistung inkonsistenter, komplexer praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten ist in Hypothese 11 kein Zusammenhang postuliert. Im PhaI-Prädiktionsmodell nimmt der Pfadkoeffizient  $\beta$  des Prädiktors Schulabschlussnote zur endogenen Variable Fertigkeitserwerb eine Ausprägung von -0.036 an ( $f^2 = 0.003$ ), der keinen Zusammenhang zwischen beiden latenten Variablen anzeigt und mit einem  $t$ -Wert von 0.44 auf dem zweiseitigen Niveau  $p < .2$  nicht signifikant ausfällt. Demzufolge ist Hypothese 11 bestätigt.

In Hypothese 12 wird erwartet, dass kein Zusammenhang zwischen dem naturwissenschaftlichem Verständnis und der Leistung inkonsistenter, komplexer praktischer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten besteht. Der Pfadkoeffizienten nimmt eine Ausprägung von  $\beta = -0.041$  an, sodass kein Zusammenhang zwischen Prädiktor und Kriterium besteht, was sich in der Effektstärke  $f^2$  mit 0.001 widerspiegelt. Das zur Signifikanzprüfung eingesetzte Bootstrapping-Verfahren ergab einen  $t$ -Wert von 0.62 für den Pfadkoeffizienten  $\beta$ , der unterhalb des kritischen Werts von  $t = 1.281$  ( $df_n = 5000$ ) für den zweiseitigen Signifikanztest auf dem Niveau von  $\alpha = 0.2$  liegt und damit nicht signifikant ist. Hypothese 12 kann aufgrund dieser Untersuchungsergebnisse bestätigt werden.

Tabelle 5.20: Inkrementelle Validitäten im PhaI-Prädiktionsmodell

Prädiktor (exogene Variable)	- R <sup>2</sup> - Prädiktor im PLS- Modell eingeschlossen	- R <sup>2</sup> - Prädiktor im PLS- Modell ausgeschlossen	ΔR <sup>2</sup>
instrumentelles, manuelles Geschick	0.356	0.248	0.108
räumliches Vorstellungsvermögen		0.246	0.106

Anm.: ΔR<sup>2</sup> = inkrementelle Validität

Die in Hypothese 13 für das räumliche Vorstellungsvermögen und in Hypothese 14 für das instrumentelle, manuelle Geschick angenommene inkrementelle Validität soll moderat ausfallen. Dafür wurden beide Fähigkeitsarten einzeln dem PLS-Modell zur Vorhersage der Leistung inkonsistenter, komplexer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten entnommen, in das ebenfalls die Prädiktoren Schulabschlussnote und naturwissenschaftliches Verständnis eingeschlossen waren, die anschließend neu berechnet wurden (vgl. Unterkapitel 4.3.3.4). Die Ergebnisse sind in Tabelle 5.20 berichtet, welche für die kognitive Fähigkeit räumliches Vorstellungsvermögen wie auch für die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick eine fast identische inkrementelle Validität mit 10.8 Prozent und 10.6 Prozent ausweisen, deren Effektstärken  $f^2$  in Höhe von 0.171 und 0.168 eine moderate Ausprägung annehmen. Die Hypothesen 13 und 14 können demzufolge angenommen werden.

### 5.3.2.3 Zusammenfassung

In Tabelle 5.21 sind die inhaltlichen Hypothesen in Kurzform sowie die zur Prüfung verwendeten Maße und Ergebnisse berichtet. Ebenfalls werden die Entscheidungen aufgeführt, ob die Hypothesen bestätigt, abgelehnt oder teilweise bestätigt bzw. abgelehnt wurden. Die erste und zweite Hypothese mussten teilweise abgelehnt werden, da keine hohe Reliabilität und diskriminante Validität für das räumliche Vorstellungsvermögen bestätigt werden konnten. Während sich für das instrumentelle, manuelle Geschick in Hypothese 4 ein positiv moderater und signifikanter Einfluss auf die vorklinischen Laborkursleistungen konsistenter, einfacher Aufgaben beim Erwerb zahnmedizinischer Dentaltechniken zeigte, konnte für das räumliche Vorstellungsvermögen zu diesen kein schwacher Zusammenhang nachgewiesen werden, weshalb Hypothese 3 zurückzuweisen war.

Dem folgend fand sich für das räumliche Vorstellungsvermögen auch keine schwache inkrementelle Validität über die Prädiktoren instrumentelles, manuelles Geschick, Schulabschlussnote und naturwissenschaftliches Verständnis, sodass Hypothese 7 ebenfalls zurückgewiesen werden musste. Hingegen konnte für das instrumentelle,

manuelle Geschick eine moderate inkrementelle Validität aufgezeigt werden und Hypothese 8 bestätigt wurde. Entgegen der in Hypothese 5 postulierten Erwartung ergab sich zwischen der Schulabschlussnote und der Leistung konsistenter, einfacher Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten ein schwach negativer und signifikanter Zusammenhang, der sich jedoch nicht in Hypothese 6 für den Prädiktor naturwissenschaftliches Verständnis bestätigte.

Tabelle 5.21: Resultate der Hypothesenprüfung

Hypothese – aggregierte Inhalte		Ergebnisse	Bestätigt
1	Hohe Reliabilität für räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick.	$DEV_{MAN} = 0.78$ $rel(\xi_j)_{MAN} = 0.97$ $DEV_{RV} = 0.27$ $rel(\xi_j)_{RV} = 0.57$	teilweise
2	Hohe diskriminante Validität für räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick.	$FLK_{MAN}$ : ja $FLK_{RV}$ : nein	teilweise
<b>Kriterium I:</b> Leistung konsistenter, einfacher Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten			
3	Signifikanter, schwach positiver Zusammenhang zwischen räumlichem Vorstellungsvermögen und Kriterium I	$\beta = -0.08 (p < .05)$	nein
4	Signifikanter, moderat positiver Zusammenhang zwischen instrumentellem, manuellen Geschick und Kriterium I	$\beta = 0.52 (p < .05)$	ja
5	Kein Zusammenhang Schulabschlussnote und Kriterium I	$\beta = -0.16 (p < .2)$ $f^2 = 0.03$	nein
6	Kein Zusammenhang naturwissenschaftliches Verständnis und Kriterium I	$\beta = -0.01 (n.s.)$ $f^2 = 0.00$	ja
7	räumliches Vorstellungsvermögen: schwache inkrementelle Validität	$\Delta R^2 = 0.03$ $f^2 = 0.00$	nein
8	instrumentelles, manuelles Geschick: moderate inkrementelle Validität	$\Delta R^2 = 0.17$ $f^2 = 0.25$	ja
<b>Kriterium II:</b> Leistung inkonsistenter, komplexer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten			
9	Signifikanter, moderat positiver Zusammenhang zwischen räumlichem Vorstellungsvermögen und Kriterium	$\beta = 0.36 (p < .05)$	ja
10	Signifikanter, moderat positiver Zusammenhang zwischen instrumentellem, manuellen Geschick und Kriterium II	$\beta = 0.37 (p < .05)$	ja
11	Kein Zusammenhang zwischen Schulabschlussnote und Kriterium II	$\beta = -0.04 (n.s.)$ $f^2 = 0.00$	ja
12	Kein Zusammenhang naturwissenschaftliches Verständnis und Kriterium II	$\beta = -0.04 (n.s.)$ $f^2 = 0.00$	ja
13	räumlichen Vorstellungsvermögen: moderate inkrementelle Validität	$\Delta R^2 = 0.106$ $f^2 = 0.17$	ja
14	instrumentelles, manuelles Geschick: moderate inkrementelle Validität	$\Delta R^2 = 0.108$ $f^2 = 0.17$	ja

Anm.: FLK = Fornell-Larcker-Kriterium; MAN = instrumentelles, manuelles Geschick; RV = räumliches Vorstellungsvermögen

Die in den Hypothesen 9 und 10 postulierten positiv moderaten, signifikanten Zusammenhänge zwischen der Leistung inkonsistenter, komplexer Aufgaben und der kognitiven Fähigkeit räumliches Vorstellungsvermögen (9) sowie der nichtkognitiven Fähigkeit instrumentelles, manuelles Geschick (10) konnten bestätigt werden, ebenso

die Annahmen über die inkrementellen Validitäten beider Fähigkeitsarten in den Hypothesen 13 und 14. Ebenfalls konnten die Hypothesen zu den Prädiktoren Schulabschlussnote (11) und das naturwissenschaftliche Wissen (12) angenommen werden, die wie hergeleitet, in keinen Zusammenhang zur Leistung inkonsistenter, komplexer Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten standen.

## **6. Diskussion**

Die Ziele der vorliegenden Arbeit sind einleitend im Unterkapitel 6.1 vorgestellt. In den nachfolgenden Unterkapiteln 6.2 und 6.3 werden die zur Überprüfung der Hypothesen herangezogenen Untersuchungsergebnisse diskutiert. Während im Abschnitt 6.4 die daraus resultierenden Erkenntnisse zur Prädiktion vorklinischer Laborkursleistungen in Diskurs zu früheren Studien gestellt werden, finden sich mögliche Einschränkungen der quasiexperimentellen Untersuchungsergebnisse im vorletzten Kapitelabschnitt 6.6. Das Kapitel abschließend, werden im Abschnitt 6.7 Empfehlungen für nachfolgende Studien sowie ein Ausblick über zukünftige Forschungsfragestellungen aufgezeigt.

### **6.1 Ziele der vorliegenden Untersuchung**

Mit Änderung des Hochschulrahmengesetzes in 2004 ist es auch Hochschulen in der Bundesrepublik Deutschland gestattet, für die Zulassung zukünftiger Studierender weitere Kriterien als die Abiturdurchschnittsnote einzubeziehen. Es bestehen kaum Zweifel an der Zuverlässigkeit der Abiturdurchschnittsnote oder auch an der von Naturwissenschaftstests als Auswahlkriterien, um sicherzustellen, dass zukünftige Studierende die Befähigung besitzen, nicht nur erfolgreich, sondern auch zeitlich schnell Studienkurse mit naturwissenschaftlichem Schwerpunkt zu absolvieren. In der Zahnmedizin sind darüber hinaus aber auch die in den vorklinischen und klinischen Laborkursen zu erbringenden Studienleistungen zu berücksichtigen, in denen die Studierenden zahnmedizinische Untersuchungs- und Behandlungsmethoden erlernen. Die in der Literatur oftmals konträren Studienergebnisse zur Prädiktion der Laborkursleistungen systematisch nachzuvollziehen, gestaltet sich schwierig, da die Definition der Beziehungen zwischen den in den Laborkursen zu erlernenden zahnmedizinischen Fertigkeiten und den dafür benötigten Fähigkeiten auf keinem theoretischen Fundament erfolgte.

Aus dieser Diskrepanz wurden in der vorliegenden Arbeit zwei Fragenkomplexe abgeleitet. Zum einen wurde überprüft, ob die Fähigkeiten räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick voneinander separat operationalisiert werden können und zum anderen, ob Ackermans (1988) Theorie des Fertigkeitserwerb geeignet ist, um zu klären, ob beide Fähigkeitsarten einen Einfluss auf die Leistung in den vorklinischen Laborkursen besitzen. Die Klärung der Frage ist nicht nur für die Ökonomie eines Auswahlverfahren in der Zahnmedizin relevant, sondern auch im Hinblick auf die soziale Akzeptanz und Fairness der zur Bewerberselektion

eingesetzten räumlichen und manuellen Tests. Darüber hinaus stellt sich auch im Kontext von Förderprogrammen für Zahnmedizinstudierende die Frage, aufgrund welcher Fähigkeiten bzw. deren defizitären Ausprägungen, Schwierigkeiten bestehen könnten, die Aufgaben in den vorklinischen Laborkursen erfolgreich zu absolvieren. Ohne eine rechtzeitige Förderung der nicht ausreichend entwickelten Fähigkeiten ist für die betroffenen Studierenden damit nicht nur das Erleben von Gefühlen der Unzufriedenheit und Unzulänglichkeit verbunden, sondern führt auch zu längeren Studienzeiten und möglicherweise zu einem sogar ein Studienabbruch (Smith, 1989), obwohl ihre akademischen Studienleistungen ausreichend wären. Das gewählte quasiexperimentelle Untersuchungsdesign der vorliegenden Arbeit wurde einem randomisiert kontrollierten Studiendesign vorgezogen, um die psychometrischen Tests von Bewerbern absolvieren zu lassen, die aufgrund des Ziels, einen Studienplatz in der Zahnmedizin zu erhalten, zu Bestleistungen in der Prüfungssituation eines Hochschulauswahlverfahrens motiviert waren.

## 6.2 Diskriminanz der Fähigkeitsart

Die Annahmen der ersten beiden Hypothesen griffen die Überlegung auf, dass die kontroversen Studienergebnisse zur Prädiktion der vorklinischen Laborkursleistungen in der Operationalisierungsgüte der räumlichen und manuellen Fähigkeiten begründet sein könnten. Die Konstruktvalidierung der beiden Fähigkeitsarten diene somit in der vorliegenden Untersuchung nicht nur zur Beurteilung der Messmodellgüte, sondern gleichzeitig auch zur Hypothesenprüfung. Aufgrund der konsistenteren Schätzer, der Möglichkeit konkurrierende Modelle zu testen und der Verfügbarkeit globaler Gütemaße zur ganzheitlichen Modelleignung, erfolgte für die PLS-Prädiktionsmodelle die Prüfung der Operationalisierungsgüte mittels konfirmatorischer Faktorenanalyse. Hierfür musste jedoch die exogene Variable Schulabschlussnote aufgrund des formativen Single-Item-Messmodells exkludiert werden wie auch die exogene Variable naturwissenschaftliches Verständnis, die zwar über ein reflektives Single-Item-Messmodell verfügt, das aber in der KFA zum Auftreten negativer Varianzen hätte führen können (vgl. Unterkapitel 4.3.2.3).

### 6.2.1 Reliabilität und diskriminante Validität der Fähigkeitsarten

Die Frage, ob die kognitive Fähigkeit räumliches Vorstellungsvermögen und die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick als autonome Konstrukte in einem Modell operationalisierbar sind, konnte nicht eindeutig in der vorliegenden Untersuchung beantwortet werden. Während die in Hypothese 1 postulierte hohe Reliabilität nur für das instrumentelle, manuelle Geschick bestätigt werden konnte, musste die Hypothese im Fall der kognitiven Fähigkeit zurückgewiesen werden. Die durch das Fornell-Larcker-Kriterium angezeigte diskriminante Validität bestätigte sich zwar für das latente Konstrukt instrumentelles, manuelles Geschick, deren  $DEV(\xi_j)$  die quadrierte Korrelation zum räumlichen Vorstellungsvermögen von 0.276 bei weitem überstieg, aber die von Fornell und Larcker 1981 aufgestellte Forderung konnte nicht von der latenten Variable räumliches Vorstellungsvermögen erfüllt werden, sodass auch Hypothese 2 teilweise zurückgewiesen werden musste. In den nachfolgenden Unterkapiteln werden verschiedene Ursachen und mögliche Gründe diskutiert.

## 6.2.2 Operationalisierungsgüte des räumlichen Vorstellungsvermögens

### 6.2.2.1 Statistischer Erklärungsansatz

Als statistische Erklärung für die geringen Reliabilitäten des räumlichen Vorstellungsvermögens sind eindeutig die schwachen Faktorladungen  $\lambda_{ij}^2$  und die hohen Messfehlervarianzen  $\theta_{ii}$  der zur Operationalisierung des hypothetischen Konstrukts verwendeten Indikatoren anzuführen (s. Tabelle 5.05). Die zur Messung des räumlichen Vorstellungsvermögens verwendete Skala technische Begabung des LPS (Horn, 1983) beinhaltet vier Untertests, von denen jedoch nur der LPS UT 9 als Indikator im KFA-Modell zufriedenstellende Gütemaße erreichte. Demzufolge fiel nicht nur die Durchschnittlich Extrahierte Varianz der kognitiven Fähigkeit gering aus, sondern auch die Faktorreliabilität  $rel(x_i)$  mit 0.572. Die interne Konsistenz der Skala technisches Verständnis betrug  $\alpha = 0.58$  und ist nur noch als bedingt akzeptabel einzustufen.

Die Itemschwierigkeiten  $p_m$  der Indikatoren bzw. Untertests der LPS-Skala technische Begabung in Tabelle 4.04 deuteten mit Werten zwischen 0.4 und 0.8 nicht auf zu einfache oder zu schwere Aufgaben hin, sodass in der Schwierigkeit der Aufgaben keine Ursache für die insgesamt bedingt akzeptable Messungsgüte der räumlichen Fähigkeit zu vermuten ist. Aus rein inhaltlichem Blickwinkel ist es außerdem nicht nachzuvollziehen, weshalb für den räumliche Orientierung messenden Untertest 9 des LPS als Indikator eine zufriedenstellende Ladung zur latenten Variable bestand, aber nicht für die übrigen Untertests des LPS, die räumliche Beziehungen, räumliche Vorstellungsfähigkeit und die Flexibilität der Gestaltbildung erfassen, welche nachweislich Faktoren des räumlichen Vorstellungsvermögens darstellen (vgl. Linn & Peterson, 1985; Quaiser-Pohl, 2001; Thurstone & Thurstone, 1941).

### 6.2.2.2 Psychometrische Güte des LPS

Bezieht man in die Überlegungen weiterer, möglicher Ursachen für die nur bedingt akzeptable Operationalisierungsgüte des räumlichen Vorstellungsvermögens den LPS als Testinstrument ein, der in der 2. Auflage von 1983 vorlag, kann nicht ausgeschlossen werden, dass die Testnormierung und die Aufgaben der Skala technische Begabung seitdem an psychometrischer Qualität eingebüßt haben. Bedingt durch theoretische und methodische Weiterentwicklungen in der psychologischen Diagnostik, erscheint die dem LPS zugrundeliegende Primärfaktoretheorie von Thurstone und Thurstone zur Intelligenzdiagnostik aus dem Jahr 1941 und das darauf

zurückzuführende, von Horn operationalisierte einfaktorielles Messdesign der Skala technische Begabung fraglich (Kreuzpointner, 2010).

Die schwachen Faktorladungen als auch die hohen Messfehlervarianzen der Indikatoren im Messmodell der latenten Variable räumliches Vorstellungsvermögen (s. Tabelle 5.05 in Unterkapitel 5.2.3) könnten sowohl durch die veralteten Testnormen des LPS als auch durch die Aufgabenarten der Untertests 7 bis 10 verursacht worden sein, die zwar die Fähigkeitsausprägungen in der Normierungsstichprobe des LPS von 1983 reflektieren, aber nicht den der heutigen Bevölkerung entsprechen müssen. In 2013 veröffentlichten Kreuzpointner et al. den LPS-2, eine revidierte Fassung der 2. Auflage des LPS von 1983, die neben einer aktualisierten Normierung auch eine Neukonzeption der Faktorenstruktur umfasste (vgl. Unterkapitel 4.1.2.5). In Übereinstimmung mit Colom et al. (2002), die sich auf Basis ihrer Studienergebnisse gegen eine Operationalisierung des räumlichen Vorstellungsvermögens als Mehrfaktorenmodell aussprachen, konnten Kreuzpointner et al. auch im LPS-2 ein einfaktorielles Messmodell zur Erfassung der kognitiven Fähigkeit räumliches Vorstellungsvermögen bestätigen. Lediglich der Untertest 8 zum räumlichen Vorstellen wurde der Skala technische Begabung entnommen, ohne dass die verbleibenden Aufgabentypen aber eine grundlegende Überarbeitung erfuhren. Demzufolge sind die Ursachen für die schwache Güte des räumlichen Konstrukts in der vorliegenden Arbeit weniger im Messdesign und in der Aufgabenbeschaffenheit des LPS zu finden, sondern eher in der veralteten Testnormierung von 1983.

Ebenfalls könnte in der Beschaffenheit der Untersuchungsstichprobe in der vorliegenden Untersuchung ein Grund für die bedingt akzeptable Konstruktgüte der kognitiven Fähigkeit räumliches Vorstellungsvermögen vermutet werden. Aufgrund des quasiexperimentellen Untersuchungsdesigns erfolgte in der vorliegenden Arbeit die Auswahl der Stichprobe nicht nach dem Zufallsprinzip. Darüber hinaus erfüllt die Untersuchungskohorte eher die Mindestanforderungen des ML-Schätzverfahrens mit 129 Fällen (vgl. Unterkapitel 4.3.2.3). Aufgrund der Einladung zum Auswahlverfahren auf Basis des Abiturnotendurchschnitts weist die Stichprobe zudem eine selektive Charakteristik auf (s. Tabelle 5.01 im Unterkapitel 5.1.1), für welche die Fähigkeitsmessung durch die veraltete Testnormierung des LPS verzerrt wurde.

### 6.2.2.3 Zusammenfassende Bewertung der Messungsgüte

Die Messqualität des räumlichen Vorstellungsvermögens ist in der vorliegenden Arbeit als bedingt akzeptabel einzustufen, da das Unterschreiten der geforderten Akzeptanzniveaus in den lokalen Gütemaßen der konfirmatorischen Faktorenanalyse noch tolerierbar ist (Rossiter, 2002). Die Ursachen sind sehr wahrscheinlich auf die veraltete Normierung des LPS von 1983 und die Beschaffenheit der quasiexperimentellen Untersuchungsstichprobe zurückzuführen. Zudem wurden durch die Gütekriterien der KFA keine eindeutigen Fehlspezifikationen im Messmodell der räumlichen Fähigkeit angezeigt, d.h. die Indikatorreliabilität  $rel(x_i)$  nimmt keine Wertausprägung größer eins an und die Kriterien *CFI* sowie *NFI* keinen Wert kleiner 0.90 (Bentler, 1990).

### 6.2.3 Operationalisierungsgüte des instrumentellen, manuellen Geschicks

#### 6.2.3.1 Gütekriterien der Klassischen Testtheorie

Die nach Kriterien der Klassischen Testtheorie ermittelten Reliabilitäten des Drahtbiegetests HAM-Man zur Operationalisierung des instrumentellen, manuellen Geschicks sind insgesamt als sehr gut zu bezeichnen, da die Interrater-Reliabilität für die zwei Juroren und die interne Konsistenz der einzelnen Drahtfiguren exzellent ausfielen. Im Hinblick auf die hohe korrelative Übereinstimmung der Bewertungskriterien Deckungsgleichheit und Qualität der Biegung im HAM-Man (s. Unterkapitel 5.2.1) sind aber die internen Konsistenzen der einzelnen Figuren kritisch zu betrachten. Lediglich die Retest-Reliabilität weicht von dem ansonsten exzellenten Reliabilitätsniveau mit einer Übereinstimmung in Höhe von  $ICC = 0.70$  ab, die jedoch als akzeptabel einzustufen ist (vgl. Wirtz & Caspar, 2002). Hierfür könnte sehr wahrscheinlich ein „testing effect“ (Gates, 1917) als Grund in Frage kommen, da die 38 Erstsemesterstudierenden aufgrund ihrer Vorerfahrungen im Auswahlverfahren über bessere Bearbeitungsstrategien und -methoden im HAM-Man verfügten, die wiederherum zu einem höheren Testergebnis führten und damit zu einer Verringerung der Übereinstimmung beider Testwerte (s. dazu auch Unterkapitel 6.4.5).

#### 6.2.3.2 Bi-Faktor-Modell der manuellen Fähigkeit

Das Bi-Faktor-Modell des Drahtbiegetests HAM-Man erwies sich als sehr gut zur Operationalisierung des instrumentellen, manuellen Geschicks geeignet, wie die lokalen Gütekriterien der Indikatoren und Faktoren als Ergebnisse der konfirmatorischen

Faktorenanalyse aufzeigten (s. Tabellen 5.05 und 5.06). Die Itemschwierigkeiten  $p_m$  der Indikatoren in Tabelle 5.04 indizierten eine Verringerung des Anteils an Teilnehmern, die über die Figuren Dreieck, Dritte Dimension und Rundbögen sehr gute Leistungen in den Kriterien Deckungsgleichheit, Qualität der Biegung und planes Aufliegen erreichten. Während noch ungefähr 40 Prozent der Teilnehmer beim Biegen der Drahtfigur Dreieck sehr gute Leistungen erzielten, konnte dies nur von circa zehn Prozent bei der Figur Dritte Dimension und von weniger als fünf Prozent bei der Figur Rundbögen erreicht werden. Die unterschiedliche Schwierigkeit über die einzelnen Drahtfiguren kann aber nicht ausschließlich auf die zunehmende Komplexität der zu biegenden Formen zurückgeführt werden (vgl. Unterkapitel 4.1.1.1), da es im Drahtbiegetest HAM-Man freigestellt ist, ob und wie viel Zeit für Übungen mit den drei Probedrähten verwendet wird, in welcher Reihenfolge die Figuren bearbeitet werden und ob im Sinne des „speed-accuracy tradeoff“ die Aufgaben eher schnell, aber auf Kosten der Genauigkeit oder eher genau und zu Lasten der Schnelligkeit von den Probanden bearbeitet werden (Wickelgren, 1977).

Das Bi-Faktor-Modell bewährte sich auf statistischer Ebene eindeutig gegenüber einem einfaktoriellen Modell zur Operationalisierung des instrumentellen, manuellen Geschicks, wie die für den Modellvergleich herangezogenen globalen Gütemaße der konfirmatorischen Faktorenanalyse aufzeigten (s. Unterkapitel 5.2.2). Der Vergleich zur Messstruktur eines Drei-Faktoren-Modells konnte hingegen nur auf inhaltlicher Ebene entschieden werden. Die Vorteile das instrumentelle, manuelle Geschick über einen allgemeinen Faktor zu erfassen und zusätzlich den Einfluss der einzelnen Drahtfiguren als eigenständige Faktoren auf die ihnen zugehörige Gruppe von Indikatoren abzubilden, ließen die Entscheidung zugunsten des Bi-Faktor-Modells ausfallen. Auffällig waren zudem die hohen positiven und signifikanten Korrelationen ( $r > 0.85$ ) zwischen den Indikatoren Deckungsgleichheit und Qualität der Biegung in allen drei Drahtfiguren sowie zwischen den Indikatoren Deckungsgleichheit und planes Aufliegen der Drahtfigur Dritte Dimension (s. Unterkapitel 5.2.1), die auf Multikollinearität hindeuteten.

Eine starke Verzerrung der Parameterschätzungen im Verfahren der konfirmatorischen Faktorenanalyse war jedoch nicht zu vermuten, da die Reliabilitäten  $rel(x_i)$  der betroffenen Indikatoren mit  $rel(x_i) \geq 0.6$  als zuverlässig eingestuft werden konnten, die Grewal et al. (2004) zufolge eine der wichtigsten Determinanten zur Ermittlung der Koeffizienten in der konfirmatorischen Faktorenanalyse darstellen. Als Ursache für das

Auftreten von Multikollinearität zwischen den Indikatoren ist die Gleichartigkeit der Bewertungskriterien Deckungsgleichheit und Qualität der Biegung zu vermuten. Die Übereinstimmung zwischen der gebogenen Drahtfigur mit der Zeichnungsvorlage scheint nicht nur die Bewertung der Deckungsgleichheit abzubilden, sondern zugleich auch die Qualität der Biegung einzubeziehen. Wenig akkurate Biegungen der Drähte führen zwangsläufig zu Abweichungen mit der Vorlage, die sich entsprechend in den Beurteilungen beider Kriterien widerspiegeln. Ein möglicher Lösungsvorschlag wird in Unterkapitel 6.4.3.1 diskutiert.

### **6.2.3.3 Konkurrente Validität**

Die zur Validierung des Drahtbiegetests HAM-Man eingesetzte Drahtbiegeprobe von Lienert (1967) wies insgesamt eine mittlere positive und signifikante Korrelation zwischen beiden manuellen Testverfahren auf ( $r = 0.33$ ,  $p < .001$ ). Die Stärke des korrelativen Zusammenhangs ist als Indiz für eine eher schwache Kriteriumsvalidität innerhalb der Konstruktvalidierung zu interpretieren, bestätigt aber im vorliegenden Fall die spezifische, durch den HAM-Man gemessene Art des manuellen Geschicks. Während im HAM-Man harte Drähte mittels einer Spitzzange gebogen werden müssen, ist es in der Lienertschen Drahtbiegeprobe erforderlich, einen weichen Draht mit den Fingern zu biegen (vgl. Unterkapitel 2.3.1.3). Daher sollte in der DBP eher Fingergeschicklichkeit als spezifische Form des manuellen Geschicks erfasst werden, hingegen instrumentelles, manuelles Geschick im HAM-Man, das manuelles Geschick im manipulativen Umgang mit Handinstrumenten bezeichnet. Die schwache Überschneidung beider Testverfahren scheint demzufolge mehr in der Grundform des manuellen Geschicks zu bestehen. Zudem bestehen in der Durchführung und Auswertung beider Tests elementare Unterschiede, wodurch der schwach positive Zusammenhang mitunter bedingt sein könnte. So gibt es in Lienerts Drahtbiegeprobe von 1967 kein Zeitlimit, es werden keine Probedrähte für das Biegen der Figur zur Verfügung gestellt und Fehler können von den Probanden aufgrund des weichen Drahts einfacher korrigiert werden. Außerdem kann das im Verhältnis sieben zu fünf vergrößerte Figurenmuster in der DBP nicht als maßstabsgetreue Vorlage zum Biegen des Drahtes verwendet werden und die Auswertung erfolgt in Lienerts Test über insgesamt 29 Kriterien, indem einzelne Streckenabschnitte der gebogenen Figur mittels Schablone auf Abweichungen im Millimeterbereich kontrolliert werden.

#### 6.2.4 Beziehung zwischen räumlicher und manueller Fähigkeit

Die moderate Stärke des linearen Zusammenhangs zwischen den latenten Variablen beider Fähigkeitsarten fiel im konfirmatorischen Faktorenmodell mit  $r = 0.525$  höher als erwartet aus, da in der Literatur überwiegend schwach positive Korrelationen berichtet wurden (z.B. Boyle & Santelli, 1986; Gansky et al., 2004). Der im Drahtbiegetest HAM-Man erfasste manipulative Umgang mit Instrumenten stellt zwar eine spezifische Form des manuellen Geschicks dar, jedoch kann darin keine Erklärung für die moderate Stärke der Beziehung zum räumlichen Vorstellungsvermögen gesehen werden. Die in den zitierten Studien eingesetzten psychometrischen Testverfahren erforderten ebenfalls den Einsatz eines Instruments, sei es ein Schraubendreher oder ein Schnitzmesser, um Objekte zu manipulieren, d.h., zu bearbeiten, wie es beispielsweise in dem von Gansky et al. (2004) verwendeten „Block Carving“-Test der Fall ist. Eher sind hierfür Erklärungen in der Stichprobenbeschaffenheit und dem statistischen Analyseverfahren zu vermuten.

In der Bewerberkohorte der vorliegenden Untersuchung sollten stärker variierende Merkmalsausprägungen der manuellen, aber vor allem räumlichen Fähigkeit auftreten, da im Vergleich zu den zitierten Studien keine Vorselektion der Probanden durch die im Studierfähigkeitstest DAT erreichten Ergebnisse erfolgte, zu denen auch die Leistungen im Subtest PAT zum räumlichen Vorstellungsvermögen zählten. Demzufolge sollten in der vorliegenden Untersuchung keine verringerten Effektstärken in der Beziehung zwischen räumlichen Vorstellungsvermögen und instrumentellem, manuellen Geschick auftreten, die durch „restrictions in range“ bedingt wären. Die konfirmatorische Faktorenanalyse erlaubt in der vorliegenden Untersuchung als statistisches Verfahren der zweiten Generation eine zuverlässigere Prüfung der Beziehung zwischen den Fähigkeitsarten als die Korrelationsanalyse der zitierten Studien (vgl. Unterkapitel 4.3.2) allein aufgrund der Modellierung latenter Variablen über Indikatoren, dem ML-Parameterschätzverfahren und der Berücksichtigung von Messfehlern. Gestützt wird diese Überlegung durch die bivariaten Korrelationen in der vorliegenden Arbeit, die für das räumliche Vorstellungsvermögen zur Drahtbiegeprobe von Lienert (1967) als auch zum HAM-Man schwach positiv in einer Höhe von 0.29 ( $p < .01$ ) bzw. 0.28 ( $p < .01$ ) (s. Tabelle A02 im Anhang) und somit vergleichbar zu den zitierten Studien ausfielen. Der Unterschied in der Beziehungsstärke zwischen den Fähigkeitsarten ist in der vorliegenden Untersuchung demzufolge auf das statistische Analyseverfahren zurückzuführen.

Der Korrelationsanalyse gegenüber sind jedoch die Ergebnisse der konfirmatorischen Faktorenanalyse als statistisches Verfahren der zweiten Generation zu präferieren, die einen positiv moderaten, linearen Zusammenhang zwischen dem instrumentellen, manuellen Geschick und dem räumlichen Vorstellungsvermögen anzeigten. Die Höhe der Beziehung zwischen beiden Fähigkeitsarten lässt darauf schließen, dass Bewerber, die über ein gut oder gering ausgeprägtes instrumentelles, manuelles Geschick verfügen, auch ein entsprechendes, d.h., gutes oder geringes räumliches Vorstellungsvermögen besitzen. Über mögliche Ursachen oder Zusammenhänge kann in der vorliegenden Arbeit jedoch nur spekuliert werden. Beide Fähigkeiten könnten einander sehr stark bedingen, vor allem im Hinblick auf den manipulativen Umgang mit Instrumenten, da räumliches Vorstellungsvermögen sowie manuelles Geschick zur Koordination der Motorik und damit zur Ausführung zahnmedizinischer Untersuchungs- und Behandlungstätigkeiten gemeinsam benötigt werden.

## **6.3 Vorhersage der Leistung in den vorklinischen Laborkursen**

### **6.3.1 Prädiktion der Leistung konsistenter, einfacher Aufgaben**

Ackermans (1988) Theorie des Fertigkeitserwerbs zufolge erfordert das erstmalige Bearbeiten einer Aufgabe zu Beginn ein hohes Maß an räumlichem Vorstellungsvermögen. Um beispielsweise einen Zahn aus Wachs zu schnitzen oder eine Vollgusskrone herzustellen, ist es zunächst erforderlich, die Aufgabeninstruktionen zu verstehen, wofür die anzufertigen Objekte visualisiert und als mentales Modell vorgestellt werden müssen (vgl. Unterkapitel 4.1.5.1). Im nächsten Schritt gilt es, die benötigten Techniken zu erlernen, um beispielsweise die Zahnformen aus Wachs herausarbeiten zu können. Diese Phase stellt in Ackermans (1988) Modell die kognitive Phase des Fertigkeitserwerbs dar, in der ein hohes Korrelat zwischen der kognitiven Fähigkeit räumliches Vorstellungsvermögen und der Leistung konsistenter, einfacher Aufgaben besteht, welche in diesem Stadium beispielsweise das Herausstellen erster Zahnstrukturen beim Schnitzen eines Wachszahns widerspiegelt.

Mit zunehmender Übung der Tätigkeiten verringert sich aber der Einfluss des räumlichen Vorstellungsvermögens, da für die anzufertigenden Objekte bzw. mit Bezug auf das angeführte Beispiel die Wachszähne als mentale Modelle umgesetzt wurden und nicht mehr die Notwendigkeit besteht, diese fortwährend für die weitere Aufgabenbearbeitung abzurufen, da die Arbeitsstücke jederzeit von allen Seiten frei betrachtet werden können. In der anschließenden, autonomen Phase des Fertigkeitserwerbs wurden die zu erlernenden Techniken weitestgehend automatisiert, sodass die individuellen Leistungsunterschiede zur Fertigstellung der Wachszähne bzw. Werkstücke durch die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick bedingt sein sollten, hingegen aber nicht mehr von der kognitiven Fähigkeit räumliches Vorstellungsvermögen.

#### **6.3.1.1 Einfluss der räumlichen und manuellen Fähigkeit**

Entgegen den Resultaten von Beck et al. (1978), Evans und Dirks (2001), Gray und Deem (2002a), Hegarty et al. (2009) und Kao et al. (1990) ergab in der vorliegenden Untersuchung die Ergebnisauswertung keinen Zusammenhang zwischen der räumlichen Fähigkeit und den Leistungen konsistenter, einfacher Aufgaben beim Erlernen zahnmedizinischer Fertigkeiten, zu denen das Schnitzen und Aufwachsen von Wachszähnen und die Herstellung einer Vollgusskrone sowie Interimsprothese

gehörten, weshalb Hypothese 3 zurückgewiesen werden musste. Obwohl kein schwach positiver Zusammenhang ermittelt werden konnte, stützt das Untersuchungsergebnis die auf Ackermans (1988) Modell des Fertigkeitserwerbs zurückgehenden, inhaltlichen Überlegungen, denen zufolge der Einfluss der kognitiven Fähigkeit räumliches Vorstellungsvermögen beim Erlernen zahnmedizinischer Techniken in der autonomen Phase nicht mehr ausschlaggebend für die Leistungen konsistenter, einfacher Aufgaben ist. Individuelle Leistungsunterschiede in den anzufertigen Werkstücken sind in diesem Stadium des Fertigkeitserwerbs primär durch die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick beeinflusst, was entsprechend in Hypothese 4 der vorliegenden Arbeit postuliert und bestätigt werden konnte.

### **6.3.1.2 Einfluss der komplementären Prädiktoren**

Überraschend gestaltete sich der schwach negative und signifikante Einfluss des komplementären Prädiktors Schulabschlussnote auf die Leistungen konsistenter, einfacher Aufgaben beim Fertigkeitserwerb, da hierfür kein Zusammenhang erwartet wurde. Das damit einhergehende Zurückweisen von Hypothese 5 wirft nicht nur Fragen nach möglichen Ursachen auf, sondern erfordert auch den Einbezug des komplementären Prädiktors naturwissenschaftliches Verständnis, der hingegen wie in Hypothese 6 angenommen, keinen Zusammenhang zum Fertigkeitserwerb im TPK-Laborkurs aufwies. Inhaltlich sind beide Ergebnisse schwierig nachzuvollziehen, da in der Literatur einheitlich argumentiert wird, dass Schulabschlussnoten oder naturwissenschaftliches Verständnis nicht die zur Anfertigung der Dentalarbeiten benötigten manuellen und räumlichen Fähigkeiten reflektieren, sodass kein Zusammenhang existieren sollte (Bellanti et al., 1972; Gray & Deem, 2002a; Kao et al., 1990; Lundergan et al., 2007; Ranney et al., 2005; Walcott et al., 1986; Weinstein & Kiyak, 1981; Wong et al., 1979).

Zur Vermeidung von „reverse coding“-Effekten wurden alle invers kodierten Indikatoren in den PLS-Prädiktionsmodellen, d.h., die Abiturdurchschnittsnote als auch die Aufgabennoten der vorklinischen Laborkurse, umkodiert und somit einheitlich ausgerichtet (vgl. Unterkapitel 4.3.5). Daher ist zusätzlich die negative Richtung des schwachen Zusammenhangs zwischen der Schulabschlussnote und der Leistung konsistenter, einfacher Aufgaben beim Erwerb zahnmedizinischer Fertigkeiten überraschend. Ein klassischer Suppressoreffekt ist als mögliche Ursache eher auszuschließen, da zu keinem anderen Prädiktor im Strukturmodell eine zumindest

schwache lineare Beziehung besteht (s. Tabelle 5.16) und anstatt eine Korrelation nahe null zur Leistung konsistenter, einfacher Aufgaben aufzuweisen, zu dieser über einen Korrelationskoeffizienten von  $r = -0.21$  verfügt. Aufgrund dieser Relationen ist auch nicht das Vorliegen eines negativen oder reziproken Suppressionseffekts anzunehmen (vgl. Maassen & Bakker, 2001).

Die Ergebnisse der komplementär zur PLS-Pfadanalyse berechneten robusten Regression deuten eher auf die Stichprobenbeschaffenheit als wahrscheinlichste Ursache für das Zurückweisen der fünften Hypothese hin. Der MM-Schätzer nimmt als robuster Regressionskoeffizient für die Schulabschlussnote mit  $\beta^{MM} = 0.172$  ( $t = 1.217$ ) kein negatives, sondern ein positives Vorzeichen in der Beziehung zur Leistung konsistenter, einfacher Aufgaben beim Erlernen zahnmedizinischer Fertigkeiten an. Vor allem gegenüber Extremwertkonstellationen in kleinen Stichproben erweist sich der MM-Schätzer als resistent (vgl. Unterkapitel 4.2.2), die in der vorliegenden Untersuchung durch das Zulassungsverfahren zum Studium der Zahnmedizin an der Medizinischen Fakultät der Universität Hamburg bedingt sein könnten. Die Vergabe der 200 Teilnahmeplätze am Auswahlverfahren erfolgte für die 467 Bewerber auf Basis der Abiturdurchschnittsnote, sodass deren Spanne in der Untersuchungsstichprobe mit 1.5 bis 2.3 nur gering ausfiel (s. Tabelle 5.01 in Unterkapitel 5.1.1). Die Zulassungen zum Studium über die zwei unabhängigen Ranglisten könnten einerseits auch dazu geführt haben, dass schlechtere Ergebnisse im Drahtbiegetest HAM-Man aufgrund einer Zulassung über die erste Rangliste, d.h., den Ergebnissen des Naturwissenschaftstests HAM-Nat und der Abiturdurchschnittsnote, kompensiert und somit in der Studierendenauswahl nicht weiter zum Tragen kamen. Andererseits bestünde auch die Möglichkeit, dass sehr gute Testergebnisse im Drahtbiegen zu einer Zulassung über die zweite Rangliste geführt haben könnten, mit denen weniger gute Abiturdurchschnittsnoten im Auswahlverfahren der Zahnmedizin ausgeglichen wurden.

### **6.3.1.3 Inkrementelle Validitäten der räumlichen und manuellen Fähigkeit**

Die Frage nach den inkrementellen Validitäten ist vor allem unter ökonomischen Gesichtspunkten im Studierendenauswahlverfahren interessant und erlaubt die Beurteilung, wie hoch der Gewinn an zusätzlich aufgeklärter Leistungsvarianz unter Berücksichtigung des Durchführungsaufwandes eines Testverfahrens ausfällt. Die Aufnahme der kognitiven Fähigkeit räumliches Vorstellungsvermögen in ein Modell zur Vorhersage der Leistung konsistenter, einfacher Aufgaben beim Erlernen

zahnmedizinischer Dentaltechniken, in das bereits instrumentelles, manuelles Geschick, die Schulabschlussnote und das naturwissenschaftliche Verständnis eingebunden sind, führte in der vorliegenden Untersuchung zu einer Erhöhung der aufgeklärten Leistungsvarianz von drei Prozent (s. Tabelle 5.19). Deshalb kann nicht wie in Hypothese 7 erwartet, von einer schwachen inkrementellen Validität für die räumliche Fähigkeit gesprochen werden. Die Ursache ist hierfür in dem nicht vorhandenen Zusammenhang zwischen der räumlichen Fähigkeit und der Leistung konsistenter, einfacher Aufgaben beim Fertigkeitserwerb zu suchen, weshalb bereits Hypothese 3 zurückzuweisen war.

Die inkrementelle Validität der nichtkognitiven Fähigkeit instrumentelles, manuelles Geschick ist hingegen, wie in Hypothese 8 angenommen, als moderat mit  $f^2 = 0.25$  einzustufen (s. Tabelle 5.19). Das Hinzufügen von instrumentellem, manuellem Geschick in ein PLS-Modell zur Vorhersage der Leistung konsistenter, einfacher Aufgaben beim Fertigkeitserwerb, in dem bereits die Prädiktoren räumliches Vorstellungsvermögen, Schulabschlussnote und das naturwissenschaftliche Verständnis eingeschlossen waren, führte zu einer Erhöhung von 17 Prozent an aufgeklärter Leistungsvarianz. Aus diesen Ergebnissen geht eindeutig hervor, dass zur Vorhersage der vorklinischen Laborkursleistung im Fall konsistenter, einfacher Aufgaben eher das instrumentelle, manuelle Geschick einzubeziehen ist als das räumliche Vorstellungsvermögen. In diesem speziellen Kontext empfiehlt sich daher kein Einsatz räumlicher Tests zur Studierendenauswahl in der Zahnmedizin, sodass die mit einer Testdurchführung verbundenen Kosten und Zeit eingespart werden können. In den Überlegungen zur Testauswahl ist aber hierbei die Tatsache zu berücksichtigen, dass im weiteren Kursverlauf der Anteil inkonsistenter, komplexer Aufgaben sehr stark zunimmt.

### **6.3.2 Prädiktion der Leistung inkonsistenter, komplexer Aufgaben**

Die Einführung des Phantomkopfs in den vorklinischen Laborkursen bedeutet zum einen eine Erhöhung der Aufgabenkomplexität, da die anzufertigen Werkstücke in die Nachbildung eines menschlichen Kopfes eingepasst werden müssen, wodurch sich automatisch die Anzahl der Aufgabenschritte erhöht. Zum anderen verhindert der Phantomkopf nicht nur die direkte Sicht auf die Zahnanatomie, sodass der Einsatz von Dentalspiegeln erforderlich ist. Ebenfalls ist auch der direkte und uneingeschränkte Zugriff auf die Zahnstrukturen nicht mehr möglich, wie beispielsweise beim Aufbau

eines tief zerstörten Zahns mit einem Titanstift oder der Präparation (Abschleifen) von Zähnen zum Einsetzen einer Füllung. Daher ist es für Zahnmedizinstudierende nicht nur notwendig, beim Herstellen einer Totalprothese und der Versorgung eines Lückengebisses mit einer Tangentialbrücke (Vollguss) mentale, anatomische Modelle der Zähne bzw. des Gebisses zu entwickeln, sondern diese auch fortwährend zu adaptieren, da mit der Planung, Herstellung und Eingliederung von Zahnersatz, insbesondere im Hinblick auf die individuelle Passung, ständige Kontroll- und Korrekturmaßnahmen einhergehen.

### **6.3.2.1 Einfluss der räumlichen und manuellen Fähigkeit**

Die Charakteristik inkonsistenter Aufgaben führt zu einem Stopp der Prozeduralisierung in der kognitiven Phase des Fertigkeitserwerbs (Ackerman, 1988), was sich in den Ergebnissen der vorliegenden Untersuchung widerspiegelt, sodass nicht nur instrumentelles, manuelles Geschick als nichtkognitive Fähigkeit ( $\beta = 0.36, p < .05$ ), sondern ebenfalls die kognitive Fähigkeit räumliches Vorstellungsvermögen ( $\beta = 0.37, p < .05$ ) einen moderaten, positiven Einfluss auf die Leistung inkonsistenter, komplexer Aufgaben beim Erlernen von Dentaltechniken besaßen, weshalb die Hypothesen 9 und 10 anzunehmen waren. Als überraschend erwies sich die fast identische Stärke der Pfadkoeffizienten  $\beta$  beider Fähigkeitsarten im PLS-Prädiktionsmodell zur Leistung inkonsistenter, komplexer Aufgaben, die sich zwar auch in den Ergebnissen früherer Studien vergleichbar wiederfinden, ohne aber von den Autoren inhaltlich erklärt oder zu mindestens diskutiert zu werden (z.B. Boyle & Santelli, 1986; Derevere, 1961).

Die fast einheitlich ausfallenden Pfadstärken des räumlichen Vorstellungsvermögens und des instrumentellen, manuellen Geschicks stützen die Annahme, dass beide Fähigkeitsarten in gleichem Maße in den vorklinischen Laborkursen der Zahnmedizin zur Bewältigung inkonsistenter, komplexer Aufgaben benötigt werden. Ausschlaggebend könnte hierbei die benötigte sensomotorische Koordination von Instrumenten in der Mundhöhle sein, um beispielsweise einen Zahn für die Versorgung mit einer Krone zu präparieren. Das Zusammenspiel manueller und räumlicher Fähigkeiten ist bereits bei den ersten Arbeitsschritten erforderlich, um den Zahnbohrer exakt im begrenzten Raum der Mundhöhle zu steuern, am Zahn anzusetzen und anschließend zielgerichtet die Zahnschicht abzutragen. Auch das in späteren Arbeitsschritten folgende Entfernen des Provisoriums sowie das Anpassen und Aufsetzen der Krone erfordern sowohl den geschickten Umgang mit zahnmedizinischen

Instrumenten als auch räumliches Vorstellungsvermögen, um den Arbeitsverlauf visuell zu kontrollieren, koordinieren und entsprechend den Einsatz der benötigten Instrumente adaptiv zu gestalten.

### **6.3.2.2 Einfluss komplementärer Prädiktoren**

Konform zu den inhaltlichen Herleitungen der Hypothesen 11 und 12 konnte in der vorliegenden Untersuchung ein Zusammenhang weder für die Schulabschlussnote noch für das naturwissenschaftliche Verständnis zur Leistung inkonsistenter, komplexer Aufgaben beim Erlernen zahnmedizinischer Fertigkeiten identifiziert werden. Die Schulabschlussnote und das naturwissenschaftliche Verständnis reflektieren eher kognitive Fähigkeiten, wie logisches Schlussfolgern und allgemeine Intelligenz (z.B. Krumm et al., 2008; Stemmler et al., 2011, S. 196; Asendorpf & Neyer, 2012, S. 150f.), die aber beim Erlernen zahnmedizinischer (wie auch allgemein motorischer Fertigkeiten) eine untergeordnete Rolle spielen (Deubert et al., 1975), da für die Erzeugung von Bewegungen auf analoges Wissen zurückgegriffen werden kann, dem keine Abstraktion der Bedeutung durch sprachliche, logische oder allgemein symbolische Aussagen zugrunde liegt (Zimmer, 1990, S. 153).

### **6.3.2.3 Inkrementelle Validitäten der räumlichen und manuellen Fähigkeit**

Die in den Hypothesen 13 und 14 postulierten Annahmen über die inkrementellen Validitäten bestätigten sich für beide Fähigkeitsarten. Sowohl für das räumliche Vorstellungsvermögen als auch für das instrumentelle, manuelle Geschick besteht eine moderate inkrementelle Validität ( $f^2 = 0.17$ ) über die Schulabschlussnote, das naturwissenschaftliche Verständnis und jeweils entgegengesetzt die manuelle bzw. räumliche Fähigkeit in Bezug auf die Leistung inkonsistenter, komplexer Aufgaben beim Erwerb zahnmedizinischer Dentaltechniken. Die Höhe der inkrementellen Validitäten gestaltet sich mit 10.6 Prozent für die kognitive und mit 10.08 Prozent für die nichtkognitive Fähigkeit beinahe identisch, sodass sich eine Kombination der Fähigkeitsarten im Studierendenauswahlverfahren als gleichwertig im Hinblick auf den Informationsgewinn an zusätzlich aufgeklärter Leistungsvarianz inkonsistenter, komplexer Aufgaben erweist. Ursache ist hierfür nach Ackerman (1988) die ständige Neubildung von Reiz-Reaktionsprodukten aufgrund der fortwährenden Änderungen in der Beschaffenheit und Darbietung der Reize, wie es beispielsweise beim Einpassen von Stiftzähnen und Zahnbrücken der Fall ist, weshalb das räumliche Vorstellungsvermögen

als kognitive Fähigkeit in allen drei Phasen einen Einfluss auf den Erwerb zahnmedizinischer Fertigkeiten besitzt.

### **6.3.3 Resümee**

Die Unterteilung der vorklinischen Laborkursarbeiten in konsistente, einfache und inkonsistente, komplexe Aufgaben ermöglicht mit Bezug auf Ackermans (1988) Theorie des Fertigkeitserwerbs eine theoretisch gestützte Charakterisierung der Beziehungen zwischen räumlichen sowie manuellen Fähigkeiten und den zu erlernenden zahnmedizinischen Untersuchungs- und Behandlungsmethoden in der Vorklinik der Zahnmedizin. Die von Ackerman (1988) postulierten Korrelate zwischen Fähigkeiten und individuellen Leistungsunterschieden in den Phasen des Fertigkeitserwerbs ermöglichen somit auch im Kontext der Studierendenauswahl zuverlässige Aussagen über die im Auswahlverfahren eingesetzten Testverfahren bzw. den zugrundeliegenden Fähigkeiten und der Aufgabenleistungen in den vorklinischen Laborkursen. Trotz verschiedener Fähigkeitsausprägungen der Bewerber, die von Studierenden unterschiedlich in Anspruch genommenen Übungszeiten im Labor und vor allem, des zeitlichen Abstands zwischen Fähigkeitsmessung im Auswahlverfahren und den Aufgabenbeurteilungen in den vorklinischen Laborkursen erwiesen sich diese Korrelate als stabil (vgl. Ackerman, 1992; Ackerman & Cianciolo, 2000).

Den Ergebnissen der vorliegenden Untersuchung zufolge, insbesondere aufgrund der ermittelten inkrementellen Validitäten, ist ein Einsatz manueller und räumlicher Tests in Studierendenauswahlverfahren der Zahnmedizin zu empfehlen, um für die Bewerber bzw. zukünftigen Studierenden zu prüfen, ob deren Fähigkeitsausprägungen den Anforderungen in den vorklinischen Laborkursen entsprechen. Zur Ermittlung der prädiktiven Validitäten, die zugleich als Gütekriterien der im Auswahlverfahren eingesetzten Testverfahren herangezogen werden, ist die Charakteristik der Laborkursaufgaben zu berücksichtigen, die zur Operationalisierung der praktischen Studienleistung dienen. Vor allem die individuellen Leistungen konsistenter, einfacher Aufgaben beim Erlernen erster zahnmedizinischer Fertigkeiten, wie das Schnitzen von Wachs Zähnen, sind nicht durch das räumliche Vorstellungsvermögen beeinflusst, falls die Leistungsbeurteilung an einem fertiggestellten Werkstück erfolgt.

## **6.4 Applikation der Untersuchungsergebnisse**

### **6.4.1 Prädiktion vorklinischer Laborkursleistungen**

#### **6.4.1.1 Prädiktion konsistenter-einfacher Laborkursaufgaben (Vorklinik)**

Die Klassifizierung vorklinischer Laborkursaufgaben in den Typ konsistent-einfach ermöglicht unter Anwendung von Ackermans (1988) Modell des Fertigkeitserwerbs eine theoretisch fundierte Beschreibung der Beziehung zwischen Fähigkeiten und den zu erwerbenden, grundlegenden zahnmedizinischen Fertigkeiten, die eine Betrachtung bestehender empirischer Ergebnisse in einem erweiterten Blickwinkel erlaubt. So erscheint beispielsweise die Argumentation von Bellanti et al. (1972) in der Debatte um den Verbleib des Kalkschnitttests im Studierfähigkeitstest DAT nicht sehr schlüssig, betrachtet man die als Leistungskriterien in die Studie einbezogenen Laborkursaufgaben (vgl. Unterkapitel 2.3.3.1). Die Autoren sprachen sich für den Kalkschnitttest aus, da der konkurrierende, räumliche DAT-Subtest „Space Visualization“ über den Schulabschlussnotendurchschnitt GPA und den Kalkschnitttest nur eine geringe inkrementelle Validität von 1.79 Prozent aufwies, während die des Kalkschnitttests über den GPA und räumlichen DAT-Subtest hingegen 13.83 Prozent betrug. Die Unterschiede in den Ergebnissen zum Inkrement beider DAT-Untertests sind jedoch wenig überraschend, berücksichtigt man, dass die zur Operationalisierung der praktischen Studienleistung verwendeten Laborkurstätigkeiten als konsistente, einfache Aufgaben zu klassifizieren sind, die beispielsweise das Präparieren von Zähnen zum Einsetzen von Kronen sowie das Anfertigen von Zahnkronen aus Wachs und Porzellan umfassten, die nicht an einem Phantomkopf durchgeführt wurden.

Das räumliche Vorstellungsvermögen besitzt als kognitive Fähigkeit den Ergebnissen der vorliegenden Arbeit zufolge in der autonomen Phase des Erwerbs zahnmedizinischer Fertigkeiten keinen Einfluss auf die Leistung konsistenter, einfacher Aufgaben, sodass Bellanti et al. für ihre Argumentation über den Verbleib des Kalkschnitttests im DAT, auch inkonsistente, komplexe Aufgaben zur Operationalisierung der vorklinischen Laborkursleistung hätten einbeziehen müssen, um objektiv die prädiktive Güte der vorklinischen Laborkursleistung beider Subtests miteinander vergleichen zu können. Auch die Ergebnisse von Kao et al. (1990) und Walcott et al. (1986) lassen sich mit der in Ackermans (1988) Theorie postulierten Beziehung zwischen Fähigkeiten und der Leistung konsistenter, einfacher Aufgaben beim Erlernen erster zahnmedizinischer Techniken nachvollziehen. In beiden Studien

wurden für Testverfahren, die manuelles Geschick im Umgang mit Instrumenten erfassen, deutlich höhere inkrementelle Validitäten über den GPA und die naturwissenschaftliches Wissen sowie räumliches Vorstellungsvermögen erfassenden DAT-Subtests „Academic Average“ und PAT berichtet (vgl. Unterkapitel 2.2.3.3). Jedoch erfolgte die Prädiktion der Leistungen nur für konsistente, einfache Aufgaben in den vorklinischen Laborkursen der Zahnmedizin, die beispielsweise das Schnitzen von Wachsähnen und das Präparieren, Herstellen und Einsetzen von Amalgam-, Gold- und Kunstharzfüllungen umfassten.

Ebenfalls sind ex post die Studien von Wong et al. (1979) und Wood (1979) zu beanstanden, in denen die praktische Studienleistung ausschließlich über konsistente, einfache Laborkursaufgaben operationalisiert wurde. Die Autoren beider Studien empfahlen aber aufgrund ihrer empirischen Ergebnisse, den manuellen Fähigkeiten erfassenden Kalkschnitttest im kanadischen Studierfähigkeitstest DAT einzusetzen, da dieser besser die vorklinische Laborkursleistung prädizierte als der PMAT, dem Vorläufer des DAT-Subtests PAT zum räumlichen Vorstellungsvermögen (vgl. Unterkapitel 2.3.3.4). Die Entscheidung über den Einsatz psychometrischer Testverfahren sollte demzufolge nicht nur in Abhängigkeit von konsistenten, einfachen Aufgaben, in denen überwiegend die Herstellung von Zahnersatz, wie Wachsähnen, Zahnkronen oder Prothesen, getroffen werden, sondern auch inkonsistente, komplexe Aufgaben einbeziehen, in denen zahnmedizinische Untersuchungs- und Behandlungstechniken am Phantomkopf vermittelt werden, um die in den vorklinischen Laborkursen zu erlernenden zahnmedizinischen Fertigkeiten realistisch abzubilden.

#### **6.4.1.2 Prädiktion inkonsistenter-komplexer Laborkursaufgaben (Vorklinik)**

Sollten praktische Arbeiten in den vorklinischen Laborkursen sowohl konsistente als auch inkonsistente Aufgabenbestandteile enthalten, ist der Empfehlung Ackermans (1992) zu folgen, der ab einem inkonsistenten Aufgabenanteil von 30 Prozent dazu rät, den Einfluss von Fähigkeiten auf die Aufgabenleistung beim Erlernen von Fertigkeiten wie für inkonsistente Aufgaben zu interpretieren. Die Kenntnisse über die Arbeitsprozesse und -bedingungen der als Leistungskriterium einbezogenen Laborkursaufgaben vorausgesetzt, können frühere Studienergebnisse, die kontroverse Ergebnisse über die Beziehungen zwischen Fähigkeiten und der praktischen Leistung in den vorklinischen Laborkursen berichteten, ausführlicher analysiert werden. Neben methodischen Aspekten (vgl. Unterkapitel 2.3.5) kann nun auch die Charakteristik der

Beziehungen zwischen kognitiven sowie nichtkognitiven Fähigkeiten und individuellen Leistungsunterschieden beim Erlernen zahnmedizinischer Fertigkeiten als ein weiterer Faktor in die Ursachensuche einbezogen werden.

Studien, in denen die praktische Studienleistung über Durchschnittswerte aus mehreren Laborkursen bzw. den des gesamten vorklinischen Studienabschnitts operationalisiert war, beziehen zwangsläufig die Leistungen inkonsistenter, komplexer Aufgaben durch die Übungen am Typodonten bzw. Phantomkopf in der Vorklinik ein. Aus diesem Grund finden sich beispielsweise in den Studien von Coy et al. (2003), Deubert et al. (1975), Dirks und Evans (2001), Gray und Deem (2002a), Hegarty et al. (2009), Kramer et al. (1989) und Lundergan et al. (2007) zwangsläufig Ergebnisse, die positiv moderate bis schwache lineare Zusammenhänge zwischen der kognitiven Fähigkeit räumliches Vorstellungsvermögen und den gemittelten Aufgabenleistungen eines oder mehrerer Laborkurse im vorklinischen Studienabschnitt aufzeigten, ohne jedoch dazu im Widerspruch stehende Studienergebnisse zu erklären (z.B. Kao et al., 1990; Walcott et al., 1986), die räumliche Tests als nicht zuverlässige Prädiktoren praktischer Studienleistung einstufen, welche aber ausschließlich über konsistente, einfache Aufgaben aus den vorklinischen Laborkursen operationalisiert waren.

#### **6.4.2 Prädiktive Stärke der räumlichen und manuellen Fähigkeit**

Die Studienergebnisse in der Literatur zur Vorhersage der Aufgabenleistungen in den vorklinischen Laborkursen durch räumliche und manuelle Fähigkeiten werden von Autoren sehr gegensätzlich interpretiert, obwohl die einen linearen Zusammenhang anzeigenden Korrelationen im Durchschnitt vergleichbare Ausprägungen um 0.3 bis 0.4 annehmen (vgl. Unterkapitel 2.3.4). Solche Koeffizientenstärken sahen beispielsweise Evans und Dirks (2001) und Kramer et al. (1989) als ausreichend an, um von einem zufriedenstellenden Einfluss des räumlichen Vorstellungsvermögens auf die vorklinische Laborkursleistungen sprechen zu können, während Coy et al. (2003) auf Basis derartiger Beziehungen keinen Einsatz räumlicher Tests zur Studierendenauswahl in der Zahnmedizin empfahlen.

Ähnlich unterschiedliche Ergebnisinterpretationen finden sich für das instrumentelle, manuelle Geschick, das Oudshoorn (2003) im Kontext des kanadischen DAT-Subtests „Carving Dexterity“ nur als schwachen Prädiktor der Aufgabenleistungen in den Laborkursen des ersten Studienjahrs klassifizierte und demzufolge nach Ansicht des Autors, auch nur von geringem Nutzen in der Studierendenauswahl sei. Beispielsweise

Kao et al. (1990) stuften aber hingegen vergleichbare Korrelationen zwischen 0.3 bis 0.4 als moderat ein und forderten den Einsatz eines Drahtbiegetests für die Selektion von Studienbewerbern in der Zahnmedizin. Bezieht man in die Diskussion um die prädiktiven Validitäten der Testverfahren zum räumlichen Vorstellungsvermögen und instrumentellen, manuellen Geschick den „whole-person“ Ansatz von Snow und Farr (1987) aus der Berufseignungsdiagnostik ein (vgl. Unterkapitel 2.4.3.2), so ist ersichtlich, dass eine einzelne Fähigkeit höchstens 30 bis 40 Prozent der Leistung einer komplexen Aufgabe präzisieren kann, was in der Literatur als „Silo“-Problem bezeichnet wird (Ackerman et al., 1995). Hierbei ist allerdings zu berücksichtigen, dass ein direkter Vergleich der Leistungsprädiktionen über beide Forschungsfelder, vor allem in der Höhe der vorhergesagten Leistung, nur bedingt zulässig ist, da in der Studieneignungsdiagnostik überwiegend neu zu erlernende Fertigkeiten zur Operationalisierung der Leistung herangezogen werden und kaum Vorerfahrungen in einem Studiengang seitens der Studierenden bestehen, die vergleichbar zur Berufserfahrung von Arbeitnehmern ausfallen.

#### **6.4.2.1 „Silo“-Problem**

Die als „Silo“-Problem bezeichnete Limitation in der Leistungsprognostik wäre durch den „whole-person“ Ansatz bei der Vorhersage der vorklinischen Laborkursleistung insofern zu umgehen, indem nicht nur Fähigkeiten der Bewerber einbezogen werden, sondern ein individuelles Gesamtbild der Person erstellt wird. Neben dem räumlichen Vorstellungsvermögen und dem instrumentellen, manuellen Geschick wären diesem Ansatz folgend (vgl. Kanfer et al., 2010), auch individuelle Dispositionen der Bewerber in die Prädiktionsmodelle der praktischen Aufgabenleistungen einzubeziehen, die Interessen, Persönlichkeitseigenschaften, aber auch Unterschiede in der Lernfähigkeit und im Lernstil widerspiegeln (vgl. Ackerman & Cianciolo, 2000). Die in Studien kritisierte prädiktive Stärke der vorklinischen Laborkursleistungen von räumlichen und manuellen Fähigkeiten ist somit zu relativieren, zumal eine Fähigkeit letztlich immer nur die Kapazität widerspiegelt, etwas zu tun (van der Maas et al., 2011). Darüber hinaus sind für das erfolgreiche Erlernen zahnmedizinischer Untersuchungs- und Behandlungsmethoden in den Laborkursen auch didaktische Faktoren entscheidend. Vor allem zu Beginn des Fertigkeitserwerbs sind Zahnmedizinierende auf spezifisches und sofortiges Feedback der Lehrenden angewiesen (Chambers & Glassman, 1997) und benötigen detaillierte, transparente Aufgabeninstruktionen

(Hauser & Bowen, 2009), die zudem in einer Lernsituation vermittelt werden sollten, welche den Anforderungen der Anwendungssituation entspricht (Knight et al., 1990). Ein Einbezug didaktischer Faktoren in Regressionsmodelle zur praktischen Studienleistung sollte demzufolge zu einem substantiellen Anstieg der prädierten Aufgabenleistung führen.

#### **6.4.2.2 Beurteilung der Einflussstärken in der vorliegenden Untersuchung**

Die zuvor aufgeführten Gründe erklären, warum in der vorliegenden Untersuchung die Determinationskoeffizienten  $R^2$  der PLS-Prädiktionsmodelle nicht höher ausfallen können. Neben dem schwachen Einfluss der Schulabschlussnote wurden primär durch die nichtkognitive Fähigkeit instrumentelles, manuelles Geschick insgesamt 32.3 Prozent der Varianz konsistenter, einfacher Aufgabenleistungen beim Erlernen zahnmedizinischer Fertigkeiten erklärt (s. Tabelle 5.14). Somit verblieben gerundet 68 Prozent an nicht aufgeklärter Varianz, was als ein klares Indiz für den Einfluss weiterer Faktoren auf die Leistung konsistenter, einfacher Aufgaben zu interpretieren ist. Der Anteil erklärter Varianz für die Leistung inkonsistenter, komplexer Aufgaben beim Erlernen zahnmedizinischer Fertigkeiten fiel mit insgesamt 35.6 Prozent im Vergleich zu den konsistenten, einfachen Aufgaben nur geringfügig höher aus, obwohl nun auch die kognitive Fähigkeit räumliches Vorstellungsvermögen einen moderaten Einfluss auf die Aufgabenleistung besaß, der fast identisch zu dem der nichtkognitiven Fähigkeit instrumentelles, manuelles Geschick war (s. Tabelle 5.15). Trotzdem verblieb aber auch bei der Prädiktion inkonsistenter, komplexer Aufgaben ein mehrheitlicher Anteil von abgerundet 64 Prozent an nicht aufgeklärter Leistungsvarianz, sodass der Einbezug weiterer Fähigkeiten zur Vorhersage vorklinischer Laborkursleistungen wenig vielversprechend erscheint. Die hohen Anteile nicht erklärter Varianz beider Aufgabentypen in den vorklinischen Laborkursen verdeutlichen die Notwendigkeit für den Einschluss weiterer Prädiktoren zur Vorhersage der Laborkursleistungen, die sowohl Persönlichkeitseigenschaften als auch komplementär didaktische Lehrfaktoren umfassten sollten.

Gleichzeitig ist damit die Argumentation, räumliche und manuelle Fähigkeiten nur zur Auswahl von Zahnmedizinstudierenden einzusetzen, falls für diese ein substantieller Einfluss auf die Aufgabenleistungen in den vorklinischen Laborkursen bestünde, zurückzuweisen und die Interpretation der prädiktiven Gütemaße zu relativieren. Testverfahren, die zur Aufgabenbearbeitung instrumentelles, manuelles Geschick

erfordern, können somit in Regressionsmodellen zur Vorhersage der vorklinischen Aufgabenleistungen, unabhängig vom Aufgabentyp, nur Pfadkoeffizienten aufweisen, die höchstens über moderate prädiktive Stärken verfügen. Mit Bezug auf die Ergebnisse der vorliegenden Untersuchung ist für die kognitive Fähigkeit räumliches Vorstellungsvermögen nur in Regressionsmodellen zur Vorhersage der Leistung inkonsistenter, komplexer Kursaufgaben zu erwarten, dass die Pfadkoeffizienten moderate prädiktive Stärken annehmen. Demzufolge sollten für Vorhersagemodelle vorklinischer Laborkursleistungen in der Zahnmedizin, in Abhängigkeit des Aufgabentyps und der als Prädiktoren eingebundenen räumlichen und manuellen Fähigkeiten, nur höchstens Determinationskoeffizienten  $R^2$  mittlerer Stärke zu erwarten sein. Als Prämisse ist hierbei zu beachten, dass die erfasste Laborkursleistungen sich nur auf die zur Prüfung eingereichte, d.h. fertiggestellte prothetische Zahnarbeit bzw. vorgeführte Demonstration einer Untersuchungstechnik bezieht (s. dazu auch Unterkapitel 6.5.2).

#### **6.4.2.3 Geschlechtsspezifität des räumlichen Vorstellungsvermögens**

In Metaanalysen (z.B. Linn & Petersen, 1985; Maeda & Yoon, 2013; Peters et al., 2006) sowie in Primärstudien zur Prädiktion der vorklinischen Laborkursleistungen wurde für das räumliche Vorstellungsvermögen ein Geschlechtseffekt nachgewiesen (z.B. Coy et al., 2003; Hegarty et al. 2009), der sich insbesondere bei Aufgaben zeigte, die das mentale Rotieren von dreidimensionalen Objekten erfordern. Als Gründe werden in der Literatur neben biologischen und affektiven Faktoren, kognitive Strategien, persönliche Erfahrungen und Durchführungsbedingungen der Testverfahren vermutet (s. dazu Yoon, 2011). Zwischen Männern und Frauen zeigte sich in der vorliegenden Arbeit aber kein Unterschied in der räumlichen Fähigkeit, weder bei den Bewerbern im Auswahlverfahren ( $t = 1.515$ ,  $df = 126$ ,  $n.s.$ ), noch für die über das Auswahlverfahren zugelassenen Studierenden ( $t = -0.402$ ,  $df = 36$ ,  $n.s.$ ).

Ein Grund ist hierfür in der quasiexperimentellen Stichprobenbeschaffenheit zu vermuten. Die Vorauswahl der Bewerber durch die Abiturdurchschnittsnote (vgl. Unterkapitel 4.2.2) oder der überwiegende Anteil von 81 Frauen gegenüber 48 Männern (s. Unterkapitel 5.1.1) könnten als Ursachen für das Nichtauftreten des Geschlechtseffekts in Frage kommen. Gegen den zuletzt genannten Grund spricht aber die Gleichverteilung des Geschlechts der 38 über das Auswahlverfahren zugelassenen und am ersten vorklinischen Laborkurs teilnehmenden Studierenden. Somit ist die

Ursache eher im LPS als Messinstrument zu vermuten, für das Horn in der Normierung von 1983 keine geschlechtsspezifischen Leistungsunterschiede in den einzelnen Subtests berichtete.

Ausschlaggebend sind wahrscheinlich die zur Operationalisierung des räumlichen Vorstellungsvermögens verwendeten Aufgabentypen der Skala technische Begabung, die neben mentalem Rotieren (UT7) auch die Faktoren räumliche Vorstellungsfähigkeit (UT8) und Orientierung (UT9) sowie die Flexibilität der Gestaltbildung (UT10) beinhalten (s. Tabelle 4.01), für die generell nur minimale Geschlechtsunterschiede bestehen (Linn & Petersen, 1985; Voyer et al., 1995). Coy et al. (2003) sahen in den geschlechtsspezifischen Unterschieden im räumlichen Vorstellungsvermögen eine Erklärung für die geringe prognostische Stärke der vorklinischen Laborkursleistungen, die somit in der vorliegenden Untersuchung ausgeschlossen werden können.

#### **6.4.3 Operationalisierung des manuellen Geschicks**

Aufgrund des breiten Spektrums und der Fülle an unterschiedlichen Definitionen, werden eine Vielzahl einzelner Fähigkeiten dem manuellen Geschick zugeordnet, die beispielsweise allein in der Feinmotorik die Geschicklichkeit der Finger, der Hand sowie präziser Bewegungen von Händen und Fingern umfasst (vgl. Unterkapitel 2.3.1.3). Daher besteht die Auflage, die zur Operationalisierung des manuellen Geschicks eingesetzten Testverfahren nicht nur hinsichtlich ihrer Reliabilität und Validität zu prüfen (Ackerman & Cianciolo, 2000; Yancosek & Howell, 2009), sondern vor allem für das letztgenannte Gütekriterium auch bewährte Definitionen heranzuziehen. So ist bei der Interpretation von Untersuchungsergebnissen zu vermeiden, dass von einer speziellen Form des manuellen Geschicks auf die Fähigkeit im Allgemeinen geschlussfolgert wird und somit generell die Eignung der nichtkognitiven Fähigkeit bzw. von Testverfahren zum manuellem Geschick zur Bewerberauswahl in der Zahnmedizin in Frage gestellt werden.

In Studien, die Testverfahren als nicht geeignete Instrumente zur Studierendenauswahl ermittelten, in denen manuelles Geschick zur Handhabung kleiner Stifte mit Pinzetten gefordert war (de Andres et al., 2004; Lundergan et al., 2007; Raybould et al., 1983; Waldman et al., 1995), ist die mangelnde Berücksichtigung der Empfehlung von Weinstein und Kiyak (1981) zu kritisieren. Die Abbildung zahnmedizinisch relevanter Tätigkeiten und Behandlungsbedingungen erwies sich als essentiell in manuellen Testverfahren zur Auswahl von Studienbewerbern in der Zahnmedizin (vgl.

Unterkapitel 2.3.5.4.2). Konform zu den Schlussfolgerungen von Kao et al. (1990) konnte in der vorliegenden Untersuchung für einen Drahtbiegetest aufgezeigt werden, dass manuelles Geschick im manipulativen Umgang mit Instrumenten eine moderate prognostische Validität in Bezug auf die Leistungen in den vorklinischen Laborkursen besitzt, womit die inhaltliche Eignung des Tests im Studierendenauswahlverfahren unterstrichen wird.

Das Erlernen der Untersuchungs- und Behandlungstechniken in der Zahnmedizin erfordert primär die Anwendung von Instrumenten, welche sich im Umgang mit der Zange reflektiert, die zum Biegen der Drähte sowohl bei Kao et al. (1990) als auch im Drahtbiegetest HAM-Man der vorliegenden Untersuchung benötigt wurde. Darüber hinaus werden zahnmedizinisch relevante Fertigkeiten im HAM-Man erfordert, die unter anderem das Maßnehmen von Strecken und Winkeln der anzufertigen Drahtfiguren umfasst sowie die eigenständige Beurteilung der Arbeitsqualität. Diese beinhaltet das Erkennen und Korrigieren von Fehlern bzw. Ungenauigkeiten in den Drahtfiguren. Ebenfalls ist im HAM-Man die Planung und zeitliche Koordination der Tätigkeiten bzw. einzelnen Arbeitsschritte erforderlich, da im Drahtbiegetest durchschnittlich pro anzufertigende Figur 15 Minuten Bearbeitungszeit zur Verfügung stehen.

#### **6.4.3.1 Weiterentwicklung des Drahtbiegetests HAM-Man**

Der Drahtbiegetest HAM-Man erwies sich in der vorliegenden Untersuchung als ökonomisches Testverfahren, dessen Bereitstellung, Durchführung und Auswertung mit geringem Aufwand verbunden war (vgl. Unterkapitel 4.1.1.1). Zukünftig wäre zu überlegen, ob die eingesetzte, handelsübliche Spitzzange nicht durch professionellere, in der Zahntechnik verwendete Zangen zu ersetzen ist. Hier bieten sich als Alternative verschiedene Modelle an, beispielsweise Draht- und Klammerbiegezangen, Schlaufenbiegezangen oder Rundzangen, die auch in Kombination für das Biegen der Drahtfiguren zu verwenden wären. Der Einsatz hochwertigerer Instrumente im HAM-Man würde zugleich qualitativ bessere Biegungen erlauben, wodurch das Bewertungskriterium Qualität der Biegung revidiert werden könnte, für das sich in allen Drahtfiguren Dreieck, Dritte Dimension und Rundbögen zu hohe Übereinstimmung mit dem Auswertungskriterium Deckungsgleichheit ( $r \geq 0.85$ ) zeigte. Somit könnte nicht nur die Form der Biegungen, d.h. die Genauigkeit der Winkel beurteilt werden, sondern eher die Sauberkeit der Biegung, d.h., das optische Bild, welches durch Abdrücke am

Draht, die durch Abrutschen, oder falsches Ansetzen der Zange entstehen und mehrfache Biegungen eines Winkels gemindert wird. Eine schriftliche Instruktion sowie Zeit zum Üben mit den professionellen Zangen sollten entsprechend im Test und in den Vorbereitungsempfehlungen repräsentiert werden, wodurch aber keine Beeinträchtigung der Ökonomie des HAM-Man zu erwarten wäre.

Die Normierung anhand von geeigneten Eichstichproben stellt eine zukünftige Entwicklungsaufgabe für den Drahtbiegetest HAM-Man dar, um die Einsatzfelder des Testverfahrens zu erweitern. Vor allem die Definition einer „cut-off“-Grenze, d.h., die Festlegung einer Mindestbestehensgrenze im HAM-Man auf Basis umfangreicher empirischer Daten ermöglicht einen Einsatz zur Identifikation Erstsemesterstudierender in der Zahnmedizin, deren instrumentelles, manuelles Geschick nicht hinreichend ausgeprägt ist, um die vorklinischen Laborkurse erfolgreich zu bestehen und somit Bedarf an zusätzlichen Förderkursen haben. Weiterführend wäre für den HAM-Man zu prüfen, ob das Zeitlimit von 45 Minuten nicht auf jede einzelne Drahtfigur von je 15 Minuten zu übertragen ist. Nach Ablauf des Zeitlimits müssten die Arbeiten an der gegenwärtigen Figur eingestellt und mit der nächsten begonnen werden. Die daraus resultierende Reihenfolge sollte konform zum ansteigenden Schwierigkeitsgrad der einzelnen Drahtfiguren ausfallen, der aber zuvor unter dem Prüfungsaspekt der figurenbezogenen Zeitlimits zu ermitteln wäre. Diese Modifikation im HAM-Man würde zwar die mit dem Zeitmanagement des Teilnehmers und der Wahl der strategischen Vorgehensweise verbundenen Einflüsse auf die Testleistung eliminieren. Zukünftig könnte aber ausgeschlossen werden, dass beispielsweise ein Bewerber gezielt nur vereinzelte Figuren bearbeitet (vgl. Unterkapitel 6.2.3.2) oder zu viel Zeit für das Üben mit den Probedrähten bzw. für das Biegen einer Figur aufwendet, wodurch die tatsächliche Ausprägung seines instrumentellen, manuellen Geschicks verzerrt würde.

#### **6.4.3.2 Trainierbarkeit manueller Fähigkeiten**

Eine Vielzahl an früheren und aktuelleren Primärstudien zeigte zweifelsfrei eine Verbesserung der manuellen Fähigkeiten von Studierenden im Verlauf des Zahnmedizinstudiums auf, die auf die in den Laborkursen erlernten zahnmedizinischen Untersuchungs- und Behandlungsfertigkeiten zurückgeführt wurden (z.B. Boyle & Santelli, 1986; Giuliani et al., 2007; Luck et al., 2000; Silvestri et al., 1979; Wilson & Husak, 1988). Aufgrund der Trainierbarkeit des manuellen Geschicks wird vor allem in aktuelleren Studien gefordert, zukünftige Studierende nicht auf Basis von Defiziten in

einer Fähigkeit vom Studium auszuschließen, die durch Übung ausgeglichen werden könnten. Eher seien manuelle Testverfahren zur frühzeitigen Diagnostik im Zahnmedizinstudium einzusetzen, um Studierende mit geringen manuellen Fähigkeiten rechtzeitig zu identifizieren und durch Förderkurse zielgerichtet zu trainieren (Gansky et al., 2004; Giuliani et al., 2007; Luck et al., 2000).

Der Argumentation ist aber nur unter der Prämisse eines ausbalancierten Verhältnisses von Studienplätzen und Bewerberzahlen zuzustimmen. In 2010 bewarben sich allein im Auswahlverfahren der Medizinischen Fakultät an der Universität Hamburg 467 Bewerber, von denen lediglich 39 einen Studienplatz erhielten, sodass pro verfügbaren Studienplatz fast zwölf Bewerbungen eingingen (vgl. Unterkapitel 5.1.2). Die Notwendigkeit eines Auswahlverfahrens kann daher nicht in Frage gestellt werden, um einerseits die Fairness im Bewerbungsprozess zu garantieren, indem jeder Bewerber die gleiche Chance auf eine Zulassung zum Studium der Zahnmedizin in Abhängigkeit seiner individuellen Fähigkeiten erhält und andererseits zugleich den Studienerfolg zukünftiger Studierender sicherstellt, die auf Basis studienrelevanter Fähigkeiten einen Nachweis über die Eignung zum Zahnmedizinstudium erbringen, sodass nicht geeignete Bewerber keine Zulassung erhalten.

#### **6.4.3.3 Anforderungen an Aufgaben manueller Tests**

An Aufgaben manueller Testverfahren zur Auswahl Zahnmedizinstudierender sollte der Anspruch gestellt werden, mehrere, in den vorklinischen Laborkursen zu erlernende Fertigkeiten in Form von Arbeitsproben abzubilden. Damit einhergehende Einschränkungen in der Testökonomie, wie beim DAT-Untertest „Carving Dexterity“ (vgl. Unterkapitel 2.3.3.3), könnten kompensiert werden, indem fokussiert elementare zahnmedizinische Tätigkeiten über die manuellen Testaufgaben erfasst werden. Das Gießen oder Brennen von Zahnersatzformen wären als Fertigkeiten somit aus testökonomischer Sicht eher weniger als Anforderungen an manuelle Testaufgaben zu richten, aber hingegen das Schleifen, Fräsen, Bohren und Polieren. Beispiele finden sich bereits in universitären Auswahlverfahren (s. Unterkapitel 2.2.7), in denen Bewerber neben dem Biegen von Drähten, einen Acrylblock polieren oder mit einem Dentalbohrer vorgegebene Löcher ausbohren müssen (z.B. Beier et al., 2010).

Insbesondere die Anforderungen im Umgang mit Instrumenten, wie beispielsweise Gracey-Küretten, Dentalbohrern und -spiegeln, stellen Veränderungen in den Werkzeugschnittstellen der manuellen Aufgaben dar, die nicht stellvertretend durch den

Umgang mit einem Instrument, wie der Spitzzange im HAM-Man, vollständig simuliert werden können. Vergleichbar zur Spiegelprobe im BATZ (s. Unterkapitel 2.2.3.2) sind folglich Aufgaben in manuelle Testverfahren zu implementieren, die eine höhere Bandbreite zahnmedizinischer Instrumente repräsentieren.

Innovationen in Behandlungsformen, sei es durch technologische Weiterentwicklungen der Behandlungsinstrumente oder neue Untersuchungstechniken, führen gleichzeitig zu Veränderungen in den zu erlernenden zahnmedizinischen Fertigkeiten der (vor)klinischen Laborkurse. Testverfahren, die verschiedene Fassetten des manuellen Geschicks über unterschiedliche Aufgaben operationalisieren, könnten Veränderungen in den Behandlungsformen durch eine Adaption einzelner Aufgaben oder der Entwicklung neuer Teilaufgaben entsprechend kompensieren.

#### **6.4.4 Operationalisierung der praktischen Studienleistung**

Die Operationalisierung der praktischen Studienleistung erfolgte in der vorliegenden Untersuchung über das Erlernen zahnmedizinischer Untersuchungs- und Behandlungstechniken als latente Variable (vgl. Unterkapitel 4.1.5.2), wofür die mit Noten bewerteten Aufgabenleistungen in den vorklinischen Laborkursen als Indikatoren herangezogen wurden. Die Güte der latenten, endogenen Variable erwies sich in den varianzbasierten Prädiktionsmodellen insgesamt als akzeptabel, sowohl für die konsistenten, einfachen Aufgaben als auch für die inkonsistenten, komplexen. Der generellen Kritik an Noten zur Messung der praktischen Studienleistung, unter anderem von Natkin und Guild (1967), Phipps et al. (1968), Spratley (1992) und Thompson (1947), kann somit nur bedingt zugestimmt werden. Vermutlich ist die Kritik größtenteils auf das Fehlen zuverlässiger Beurteilungskriterien für die Benotung der Laborkursarbeiten zurückzuführen, was aber in der vorliegenden Untersuchung aufgrund des vorhandenen Fachkriterienkatalogs ausgeschlossen werden konnte (vgl. Tabelle A22 im Anhang).

Ansätze von beispielsweise Evans und Dirk (2001), anstatt von Noten eine Punkteskala von null bis einhundert zur Beurteilung der Aufgabenleistungen in den Laborkursen des ersten Semesters einzusetzen, sollten aufgrund der höheren Auflösung der Punkteskala zu einer Verbesserung der Messgenauigkeit führen. Die Wirksamkeit dieser Maßnahme ist im Kontext der Vorhersage praktischer Studienleistungen in der Zahnmedizin jedoch nicht nachzuvollziehen, da die für eine Evaluation benötigten Gütekriterien von den Autoren nicht berichtet wurden. Die Operationalisierungsgüte der praktischen

Studienleistungen sollte zudem weniger eine Frage des verwendeten Skalentyps sein, sondern eine valider, der Leistungsbeurteilung zugrundeliegender Fachkriterien.

#### **6.4.5 Motivation als Störvariable**

Hegarty et al. (2009) vermuteten in der Motivation der Studienteilnehmer, einen Fähigkeitstest bestmöglich zu bearbeiten, den ausschlaggebenden Grund, weshalb in ihrer Untersuchung lediglich der räumliches Vorstellungsvermögen erfassende PAT eine positiv moderate, signifikante Beziehung zu den durchschnittlichen Aufgabenleistungen in den vorklinischen Laborkursen der Zahnerhaltungskunde aufwies (s. Unterkapitel 2.3.4.1.2). Während der PAT als Subtest des Studierfähigkeitstests DAT zur Bewerberauswahl in der Zahnmedizin eingesetzt wird und das Abschneiden somit von Relevanz für die Teilnehmer ist, erfolgte in der Studie von Hegarty et al. die Durchführung des „Novel Object Cross Section-tests“, des „Vandenberg Mental Rotation tests“ und der modifizierten Version von Guays (1976) „Visualization of Views Test“ in einer kontrollierten Laborsituationen, ohne somit für die teilnehmenden Studierenden ein Anreiz oder Konsequenz zu besitzen. Interessant ist zudem der Hinweis der Autoren, dass die Verhaltensbeobachtung bei der Testdurchführung im Studium keinen Anlass zur Vermutung gab, die Studierenden wären gering motiviert gewesen, die Testverfahren zu absolvieren.

Es stellt sich daher die Frage, inwieweit die Fähigkeitsmessungen in Studien zur Vorhersage der vorklinischen Laborkursleistungen durch die Motivation der Studierenden, räumliche und manuelle Tests bestmöglich zu bearbeiten, beeinflusst sein könnte, falls für die Teilnehmer keine Anreize, wie eine Zulassung zum Studium, bestanden. Daher ist es mit Bezug auf das „Resource Allocation Model“ von Kanfer und Ackerman (1989) nicht auszuschließen (vgl. Unterkapitel 2.4.2.3), dass Studierende mit einer hohen Fähigkeitsausprägungen in einer Laborsituation nur wenig motiviert waren und demzufolge geringere Testergebnisse erzielten. Die beispielsweise von Gansky et al. (2004) geäußerte Kritik an der Sensitivität von manuellen Testverfahren, die Smith (1989) für räumliche Tests übte, erscheint somit post hoc fraglich, da in den Studien die Bearbeitung der Verfahren keine Konsequenzen hatte, weshalb nicht auszuschließen ist, dass eine geringe Motivation der teilnehmenden Studierenden zu einer Verzerrung der Testleistungen führte.

#### **6.4.5.1 Einfluss der Motivation auf die Fähigkeitsmessungen**

In der vorliegenden Untersuchung kann eine Motivation der Bewerber, die zur Messung des räumlichen Vorstellungsvermögens und des instrumentellen, manuellen Geschicks absolvierten Testverfahren nicht bestmöglich zu absolvieren, ausgeschlossen werden. Die Durchführung der Tests fand im Studierendenauswahlverfahren der Zahnmedizin statt, für das die Bewerbung und Teilnahme freiwillig erfolgte, mit dem Ziel einen Studienplatz zu erhalten. Daher ist nicht anzunehmen, dass die Bewerber nicht motiviert waren, bestmögliche Ergebnisse in den die Fähigkeiten messenden Leistungstests zu erzielen. Die deduktiv hergeleitete Annahme bestätigte sich in der Auswertung der Untersuchungsdaten für den HAM-Man (s. Unterkapitel 5.1.2).

Die Wiederholung des Drahtbiegetests in der ersten Studienwoche führte zwar für die 37, über das Auswahlverfahren zugelassenen Studierenden, zu einer signifikanten Verbesserung ihrer Testleistung. Als Ursache für die Leistungsverbesserung ist jedoch weniger eine Steigerung ihres instrumentellen, manuellen Geschicks anzunehmen, da der zeitliche Abstand zwischen den zwei Testungen lediglich circa acht Wochen betrug, ohne dass in diesem Zeitraum ein intensives Training der manuellen Fähigkeit erfolgte. Eher scheinen hierfür die Erfahrungen in der Anwendung von Strategien zur Bearbeitung des Drahtbiegetests und die Vertrautheit im Umgang mit den Testmaterialien ausschlaggebend zu sein.

Bei der Datenanalyse ist jedoch die Installation einer kompetitiven Bedingung für die Testung in der ersten Studienwoche einzubeziehen. Die 63 Erstsemesterstudierenden wurden nicht nur verbal instruiert, den Drahtbiegetest vergleichbar motiviert wie in einer Prüfungssituation zu absolvieren, sondern erhielten den zusätzlichen Anreiz, eine bestmögliche Leistung im HAM-Man zu erzielen, indem die zehn höchsten Ergebnisse mit einem Büchergutschein im Wert von je 50,- Euro prämiert wurden. Die 26 Zahnmedizinistudierenden, die den HAM-Man ohne Vorerfahrung in der ersten Studienwoche absolvierten, erreichten im Vergleich zu den 37 Testwiederholenden zwar signifikant schlechtere Resultate, jedoch unterschied sich ihre Testleistung nicht signifikant im Vergleich zu den 129 Bewerbern im Auswahlverfahren, die den HAM-Man zum ersten Mal bearbeitet hatten.

Unterschiede in der Motivation, den Drahtbiegetest bestmöglich im Auswahlverfahren als auch in der ersten Studienwoche zu bearbeiten, können in der vorliegenden Untersuchung mit Verweis auf die kompetitive Bedingung für die Testung in der ersten Studienwoche ausgeschlossen werden. Demzufolge ist nicht anzunehmen, dass nur

gering motivierte Studierende, die beispielsweise über ein gut ausgeprägtes instrumentelles, manuelles Geschick verfügen, lediglich geringe Leistungen im HAM-Man erzielt hätten. Aufgrund der Stichprobengröße konnte für die Testung in der ersten Studienwoche die positive Verstärkung der Motivation mittels Belohnung nicht durch eine Variation der kompetitiven Bedingung kontrolliert werden, ohne die Bestimmung der Retest-Reliabilität des HAM-Man zu unterminieren.

## 6.5 Limitationen

Die aus dem Verfahren der „Partial Least Squares-Pfadanalyse“ resultierenden Einschränkungen der vorliegenden Untersuchungsergebnisse wurden bereits im Unterkapitel 4.3.3.1 berichtet, sodass keine erneute Aufführung in diesem Abschnitt erfolgt. In Unterkapitel 6.6.1.2 wird für zukünftige Studien eine Empfehlung aufgezeigt, um kovarianzbasierte Strukturgleichungsmodelle zur statistischen Datenauswertung einsetzen zu können.

### 6.5.1 Untersuchungsstichprobe

Die über die vorklinischen Laborkurse immer geringer ausfallende Stichprobe ist zwar repräsentativ für die Studierendenauswahl in der Zahnmedizin, jedoch kann aufgrund der geringen Fallzahlen und der normorientierten Zulassung im Auswahlverfahren, d.h., Bewerber mit höheren Ergebniswerten erhielten einen Studienplatz, nicht ausgeschlossen werden, dass die einbezogenen Untersuchungsvariablen nur über eingeschränkte Wertebereiche verfügten. Diese Beschränkungen des Wertebereichs einer oder mehrerer Variablen könnten zu einer Verringerung statistischer Effektstärken geführt haben, weshalb das Auftreten von „restriction of range“-Effekten für die Ergebnisse der vorliegenden Arbeit nicht auszuschließen ist. Insbesondere der Wertebereich der latenten Variable Schulabschlussnote könnte hiervon betroffen sein, da nur Bewerber mit einer Abiturdurchschnittsnote bis zu 2.3 zum Auswahlverfahren der Medizinischen Fakultät an der Universität Hamburg eingeladen wurden.

Kritisch ist das Fehlen von sieben Studierenden im Phantomkurs I zu betrachten, die am vorherigen TPK-Laborkurs im ersten Semester noch teilgenommen hatten. Zwar ist die Stichprobengröße mit 31 Studierenden für die Auswertung des Prädiktionsmodells über die inkonsistenten, komplexen Aufgaben für das PLS-Verfahren noch ausreichend, aber die fehlende Kenntnis über die Gründe der Nichtteilnahme am Phantomkurs I stellt eine inhaltliche Limitation dar. An dem im fünften Semester zu absolvierenden Phantomkurs II nahmen nur noch 28 der über das Auswahlverfahren zugelassenen Studierenden teil, wodurch die Auflage der Mindeststichprobengröße für die Auswertung des Prädiktionsmodells mit der PLS-Methode verletzt wurde (vgl. Unterkapitel 5.3.1). Vor allem die zeitlich versetzte Anordnung des zweiten Phantomkurses hätte Aufschluss darüber geben können, inwieweit der Abstand von einem Jahr Auswirkungen auf die in den Laborkursen des ersten und zweiten Semesters zu erlernenden zahnmedizinischen

Fertigkeiten gehabt hätte. So hätte einerseits in Erfahrung gebracht werden können, ob eine Prädiktion inkonsistenter, komplexer Aufgabenleistungen auch zweieinhalb Jahre nach Messung der Fähigkeiten im Studierendenauswahlverfahren möglich wäre und andererseits, ob die Aufgabencharakteristik im letzten vorklinischen Laborkurs gegenüber dem Phantomkurs I bestehen bleibt.

### **6.5.2 Messzeitpunkt der praktischen Studienleistung**

Die Benotung der Aufgabenleistungen erfolgte in der vorliegenden Untersuchung primär am finalen Arbeitsstück, dessen überwiegende Anfertigung jedoch in der autonomen Phase des Fertigkeitserwerbs geschah (vgl. Ackerman, 1988). Daher ist fraglich, inwieweit die die Aufgabenleistung ausschlaggebend beeinflussenden Fähigkeiten in der kognitiven und assoziativen Phase noch repräsentiert waren. Bereits das Schnitzen von Wachsähnen erforderte als erste Aufgabe im TPK-Kurs (vgl. Tabelle 4.02 in Unterkapitel 4.1.5.1) mehrere Fertigkeiten, die außer Schnitzen auch das Polieren und Kontrollieren von Maßen zum Anfertigen zweier Wachsähne umfassten. Die zur Beurteilung der Aufgabenleistung herangezogenen Kriterien Morphologie der Krone, Proportionen, Wachsverarbeitung und Politur (s. Tabelle A22 im Anhang) bewerteten aber lediglich den Endzustand der angefertigten Wachsähne, die zudem nur in die Bildung einer einzigen Note eingingen. Würde aber beispielsweise auch das Übertragen der Größe des Zahns mit Hilfe der Schublehre auf den Wachsblock bewertet werden, so wäre zu erwarten gewesen, dass die Leistung in diesem ersten Aufgabenschritt maßgeblich durch das räumliche Vorstellungsvermögen als Fähigkeit der kognitiven Phase beim Erlernen des Schnitzens von Wachsähnen beeinflusst wäre.

### **6.5.3 Güte der latenten Variable räumliches Vorstellungsvermögen**

Die unterschrittenen Güteanforderungen der Skala technische Begabung des LPS von Horn (1983) zur Messung des räumlichen Vorstellungsvermögen, die sich in der Prüfung des Messmodells sowohl in der konfirmatorischen Faktorenanalyse (vgl. Unterkapitel 5.2.3) als auch im PLS-Verfahren zeigten, stellen eine Einschränkung der Untersuchungsergebnisse dar (vgl. Unterkapitel 5.3.1). Im letztgenannten Verfahren führte die konfirmatorische Prüfung der Faktorladungen zu einem Ausschluss des Untertests 10 der Skala technische Begabung im LPS, um die Güte des latenten Konstrukts zu erhöhen, welche aber immer noch nicht den in der Literatur geforderten Akzeptanzniveaus entsprach (s. Unterkapitel 5.3.1.1). Zwar ist die Messqualität noch als

bedingt akzeptabel zu bezeichnen (vgl. Unterkapitel 6.2.2.3), stellt aber eine Einschränkung der vorliegenden Ergebnisse über den prädiktiven Einfluss der kognitiven Fähigkeit räumliches Vorstellungsvermögen auf die Aufgabenleistungen in den vorklinischen Laborkursen dar, vor allem im Hinblick auf die Reproduzierbarkeit der Ergebnisse. Zukünftig sollten zur Messung des räumlichen Vorstellungsvermögens aktuellere Instrumente eingesetzt werden, die den Güteanforderungen eines reliablen und validen Verfahrens besser entsprechen. Als Beispiel ist hier der 3-DW von Gittler (1990) anzuführen, der als bewährter Test zur Messung des räumlichen Vorstellungsvermögens gilt, über eine ökonomische Kurzform verfügt und keine geschlechtsspezifischen Unterschiede aufweist.

#### **6.5.4 Interkulturelle Disparität**

Die zur theoretischen und empirischen Herleitung der Untersuchungshypothesen berichteten Studien zur Vorhersage praktischer Studienleistungen in der Zahnmedizin entstammen überwiegend aus dem angloamerikanischen Raum, beinhalten aber auch Untersuchungen aus Europa, insbesondere aus Großbritannien. Ohne eine fundierte Prüfung, ob länderspezifische Normierungen von Testverfahren auf die deutsche Population zu übertragen sind, stellt dieser Punkt eine Limitation der vorliegenden Arbeit dar. Folglich sind ebenfalls die gewonnenen Erkenntnisse nur mit Einschränkungen auf andere Kulturkreise zu übertragen.

Es ist fraglich, inwieweit die Entwicklung von individuellen Fähigkeiten und Fertigkeiten durch interkulturell abweichende soziale Normen oder Unterschiede in der Sozialisation beeinflusst ist (vgl. Trommsdorff, 2003). Ein Lösungsansatz bestünde in der Etablierung länderübergreifender Normierungen von Testverfahren zur Erfassung manueller und räumlicher Fähigkeiten, sodass die Testergebnisse transformierbar und somit im interkulturellen Kontext vergleichbar wären. Interkulturell bedingte Unterschiede in den Ausbildungsinhalten und -formen der vorklinischen Laborkursaufgaben werden in Unterkapitel 6.6.1.5 diskutiert.

## **6.6 Ausblick und zukünftige Fragestellungen**

### **6.6.1 Empfehlungen für zukünftige Studien**

#### **6.6.1.1 Operationalisierung von Fähigkeiten**

In Studien zur Vorhersage der Aufgabenleistungen in den vorklinischen Laborkursen ist es von entscheidender Bedeutung, nicht nur die zur Operationalisierung einer Fähigkeit herangezogenen Testverfahren zu berichten, sondern unter Berücksichtigung etablierter Definitionen der zu messenden Fähigkeit(en) sicherzustellen, ob die erhobenen Daten auch der vorgegebenen Zielgröße entsprechen. Die Forderung nach einer wissenschaftlichen Überprüfung der Validität wird insbesondere durch die Tatsache unterstrichen, da in der Vergangenheit in Studien zur Messung des manuellen Geschicks und räumlichen Vorstellungsvermögens nicht nur unterschiedliche Testverfahren eingesetzt, sondern ein und dieselben Testverfahren verwendet wurden, um unterschiedliche Fähigkeiten zu messen. Die Ursachen lagen dafür oftmals im Fehlen einer zuverlässigen Konstruktvalidierung in Studien zur Vorhersage der vorklinischen Laborkursleistungen, sodass für die Operationalisierung einer Fähigkeit nicht nur unterschiedliche Definitionen herangezogen wurden, sondern die Prüfung der Güte oftmals nur mittels Augenscheinvalidität erfolgte. Daher ist nicht auszuschließen, dass unterschiedliche Faktoren einer Fähigkeit erfasst, sondern auch unterschiedlich bezeichnet wurden, wodurch die in Unterkapitel 2.3 berichteten widersprüchlichen Studienergebnisse zur Vorhersage der vorklinischen Laborkursleistung bedingt sein könnten.

Manuelles Geschick wurde beispielsweise von Luck et al. (2000) unter anderem mit einem Tremometertest und einem Zweihandsinusoidtest erfasst, während Ganksy et al. (2004) einen Wachsschnitztest einsetzten, Walcott et al. (1986) einen Aufwachstest, Kao et al. (1990) einen Drahtbiegetest und Weinstein und Kiyak (1981) den selbstentwickelten DDT. Der im DAT-Subtest „Carving Dexterity“ eingesetzte Kalkschnitztest misst Zullo (1971) zufolge manuelles Geschick im Umgang mit Handinstrumenten als eine spezifische Fähigkeit, wurde aber ohne Erklärung von Oudshoorn (2003) zusammen mit dem DAT-Subtest PAT zur Operationalisierung psychomotorischer Fähigkeiten herangezogen, für die Boyle und Santelli (1986) den „Crawford Small Parts Dexterity Test“ verwendeten und Raybould et al. (1983) das „Purdue Pegboard“. Auch Yancosek und Howell (2009) stuften in ihrem Review den „Minnesota Rate of Manipulation Test“, das „Purdue Pegboard“, den „O' Connor Finger

Dexterity Test” sowie den „Crawford Small Parts Dexterity Test” und das „Grooved Pegboard” in Testverfahren zur Erfassung psychomotorischer Fähigkeiten ein.

Evans und Dirks (2001) verwendeten zur Messung der räumlichen Fähigkeit den „Revised Minnesota Paper Form Board“, den „Blocks test“ und den „Dimensions test“, hingegen Coy et al. (2003) lediglich den DAT-Subtest PAT, während Hegarty et al. (2009) sehr umfassend in ihrer Untersuchung den „Novel Object Cross-Section test“, den „Tooth-Cross-Section-test“, den „Vandenberg Mental Rotation test“, den DAT-Subtest PAT und den „Visualization of Views Test“ einsetzten. Die Operationalisierung der kognitiven Fähigkeit räumliches Vorstellungsvermögen sollte zudem im gleichen Umfang über Aufgaben zur räumlichen Wahrnehmung, mentalen Rotation und räumlichen Veranschaulichung erfolgen, unabhängig vom Kontext, sei es zur Erfassung der räumlichen Fähigkeit im Auswahlverfahren oder in empirischen Studien zur Prädiktion der vorklinischen Laborkursleistung in der Zahnmedizin, um eine Konfundierung mit dem Geschlecht zu vermeiden (vgl. Unterkapitel 6.4.2.3).

#### **6.6.1.2 Nachverfolgung von Studienabbrüchen**

Zukünftig empfiehlt sich eine Nachverfolgung der Studierenden, die an Laborkursen nicht teilnehmen, um identifizieren zu können, ob hierfür Schwierigkeiten in der Aufgabenbearbeitung im vorherigen praktischen Kurs, die persönliche Studienplanung des Studierenden oder sogar ein Studienabbruch als Ursache in Frage kommen. Im zuletzt genannten Fall wäre es im Hinblick auf den in der vorliegenden Untersuchung eingesetzten Drahtbiegetest von besonderem Interesse gewesen zu erfahren, ob die Anforderungen in den vorklinischen Laborkursen ausschlaggebend für die Nichtteilnahme waren, um die Testergebnisse dieser Studierenden für die Ermittlung von Mindestgrenzen zum erfolgreichen Bestehen des Drahtbiegetests verwenden zu können. Eine somit ermöglichte kriterienorientierte Auswertung der Leistungen im HAM-Man würde einen diagnostischen Einsatz des psychometrischen Verfahrens zu Studienbeginn erlauben, um frühzeitig Studierende zu identifizieren, die aufgrund einer geringen Ausprägung ihres instrumentellen, manuellen Geschicks Schwierigkeiten bei der Aufgabenbearbeitung in den vorklinischen Laborkursen haben könnten, sodass ihnen rechtzeitig gezielte Fördermaßnahmen angeboten würden. Des Weiteren wäre die Installation einer Mindestbestehensgrenze nicht nur im Drahtbiegetest HAM-Man zweckmäßig, sondern generell für alle in der Studierendenauswahl eingesetzten Tests. In kompensatorischen Auswahlverfahren wäre es dann möglich, zukünftige Studierende

mit einer geringen Fähigkeitsausprägung zu identifizieren, welche diese aufgrund von Leistungen in anderen, beispielsweise auf kognitiven Fähigkeiten basierenden Studieneignungstests, ausgleichen konnten. In der Zahnmedizin bieten sich zur Nachverfolgung von Studienabbrechern bzw. der Nichtteilnahme an Laborkursen, aufgrund der überschaubaren Studierendenzahlen halbstrukturierte Telefoninterviews oder per E-Mail zugestellte internetbasierte Fragebögen an, die im Auswahlverfahren oder zu Studienbeginn erteilte Einverständniserklärung der Studierenden vorausgesetzt.

### **6.6.1.3 Erhöhung der Stichprobengröße**

Die Aggregation von Zulassungsdaten im Auswahlverfahren der Hochschule aus mehreren aufeinander folgenden und vergleichbaren Jahren oder Kooperationen mehrerer zahnmedizinischer Fakultäten wären Vorgehen, die Stichprobengröße auf 200 bzw. 250 Untersuchungsteilnehmern zu erhöhen, sodass der Einsatz kovarianzbasierter Strukturgleichungsmodelle zur Prädiktion der Leistungen in den vorklinischen Laborkursen der Zahnmedizin möglich wäre (vgl. Potter & McDonald, 1985). Die Parameterschätzungen eines kovarianzbasiertes Verfahrens weisen eine höhere Konsistenz gegenüber der PLS-Methode auf, in der zudem Messfehler berücksichtigt werden können (Christophersen & Grape, 2006), wodurch eine Limitation der vorliegenden Untersuchungsergebnisse entkräftet wäre. Zugleich können Rückschlüsse auf Abhängigkeitsbeziehungen zwischen zugrundeliegenden latenten Variablen in kovarianzbasierten Modellen gezogen werden (vgl. Chin, 1998a), was im Kontext der vorliegenden Untersuchung vor allem für die Beziehung zwischen den Fähigkeitsarten räumliches Vorstellungsvermögen und instrumentelles, manuelles Geschick interessant wäre, um klären zu können, ob beide Fähigkeiten einander bedingen und somit der Psychomotorik zugeordnet werden könnten.

### **6.6.1.4 Klassifikation der Aufgabencharakteristik**

Im Gegensatz zu den Studien von Gray und Deem (2002a, 2002b), in denen alle Aufgaben der vorklinischen Laborkursen als konsistent bzw. alle der klinischen als inkonsistent eingestuft wurden, erfolgte in der vorliegenden Untersuchung eine Klassifizierung der Aufgabenanforderung als inkonsistent, insofern die Arbeiten am Phantomkopf ausgeführt wurden (vgl. Unterkapitel 4.1.5.2). Die Ausarbeitung eines zuverlässigen Kriterienkatalogs ist für zukünftige Studien empfohlen, um eine zuverlässige Einstufung vorklinischer sowie klinischer Laborkursaufgaben hinsichtlich

des Aufgabentyps konsistent-inkonsistent sowie der Komplexität zu ermöglichen, wofür auch die Einschätzungen Zahnmedizinstudierender heranzuziehen wären. Diese Selbsteinschätzungen würden eine Abstufung der Aufgabenanforderungen ermöglichen, sodass Komplexität und Konsistenz in geringe, moderate und hohe Ausprägungen unterteilt werden könnten, um den Einfluss auf die Beziehungen zwischen Fähigkeiten und der Aufgabenleistung in den einzelnen Phasen des Erlernens von Fertigkeiten genauer zu bestimmen.

#### **6.6.1.5 Operationalisierung der praktischen Studienleistung**

Für nachfolgende Studien empfiehlt sich, eher einzelne und vor allem erste Zwischenschritte der Aufgabenbearbeitung mittels definierter Kriterien in die Bewertung separiert einzubeziehen, um beispielsweise den Einfluss des räumlichen Vorstellungsvermögens auf die Leistung konsistenter Aufgaben in der kognitiven Phase des Erlernens zahnmedizinischer Fertigkeiten ermitteln zu können. Zudem sollten die durch die Bearbeitung einer Laborkursaufgabe zu erlernende(n) zahnmedizinische(n) Fertigkeit(en) in zukünftigen Studien eindeutiger definiert werden, damit nachvollziehbar ist, welche spezifische Fähigkeit zu welchem Zeitpunkt auf welche zu erlernende Fertigkeit einen Einfluss ausübt (vgl. Waldman et al., 1995). So sind Überschneidungen in der Zuordnung zwischen Fähigkeiten und zahnmedizinischen Fertigkeiten zu vermeiden, was in der Vergangenheit zu verzerrten Ergebnisbildern geführt haben könnte, deren Aufklärung nur bedingt möglich war und über die Jahre die widersprüchlichen Studienresultate zur Vorhersage der praktischen Studienleistungen in der vorklinischen Zahnmedizin mit bedingte (Ranney et al., 2005).

Setzt man dieses Vorgehen konsequent für alle Aufgaben in den vorklinischen Laborkursen fort, würde zugleich die Vielfalt der zu erlernenden Untersuchungs- und Behandlungstechniken in den vorklinischen Laborkursen eine Strukturierung erfahren (vgl. Wilson & Husak, 1988), sodass zur Ermittlung prädiktiver Validitäten zukünftig Fehlspezifikationen zwischen Fähigkeiten und zahnmedizinischen Fertigkeiten vermieden werden, die in der Entnahme bzw. dem Abweisen von Instrumenten zur Studierendenauswahl resultieren könnten. In einem weiterführenden Schritt wären die zu erlernenden zahnmedizinischen Fertigkeiten des vorklinischen und klinischen Studienabschnitts in Stufen zu unterteilen, um eine transparente Übersicht zu erhalten, wofür die von Chambers vorgeschlagenen Kompetenzphasen des Erlernens

zahnmedizinischer Fertigkeiten von Zahnmedizinierenden herangezogen werden könnten (vgl. Unterkapitel 2.3.6.3).

Entscheidend ist zudem, zukünftig nicht mehr Mittelwerte zur Operationalisierung der praktischen Studienleistung in der zahnmedizinischen Vorklinik einzusetzen, die über die Beurteilungen mehrerer Laborkursaufgaben oder Kurse gebildet werden. Eine derartige Aggregation von Leistungsdaten steht aus inhaltlicher Sicht im Widerspruch zum Vorgehen etablierter Forschungsmethoden. Es kann nicht ausgeschlossen werden, dass nicht nur in unterschiedlichen Stadien der Entwicklung befindende, sondern auch voneinander unterschiedliche Dentaltechniken zusammengefasst werden, beispielsweise das Aufwachsen von Zähnen und das Präparieren eines Zahns zum Einsetzen einer Füllung. Demzufolge sollte in zukünftigen Untersuchungen der praktische Studienerfolg in Abhängigkeit der einbezogenen Laborkursaufgaben über eine oder mehrere latente endogene Variable(n) operationalisiert werden (vgl. Potter & McDonald, 1985), welche jeweils die in den Laborkursaufgaben zu erlernenden Fertigkeiten repräsentieren, deren manifeste Variablen aus den entsprechend zugehörigen Bewertungskriterien bestehen. Offen ist in diesem Kontext die Frage nach der Struktur der Messmodelle, um den Fertigkeitserwerb in den vorklinischen Laborkursen der Zahnmedizin abzubilden. Allein die Phasen des Fertigkeitserwerbs in Ackermans (1988) Theorie deuten auf eher mehrdimensionale Konstrukte hin, beispielsweise auf ein Drei-Faktoren-Modell oder ein Konstrukt zweiter Ordnung.

Zur Gewährleistung einer interkulturellen Vergleichbarkeit der Studienergebnisse sind in zukünftigen Untersuchungen detaillierter die Ausbildungsinhalte und -methoden der (vor)klinischen Laborkurse einzubeziehen. Kulturell bedingte Unterschiede in der Didaktik, welche Prüfungsbedingungen, das Unterrichten der Untersuchungs- und Behandlungsmethoden und die Integration der (vorklinischen) Laborkurse in den Studienverlauf umfassen, könnten den Erwerb zahnmedizinischer Fertigkeiten von Studierenden unterschiedlich beeinflussen (vgl. Unterkapitel 6.5.4). In früheren, aber auch aktuelleren Studien sind die Laborkursaufgaben zur Operationalisierung der praktischen Studienleistung überwiegend rudimentär beschrieben, sodass interkulturell bedingte Unterschiede und Gemeinsamkeiten in den zu erlernenden zahnmedizinischen Fertigkeiten nicht eindeutig ermittelbar sind. Dadurch wird die Generalisierbarkeit von Studienergebnissen zur Beziehung zwischen Fähigkeiten und dem zahnmedizinischen Fertigkeitserwerb nicht nur auf internationaler Ebene limitiert, sondern auch für die Fakultäten eines Landes.

## **6.6.2 Zukünftige Fragestellungen**

### **6.6.2.1 Phasenverlauf des Fertigkeitserwerbs in Ackermans Modell**

Für die Zuordnung der kognitiven und nichtkognitiven Fähigkeiten auf die einzelnen Lernphasen beim Erwerb von zahnmedizinischen Fertigkeiten sind die Ergebnisse der Untersuchung von Keil und Cortina (2001) weiter zu verfolgen. Die Autoren konnten den von Ackerman (1988) postulierten linearen Verlauf für die Korrelate zwischen Fähigkeiten und Aufgabenleistungen in den Phasen des Fertigkeitserwerbs nicht bestätigen, sondern fanden, dass die Abnahme in der korrelativen Beziehung dem Muster einer Cusp-Katastrophe folgte, die unstete, sprunghafte Veränderungen zwischen mehreren Extremen darstellt. Eine genaue Kenntnis der Phasenübergänge beim Erlernen einer Fähigkeit ist aber Voraussetzung, um theoretische Überlegungen empirisch überprüfen zu können, sodass in diesem Forschungsfeld weitere Studien zur inhaltlichen Klärung der Thematik erforderlich sind.

### **6.6.2.2 Einfluss der Motivation auf den Erwerb zahnmedizinischer Fertigkeiten**

In zukünftigen Studien zur Vorhersage der vorklinischen Laborkursleistungen wäre der Einfluss der Motivation zu überprüfen, um dazu beizutragen, die Beziehungen kognitiver Fähigkeiten auf die Leistung beim Erlernen zahnmedizinischer Untersuchungs- und Behandlungstechniken zuverlässiger analysieren zu können. Suksudaj et al. gingen 2011 in einer Studie der Frage nach, ob die Motivation beim Erlernen zahnmedizinischer Fertigkeiten von Bedeutung ist, indem sie eine direkte Wirkung der Motivation auf die Laborkursleistungen vermuteten. In ihren Untersuchungsergebnissen konnten die Autoren aber keinen korrelativen Zusammenhang zwischen der Motivation Zahnmedizinstudierender und ihrer Leistung beim Präparieren (Bohren) eines Zahnes im Phantomkopf identifizieren (vgl. Unterkapitel 2.4.2.4). Diese Resultate lassen sich aber in das „Resource Allocation Model“ von Kanfer und Ackerman (1989) integrieren, demzufolge Motivation einen mediierenden Einfluss zwischen kognitiven Fähigkeiten und der Leistung beim Fertigkeitserwerb besitzt, indem die zur Bearbeitung einer Aufgabe aufgewendete individuelle Aufmerksamkeitsleistung über distale (zielsetzende) und proximale (selbstregulierende) Motivationsprozesse gesteuert wird (vgl. Unterkapitel 2.4.2.3). Folglich sollte die Motivation keine direkte Kausalität auf die Leistung beim Erwerb zahnmedizinischer Fertigkeiten in den vorklinischen Laborkursen aufweisen, sodass ein

Einschluss als separater Prädiktor in ein Modell zur Vorhersage der praktischen Studienleistung nicht vielversprechend erscheint.

## 7. Literatur

- Ackerman, P.L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, 117(3), 288-318.
- Ackerman, P.L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(5), 883-901.
- Ackerman, P.L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology*, 77(5), 598-614.
- Ackerman, P.L. & Beier, M.E. (2007). Further explorations of perceptual speed abilities in the context of assessment methods, cognitive abilities, and individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, 13(4), 249-272.
- Ackerman, P.L. & Cianciolo, A.T. (2000). Cognitive, perceptual speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, 6(4), 259-290.
- Ackerman, P.L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied*, 1(4), 270-304.
- Albers, S. & Hildebrandt, L. (2006). Methodische Probleme bei der Erfolgsfaktorenforschung - Messfehler, formative versus reflexive Indikatoren und die Wahl des Strukturgleichungs-Modells. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 58(2), 2-33.
- Amthauer, R. (1970). *Intelligenz-Struktur-Test (I-S-T 70)*. Göttingen: Hogrefe.
- Andersen, R. (2008). *Modern methods for robust regression*. Thousand Oaks: Sage.
- Anderson, J.R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22(3), 261-295.
- Arbuckle, J. (2009). Amos 18. Computer software. Chicago: Small Waters.
- Arnold, W.H., Gonzalez, P. & Gaengler, P. (2011). The predictive value of criteria for student admission to dentistry. *European Journal of Dental Education*, 15(4), 236-243.
- Asendorpf, J.B., & Neyer, F.J. (2012). *Psychologie der Persönlichkeit*. 5., vollständig überarbeitete Auflage. Berlin, Heidelberg: Springer.
- Backhaus, K., Erichson, B. & Weiber, R. (2010). *Fortgeschrittene Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. 13. Auflage. Berlin, Heidelberg: Springer.

- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2011). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 13. Auflage, Berlin: Springer.
- Bandura, A. (1986). *Social foundations of thought and action: A cognitive social theory*. Englewood Cliffs (NY): Prentice-Hall.
- Bandura, A. (1990). Self-Regulation of Motivation Through Anticipatory and Self-Reactive Mechanisms. In R.A. Dienstbier (Hrsg.), *Perspectives on Motivation, Nebraska Symposium on Motivation 1990* (S. 69 - 164). Lincoln: University of Nebraska Press.
- Barkowski, D. (1986). *Vorhersage von Studien- und Berufserfolg bei Zahnmedizinern. Längsschnittvalidierung einer apparativen Testbatterie zur Eignungsdiagnose für das Studium der Zahnmedizin*. Dissertation, Friedrich-Wilhelms-Universität, Bonn, Philosophische Fakultät. München: Profil.
- Bartko, J.J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, 19(1), 3-11.
- Beaton, A.E., & Tukey, J.W. (1974). The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data. *Technometrics*, 16(2), 147-185.
- Beck, M., Taft Jr., T.B. & Zimmer, S.E. (1978). Predicting clinical performance: a two-step approach. *Journal of Dental Education*, 42(9), 524-527.
- Beier, U.S., Kapferer, I., Ostermann, H., Staudinger, R. & Dumfahrt, H. (2010). Impact of a Novel Dental School Admission Test on Student Performance at Innsbruck Medical University, Austria. *Journal of Dental Education*, 74(5), 531-538.
- Bellanti, N.D., Mayberry, W.E. & Tira, D.E. (1972). Relation between selected predictor variables and grades in fixed prosthodontics laboratory. *Journal of Dental Education*, 36(12), 16-21.
- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246.
- Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588.
- Bergkvist, L. & Rossiter, J.R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, 44(2), 175-184.
- Bhattacharya, S.K., Roy, A. & Chatterjee, S.K. (1978). Comparison of Motor-Manual Performance in Quiet and Noisy Environments. *Indian Journal of Medical Research*, 68(1), 306-311.
- Bliemel, F., Eggert, A., Fassott, G. & Henseler, J. (Hrsg.) (2005). *Handbuch PLS-Pfadmodellierung: Methoden, Anwendung, Praxisbeispiele*. Stuttgart: Schäffer-Poeschel.

- Blommers, P. (1956). Prediction of success in the first and second years of the study of dentistry on basis of certain selected variables. *Journal of Dental Education*, 20(1), 5-16.
- Bollen, K.A. & Lennox, R. (1991). Conventional Wisdom on Measurement: A Structural Equation Perspective. *Psychological Bulletin*, 110(2), 305-314.
- Bollen, K.A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1), 605-634.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation*. 4., überarbeitete Auflage. Heidelberg: Springer.
- Boyd, M.A. & Teteruck, W.R. (1979). The Dental Aptitude Test - its application in admissions. *Journal of the Canadian Dental Association*, 45(8), 414-419.
- Boyd, M.A., Teteruck, W.R. & Thompson, G.W. (1980). Interpretation and use of the dental admission and aptitude tests. *Journal of Dental Education*, 44(5), 275-278.
- Boyle, A. & Santelli, J. (1986). Assessing psychomotor skills: the role of the Crawford Small Parts Dexterity Test as a screening instrument. *Journal of Dental Education*, 50(3), 176-179.
- Brandstätter, H. (2008). Validität kognitiver Tests als Funktion der Valenz von Test- und Studiensituation. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 43-54). Göttingen: Hogrefe.
- Brandstätter, H. & Farthofer, A. (2002). Studienerfolgsprognose - konfigurativ oder linear additiv?. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 23(4), 381-391.
- Breivik, E. & Olsson, U.H. (2001). Adding variables to improve fit: the effect of model size on fit assessment in LISREL. In R. Cudeck, S. du Toit & D. Sörbom (Hrsg.), *Structural equation modeling: Present and future. A Festschrift in honor of Karl Jöreskog* (S. 169-194). Lincoln Wood (IL): Scientific Software International.
- Brigante, R. & Lamb, R. (1968). Perception and control test: the dental technical aptitude test of the future? *Journal of Dental Education*, 32(3), 340-354.
- Brosius, F. (2011). *SPSS 19*. Heidelberg u.a.: mitp.
- Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37(1), 62-83.
- Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Hrsg.), *Testing Structural Equation Models* (S. 136-162). Beverly Hills (CA): Sage.
- Buffington, C.W., MacMurdo, S.D. & Ryan, C.M. (2006). Body position affects manual dexterity. *Anesthesia and Analgesia*, 102(6), 1879-1883.

- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. 3., überarbeitete Auflage. München: Pearson.
- Bundesministerium für Bildung und Forschung (2007). *Hochschulrahmengesetz in der Fassung der Bekanntmachung vom 19. Januar 1999* (BGBl. I, S. 18), zuletzt geändert durch Artikel 2 des Gesetzes vom 12. April 2007 (BGBl. I S. 506). Berlin: Bundesministerium für Bildung und Forschung.
- Buyse, T. & Lievens, F. (2011). Situational judgment tests as a new tool for dental student selection. *Journal of Dental Education*, 75(6), 743-749.
- Cassel, C.M., Hackl, P. & Westlund, A.H. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics*, 26(4), 435-446.
- Chamberlain, T.C., Catano, V.M. & Cunningham, D.P. (2005). Personality as a predictor of professional behavior in dental school: comparisons with dental practitioners. *Journal of Dental Education*, 69(11), 1222-1237.
- Chambers, D.W. (1972). Discrimination indexes for improving dental school admissions. *Journal of Dental Education*, 36(1), 32-41.
- Chambers, D.W. (1993). Toward a competency-based curriculum. *Journal of Dental Education*, 57(11), 790-793.
- Chambers, D.W. & Geissberger, M. (1997). Toward a competency analysis of operative dentistry technique skills. *Journal of Dental Education*, 61(10), 795-803.
- Chambers, D.W. & Glassman, P. (1997). A primer on competency-based evaluation. *Journal of Dental Education*, 61(8), 651-666.
- Chen, M.K. & Podshadley, D.W. (1967). A factorial study of some psychological, vocational interest, and mental ability variables as predictors of success in dental school. *Journal of Applied Psychology*, 51(3), 236-241.
- Chin, W.W. (1998a). Issues and opinion on structural equation modelling. *MIS Quarterly*, 22(1), vii-xvi.
- Chin, W.W. (1998b). The partial least squares approach to structural equation modelling. In G.A. Marcoulides (Hrsg.), *Modern methods for business research* (S. 295-336). Mahwah (NJ): Lawrence Erlbaum Associates.
- Chin, W.W. & Newsted, P.R. (1999). Structural equation modelling analysis with small samples using partial least squares. In R.H. Hoyle (Hrsg.), *Statistical strategies for small sample research* (S. 307-341). Thousand Oaks (CA): Sage.
- Chin, W.W., Marcolin, B.L. & Newsted, P.R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research*, 14(2), 189-217.

- Christophersen, T. (2007). *Usability im Online-Shopping: Entwicklung eines Fragebogeninstrumentes (ufosV2) unter Berücksichtigung formativer und reflektiver Messmodelle*. Dissertation, Christian-Albrechts-Universität zu Kiel (CAU), Institut für Betriebswirtschaftslehre.
- Christophersen, T. & Grape, C. (2006). Die Erfassung latenter Konstrukte mit Hilfe formativer und reflektiver Messmodelle. In S. Albers, D. Klapper, U. Konradt, A. Walter, J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (S. 311-326). Wiesbaden: Deutscher Universitäts-Verlag.
- Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290.
- Clauß, G. (Hrsg.) (1995). *Fachlexikon ABS Psychologie*. 5., völlig überarbeitete Auflage. Thun, Frankfurt am Main: Deutsch.
- Colom, R., Contreras, M.J., Botella, J. & Santacreu, J. (2002). Vehicles of spatial ability. *Personality and Individual Differences*, 32(5), 903-912.
- Cortina, K.S., Baumert, J., Leschinsky, A., Mayer, K.U. & Trommer, L. (2008). *Das Bildungswesen in der Bundesrepublik Deutschland*. Reinbek: Rowohlt.
- Coy, K., McDougall, H. & Sneed, M. (2003). Issues regarding practical validity and gender bias of the Perceptual Abilities Test (PAT). *Journal of Dental Education*, 67(1), 31-37.
- Curtis, D.A., Lind, S.L., Plesh, O. & Finzen, F.C. (2007). Correlation of admissions criteria with academic performance in dental students. *Journal of Dental Education*, 71(10), 1314-1321.
- deAndres, A.G., Sanchez, E., Hilgado, J.J. & Diaz, M.J. (2004). Appraisal of psychomotor skills of dental students at University Complutense of Madrid. *European Journal of Dental Education*, 8(1), 24-30.
- Deary, I.J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, 11(3), 201-211.
- Derevere, R.E. (1961). Comparison of dental aptitude tests with achievement in operative dentistry. *Journal of Dental Education*, 25(1), 50-56.
- Desrosiers, J., Hebert, R., Bravo, G. & Dutil, E. (1995). The Purdue Pegboard test: normative data for people aged 60 and over. *Disability and Rehabilitation*, 17(5), 217-224.
- Deubert, L.W., Smith, M.C., Jenkins, C.B. & Berry, D.C. (1975). The selection of dental students. A pilot study of an assessment of potential manual ability by psychometric tests. *British Dental Journal*, 139(5), 167-170.
- Diamantopoulos, A. (1994). Modelling with LISREL: A guide for the uninitiated. *Journal of Marketing Management*, 10(1-3), 105-136.

- Diamantopoulos, A. (2012). Formative indicators: Introduction to the special issue. *Journal of Business Research*, 61(12), 1201-1202.
- Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P. & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, 40(3), 434-449.
- Dillon, W.R., Kumar, A. & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin*, 101(1), 126-135.
- Downey, M.C., Collins, M.A. & Browning, W.D. (2002). Predictors of success in dental hygiene education: a six-year review. *Journal of Dental Education*, 66(11), 1269-1273.
- Dworkin, S.F. (1970). Dental aptitude test as performance predictor over four years of dental school: analysis and interpretation. *Journal of Dental Education*, 34(1), 28-38.
- Ekstrom, R.B., French, J.W., Harman, H.H. & Dermen, D. (1976). *Kit of factor-referenced cognitive tests*. Princeton (NJ): Educational Testing Service.
- Eichner, K. & Kappert, H.F. (2005). *Zahnärztliche Werkstoffe und ihre Verarbeitung. 1. Grundlagen und Verarbeitung*. Stuttgart: Thieme.
- Evans, J.G. & Dirks, J.D. (2001). Relationships of admissions data and measurements of psychological constructs with psychomotor performance of dental technology students. *Journal of Dental Education*, 65(9), 874-882.
- Evermann, J. & Tate, M. (2010). Testing Models or Fitting Models? Identifying Model Misspecification in PLS. *ICIS 2010 Proceedings*, Papier 21, verfügbar unter [http://aisel.aisnet.org/icis2010\\_submissions/21](http://aisel.aisnet.org/icis2010_submissions/21).
- Eyring, J.D., Johnson, D.S. & Francis, D.J. (1993). A cross-level units-of-analysis approach to individual differences in skill acquisition. *Journal of Applied Psychology*, 78(5), 805-814.
- Farber, S.D. (1991). Assessing neuromotor performance enablers. In C. Christiansen & C. Baum (Hrsg.), *Occupational Therapy. Overcoming Human Performance*. Thorofare (NJ): Slack.
- Farrell, J.N. & McDaniel, M.A. (2001). The stability of validity coefficients over time: Ackerman's (1988) model and the General Aptitude Test Battery. *Journal of Applied Psychology*, 86(1), 60-79.
- Fenlon, M.R., McCartan, B.E., Sheriff, M. & Newton, T. (2001). Personality of dental students in two dental schools in the United Kingdom and in Ireland. *European Journal of Dental Education*, 5(4), 173-176.

- Fernandez-Pabon, J.J. (1968). Prediction of success in dental school on the basis of dental aptitude test scores and other variables. *Journal of Dental Education*, 32(3), 261-271.
- Fissini, H. (2004). *Lehrbuch der Psychologischen Diagnostik*. 3. vollständig überarbeitete und erweiterte Auflage. Göttingen: Hogrefe.
- Fitts, P.M. (1964). Perceptual-motor skill learning. In A.W. Melton (Hrsg.), *Categories of Human Learning* (S. 243-285). London: Academic Press.
- Fitts, P.M. & Posner, M.I. (1967). *Human Performance*. Belmont (CA): Brooks-Cole.
- Fleishman, E.A. (1972). On the relation between abilities, learning, and human performance. *American Psychologist*, 27(11), 1017-1032.
- Fleishman, E.A. & Ellison, G.D. (1962). A factor analysis of fine manipulative tests. *Journal of Applied Psychology*, 46(2), 96-105.
- Fleishman, E.A. & Hempel Jr., W.E. (1955). The relation between abilities and improvement with practice in a visual discrimination reaction task. *Journal of Experimental Psychology*, 49(5), 301-312.
- Fleishman, E.A. & Reilly, M.E. (1992). *Handbook of human abilities: definitions, measurements, and job task requirements*. Palo Alto (CA): Consulting Psychologists Press.
- Fornell, C., & Bookstein, F.L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, 19(4), 440-452.
- Fornell, C. & Cha, J. (1994). Partial Least Squares. In R.P. Bagozzi (Hrsg.), *Advanced Methods of Marketing Research* (S. 52-78), Cambridge (MA): Blackwell.
- Fornell, C. & Larcker, D.F. (1981): Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39-50.
- Gagne, P. & Hancock, G.R. (2006). Measurement Model Quality, Sample Size, and Solution Propriety in Confirmatory Factor Models. *Multivariate Behavioral Research*, 41(1), 65-83.
- Gallagher, J.E., Niven, V., Donaldson, N. & Wilson, N.H.F. (2009). Widening access? Characteristics of applicants to medical and dental schools, compared with UCAS. *British Dental Journal*, 207(9), 433-445.
- Gansky, S.A., Pritchard, H., Kahl, E., Mendoza, D., Bird, W., Miller, A.J. & Graham, D. (2004). Reliability and validity of a manual dexterity test to predict preclinical grades. *Journal of Dental Education*, 68(9), 985-994.
- Gates, A.I. (1917). Recitation as a factor in memorizing. In R.S. Woodworth (Hrsg.), *Archives of psychology* (S. 1-104). New York: Science Press.

- Gayen, A.K. (1951). The Frequency Distribution of the Product-Moment Correlation Coefficient in Random Samples of Any Size Drawn from Non-Normal Universes. *Biometrika*, 38(1/2), 219-247.
- Gefen, D., Straub, D.W. & Boudreau, M.C. (2000). Structural equation modelling and regression: Guidelines for research practice. *Communications of the Association for Information Systems*, 4(7), 1-79.
- Geisser, S. (1974). A predictive approach to the random effect model. *Biometrika*, 61(1), 101-107.
- Gentsch, S. & Gold, A. (2008). Studierendenauswahl durch Interviews - Ergebnisse einer Pilotstudie. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 156-167). Göttingen: Hogrefe.
- Gibbons, R.D. & Hedeker, D.R. (1992). Full-information item bi-factor analysis. *Psychometrika*, 57(3), 423-436.
- Ginley, T. (1966). Present status and future plans of the dental aptitude testing program. *Journal of Dental Education*, 30(2), 163-174.
- Gittler, G. (1990). *Dreidimensionaler Würfeltest (3DW) - Ein Rasch-skaliertes Test zur Messung des räumlichen Vorstellungsvermögens*. Weinheim: Beltz Test.
- Giuliani, M., Lajolo, C., Clemente, L., Querqui, A., Viotti, R., Boari, A. & Miani, C.M. (2007). Is manual dexterity essential in the selection of dental students?. *British Dental Journal*, 203(3), 149-155.
- Gottfredson, L.S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24(1), 13-23.
- Graham, J.W. (1972). Substitution of perceptual-motor ability test for chalk carving in dental admission testing program. *Journal of Dental Education*, 36(1), 9-14.
- Graham, J.W. (1974). Factor analysis of the perceptual-motor ability test. *Journal of Dental Education*, 38(1), 16-19.
- Gray, S.A. & Deem, L.P. (2002a). Predicting student performance in preclinical technique courses using the theory of ability determinants of skilled performance. *Journal of Dental Education*, 66(6), 721-727.
- Gray, S.A. & Deem, L.P. (2002b). Are traditional cognitive tests useful in predicting clinical success? *Journal of Dental Education*, 66(11), 1241-1245.
- Grewal, R., Cote J.A. & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science*, 23(4), 519-529.
- Guay, R.B. (1976). *Purdue Spatial Visualization Test*. West Lafayette (IN): Purdue Research Foundation.
- Haenlein, M. & Kaplan, A.M. (2004). A beginner's guide to partial least squares (PLS) analysis. *Understanding Statistics*, 3(4), 283-297.

- Hair, J.F., Ringle, C.M. & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, 18(2), 139-152.
- 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). *Zeitschrift für Medizinische Ausbildung*, 25(2), doc82.
- Harris, A.J. (1937). The relative significance of measures of mechanical aptitude, intelligence, and previous scholarship for predicting achievement in dental school. *Journal of Applied Psychology*, 21(5), 513-521.
- Hartigan, J.A. & Wigdor, A.K. (1989). *Fairness in Employment Testing: Validity Generalization, Minority Issues, and the General Aptitude Test Battery*. Washington (DC): National Academy Press.
- Hasselhorn, M. & Gold, A. (2009). *Pädagogische Psychologie: Erfolgreiches Lernen und Lehren*. 3. Auflage. Stuttgart: Kohlhammer.
- Hauser, A.M. & Bowen, D.M. (2009). Primer on preclinical instruction and evaluation. *Journal of Dental Education*, 73(3), 390-398.
- Hegarty, M., Keehner, M., Khooshabeh, P. & Montello, D.R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences*, 19(1), 61-70.
- Heintze, U., Radeborg, K., Bengtsson, H. & Stenlås, A. (2004). Assessment and evaluation of individual prerequisites for dental education. *European Journal of Dental Education*, 8(4), 152-160.
- Heldmann, W. (1984). *Studierfähigkeit: Ergebnisse einer Umfrage*. Göttingen: Schwarz
- Hell, B., Trapmann, S. & Schuler, H. (2007). Eine Metaanalyse der Prognosekraft von Studierfähigkeitstests. *Empirische Pädagogik*, 21(3), 251-270.
- Hell, B., Trapmann, S. & Schuler, H. (2008). Synopse der Hohenheimer Metaanalysen zur Prognostizierbarkeit des Studienerfolgs und Implikationen für die Auswahl- und Beratungspraxis. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 43-54). Göttingen: Hogrefe.
- Henry, R.A. & Hulin, C.L. (1989). Changing Validities: Ability-Performance Relations and Utilities. *Journal of Applied Psychology*, 74(2), 365-367.
- Henseler, J., Ringle, C. & Sinkovics, R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing (AIM)*, 20, 277-320.
- Herrmann, A., Huber, F. & Kressmann, F. (2006). Varianz- und kovarianzbasierte Strukturgleichungsmodelle - ein Leitfaden zu deren Spezifikation, Schätzung und Beurteilung. *Zeitschrift für betriebswirtschaftliche Forschung*, 58(2), 34-66.

- Heublein, U., Schmelzer, R., Sommer, D. & Wank, J. (2008). *Die Entwicklung der Schwund- und Studienabbruchquoten an den deutschen Hochschulen*. HIS Projektbericht. Berlin: Bundesministerium für Bildung und Forschung.
- Heublein, U., Hutzsch, C., Schreiber, J., Sommer, D. & Besuch, G. (2010). *Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen. Ergebnisse einer bundesweiten Befragung von Exmatrikulierten des Studienjahres 2007/08*. HIS Projektbericht. Berlin: Bundesministerium für Bildung und Forschung.
- Heuer, H. (2006). Arbeitsbewegungen und motorische Fertigkeiten. In B. Zimolong & U. Konradt (Hrsg.), *Ingenieurpsychologie* (S. 71-103). Göttingen: Hogrefe.
- Hissbach, J., Klusmann, D. & Hampe, W. (2011). Reliabilität des Hamburger Auswahlverfahrens für Medizinische Studiengänge, Naturwissenschaftsteil (HAM-Nat). *Zeitschrift für Medizinische Ausbildung*, 28(3), doc44.
- Hitpass, J. (1978). Hochschulzulassung - Besonderer Auswahltest Zahnmedizin (BATZ). *Zeitschrift für experimentelle und angewandte Psychologie*, 25(1), 75-94.
- Holliday, H. & Koepchen, E.E. (1954). Visual Tests as a Criterion for the Selection of Dental Students. *Journal of Dental Education*, 18, 7-17.
- Holmes, D.C., Doering, J.V. & Spector, A. (2008). Associations among pre dental credentials and measures of dental school achievement. *Journal of Dental Education*, 72(2), 142-152.
- Homburg, C. & Giering, A. (1996). Konzeptualisierung und Operationalisierung komplexer Konstrukte - Ein Leitfaden für die Marketingforschung. *Marketing - Zeitschrift für Forschung und Praxis*, 18(1), 5-24.
- Homburg, C., Klarmann, M. & Pflesser, C. (2008). Konfirmatorische Faktorenanalyse. In A. Herrmann, C. Homburg, & M. Klarmann (Hrsg.), *Handbuch Marktforschung: Methoden - Anwendungen - Praxisbeispiele* (S. 271-304). 3. Auflage. Wiesbaden: Gabler-Verlag.
- Hood, A.B. (1963). Predicting achievement level in dental school. *Journal of Dental Education*, 27(2), 148-155.
- Horn, W. (1956). *B-T-S Begabungstestsystem*. 1. Auflage. Göttingen: Hogrefe.
- Horn, W. (1962). *Das Leistungsprüfsystem (L-P-S)*. Göttingen: Hogrefe.
- Horn, W. (1983). *L-P-S Leistungsprüfsystem*. 2. Auflage. Göttingen: Hogrefe.
- Hoyos, C.G. (1974). *Arbeitspsychologie*. Stuttgart: Kohlhammer.
- Huber, P.J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1), 73-101.

- Hui, B.S. & Wold, H. (1982). Consistency and Consistency at Large of Partial Least Squares Estimates. In K.G. Jöreskog & H. Wold (Hrsg.), *Systems under Indirect Observations: Causality, Structure, Prediction, Part 2* (S. 119-130), Amsterdam, New York, u.a.: North-Holland.
- Hulland, J. (1999). Use of Partial Least Squares (PLS) in Strategic Management Research: A Review of Four Recent Studies. *Strategic Management Journal*, 20(2), 195-204.
- Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29(3), 340-362.
- Jann, B. (2010). Robuste Regression. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 707-740). Wiesbaden: Verlag für Sozialwissenschaften.
- Jarvis, C., Mac Kenzie, S.B. & Podsakoff, P.M. (2005). The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions. *Journal of Applied Psychology*, 90(4), 710-730.
- Johnston, P.J. & Catano, V.M. (2002). Psychomotor Abilities Tests as Predictors of Training Performance. *Canadian Journal of Behavioural Science*, 34(2), 75-83.
- Jöreskog, K.G. & Sorbom, D. (1981). *LISREL V: Analysis of linear structural relations by the method of maximum likelihood*. Chicago: International Educational Services.
- Kanfer, R. & Ackerman, P.L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, 74(4), 657-690.
- Kanfer, R., Wolf, M.B., Kantrowitz, T.M. & Ackerman, P.L. (2010). Ability and trait complex predictors of academic and job performance: A person-situation approach. *Applied Psychology: An International Review*, 59(1), 40-69.
- Kao, E.C., Ngan, P.W., Wilson, S. & Kunovich, R. (1990). Wire-bending test as a predictor of preclinical performance by dental students. *Perceptual and Motor Skills*, 71(2), 667-673.
- Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research*, 23(1), 69-86.
- Kay, E., Bennett, J., Allison, P. & Coombes, L.R. (2010). Evidence-informed dental student recruitment techniques. *British Dental Journal*, 208(3), 127-131.
- Keil, C.T. & Cortina, J.M. (2001). Degradation of validity over time: A test and extension of Ackerman's model. *Psychological Bulletin*, 127(5), 673-697.
- Keith, N., Richter, T. & Naumann, J. (2010). Active/exploratory training promotes transfer even in learners with low motivation and cognitive ability. *Applied Psychology: An International Review*, 59(1), 97-123.

- Kline, R.B. (2010). *Principles and practice of structural equation modeling*. 3. Auflage. New York (NY): The Guilford Press.
- Knight, G.W., Guenzel, P.J. & Fitzgerald, M. (1990). Teaching recognition skills to improve products. *Journal of Dental Education*, 54(12), 739-742.
- Konegen-Grenier, C. (2002). Studierfähigkeit und Hochschulzugang. *Forschung & Lehre*, 9/2002, 481-483.
- Kramer, G.A. (1986). Predictive validity of the Dental Admission Test. *Journal of Dental Education*, 50(9), 526-531.
- Kramer, G.A., Kubiak, A.T. & Smith, R.M. (1989). Construct and predictive validities of the Perceptual Ability Test. *Journal of Dental Education*, 53(2), 119-125.
- Kreit, L.H. & McDonald, R.E. (1968). Preprofessional grades and the dental aptitude test as predictors of student performance in dental school. *Journal of Dental Education*, 32(4), 452-457.
- Kreuzpointner, L. (2010). *Bedingungen für die Äquivalenz von Papier-Bleistift-Version und Computerversion bei Leistungstests*. Dissertation, Universität Regensburg, Institut für Psychologie, Lehrstuhl für Psychologie VI (Pädagogische Psychologie und Medienpsychologie).
- Kreuzpointner, L., Lukesch, H. & Horn, W. (2013). *Leistungsprüfsystem 2. LPS-2. Manual*. Göttingen: Hogrefe.
- Krumm, S., Ziegler, M. & Bühner, M. (2008). Reasoning and working memory as predictors of school grades. *Learning and Individual Differences*, 18(2), 248-257.
- Lehmann, K.M., Hellwig, E. & Wenz, H.J. (2008). *Zahnärztliche Propädeutik: Einführung in die Zahnheilkunde*. 12., vollständig überarbeitete und erweiterte Auflage. Köln: Deutscher Ärzteverlag.
- Lewin, D. & Lischka, I. (2004). Passfähigkeit - ein neuer Ansatz für den Hochschulzugang. *Die Hochschule*, 2/2004, 81-95.
- Lienert, G.A. (1958). *F-L-T, Formlegetest, Handanweisung*. Hogrefe: Göttingen.
- Lienert, G.A. (1961). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lienert, G.A. (1967). *D-B-P Die Drahtbiegeprobe*. 2. Auflage. Hogrefe: Göttingen.
- Lienert, G.A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. 6. Auflage. Weinheim: Beltz.
- Lievens, F., Buyse, T. & Sackett, P.R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, 90(3), 442-452.
- 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(4), 245-257.

- Lievens, F., Peeters, H. & Schollaert, E. (2008). Situational judgment tests: a review of recent research. *Personnel Review*, 37(4), 426-441.
- Linder, F., Janus, C.E., Bauer, C. & Dishman, M.V. (1991). *The Relationship of Cognitive Style to Academic Performance among Dental Students*. Artikel präsentiert auf dem jährlichen Treffen der „Eastern Educational Research Association“, Boston (MA).
- Linn, M.C. & Peterson, A.C. (1985). Emergence and characterization of sex-differences in spatial ability: a meta-analysis. *Child Development*, 56(6), 1479-1498.
- Little, T.D., Cunningham, W.A., Shahar, G. & Widaman, K.F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9(2), 151-173.
- Lohmöller, J.B. (1989). *Latent variable path modelling with partial least squares*. Heidelberg: Physica-Verlag.
- Luck, O., Reitemeier, B. & Scheuch, K. (2000). Testing of fine motor skills in dental students. *European Journal of Dental Education*, 4(1), 10-14.
- Lundergan, W.P., Soderstrom, E.J. & Chambers, D.W. (2007). Tweezer dexterity aptitude of dental students. *Journal of Dental Education*, 71(8), 1090-1097.
- Lynch, C.D., McConnell, R.J. & Hannigan, A. (2006). Dental school admissions in Ireland: can current selection criteria predict success? *European Journal of Dental Education*, 10(2), 73-79.
- Maassen, G.H. & Bakker, A.B. (2001). Suppressor variables in path models definitions and interpretations. *Sociological Methods & Research*, 30(2), 241-270.
- MacCallum, R.C. & Browne, M.W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, 114(3), 533-541.
- Mack, W. (1996). Expertise und Intelligenz. In H. Gruber & A. Ziegler (Hrsg.), *Expertiseforschung: Theoretische und empirische Grundlagen* (S. 92-114). Wiesbaden: Verlag für Sozialwissenschaften.
- Maeda, Y. & Yoon, S.Y. (2013). A Meta-Analysis on Gender Differences in Mental Rotation Ability Measured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT: R). *Educational Psychology Review*, 25(1), 69-94.
- Manhold, J.H. & Manhold, B.S. (1965). Final report of an eight-year study of the efficacy of the dental aptitude test in predicting four-year performance in a new dental school. *Journal of Dental Education*, 29(1), 41-45.
- Manhold, J.H. & Manhold, B.S. (1967). Predictive value for four-year performance of individual parts of the dental aptitude test. *Journal of Dental Education*, 31(1), 62-70.

- Maronna, R.A., Martin, D.R. & Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*. Chichester: Wiley.
- Marsh, H.W., Balla, J.R. & McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391-410.
- Mathieson, K., Peacock, E., & Chin, W.W. (2001). Extending the technology acceptance model: the influence of perceived user resources. *ACM SIGMIS Database - Special issue on adoption, diffusion, and infusion of IT*, 32(3), 86-112
- McDaniel, M.A. (1985). *The evaluation of a causal model of job performance: The interrelationships of general mental ability, job experience, and job performance*. Unveröffentlichte Dissertation, George Washington Universität, Washington (DC).
- McDonald, R.P. (1996). Path analysis with composite variables. *Multivariate Behavioral Research*, 31(2), 239-270.
- McGraw, K.O. & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46.
- McPhee, S.D. (1987). Functional hand evaluations: a review. *The American Journal of Occupational Therapy*, 41(3), 158-163.
- Meade, A.W. & Kroustalis, C.M. (2006). Problems With Item Parceling for Confirmatory Factor Analytic Tests of Measurement Invariance. *Organizational Research Methods*, 9(3), 369-403.
- Moosbrugger, H. & Kelava, A. (2008). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.
- Moosbrugger, H. & Schermelleh-Engel, K. (2008). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 307-324). Heidelberg: Springer.
- Muthén, L.K. & Muthén, B.O. (2002). How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4), 599-620.
- Natkin, E. & Guild, R.E. (1967). Evaluation of preclinical laboratory performance: a systematic study. *Journal of Dental Education*, 31(2), 152-161.
- Newton, J.T., Buck, D., Thorogood, N. & Skelly, A.M. (2003). Admission as a dental student to the former UMDS and its relationship to socio-demographic characteristics. *British Dental Journal*, 195(4), 211-217.
- Nijenhuis, J.T. & van der Flier, H. (2002). The correlation of g with attentional and perceptual-motor ability tests. *Personality and Individual Differences*, 33(2), 287-297.

- Oudshoorn, W.C. (2003). The utility of Canadian DAT Perceptual Ability and Carving Dexterity scores as predictors of psychomotor performance in first-year operative dentistry. *Journal of Dental Education*, 67(11), 1201-1208.
- Oswald, F.L., Schmitt, N., Kim, B.H., Ramsay, J. & Gillespie, M.A. (2004). Developing a biodata measure and a situational judgment inventory as predictors of college students' performance. *Journal of Applied Psychology*, 89(2), 187-207.
- Panten, G. & Boßow-Thies, S. (2006). Analyse kausaler Wirkungszusammenhänge mit Hilfe von Partial Least Squares (PLS). In S. Albers, D. Klapper, U. Konradt, A. Walter & J. Wolf (Hrsg.), *Methodik der empirischen Forschung* (S. 311-326). Wiesbaden: Deutscher Universitäts-Verlag.
- Pawlik, K. (1979). Hochschulzulassungstests: Kritische Anmerkungen zu einer Untersuchung von Hitpass und zum diagnostischen Ansatz. *Psychologische Rundschau*, 30(1), 19-33.
- Peters, M., Lehmann, W., Takahira, S., Takeuchi, Y. & Jordan, K. (2006). Mental Rotation Test Performance in Four Cross-Cultural Samples (N = 3367): Overall Sex Differences and the Role of Academic Program in Performance. *Cortex*, 42(7), 1005-1014.
- Peterson, S. (1974). The ADA chalk carving test. *Journal of Dental Education*, 38(1), 11-15.
- Phipps, G., Fishman, R. & Scott, R.H. (1968). Prediction of success in a dental school. *Journal of Dental Education*, 32(2), 161-170.
- Pixner, J. & Schüpbach, H. (2008). Chancen und Grenzen von Situational Judgment Inventories für die Studierendenauswahl. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 139-146). Göttingen: Hogrefe.
- Poirier, F. (1987). Dexterity as a valid measure of hand function: A pilot study. *Occupational Therapy in Health Care*, 4(3-4), 69-83.
- Potter, R., McDonald, R. & Sagraves, R.E. (1982a). Professional performance among dental students: a factor analysis. *Journal of Dental Education*, 46(4), 216-220.
- Potter, R., McDonald, R. & Sagraves, R.E. (1982b). A derived basic ability criterion for predicting dental students' performance. *Journal of Dental Education*, 46(11), 634-638.
- Potter, R. & McDonald, R. (1985). Use and application of structural models in dental education research. *Journal of Dental Education*, 49(3), 145-153.
- Poole, A., Catano, V.M. & Cunningham, D.P. (2007). Predicting performance in Canadian dental schools: the new CDA structured interview, a new personality assessment, and the DAT. *Journal of Dental Education*, 71(5), 664-676.
- Quaiser-Pohl, C. (2001). Räumliches Denken - Entwicklung, Erfassung und praktische Bedeutung. *Psychologie in Erziehung und Unterricht*, 48(4), 241-248.

- Quaiser-Pohl, C., Lehmann, W. & Schirra, J. (2001). Sind Studentinnen der Computervisualistik besonders gut in der Raumvorstellung? Psychologische Aspekte bei der Wahl eines Studienfachs. *FIF Kommunikation*, 18, 42-46.
- Ranney, R.R., Wilson, M.B. & Bennet, R.B. (2005). Evaluation of applicants to predoctoral dental education programs: review of the literature. *Journal of Dental Education*, 69(10), 1095-1106.
- Raybould, T.P., Raggard, D.C. & Norton, J.C. (1983). Psychomotor skills and technical ability in dental school. *Journal of Dental Education*, 47(9), 594-598.
- Reise, S.P., Moore, T.M. & Haviland, M.G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544-559.
- Reise, S.P., Morizot, J. & Hays, R.D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(1), 19-31.
- Ringle, C.M. (2004). Gütemaße für den Partial-least-squares-Ansatz zur Bestimmung von Kausalmodellen. Universität Hamburg, Institut für Industriebetriebslehre und Organisation, *Industrielles Management*, Arbeitspapier 16, 1-30.
- Ringle, C.M., Wende, S. & Will, S. (2005). SmartPLS 2.0 (M3) Beta, Hamburg 2005, <http://www.smartpls.de>.
- Robbins, S.B., Lauver, K., Le, H., Davis, D., Langley, R. & Carlstrom, A. (2004). Do psychological and study skill factors predict college outcomes? *Psychological Bulletin*, 130(2), 261-288.
- Rossiter, J.R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4), 305-335.
- Rousseeuw, P. & Yohai, V. (1984). Robust regression by means of S-estimators. In J. Franke, W. Hardle & R.D. Martin (Hrsg.), *Robust and nonlinear time series analysis* (Vol. 26, S. 256-272). New York: Springer.
- Sackett, P.R., Lievens, F., Berry, C.M. & Landers, R.N. (2007). A cautionary note on the effects of range restriction on predictor intercorrelations. *Journal of Applied Psychology*, 92(2), 538-544.
- Salvendy, G., Boydston, L., Ferguson, G.W., Birch, S. & Cunningham, P.R. (1977). Use of personality, anthropometry and vision scores to select students for psychomotor dental performance. *The Journal of Nihon University School of Dentistry*, 19(4), 174-183.
- Sandow, P.L., Jones, A.C., Peek, C.W., Courts, F.J. & Watson, R.E. (2002). Correlation of admission criteria with dental school performance and attrition. *Journal of Dental Education*, 66(3), 385-392.

- Schmidt, F.L. & Hunter, J.E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86(1), 162-174.
- Schmidt, R.A. & Wrisberg, C.A. (2004). *Motor Learning and Performance*. Champaign (IL): Human Kinetics.
- Schuler, H. (2007). *Lehrbuch Organisationspsychologie*. 4., aktualisierte Auflage. Bern: Verlag Hans Huber.
- Schuler, H. & Hell, B. (2008). Studierendenauswahl und Studienentscheidung aus eignungsdiagnostischer Sicht. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 11-17). Göttingen: Hogrefe.
- Shapiro, S.S. & Wilk, M.B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52 (3/4), 591-611.
- Shoukri, M.M., Asyali, M.H. & Donner, A. (2004). Sample size requirements for the design of reliability study: review and new results. *Statistical Methods in Medical Research*, 13(4), 251-271.
- Silver, N.C. & Dunlap, W.P. (1987). Averaging Correlation-Coefficients - Should Fisher Z-Transformation Be Used. *Journal of Applied Psychology*, 72(1), 146-148.
- Silvestri Jr, A.R., Cohen, S. & Singh, I. (1979). The improvement of technical skills in preclinical courses. *Journal of Dental Education*, 43(12), 641-644.
- Smith, B.G.N. (1976). The value of tests of spatial and psycho-motor ability in selecting dental students. *British Dental Journal*, 141(5), 150-154.
- Smith, B.G.N. (1979). The value of a perception and control test in predicting performance in dental students and ancillary personnel. *Medical Education*, 13(4), 247-250.
- Smith, B.G.N. (1989). A longitudinal study of the value of a spatial relations test in selecting dental students. *British Dental Journal*, 167(9), 305-308.
- Smithers, S., Catano, V.M. & Cunningham, D.P. (2004). What Predicts Performance in Canadian Dental Schools? *Journal of Dental Education*, 68(6), 598-613.
- Snow, R.E. & Farr, M.J. (1987). Cognitive-conative-affective processes in aptitude, learning, and instruction: An introduction. In R.E. Snow & M.J. Farr (Hrsg.), *Aptitude, learning, and instruction: Vol.3. Conative and affective process analyses* (S. 1-10). Hillsdale: Erlbaum.
- Sosik, J.J., Kahai, S.S. & Piovosio, M.J. (2009). Silver bullet or voodoo statistics? A primer for using the partial least squares data analytic technique in group and organization research. *Group & Organization Management*, 34(1), 5-36.
- Spearman, C. (1904a). The proof and measurement of association between two things. *American Journal of Psychology*, 15(1), 72-101.

- Spearman, C. (1904b). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15(2), 201-293.
- Spratley, M.H. (1990). Aptitude testing and the selection of dental students. *Australian Dental Journal*, 35(2), 159-168.
- Spratley, M.H. (1992). Regression analysis of dexterity tests and dental students' practical examination results. *Australian Dental Journal*, 37(6), 461-466.
- Statistisches Bundesamt (Hrsg.) (2010). Statistisches Jahrbuch 2010. Wiesbaden: Statistisches Bundesamt.
- Stemmler, G., Hagemann, D., Amelang, M. & Bartussek, D. (2011). *Differentielle Psychologie und Persönlichkeitsforschung*. 7. Auflage. Stuttgart: Kohlhammer.
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 111-147.
- Straub, D., Boudreau, M.C. & Gefen, D. (2004). Validation Guidelines for IS Positivist Research. *Communications of the Association for Information Systems*, 13(24), 380-427.
- Strub, J.R. (1994). *Curriculum Prothetik: Geschichte, Grundlagen, Behandlungskonzept, Vorbehandlung*. Berlin: Quintessenz-Verlag.
- Suddick, R., Yancey, J.M., Devine, S. & Wilson, S. (1982). Field dependence-independence and dental students' clinical performance. *Journal of Dental Education*, 46(4), 227-232.
- Suddick, R., Yancey, J.M. & Wilson, S. (1983). Mirror-tracing and Embedded Figures tests as predictors of dental students' performance. *Journal of Dental Education*, 47(3), 149-154.
- Suksudaj, N., Townsend, G.C., Kaidonis, J., Lekkas, D. & Winning, T.A. (2011). Acquiring psychomotor skills in operative dentistry: do innate ability and motivation matter?. *European Journal of Dental Education*, 16(1), e187-e194.
- Taatgen, N.A. (2002). A model of individual differences in skill acquisition in the Kanfer-Ackerman Air Traffic Control Task. *Cognitive Systems Research*, 3(1), 103-112.
- Temme, D., Kreis, H. & Hildebrandt, L. (2010). A Comparison of Current PLS Path Modeling Software: Features, Ease-of-Use, and Performance. In V. Esposito Vinzi, W.W. Chin, J. Henseler & H.F. Wang (Hrsg.), *Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields* (S. 737-756). Berlin: Springer.
- Tenenhaus, M., Vinzi, V.E. Chatelin, Y.M. & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, 48(1), 159-205.

- Thompson, C.E. (1942). Motor and mechanical abilities in professional schools. *Journal of Applied Psychology*, 26(1), 24-37.
- Thurstone, L.L. (1931). The measurement of change in social attitude. *The Journal of Social Psychology*, 2(2), 230-235.
- Thurstone, L.L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Thurstone, L.L. & Thurstone, T.G. (1941). Factorial studies of intelligence. *Psychometric monographs*. Nr. 2. Chicago: University of Chicago Press.
- Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21(1), 11-27.
- Trapmann, S., Hell, B., Hirn, S. & Schuler, H. (2007). Meta-Analysis of the Relationship Between the Big Five and Academic Success at University. *Zeitschrift für Psychologie*, 215(2), 132-151.
- Trommsdorff, G. (2003). Kulturvergleichende Entwicklungspsychologie. In A. Thomas (Hrsg.), *Kulturvergleichende Psychologie* (S. 139 - 169). 2., überarbeitete und erweiterte Auflage. Göttingen, u.a.: Hogrefe - Verlag für Psychologie.
- Trost, G. (1986). Die Bedeutung des Interviews für die Diagnose der Studieneignung. Darstellung der internationalen Forschungsergebnisse. In R. Lohölter, K. Hinrichsen, G. Trost & S. Drolshagen (Hrsg.), *Das Interview bei der Zulassung zum Medizinstudium* (S. 49-80). Stuttgart: Schattauer.
- Tsutsui, P.T. & Rich, S.K. (1981). An investigation of a space relations test as a predictor of dental hygiene clinical performance. *Educational Directions for Dental Auxiliaries*, 6(3), 4-9.
- Vandenberg, S.G., & Kuse, A.R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47(2), 599-604.
- van der Maas, H.L.J., Molenaar, D., Maris, G., Kievit, R.A. & Borsboom, D. (2011). Cognitive Psychology Meets Psychometric Theory: On the Relation Between Process Models for Decision Making and Latent Variable Models for Individual Differences. *Psychological Review*, 118(2), 339-356.
- Verlag, O. (Hrsg.) (2013). Approbationsordnung für Zahnärzte. 1. Auflage, Paderborn: Outlook Verlag.
- Voyer, D., Voyer, S. & Bryden, M.P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117(2), 250-270.
- Walcott, A.M., Knight, G.W. & Charlick, R.E. (1986). Waxing tests as predictors of students' performance in preclinical dentistry. *Journal of Dental Education*, 50(12), 716-721.

- Waldman, K.B., Macdonald, G., & Wilson, S.G. (1995). The relationship between standardized psychomotor tests and basic clinical dental hygiene skills. *Journal of Dental Hygiene*, 69(4), 163-168.
- Waldron, E.M. & Anton, B.S. (1995). Effects of Exercise on Dexterity. *Perceptual and Motor Skills*, 80(3), 883-889.
- Walker, J.D., Killip, D.E. & Fuller, J.L. (1985). The significance of the admission interview in predicting students' performance in dental school. *Journal of Medical Education*, 60(7), 569-571.
- Wanous, J.P. & Hudy, M.J. (2001). Single-Item Reliability: A Replication and Extension. *Organizational Research Methods*, 4(4), 361-375.
- Wanous, J.P. & Reichers, A.E. (1996). Estimating the reliability of a single-item measure. *Psychological Reports*, 78(2), 631-634.
- Wanous, J.P., Reichers, A.E. & Hudy, M.J. (1997). Overall job satisfaction: how good are single-item measures? *Journal of Applied Psychology*, 78(2), 631-634.
- Ward, S.T., Downey, M.C., Thompson, A.L. & Collins, M.A. (2010). Predictors of success in dental hygiene education: a follow-up study. *Journal of Dental Hygiene*, 84(1), 24-28.
- Wedler, B., Troche, S. & Rammsayer, T. (2008). Studierendenauswahl - Eignungsdiagnostischer Nutzen von Noten aus Schule und Studium. *Psychologische Rundschau*, 59(2), 123-125.
- Weiber, R. & Mühlhaus, D. (2009). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. Heidelberg: Springer.
- Weinstein, P. & Kiyak, H.A. (1981). Assessing manual dexterity: pilot study of a new instrument. *Journal of Dental Education*, 45(2), 71-73.
- Werts, C.E., Linn, R.L. & Jöreskog, K.G. (1974). Interclass Reliability Estimates: Testing Structural Assumptions. *Educational and Psychological Measurement*, 34(1), 24-33.
- Wickelgren, W.A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41(1), 67-85.
- Wiesendanger, M. & Serrien, J. (2001). Toward a Physiological Understanding of Human Dexterity. *Physiology*, 16(5), 228-233.
- Wilson, S. & Husak, W.S. (1988). A multivariable approach toward predicting dental motor skill performance. *Perceptual and Motor Skills*, 67(1), 211-217.
- Wilson, S., Suddick, R.P. & Shay, J.S. (1981). Correlations of scores on embedded figures and mirror tracing with preclinical technique grades and PMAT scores of dental students. *Perceptual and Motor Skills*, 53(1), 31-35.

- Wilson, S., Suddick, R.P., Shay, J.S., & Hustmyer Jr., F.E. (1981). Correlations of scores on embedded figures and mirror tracing with preclinical technique grades and PMAT scores of dental students. *Perceptual and Motor Skills*, 53(1), 31-35.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Wirtz, M. (2004). Bestimmung der Güte von Beurteilereinschätzungen mittels der Intraklassenkorrelation und Verbesserung von Beurteilereinschätzungen. *Rehabilitation*, 43(6), 384-389.
- Witkin, H.A., Lewis, H.B., Hertzman, M., Machover, K., Meissner, P.B. & Wapner, S. (1954). *Personality through perception: an experimental and clinical study*. Oxford, England: Harper.
- Wittmann, W.W. & Süß, H.M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry. In P.L. Ackerman, P.C. Kyllonen & R.D. Roberts (Hrsg.), *Learning and individual differences. Process, trait, and content determinants*. (S. 77-104). Washington: American Psychological Association.
- Wold, H. (1975). Path models with latent variables: The NIPALS approach. In H.M. Blalock, A. Aganbegian, F.M. Borodkin, R. Boudon, & V. Capecchi (Hrsg.), *Quantitative sociology: International perspectives on mathematical and statistical modeling* (S. 307-357). New York: Academic.
- Wold, H. (1982). Soft modeling: the basic design and some extensions. In K.G. Jöreskog & H. Wold (Hrsg.), *Systems under Indirect Observations: Causality, Structure, Prediction, Part 2* (S. 36-37), Amsterdam, New York, u.a.: North-Holland.
- Wong, A.Y., Watson, J.F. & Thye, R.P. (1979). Evaluation of predictor variables for a self-instructional preclinical course. *Journal of Dental Education*, 43(12), 637-640.
- Wood, W. (1979). Grade averages and DAT scores as predictors of performance in dental school. *Journal of Dental Education*, 43(12), 630-632.
- Yancosek, K.E. & Howell, D. (2009). A Narrative Review of Dexterity Assessments. *Journal of Hand Therapy*, 22(3), 258-270.
- Yohai, V.J. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics*, 15(2), 642-656.
- Yoon, S.Y. (2011). *Psychometric Properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (The Revised PSVT-R)*. Dissertation, Purdue University, West Lafayette (IN). ProQuest LLC.

- Young, W.G. (1972). Visuo-spatial ability, dental aptitude tests and multiple-choice practical examinations in oral pathology. *Journal of Dental Education*, 36(1), 48-56.
- Yung, Y.F. & Chan, W. (1999). Statistical analyses using bootstrapping: concepts and implementation. In R.H. Hoyle (Hrsg.), *Statistical strategies for small sample research* (S.81-105). Thousand Oaks (CA): Sage.
- Zimmer, A.C. (1990). Der Erwerb komplexer motorischer Fertigkeiten. In C.G. Hoyos & B. Zimolong (Hrsg.), *Enzyklopadie der Psychologie. Band 2*. Göttingen: Hogrefe.
- Zimmerhofer, A. & Trost, G. (2008). Auswahl- und Feststellungsverfahren in Deutschland - Vergangenheit, Gegenwart und Zukunft. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 32-42). Göttingen: Hogrefe.
- Zimmerman, M., Ruggero, C.J., Chelminski, I., Young, D., Posternak, M.A., Friedman, M., Boerescu, D. & Attiullah, N. (2006). Developing brief scales for use in clinical practice: the reliability and validity of single-item self-report measures of depression symptom severity, psychosocial impairment due to depression, and quality of life. *Journal of Clinical Psychiatry*, 67(10), 1536-1541.
- Zullo, T.G. (1971). A factor analysis of perceptual and motor abilities of dental students. *Journal of Dental Education*, 35(6), 356-361.

## 8. Anhang

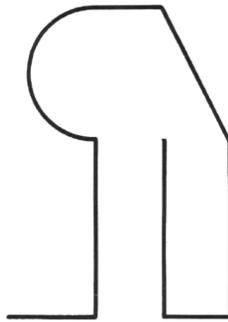


Abbildung A01: Vorlage der zu biegender Figur in der DBP (Quelle: Lienert, 1967)

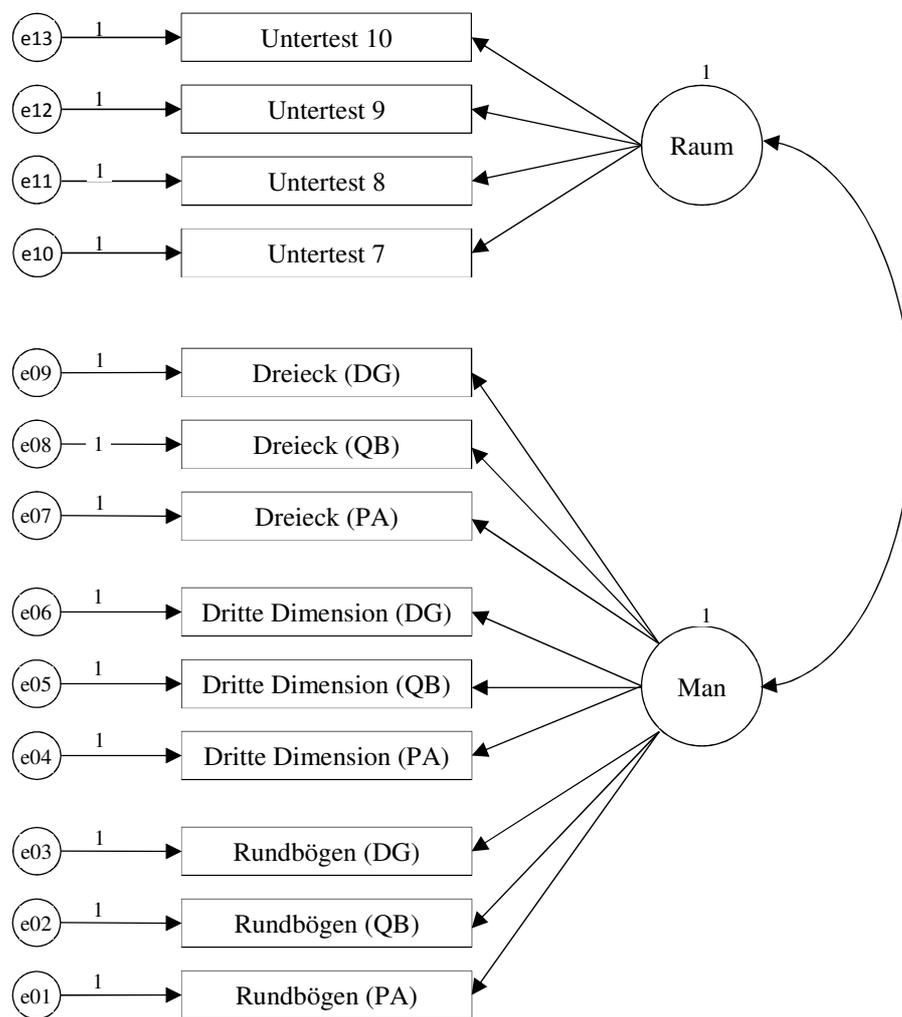


Abbildung A02: Operationalisierung des instrumentellen, manuellen Geschick als Ein-Faktor-Modell  
Anm.: Raum = räumliches Vorstellungsvermögen; Man = instrumentelles, manuelles Geschick

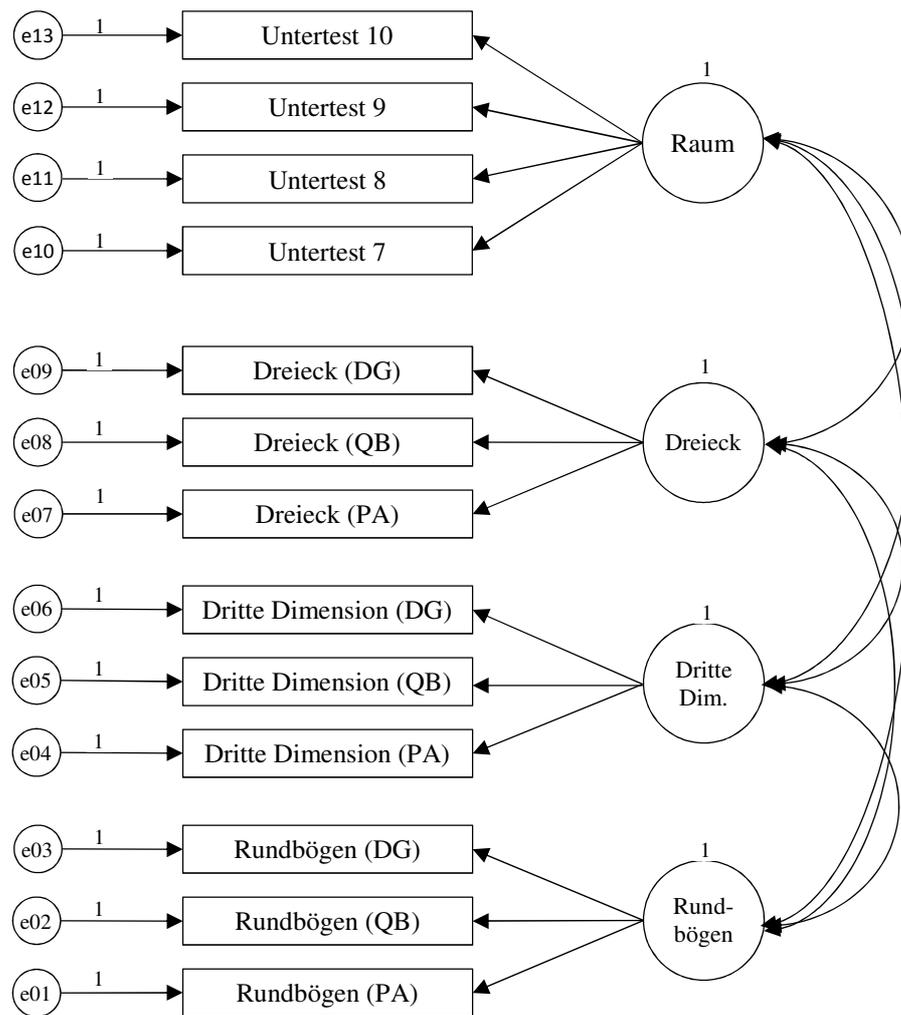


Abbildung A03: Operationalisierung des instrumentellen, manuellen Geschicks als Drei-Faktoren-Modell  
 Anm.: Raum = räumliches Vorstellungsvermögen

Tabelle A01: Deskriptive Statistik der Erhebungsinstrumente

Variable	n	M	SD	$\sigma^2$	Min	Max	Sch	Exz	p-SWT
HAM-Man	129	17.0	9.28	86.15	0	39	0.25	-0.45	.045
- Dreieck	129	9.7	3.49	12.18	0	15	-1.06	0.92	
-- DG	129	3.8	1.57	2.47	0	6	-0.78	0.09	
-- QB	129	1.9	0.74	0.56	0	3	-1.04	0.43	
-- PA	129	3.9	1.35	1.83	0	6	-1.26	1.73	
- Dritte Dim.	129	4.8	4.42	19.57	0	14	0.32	-1.26	
-- DG	129	1.9	1.87	3.50	0	6	0.47	-1.13	
-- QB	129	1.0	0.94	0.89	0	3.0	0.18	-1.43	
-- PA	129	1.9	1.76	3.11	0	5.5	0.31	-1.26	
- Rundbögen	129	2.5	3.33	11.11	0	12.8	1.31	0.78	
-- DG	129	1.0	1.50	2.26	0	6	1.33	0.84	
-- QB	129	0.4	0.64	0.41	0	2.5	1.43	1.21	
-- PA	129	1.0	1.37	1.89	0	5	1.26	0.39	
HAM-Nat	129	32.7	9.79	95.83	14	61	0.51	-0.32	.007
Schulabschluss	129	2.0	0.22	0.05	1.5	2.3	-0.65	-0.43	.000
LPS	129	207.4	30.50	930.15	128	262	-0.45	-0.01	.011
-- UT 7	129	25.4	8.43	71.04	5	40	-0.06	-0.54	
-- UT 8	129	34.8	6.61	43.70	10	40	-1.61	1.93	
-- UT 9	129	27.6	4.84	23.43	14	37	-0.74	0.37	
-- UT 10	129	31.7	5.48	30.04	13	40	-0.76	0.26	
Note TPK	38	2.9	0.28	0.08	2.5	4.0	1.40	4.04	.672
-- Note 1	38	2.9	0.45	0.20	2.0	3.9	0.06	-0.62	
-- Note 2	38	3.1	0.54	0.30	1.8	4.2	0.47	0.28	
-- Note 3	38	2.9	0.37	0.14	2.2	4.0	1.00	2.89	
-- Note 4	38	2.9	0.35	0.12	2.2	3.4	-0.09	1.32	
Note Phal	31	3.0	0.21	0.05	2.6	3.3	-0.26	-0.93	.449
-- Note 1	31	3.0	0.28	0.08	2.4	3.6	0.16	-0.50	
-- Note 2	31	3.1	0.22	0.05	2.6	3.5	-0.83	0.88	
-- Note 3	31	2.9	0.39	0.15	2.1	3.8	0.23	-0.66	
Note PhalII	28	2.5	0.32	0.10	2.0	3.3	0.52	0.06	.473
-- Note 1	28	2.5	0.37	0.14	1.6	3.3	-0.01	0.35	
-- Note 2	28	2.6	0.53	0.29	1.8	3.9	0.64	-0.03	
-- Note 3	28	2.5	0.35	0.12	1.8	3.2	0.24	0.15	
DBP	129	18.9	4.30	18.53	7	30	-0.18	0.33	.263
-- Form	129	12.99	3.353	11.242	5	22	.042	-.035	
-- Sauberkeit	129	5.91	1.598	2.554	0	8	-.966	1.094	
Geschlecht	129	1.6	0.49	0.24	1	2	-0.54	-1.74	.000
Alter	129	21.1	2.03	4.11	18	29	1.75	3.72	.000

Anm.: n = Stichprobengröße; M = Mittelwert; SD = Standardabweichung;  $\sigma^2$  = Varianz; Min / Max = Mindest- und Maximalwert; Sch = Schiefe der Verteilung; Exz = Exzess der Verteilung; p-SWT = Signifikanz des Shapiro-Wilk-Tests  
Notenskalen: 1.0 (sehr gut) bis 4.0 (ungenügend)

- HAM-Man = Drahtbiegetest (instrumentelles, manuelles Geschick, 0 bis 45 Pkt.); Dreieck = Drahtfigur (0 bis 15 Pkt.); Dritte Dim. = Dritte Dimension, Drahtfigur (0 bis 15 Pkt.); Rundbögen = Drahtfigur (0 bis 15 Pkt.); DG = Kriterium Deckungsgleichheit, siebenstufige Skala (0 bis 6 Pkt.); QB = Kriterium Qualität der Biegung, siebenstufige Skala (0 bis 3 Pkt.); PA = Kriterium planes Aufliegen, siebenstufige Skala (0 = 0 Pkt., 6 = 6 Pkt.)
- HAM-Nat = Naturwissenschaftstest (naturwissenschaftliches Verständnis, 0 bis 76 Pkt.)
- Schulabschluss. = Schulabschlussnote (Abiturdurchschnittsnote)
- LPS = Leistungsprüfungssystem von Horn (1983, räumliches Vorstellungsvermögen, C-Werte von 51 bis 272);  
UT = Untertest; Skala technische Begabung: UT 7 = spiegelbildlich gedruckte Zeiten erkennen; UT 8 = räumliches Vorstellen;  
UT 9 = Flächen bei geometrischen Körpern zählen; UT 10 = versteckte Muster erkennen; Skalen UT = 0 bis 40 Pkt.
- Note TPK = Gesamtnote Technisch-Propädeutischer-Kurs (Fertigkeitserwerb für konsistente, einfache Aufgaben);  
Note 1 = Schnitzübung; Note 2 = Aufwachsübung; Note 3 = Vollgusskrone; Note 4 = Interimsprothese
- Note Phal = Gesamtnote Phantomkurs I (Fertigkeitserwerb für inkonsistente, komplexe Aufgaben);  
Note 1 = Totalprothese; Note 2 = Stiftaufbau und Schalenprovisorium; Note 3 = Präparation und Tangentialbrücke
- Note PhalII = Gesamtnote Phantomkurs II (Fertigkeitserwerb für inkonsistente, komplexe Aufgaben);  
Teilnote 1 = Totalprothese; Teilnote 2 = Präparation und Schalenprovisorium; Teilnote 3 = Michiganschiene;
- DBP = Drahtbiegetest (Lienert, 1967); 0 bis 30 Pkt.; Skala Form (0 bis 7 Pkt.); Skala Sauberkeit (0 bis 22 Pkt.)
- Geschlecht (1 = Mann, 2 = Frau)
- Alter (Jahre)

Tabelle A02-1: Mittelwerte, Standardabweichung, Reliabilitäten und Interkorrelationen der Erhebungsinstrumente

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13
1. HAM-Man	17.0	9.28	(0.98)												
2. DE (DG)	3.8	1.57	.77***	1											
3. DE (QB)	1.9	0.74	.77***	.88***	1										
4. DE (PA)	3.9	1.35	.70***	.83***	.85***	1									
5. DE	9.7	3.49	.78***	.96***	.94***	.94***	(0.95)								
6. DD (DG)	1.9	1.87	.87***	.54***	.55***	.48***	.55***	1							
7. DD (QB)	1.0	0.94	.85***	.51***	.57***	.47***	.54***	.91***	1						
8. DD (PA)	1.9	1.76	.84***	.48***	.50***	.44***	.49***	.89***	.89***	1					
9. DD	4.8	4.42	.89***	.53***	.56***	.48***	.54***	.97***	.95***	.96***	(0.96)				
10. RB (DG)	1.0	1.50	.75***	.42***	.40***	.31***	.40***	.54***	.50***	.51***	.54***	1			
11. RB (QB)	0.4	0.64	.73***	.40***	.43***	.34***	.40***	.52***	.50***	.48***	.52***	.89***	1		
12. RB (PA)	1.0	1.37	.74***	.40***	.37***	.31***	.38***	.54***	.50***	.56***	.56***	.83***	.78***	1	
13. RB	2.5	3.33	.79***	.43***	.42***	.33***	.41***	.57***	.53***	.55***	.57***	.97***	.92***	.94***	(0.94)
14. HAM-Nat	32.7	9.79	.31***	.12	.17†	.15†	.15†	.32***	.34***	.37***	.36***	.19*	.25**	.22*	.22*
15. Abiturnote	2.0	0.22	.09	.05	.05	-0.04	.02	.05	.05	.17†	.10	.07	-.07	.18*	.09
16. LPS	207.4	30.50	.28**	.23**	.27**	.29**	.28**	.24**	.16†	.24**	.23**	.15	.13	.21*	.18*
17. LPS UT 7	25.4	8.43	.06	.09	.12	.07	.09	.05	.00	.05	.04	.03	-.01	.03	.02
18. LPS UT 8	34.8	6.61	.18*	.14	.14	.26**	.20*	.16†	.07	.20*	.16†	.05	.04	.10	.07
19. LPS UT 9	27.6	4.84	.40***	.33***	.32***	.36***	.36***	.34***	.32***	.33***	.35***	.24**	.27**	.32***	.29**
20. LPS UT10	31.7	5.48	.21*	.10	.22*	.15†	.15†	.16†	.18*	.15†	.17†	.18*	.19*	.24**	.22*
21. DBP	18.9	4.30	.33***	.21*	.23**	.30***	.26**	.28**	.31***	.32***	.31***	.17†	.22*	.24	.22**
22. DBP (S)	12.99	3.353	.36***	0.27**	.32***	.39***	.34***	.28**	.31***	.31***	.31***	.20*	.23**	.27**	.25**
23. DBP (F)	5.91	1.598	.24	.14	.14	.20	.17	.23	.25	.27	.25	.12	.18	.18	.16
24. TPK N1	2.9	0.45	-.14	-0.09	-.04	-.07	-.08	.02	.04	-.02	.01	-.21	-.19	-.35*	-.28†
25. TPK N2	3.1	0.54	-.28†	-0.19	-.30†	-.25	-.25	-.12	-.19	-.30†	-.21	-.18	-.17	-.24	-.22
26. TPK N3	2.9	0.37	-.22	-0.06	-.12	-.25	-.14	-.27	-.25	-.16	-.24	-.13	-.17	-.05	-.11
27. TPK N4	2.9	0.35	-.38*	-0.23	-.35*	-.35*	-.31	-.27	-.30	-.40*	-.34*	-.15	-.14	-.32†	-.22
28. Note TPK	2.9	0.27	-.33*	-0.10	-.18	-.20	-.16	-.17	-.25	-.30†	-.24	-.27	-.27	-.42**	-.35*
29. PhaI N 1	3.0	0.28	-.20	-0.20	-.32†	-.07	-.20	-.18	-.17	-.08	-.14	-.17	-.21	-.01	-.13
30. PhaI N 2	3.1	0.22	-.37*	-0.15	-.33†	-.26	-.25	-.20	-.23	-.22	-.22	-.32†	-.32†	-.35†	-.35†
31. PhaI N3	2.9	0.39	-.41*	-0.18	-.44*	-.41*	-.34†	-.19	-.20	-.23	-.22	-.39*	-.30	-.37*	-.39*
32. Note PhaI	3.0	0.21	-.46*	-0.25	-.52**	-.37*	-.38*	-.26	-.27	-.25	-.27	-.42*	-.38*	-.35†	-.41*
33. PhaII N1	2.5	0.38	-.22	-.28	-.39*	-.42*	-.39*	.00	-.06	-.036	-.027	-.21	-.20	-.21	-.22
34. PhaII N2	2.6	0.53	-.50**	.24	-.08	-.03	.10	-.45*	-.43*	-.62***	-.54**	-.43*	-.40*	-.51**	-.48*
35. PhaII N3	2.5	0.35	-.15	-.16	-.35†	-.18	-.23	-.02	-.05	.04	-.01	-.09	-.20	-.22	-.17
36. Note PhaII	2.5	0.32	.44*	-.04	-.33†	-.25	-.19	-.28	-.28	-.35†	-.32†	-.39*	-.40*	-.47*	-.45*
37. Geschlecht <sup>a</sup>	1.6	0.49	-.22*	-0.05	-.01	-.04	-.04	-.24**	-.19*	-.28**	-.25**	-.20*	-.17†	-.30***	-.25**
38. Alter	21.1	2.03	.13	0.09	.17†	.15†	.14	.12	.10	.10	.11	.07	.10	.07	.08

Tabelle A02-2: Mittelwerte, Standardabweichung, Reliabilitäten und Interkorrelationen der Erhebungsinstrumente (Fortsetzung)

Variable	<i>M</i>	<i>SD</i>	14	15	16	17	18	19	20	21	22	23	24	25	26	27
01. HAM-Man	17.0	9.28														
02. DE (DG)	3.8	1.57														
03. DE (QB)	1.9	0.74														
04. DE (PA)	3.9	1.35														
05. DE	9.7	3.49														
06. DD (DG)	1.9	1.87														
07. DD (QB)	1.0	0.94														
08. DD (PA)	1.9	1.76														
09. DD	4.8	4.42														
10. RB (DG)	1.0	1.50														
11. RB (QB)	0.4	0.64														
12. RB (PA)	1.0	1.37														
13. RB	2.5	3.33														
14. HAM-Nat	32.7	9.79	1													
15. Abiturnote	2.0	0.22	-.20*	1												
16. LPS	207.4	30.50	.16†	-.04	(0.58)											
17. LPS UT 7	25.4	8.43	.03	-.05	.75***	1										
18. LPS UT 8	34.8	6.61	.14	-.01	.66***	.16†	1									
19. LPS UT 9	27.6	4.84	.18*	-.05	.65***	.24*	.33***	1								
20. LPS UT10	31.7	5.48	.14	.01	.52***	.28**	.18*	.33***	1							
21. DBP	18.9	4.30	.25**	-.14	.29**	.16†	.23***	.30***	.02	(0.61)						
22. DBP (S)	12.99	3.353	.15†	-.02	.26**	.13	.25***	.20*	.12	.72***	1					
23. DBP (F)	5.91	1.598	.25	-.17	.24	.14	.18	.28	-.03	.94	.44	1				
24. TPK N1	2.9	0.45	-.06	-.08	.09	-.09	.33*	-.12	.27†	-.14	.06	-.18	1			
25. TPK N2	3.1	0.54	.10	-.32*	.10	.10	.15	.06	-.23	.04	.33*	-.07	.22	1		
26. TPK N3	2.9	0.37	-.03	.14	.01	.23	-.14	-.25	-.05	-.16	-.29†	-.09	-.18	-.27	1	
27. TPK N4	2.9	0.35	-.38*	-.19	-.03	.01	.09	-.12	-.23	-.10	.06	-.13	.16	.28†	.19	1
28. Note TPK	2.9	0.27	-.14	-.24	.08	.13	.22	-.23	-.19	-.24	.11	-.32†	.55***	.66***	.21	.68***
29. PhaI N 1	3.0	0.28	.17	.30	-.33†	-.34†	-.11	-.22	-.05	-.11	-.05	-.11	-.33†	.08	.05	-.17
30. PhaI N2	3.1	0.22	.01	-.08	-.10	.07	.01	-.51	-.08	-.34†	.00	-.39*	.16	-.04	.32†	.20
31. PhaI N3	2.9	0.39	-.16	.07	-.39*	-.14	-.49**	-.31†	-.20	-.32†	.03	-.38*	-.04	.15	.18	.38*
32. Note PhaI	3.0	0.21	-.02	.15	-.42*	-.22	-.34†	-.46**	-.17	-.36†	.00	-.41*	-.11	.11	.24	.22
33. PhaII N1	2.5	0.38	.15	-.06	.11	.19	.03	.06	-.20	-.13	.21	-.23	.09	.40**	.24	.53***
34. PhaII N2	2.6	0.53	-.35†	-.23	.07	.18	-.01	-.07	-.13	-.29	-.09	-.31	.17	.43**	.17	.55***
35. PhaII N3	2.5	0.35	-.18	-.22	.17	.12	.25	.03	-.06	-.16	.17	-.26	.26†	.47***	-.03	.27†
36. Note PhaII	2.5	0.32	-.20	-.20	.13	.21	.07	-.02	-.18	-.25	.10	-.34†	.21	.55***	.17	.56***
37. Geschlecht <sup>a</sup>	1.6	0.49	-.33***	-.06	-.12	-.04	-.09	-.14	-.06	.00	-.03	.02	-.02	-.18	-.21	-.14
38. Alter	21.1	2.03	.07	.07	.11	.03	.08	.13	.09	.09	.17†	.03	-.13	.11	-.20	-.31†

Tabelle A02-3: Mittelwerte, Standardabweichung, Reliabilitäten und Interkorrelationen der Erhebungsinstrumente (Fortsetzung)

Variable	M	SD	28	29	30	31	32	33	34	35	36	37	38	--	--
28. Note TPK	2.9	0.27	(0.48)												
29. PhaI N1	3.0	0.28	-.16	1											
30. PhaI N2	3.1	0.22	.25	.22	1										
31. PhaI N3	2.9	0.39	.28	.18	.38*	1									
32. Note PhaI	3.0	0.21	.18	.63***	.67***	0.81***	(0.46)								
33. PhaII N1	2.5	0.38	.52***	.42**	.36*	.52***	.57***	1							
34. PhaII N2	2.6	0.53	.53***	.29†	.27†	.53***	.49***	.52***	1						
35. PhaII N3	2.5	0.35	.44**	.21	.10	.35*	.30*	.36*	.39**	1					
36. Note PhaII	2.5	0.32	.62***	.38**	.30*	.59***	.57***	.77***	.85***	.72***	(0.57)				
37. Geschlecht <sup>a</sup>	1.6	0.49	-.24†	-.05	-.12	-.07	-.11	.08	.26	-.14	.14	1			
38. Alter	21.1	2.03	-.17	-.050	.23	-.32†	-.11	-.03	-.40*	-.36†	-.36†	-.06	1		

Anm.: M = Mittelwert; SD = Standardabweichung

Stichprobengröße: n = 129; TPK: n = 38; PhaI: n = 31; PhaII: n = 28

a: dichotom codierte Variable (1 = männlich, 2 = weiblich)

† $p < .10$ , \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$  (zweiseitig).

Reliabilitätskoeffizient Cronbachs Alpha ist in der Tabellendiagonalen in Klammern abgebildet.

Notenskalen: 1.0 (sehr gut) bis 4.0 (ungenügend)

1. HAM-Man = Drahtbiegetest (instrumentelles, manuelles Geschick, 0 bis 45 Pkt.)

2. - 13.: DE = Dreieck (0 bis 15 Pkt.); DD = Dritte Dimension (0 bis 15 Pkt.); RB = Rundbögen (0 bis 15 Pkt.); DG = Kriterium Deckungsgleichheit, siebenstufige Skala (0 bis 6 Pkt.);

QB = Kriterium Qualität der Biegung, siebenstufige Skala (0 bis 3 Pkt.); PA = Kriterium planes Aufliegen, siebenstufige Skala (0 bis 6 Pkt.)

14. HAM-Nat = (naturwissenschaftliches Verständnis, 0 bis 76 Pkt.)

15. Abiturnote = Abiturdurchschnittsnote (Schulabschlussnote, 1.0 bis 4.0)

16. - 20.: LPS = Leistungsprüfsystem von Horn (1983, räumliches Vorstellungsvermögen, C-Werte von 51 bis 272); UT = Untertest; Skalen UT = 0 bis 40 Pkt.;

Skala technische Begabung: UT 7 = spiegelbildlich gedruckte Zeilen erkennen; UT 8 = räumliches Vorstellen; UT 9 = Flächen bei geometrischen Körpern zählen; UT 10 = versteckte Muster erkennen;

21. - 23.: DBP = Drahtbiegetest (Lienert, 1967, manuelles Geschick, 0 bis 30 Pkt.); S = Skala Sauberkeit (0 bis 22 Pkt.); F = Skala Form (0 bis 7 Pkt.)

24. - 28.: Note TPK = Gesamtnote Technisch-Propädeutischer-Kurs (Fertigkeitserwerb für konsistente, einfache Aufgaben); Note 1 = Schnitzübung; Note 2 = Aufwachsübung; Note 3 = Vollgusskrone;

Note 4 = Interimsprothese

29. - 32.: Note PhaI = Gesamtnote Phantomkurs I (Fertigkeitserwerb für inkonsistente, komplexe Aufgaben);

Note 1 = Totalprothese; Note 2 = Stiftaufbau und Schalenprovisorium; Note 3 = Präparation und Tangentialbrücke

33. - 36.: Note PhaII = Gesamtnote Phantomkurs II (Fertigkeitserwerb für inkonsistente, komplexe Aufgaben);

Teilnote 1 = Totalprothese; Teilnote 2 = Präparation und Schalenprovisorium; Teilnote 3 = Michiganschiene

37. Geschlecht (1 = Mann, 2 = Frau)

38. Alter (Jahre)

Tabelle A03: Interrater-Reliabilität des HAM-Man

Item	ICC	99% KI		F-Wert	Sig.	$\alpha$	M R1 (SD)	M R2 (SD)	$\delta$ Rater
		u. G.	o. G.						
HAM-Man	0.96	0.94	0.98	56.35	.000	0.98	17.2 (9.97)	16.8 (8.70)	0.40
- Dreieck	0.92	0.87	0.95	22.78	.000	0.96	9.6 (3.79)	9.8 (3.32)	-0.16
-- DG	0.87	0.80	0.91	14.25	.000	0.93	3.8 (1.74)	3.9 (1.50)	-0.13
-- QB	0.91	0.86	0.94	21.14	.000	0.95	1.9 (0.81)	1.9 (0.72)	-0.04
-- PA	0.88	0.82	0.93	16.15	.000	0.94	4.0 (1.48)	3.9 (1.30)	0.02
- Dritte Dim.	0.96	0.93	0.98	55.14	.000	0.98	5.0 (4.59)	4.7 (4.23)	0.30
-- DG	0.94	0.90	0.97	35.94	.000	0.97	2.0 (1.98)	1.8 (1.81)	0.22
-- QB	0.95	0.92	0.97	43.89	.000	0.98	1.1 (1.00)	1.0 (0.90)	0.05
-- PA	0.92	0.88	0.95	25.04	.000	0.96	1.9 (1.90)	1.8 (1.69)	0.10
- Rundbögen	0.96	0.94	0.97	48.59	.000	0.98	2.6 (3.50)	2.4 (3.22)	0.17
-- DG	0.95	0.92	0.97	45.42	.000	0.98	1.1 (1.58)	1.0 (1.46)	0.14
-- QB	0.91	0.85	0.94	21.82	.000	0.95	0.5 (0.69)	0.4 (0.62)	0.08
-- PA	0.91	0.86	0.94	20.74	.000	0.95	1.0 (1.45)	1.0 (1.37)	-0.06

Anmerkungen: ICC = Intra-Klassen-Korrelationen, unjustiertes Modell: zweifach, gemischt; KI = Konfidenzintervall; Sig. = Signifikanz F-Tests;  $\alpha$  = Reliabilitätskoeffizient Cronbachs Alpha; M R1 = Mittelwert Rater1; SD = Standardabweichung; M R2 = Mittelwert Rater2;  $\delta$  Rater = Differenz der Mittelwerte Rater1 & Rater2; DG = Kriterium Deckungsgleichheit; QB = Qualität der Biegung; PA = planes Aufliegen; Dritte Dim. = Dritte Dimension; n = 129; df1 und df 2 = 128

Tabelle A04: Retest-Reliabilität des HAM-Man

Item	ICC	99% KI		F-Wert	Sig.	$\alpha$	M R1 (SD)	M R2 (SD)	$\delta$ Rater
		u. G.	o. G.						
HAM-Man	0.70	0.39	0.86	3.03	.000	0.83	21.9 (9.80)	24.3 (7.76)	-2.38
- Dreieck	0.54	0.17	0.78	3.34	.000	0.70	11.0 (3.41)	10.6 (2.67)	0.37
-- DG	0.43	0.02	0.72	2.50	.004	0.60	4.4 (1.63)	4.2 (1.28)	0.14
-- QB	0.34	-0.07	0.66	2.07	.016	0.52	2.2 (0.74)	2.0 (0.59)	0.17
-- PA	0.62	0.27	0.82	4.17	.000	0.76	4.4 (1.17)	4.4 (1.06)	0.06
- Dritte Dim.	0.59	0.24	0.81	3.89	.000	0.74	7.2 (4.73)	7.8 (3.56)	-0.60
-- DG	0.61	0.26	0.82	4.04	.000	0.75	2.9 (2.05)	2.1 (1.54)	0.81
-- QB	0.42	0.01	0.71	2.43	.005	0.60	1.5 (0.98)	1.7 (0.67)	-0.13
-- PA	0.53	0.16	0.77	3.32	.000	0.70	2.8 (1.87)	3.2 (1.47)	-0.40
- Rundbögen	0.50	0.30	0.78	3.72	.000	0.73	3.7 (3.92)	5.8 (3.54)	-2.16
-- DG	0.43	0.04	0.71	2.72	.002	0.63	1.6 (1.77)	2.4 (1.63)	-0.71
-- QB	0.33	-0.07	0.65	2.25	.009	0.56	0.7 (0.84)	1.2 (0.71)	-0.47
-- PA	0.50	-0.07	0.79	3.95	.000	0.75	1.3 (1.53)	2.3 (1.57)	-0.98

Anmerkungen: ICC = Intra-Klassen-Korrelationen, unjustiertes Modell: zweifach, gemischt; 99% KI = 99% Konfidenzintervall; Sig. = Signifikanz F-Tests;  $\alpha$  = Reliabilitätskoeffizient Cronbachs Alpha; M R1 = Mittelwert Rater1; SD = Standardabweichung; M R2 = Mittelwert Rater2;  $\delta$  Rater = Differenz der Mittelwerte Rater1 & Rater2; DG = Kriterium Deckungsgleichheit; QB = Kriterium Qualität der Biegung; PA = Kriterium planes Aufliegen; Dritte Dim. = Dritte Dimension  
Testwiederholung in der ersten Studienwoche (Orientierungseinheit); n = 38; df1 und df 2 = 37

Tabelle A05: Interkorrelationen HAM-Man (OE) und HAM-Man (AdH)

Variable OE / AdH	1. AdH	2. AdH	3. AdH	4. AdH	5. AdH	6. AdH	7. AdH	8. AdH	9. AdH	10. AdH	11. AdH	12. AdH	13. AdH
1. HAM-Man OE	.74***	.54***	.60***	.59***	.59***	.58***	.57***	.56***	.59***	.60***	.56***	.56***	.61***
2. Dreieck (DG)	.46*	.44*	.50*	.52*	.50*	.38	.41	.37	.39	.28†	.18	.19	.24
3. Dreieck (QB)	.27	.34*	.36*	.40*	.38*	.13	.20	.17	.17	.21	.00	.12	.14
4. Dreieck (PA)	.58***	.53***	.58***	.62***	.59***	.43**	.47**	.47**	.47**	.38*	.30†	.36*	.38*
5. Dreieck OE	.51**	.50**	.55***	.58***	.56***	.38*	.43†	.40*	.41*	.33*	.21	.26	.30†
6. Dritte Dim. (DG)	.67***	.37*	.53**	.48**	.46**	.63***	.59***	.57***	.62***	.52**	.55***	.45**	.53***
7. Dritte Dim. (QB)	.56***	.34*	.42**	.43**	.40*	.44**	.45**	.41*	.45**	.50**	.50*	.43**	.50**
8. Dritte Dim. (PA)	.59***	.32†	.31†	.40*	.36*	.51**	.53***	.55***	.55***	.45**	.39*	.53***	.49**
9. Dritte Dim. OE	.70***	.39*	.48**	.50**	.46**	.62***	.61***	.60***	.63***	.55***	.54***	.54***	.58***
10. Rundbögen (DG)	.57***	.45**	.50**	.40*	.46**	.46**	.45**	.43**	.46**	.46**	.44**	.41*	.47**
11. Rundbögen (QB)	.42*	.35*	.32†	.29†	.34*	.30†	.25	.27	.29†	.40*	.39*	.34*	.40*
12. Rundbögen (PA)	.54**	.38*	.38*	.37*	.39*	.34*	.27	.25	.30†	.60***	.65***	.60***	.64***
13. Rundbögen OE	.59***	.45**	.47**	.40*	.45**	.42*	.37*	.37*	.41*	.56***	.57***	.52**	.58***

Anmerkungen: M = Mittelwert; SD = Standardabweichung; n = 38 (Studierende des ersten Semesters der Zahnmedizin mit Zulassung über das Hochschulwahlverfahren)

† $p < .10$ , \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$  (zweiseitig)

1. HAM-Man = Drahtbiegetest (instrumentelles, manuelles Geschick, 0 bis 45 Pkt.)

2. - 13.: Dreieck (0 bis 15 Pkt.); Dritte Dim. = Dritte Dimension (0 bis 15 Pkt.); Rundbögen (0 bis 15 Pkt.); DG = Kriterium Deckungsgleichheit, siebenstufige Skala (0 bis 6 Pkt.); QB = Kriterium Qualität der Biegung, siebenstufige Skala (0 bis 3 Pkt.); PA = Kriterium planes Aufliegen, siebenstufige Skala (0 bis 6 Pkt.)

Tabelle A06: HAM-Man OE - Mittelwerte, Standardabweichung, Reliabilitäten und Korrelationen

Variable OE	M	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. HAM-Man OE	21.3	8.96	(0.97)												
2. Dreieck (DG)	3.8	1.46	.77	1											
3. Dreieck (QB)	1.8	0.72	.75	.79	1										
4. Dreieck (PA)	4.1	1.24	.76	.72	.75	1									
5. Dreieck OE	9.8	3.12	.83	.93	.90	.90	(0.90)								
6. Dritte Dim. (DG)	2.4	1.62	.80	.50	.49	.56	.57	1							
7. Dritte Dim. (QB)	1.4	0.76	.84	.58	.59	.60	.64	.75	1						
8. Dritte Dim. (PA)	2.8	1.60	.80	.51	.53	.53	.57	.70	.82	1					
9. Dritte Dim. OE	6.7	3.64	.88	.57	.57	.61	.64	.91	.90	.92	(0.90)				
10. Rundbögen (DG)	1.8	1.59	.81	.53	.48	.46	.54	.55	.55	.49	.58	1			
11. Rundbögen (QB)	1.0	0.77	.78	.45	.43	<b>.36**</b>	.45	.49	.60	.61	.61	.79	1		
12. Rundbögen (PA)	2.0	1.64	.77	.44	.44	.44	.48	.47	.54	.49	.54	.76	.80	1	
13. Rundbögen OE	4.9	3.70	.85	.51	.49	.47	.54	.55	.60	.56	.61	.93	.90	.93	(0.92)

Anmerkungen: M = Mittelwert; SD = Standardabweichung; n = 63 (Studierende des ersten Semesters der Zahnmedizin)

alle Korrelationen signifikant auf  $p < .001$  (zweiseitig), außer mit \*\* markierte auf  $p < .01$  signifikant (zweiseitig)

Reliabilitätskoeffizient Cronbachs Alpha ist in der Tabellendiagonalen in Klammern abgebildet

OE = Orientierungseinheit. Testdurchführung HAM-Man in der ersten Studienwoche

1. HAM-Man = Drahtbiegetest (instrumentelles, manuelles Geschick, 0 bis 45 Pkt.)

2. - 13.: Dreieck (0 bis 15 Pkt.); Dritte Dim. = Dritte Dimension (0 bis 15 Pkt.); Rundbögen (0 bis 15 Pkt.); DG = Kriterium Deckungsgleichheit, siebenstufige Skala (0 bis 6 Pkt.); QB = Kriterium Qualität der Biegung, siebenstufige Skala (0 bis 3 Pkt.); PA = Kriterium planes Aufliegen, siebenstufige Skala (0 bis 6 Pkt.)

Tabelle A07: HAM-Man OE - Deskriptive Statistik des HAM-Man in der Orientierungseinheit

Variable	n	M	SD	$\sigma^2$	Min	Max	Schiefe	Kurtosis	p-SWT
HAM-Man	63	21.3	8.96	80.28	0	37.25	-0.60	-0.01	0.038
	37	24.3	7.76	60.27	0.50	37.25	-0.88	1.50	0.084
- Dreieck	63	9.8	3.12	9.76	0	14.25	-1.25	2.44	
	37	10.6	2.67	7.11	0.50	14.25	-1.43	4.35	
-- DG	63	3.8	1.46	2.12	0	6	-0.65	0.14	
	37	4.2	1.28	1.63	0.50	6	-.76	0.71	
-- QB	63	1.8	0.72	0.51	0	2.75	-0.97	0.81	
	37	2.0	0.59	0.35	0	2.75	-1.02	1.93	
-- PA	63	4.1	1.24	1.54	0	6	-1.63	4.03	
	37	4.4	1.06	1.12	0	6	-1.84	6.96	
- Dritte Dim.	63	6.7	3.64	13.28	0	13.50	-0.36	-0.62	
	37	7.8	3.26	10.62	0	13.50	-0.69	0.33	
-- DG	63	2.4	1.62	2.63	0	6	0.05	-0.82	
	37	2.9	1.54	2.36	0	6	-0.37	-0.26	
-- QB	63	1.4	0.76	0.58	0	2.75	-0.26	-0.26	
	37	1.7	0.68	0.45	0	2.75	-0.48	0.83	
-- PA	63	2.8	1.60	2.57	0	5.5	-0.30	-0.75	
	37	3.2	1.47	2.17	0	5.5	-0.42	-0.44	
- Rundbögen	63	4.9	3.70	13.70	0	12.75	-0.01	-1.05	
	37	5.8	3.54	12.52	0	12.75	-0.19	-0.50	
-- DG	63	1.8	1.59	2.52	0	5.0	0.24	-1.09	
	37	2.4	1.63	2.65	0	5.0	-0.15	-1.08	
-- QB	63	1.0	0.77	0.06	0	2.75	0.06	-0.84	
	37	1.2	0.71	0.51	0	2.75	-0.23	-0.22	
-- PA	63	2.0	1.64	2.68	0	5.5	0.12	-1.05	
	37	2.3	1.57	2.47	0	5.5	0.12	-0.71	

Anmerkung: n = Stichprobengröße; M = Mittelwert; SD = Standardabweichung;  $\sigma^2$  = Varianz; Min / Max = Mindest- und Maximalwert; p-SWT = Signifikanz des Shapiro-Wilk-Tests; DG = Kriterium Deckungsgleichheit; QB = Kriterium Qualität der Biegung; PA = Kriterium planes Aufliegen; Dritte Dim. = Dritte Dimension  
Durchführung HAM-Man in der ersten Studienwoche (Orientierungseinheit) mit 63 Studierende, von denen 37 den Test im Hochschulauswahlverfahren absolvierten (= Testwiederholung).

Tabelle A08: HAM-Man OE - Interrater-Reliabilitäten (ICC<sub>unjustiert</sub>)

Item	ICC <sub>unjust.</sub>	99% KI		F-Wert	Sig.	$\alpha$	M R1 (SD)	M R2 (SD)	$\delta$ M R1 & R2
		u. G.	o. G.						
HAM-Man	0.95	0.57	0.99	35.40	.000	0.97	22.6 (9.51)	20.0 (8.64)	2.62
- Dreieck	0.87	0.62	0.94	17.75	.000	0.94	10.2 (3.21)	9.3 (3.21)	0.85
-- DG	0.88	0.78	0.94	15.46	.000	0.94	3.8 (1.48)	3.8 (1.53)	0
-- QB	0.53	-0.07	0.81	4.93	.000	0.80	2.1 (0.75)	1.6 (0.19)	1.64
-- PA	0.88	0.70	0.95	18.31	.000	0.95	4.3 (1.29)	4.0 (1.26)	0.29
- Dritte Dim.	0.88	0.36	0.96	26.20	.000	0.96	7.3 (4.00)	6.1 (3.40)	1.19
-- DG	0.92	0.77	0.97	30.13	.000	0.97	2.6 (1.74)	2.2 (1.55)	0.32
-- QB	0.87	-0.14	0.87	7.53	.000	0.87	1.7 (0.86)	1.1 (0.76)	0.60
-- PA	0.91	0.80	0.96	24.10	.000	0.96	3.0 (1.78)	2.7 (1.48)	0.27
- Rundbögen	0.94	0.84	0.98	41.63	.000	0.98	5.2 (3.86)	4.6 (3.63)	0.58
-- DG	0.94	0.89	0.97	34.24	.000	0.97	1.9 (1.62)	1.8 (1.60)	0.11
-- QB	0.72	0.13	0.90	9.04	.000	0.89	1.2 (0.91)	0.8 (0.71)	0.40
-- PA	0.94	0.89	0.97	33.62	.000	0.97	2.1 (1.67)	2.0 (1.65)	0.04

Anmerkung: Testdurchführung HAM-Man in der ersten Studienwoche (Orientierungseinheit) mit n = 63 Zahnmedizinstudierenden. ICC = Intra-Klassen-Korrelationen, unjustiertes Modell: zweifach, gemischt; 99% KI = 99% Konfidenzintervall; Sig. = Signifikanz F-Tests;  $\alpha$  = Reliabilitätskoeffizient Cronbachs Alpha; M R1 = Mittelwert Rater1; SD = Standardabweichung; M R2 = Mittelwert Rater2;  $\delta$  Rater = Differenz der Mittelwerte Rater1 & Rater2; DG = Kriterium Deckungsgleichheit; QB = Kriterium Qualität der Biegung; PA = Kriterium planes Aufliegen; Dritte Dim. = Dritte Dimension

Tabelle A09: Parameterschätzungen der Indikatoren im KFA-Modell (Ein-Faktor-Modell)

Indikator <-- Faktor	$\lambda_{ijn}$	S.E.	C.R.	$\lambda_{ij}$	$\theta_{ii}$	$h^2$	rel( $x_i$ )
LPS UT 7 <-- Raum	2.773	0.915	3.031	0.330	0.884	0.109	0.110
LPS UT 8 <-- Raum	2.733	0.689	3.968	0.415	0.821	0.172	0.173
LPS UT 9 <-- Raum	3.739	0.586	6.375	0.775	0.396	0.601	0.603
LPS UT 10 <-- Raum	2.478	0.607	4.084	0.454	0.788	0.206	0.207
Dreieck - DG <-- Geschick	0.928	0.128	7.246	0.593	0.643	0.352	0.354
Dreieck - QB <-- Geschick	0.465	0.060	7.697	0.623	0.608	0.388	0.390
Dreieck - PA <-- Geschick	0.727	0.112	6.474	0.540	0.703	0.291	0.293
Dritte Dim. - DG <-- Geschick	1.779	0.122	14.535	0.954	0.089	0.890	0.911
Dritte Dim. - QB <-- Geschick	0.887	0.062	14.232	0.943	0.109	0.911	0.891
Dritte Dim. - PA <-- Geschick	1.619	0.119	13.634	0.921	0.151	0.848	0.849
Rundbögen - DG <-- Geschick	0.892	0.122	7.286	0.596	0.640	0.355	0.357
Rundbögen - QB <-- Geschick	0.371	0.053	7.033	0.578	0.660	0.335	0.336
Rundbögen - PA <-- Geschick	0.820	0.112	7.344	0.599	0.636	0.359	0.361

Anm.:  $\lambda_{ijn}$  = nichtstandardisierte Faktorladungen; S.E. = Standardfehler der Schätzung; C.R. = Critical Ratio;  $\lambda_{ij}$  = standardisierte Faktorladungen;  $\theta_{ii}$  = Messfehlervarianzen;  $h^2$  = Kommunalitäten; rel( $x_i$ ) = Indikatorreliabilität; Raum = räumliches Vorstellungsvermögen; Geschick = instrumentelles, manuelles Geschick; Dritte Dim. = Dritte Dimension  
Signifikanzen: Schätzungen der Faktorladungen auf Niveau  $p < .000$  signifikant, außer für Indikator LPS UT (  $p < .002$  ); Schätzungen der Messfehlervarianzen auf Niveau  $p < .000$  signifikant, außer für Indikator LPS UT 9 (  $p < .012$  )

Tabelle A10: Faktorenkennwerte im KFA-Modell (Ein-Faktor-Modell)

Faktoren	M (SD)	$\alpha$	$r_{i(t-i)}$	DEV ( $\xi_j$ )	rel( $\xi_j$ )
räumliches Vorstellungsvermögen	29.88 (6.33)	0.58	0.347	0.274	0.574
instrumentelles, manuelles Geschick	3.2 (1.22)	0.95	0.879	0.527	0.905

Anm.: M = Mittelwert; SD = Standardabweichung;  $\alpha$  = Cronbachs  $\alpha$  für standardisierte Items;  $r_{i(t-i)}$  = mittlere korrigierte Trennschärfe; DEV ( $\xi_j$ ) = Durchschnittlich extrahierte Varianz; rel( $\xi_j$ ) = Indikatorreliabilität

Tabelle A11: Korrelationen und Validitäten im KFA-Modell (Ein-Faktor-Modell)

$r_{(\xi_j, \xi_k)}^*$			DEV ( $\xi$ )	DEV ( $\xi$ ) > $r_{\xi_j, \xi_k}^2$
LV	1	2		Fornell-Larcker-Kriterium
1. instrumentelles, manuelles Geschick $\xi_j$	1	0.441	0.527	0.527 > 0.194
2. räumliches Vorstellungsvermögen $\xi_k$	0.441	1	0.274	0.274 > 0.194

Anm.: LV = latente Variable;  $r_{(\xi_j, \xi_k)}$  = Korrelation zwischen den Prädiktoren; DEV ( $\xi_j$ ) = Durchschnittlich extrahierte Varianz; DEV ( $\xi$ ) >  $r_{\xi_j, \xi_k}^2$  = Fornell-Larcker-Kriterium;  $r_{(\xi_j, \xi_k)}$  signifikant auf dem Niveau  $p < .000$

Tabelle A12: Parameterschätzungen der Indikatoren im KFA-Modell (Drei-Faktoren-Modell)

Indikator <-- Faktor	$\lambda_{ijn}$	S.E.	C.R.	$\lambda_{ij}$	$\theta_{ii}$	$h^2$	rel( $x_i$ )
LPS UT 7 <-- Raum	2.806	0.874	3.209	0.334	0.881	0.112	0.112
LPS UT 8 <-- Raum	2.733	0.684	3.995	0.415	0.821	0.172	0.173
LPS UT 9 <-- Raum	3.718	0.550	6.757	0.771	0.402	0.595	0.597
LPS UT 10 <-- Raum	2.494	0.568	4.394	0.457	0.785	0.209	0.210
Dreieck - DG <-- Faktor Dreieck	1.450	0.106	13.674	0.926	0.141	0.858	0.859
Dreieck - QB <-- Faktor Dreieck	0.714	0.049	14.464	0.956	0.085	0.914	0.915
Dreieck - PA <-- Faktor Dreieck	1.198	0.094	12.784	0.890	0.207	0.791	0.793
Dritte Dim. - DG <-- Faktor Dritte Dim.	1.792	0.122	14.713	0.961	0.076	0.924	0.924
Dritte Dim. - QB <-- Faktor Dritte Dim.	0.896	0.062	14.478	0.953	0.092	0.907	0.908
Dritte Dim. - PA <-- Faktor Dritte Dim.	1.629	0.118	13.781	0.927	0.140	0.859	0.860
Rundbögen - DG <-- Faktor Rundbögen	1.449	0.098	14.722	0.967	0.064	0.935	0.936
Rundbögen - QB <-- Faktor Rundbögen	0.588	0.044	13.399	0.916	0.159	0.840	0.841
Rundbögen - PA <-- Faktor Rundbögen	1.179	0.097	12.125	0.862	0.256	0.742	0.744

Anm.:  $\lambda_{ijn}$  = nichtstandardisierte Faktorladungen; S.E. = Standardfehler der Schätzung; C.R. = Critical Ratio;  $\lambda_{ij}$  = standardisierte Faktorladungen;  $\theta_{ii}$  = Messfehlervarianzen;  $h^2$  = Kommunalitäten; rel ( $x_i$ ) = Indikatorreliabilität; Raum = räumliches Vorstellungsvermögen; Dritte Dim. = Dritte Dimension

Signifikanzen: Schätzungen der Faktorladungen auf Niveau  $p < .000$  signifikant; Schätzungen der Messfehlervarianzen auf Niveau  $p < .000$  signifikant. außer für die „Indikatoren Rundbögen – DG“ ( $p < .012$ ) und „LPS UT 9“ ( $p < .005$ )

Tabelle A13: Faktorenkennwerte im KFA-Modell (Drei-Faktoren-Modell)

Faktoren	M (SD)	$\alpha$	$r_{i(t-i)}$	DEV ( $\xi_j$ )	rel( $\xi_j$ )
räumliches Vorstellungsvermögen	29.88 (6.33)	0.58	0.347	0.273	0.575
Faktor Dreieck	3.2 (1.22)	0.95	0.879	0.855	0.957
Faktor Dritte Dimension	1.6 (1.52)	0.96	0.918	0.897	0.963
Faktor Rundbögen	0.9 (1.05)	0.94	0.869	0.840	0.940

Anm.: M = Mittelwert; SD = Standardabweichung;  $\alpha$  = Cronbachs  $\alpha$  für standardisierte Items;  $r_{i(t-i)}$  = mittlere korrigierte Trennschärfe; DEV ( $\xi_j$ ) = Durchschnittlich extrahierte Varianz; rel( $\xi_j$ ) = Indikatorreliabilität

Tabelle A14: Korrelationen und Validitäten im KFA-Modell (Drei-Faktoren-Modell)

r ( $\xi_{j1}$ , $\xi_{j2}$ , $\xi_{j3}$ , $\xi_k$ )					DEV ( $\xi$ )	DEV ( $\xi$ ) > $r_{\xi_j, \xi_k}^2$
LV	Faktor Dreieck $\xi_{j1}$	Faktor Dritte Dimension $\xi_{j2}$	Faktor Rundbögen $\xi_{j3}$	räumliches Vorstellungs- vermögen $\xi_k$		Fornell-Larcker- Kriterium
Faktor Dreieck $\xi_{j1}$	1				0.855	0.855 > 0.347 ( $\xi_{j2}$ ) 0.855 > 0.197 ( $\xi_{j3}$ ) 0.855 > 0.193 ( $\xi_k$ )
Faktor Dritte Dimension $\xi_{j2}$	0.589	1			0.897	0.897 > 0.347 ( $\xi_{j1}$ ) 0.897 > 0.336 ( $\xi_{j3}$ ) 0.897 > 0.171 ( $\xi_k$ )
Faktor Rundbögen $\xi_{j3}$	0.444	0.580	1		0.841	0.841 > 0.197 ( $\xi_{j1}$ ) 0.841 > 0.336 ( $\xi_{j2}$ ) 0.841 > 0.107 ( $\xi_k$ )
räumliches Vorstellungs- vermögen $\xi_k$	0.440	0.413	0.327	1	0.273	0.273 > 0.194 ( $\xi_{j1}$ ) 0.273 > 0.171 ( $\xi_{j2}$ ) 0.273 > 0.107 ( $\xi_{j3}$ )

Anm.: LV = latente Variable; r ( $\xi_j$ ,  $\xi_k$ ) = Korrelation zwischen den Prädiktoren; DEV ( $\xi_j$ ) = Durchschnittlich extrahierte Varianz; DEV ( $\xi$ ) >  $r_{\xi_j, \xi_k}^2$  = Fornell-Larcker-Kriterium; Korrelationen signifikant auf dem Niveau  $p < .000$

Tabelle A15: Faktorladungen im PhaII-Messmodell

latente Variable	Indikator	$\lambda_{ij}$	S.E.	t-Wert
räumliches Vorstellungsvermögen	LPS UT 7	0.9261	0.474	1.96
	LPS UT 8	0.6439	0.456	1.41
	LPS UT 9	0.2741	0.485	0.57
Faktor Dreieck	Dreieck - DG	0.8586	0.337	2.55
	Dreieck - QB	0.8750	0.346	2.53
	Dreieck - PA	0.9200	0.323	2.84
Faktor Dritte Dimension	Dritte Dim - DG	0.9576	0.017	56.42
	Dritte Dim - QB	0.9306	0.039	23.82
	Dritte Dim - PA	0.9612	0.020	48.90
Faktor Rundbögen	Rundbögen - DG	0.9612	0.021	46.34
	Rundbögen - QB	0.9245	0.037	25.09
	Rundbögen - PA	0.9396	0.018	52.25
allgemeiner Faktor instrumentelles, manuelles Geschick	Dreieck - DG	0.8586	0.249	1.78
	Dreieck - QB	0.8750	0.273	1.62
	Dreieck - PA	0.9200	0.236	2.50
	Dritte Dimension - DG	0.9576	0.058	14.80
	Dritte Dimension - QB	0.9306	0.100	7.54
	Dritte Dimension - PA	0.9612	0.069	11.74
	Rundbögen - DG	0.9612	0.073	10.84
	Rundbögen - QB	0.9245	0.092	8.20
	Rundbögen - PA	0.9396	0.069	11.68
Schulabschlussnote	Abiturdurchschnittsnote	1	0	0
naturwissenschaftliches Verständnis	HAM-Nat	1	0	0
Fertigkeitserwerb im PhaII für inkonsistente, komplexe Aufgaben	PhaII Note1	0.5973	0.286	2.09
	PhaII Note2	0.8899	0.229	3.88
	PhaII Note3	0.6458	0.316	2.04

Anm.:  $\lambda_{ij, un}$  = geschätzte Faktorladungen; S.E. = Standardfehler der Schätzung; t-Wert = Testprüfgröße des t-Tests, LPS = Leistungsprüfsystem (Horn, 1983); UT = Untertest; DG = Auswertungskriterium Deckungsgleichheit; QB = Auswertungskriterium Qualität der Biegung; PA = Auswertungskriterium Planes Aufliegen der Drähte; Abiturnote = Abiturdurchschnittsnote; HAM-Nat = Naturwissenschaftstest der Medizinischen Fakultät der Universität Hamburg; PhaII = Phantomkurs II

1. Stichprobe: n = 28
2. Codierung Noten: 1 = ungenügend, 6 = sehr gut
3. „Bootstrapping“-Verfahren: 5000 Zufallsstichproben

Tabelle A16: Gütekriterien des PhaII-Messmodells

latente Variable ( $\xi_j$ )	DEV ( $\xi_j$ )	$\alpha$	$H_j^2$	$\rho_c$	Fornell-Larcker-Kriterium
Fertigkeitserwerb im PhaII ( $\xi_{endo}$ ) für inkonsistente, komplexe Aufgaben	0.522	0.59	0.522	0.76	$\xi_{endo} \geq \xi_{j1}$ 0.522 > 0.0441 $\xi_{endo} \geq \xi_{j2}$ 0.522 > 0.2116 $\xi_{endo} \geq \xi_{j3}$ 0.522 > 0.0001 $\xi_{endo} \geq \xi_{j4}$ 0.522 > 0.0729
räumliches Vorstellungsvermögen ( $\xi_{j1}$ )	0.449	0.52	0.449	0.67	$\xi_{j1} \geq \xi_{j2}$ 0.449 > 0.0036 $\xi_{j1} \geq \xi_{j3}$ 0.449 > 0.0004 $\xi_{j1} \geq \xi_{j4}$ 0.449 > 0.0841 $\xi_{j1} \geq \xi_{endo}$ 0.449 > 0.0441
Schulabschlussnote ( $\xi_{j3}$ )	--	--	--	--	--
naturwiss. Verständnis ( $\xi_{j4}$ )	1	1	--	1	--
<b>Bi-Faktor-Modell</b>					
Gruppenfaktor Dreieck	0.783	0.86	0.783	0.92	$\xi_{Dreieck} \geq \xi_{Dritte Dim.}$ 0.783 > 0.0887 $\xi_{Dreieck} \geq \xi_{Rundbögen}$ 0.783 > 0.0655
Gruppenfaktor Dritte Dimension	0.902	0.95	0.902	0.97	$\xi_{Dritte Dim.} \geq \xi_{Dreieck}$ 0.902 > 0.0887 $\xi_{Dritte Dim.} \geq \xi_{Rundbögen}$ 0.902 > 0.2812
Gruppenfaktor Rundbögen	0.887	0.94	0.887	0.96	$\xi_{Rundbögen} \geq \xi_{Dreieck}$ 0.887 > 0.0655 $\xi_{Rundbögen} \geq \xi_{Dritte Dim.}$ 0.887 > 0.2812
allgemeiner Faktor instrumentelles, manuelles Geschick ( $\xi_{j2}$ )	0.504	0.87	0.504	0.91	$\xi_{j2} \geq \xi_{j1}$ 0.504 > 0.0036 $\xi_{j2} \geq \xi_{j3}$ 0.504 > 0.0114 $\xi_{j2} \geq \xi_{j4}$ 0.504 > 0.0335 $\xi_{j2} \geq \xi_{endo}$ 0.504 > 0.2116

Anm.: DEV ( $\xi_j$ ) = Durchschnittlich extrahierte Varianz;  $\alpha$  = Cronbachs Alpha;  $H_j^2$  = Kreuzvalidierungsmaß Kommunalität;

$\rho_c$  = Faktorreliaibilität; Dim. = Dimension; PhaII = Phantomkurs II; naturwiss. = naturwissenschaftlich

1. Stichprobe: n = 28

Tabelle A17: Kreuzkorrelationen im PhaII-Prädiktionsmodell

latente Variable Indikator	Raum	Man	Faktor Dreieck	Faktor Dritte Dim.	Faktor Rund- bögen	Schul.	natur. Verst.	FKE im PhaII
LPS UT 7	.93	.05	.18	.05	-.06	-.06	.35	-.21
LPS UT 8	.64	.14	.22	.08	.07	.17	.06	-.10
LPS UT 9	.27	.42	.16	.30	.46	-.12	.32	.02
Dreieck - DG	.24	.44	.86	.16	.24	.23	-.19	-.04
Dreieck - QB	.07	.44	.88	.24	.14	.26	-.22	.18
Dreieck - PA	.25	.59	.92	.37	.28	.19	-.06	.28
Dritte Dim. - DG	.05	.85	.34	.96	.56	.10	.22	.32
Dritte Dim. - QB	.11	.75	.18	.93	.48	-.17	.35	.43
Dritte Dim. - PA	.00	.81	.33	.96	.47	.05	.17	.33
Rundbögen - DG	.03	.79	.26	.47	.96	.09	.11	.39
Rundbögen - QB	-.08	.75	.18	.49	.92	-.14	.29	.48
Rundbögen - PA	-.10	.81	.27	.53	.94	.25	.18	.40
Abiturnote	.02	.11	.25	.00	.07	1	-.27	-.01
HAM-Nat	.29	.18	-.17	.26	.20	-.27	1	.27
PhaII Note1	-.16	.25	.41	.03	.22	.19	-.15	.60
PhaII Note2	-.15	.48	-.04	.52	.47	-.06	.35	.89
PhaII Note3	-.19	.18	.27	.02	.18	-.02	.17	.65

Anm.: UT = Untertest; DG = Deckungsgleichheit; QB = Qualität der Biegung; PA = planes Aufliegen der Drähte;  
Dim. = Dimension; Abiturnote = Abiturdurchschnittsnote; HAM-Nat = Naturwissenschaftstest der Medizinischen Fakultät  
Hamburg; Raum = räumliches Vorstellungsvermögen; Man = instrumentelles, manuelles Geschick; Schul. = Schulabschlussnote;  
natur. Verst. = naturwissenschaftliches Verständnis; FKE = Fertigkeitserwerb bei inkonsistenten, komplexen Aufgaben;  
Pha = Phantomkurs

3. Stichprobe: n = 28

4. Codierung Noten: 1 = ungenügend, 6 = sehr gut

Tabelle A18: Korrelationen zwischen den latenten Variablen im PhaII-Strukturmodell

latente Variable	1	2	3	4	5	6	7	8
1. instrumentelles, manuelles Geschick	(0.89)							
2. räumliches Vorstellungsvermögen	.06	(0.45)						
3. Schulabschlussnote	.11	.02	1					
4. naturwissenschaftliches Verständnis	.18	.29	-.27	1				
5. Faktor Dreieck	.57	.22	.25	.17	(0.78)			
6. Faktor Dritte Dimension	.85	.06	0	.26	.30	(0.90)		
7. Faktor Rundbögen	.83	-.05	.07	.21	.26	.53	(0.89)	
8. Fertigkeitserwerb im PhaII	.46	-.21	-.01	.27	.17	.45	.27	(0.52)

Anm.: PhaII = Phantomkurs II (inkonsistente, komplexe Aufgaben)

1. Stichprobengrößen: n = 28

2. Reliabilitätskoeffizient Cronbachs Alpha ist in der Tabellendiagonalen in Klammern abgebildet

Tabelle A19: Gütekriterien des PhaII-Strukturmodells

latente Variablen	beta (t-Wert)	R <sup>2</sup>	f <sub>j</sub> <sup>2</sup>	Q <sup>2</sup>
Fertigkeitserwerb im PhaII	--	0.345	--	-0.002
räumliches Vorstellungsvermögen	-0.325 (1.13)		0.118	
instrumentelles, manuelles Geschick	0.423 (1.86)		0.195	
Schulabschlussnote	0.035 (0.15)		0.000	
naturwissenschaftliches Verständnis	0.299 (1.16)		0.116	

Anm.:  $\beta$  = Pfadkoeffizient; t-Wert = Testprüfgröße des t-Tests (einseitig). R<sup>2</sup> = Bestimmtheitsmaß; f<sup>2</sup> = Effektstärke;

Q<sup>2</sup> = Stone-Geisser-Kriterium; PhaII = Phantomkurs II (inkonsistente, komplexe Aufgaben)

1. Stichprobe: n = 28

2. „Bootstrapping“-Verfahren: 5000 Zufallsstichproben

Tabelle A20: Inkrementelle Validitäten im PhaII-Prädiktionsmodell

Prädiktor (exogene Variable)	- R <sup>2</sup> - Prädiktor im Modell eingeschlossen	- R <sup>2</sup> - Prädiktor im Modell ausgeschlossen	ΔR <sup>2</sup>
instrumentelles, manuelles Geschick	34.5	21.7	13.8
räumliches Vorstellungsvermögen	34.5	26.8	7.7

Anm.: ΔR<sup>2</sup> = inkrementelle Validität

Tabelle A21: Koeffizienten der robusten Regressionsanalyse

Prädiktor	b <sup>MM</sup> (S.E.)	t-Wert	β <sup>MM</sup>	PLS: β (t-Wert)
<b>Kriterium inkonsistente, komplexe Aufgaben (PhaII)</b>				
instrumentelles, manuelles Geschick	0.017 (0.008)	2.071	0.426	0.423 (1.86)
räumliches Vorstellungsvermögen	-0.006 (0.005)	-1.225	-0.284	-0.325 (1.13)
Schulabschlussnote	-0.004 (0.029)	-0.155	-0.029	0.035 (0.15)
naturwissenschaftliches Verständnis	0.010 (0.009)	1.076	0.240	0.299 (1.16)

Anm.: Methode = MM-Schätzer; β = Regressionskoeffizient; S.E. = Standardfehler; beta = standardisierter Regressionskoeffizient Modell (PhaII) für β/Beta; Residuum-Standardfehler: 0.2732/ 0.8126, df = 23

Tabelle A22: Bewertungskriterien der vorklinischen Laborkursaufgaben

Laborkurs	Aufgabe	Bewertungskriterien
TPK	Schnittübung	Morphologie der Krone, Proportionen, Wachsverarbeitung, Politur
	Aufwachsübung	Zahnform, Kauflächenrelief (x2), statische Okklusion(x2), Laterotrusionskontakte, Mediotrusionskontakte (ja, nein)
	Vollgusskrone	Zahnform, Kauflächengestaltung, Ausarbeitung und Politur, okklusale Kontakte, approximale Kontakte, Mediotrusionskontakte (ja, nein), Randschluß (x2), Präparation (x3)
	Interimsprothese	Okklusion, Basiskongruenz, Verlauf der Klammern, Retention, Präparation der Auflagemulden (x2), Ausarbeitung und Politur
PhaI	Totalprothese	Frontzahnaufstellung, Seitenzahnaufstellung, Okklusion statisch (x2), Balancekontakte bei Lat. links (ja, nein), Frontzahnkontakt bei Laterotrusion links (ja, nein), Balancekontakte bei Laterotrusion rechts (ja, nein), Frontzahnkontakt bei Laterotrusion rechts (ja, nein), Ausarbeitung und Politur
	Stiftaufbau und Schalenprovisorium	Präparation (x3), approximale Kontakte, anatomische Formgebung, Randschluß (x2), Ausarbeitung und Politur
	Präparation und Tangentialbrücke	Präparation (x3), approximale Kontakte, okklusale Kontakte (x2), Mediotrusionskontakte (ja, nein), Kauflächengestaltung, Zahnform, Ausarbeitung und Politur, Randschluß (x2)
PhaII	Totalprothese	Frontzahnaufstellung, Seitenzahnaufstellung, Okklusion statisch, Balancekontakte (x3), Ausarbeitung und Politur
	Präparation und Schalenprovisorium	Anatomische Formgebung, approximale Kontakte, statische Okklusion, dynamische Okklusion, Mediotrusion (ja, nein), Ausarbeitung und Politur, Randschluß (x2), Präparation (x3)
	Michiganschiene	Passung, statische Okklusion, Laterotrusion, Protrusion, Retention, Ausarbeitung und Politur

Anm.: x = mehrfache Gewichtung des Kriteriums

## **Lebenslauf**

Christian Kothe

geboren am 5. April 1978 in Wismar  
deutsche Staatsangehörigkeit

### **Akademischer Werdegang**

---

08/2010 - bis heute	Christian-Albrechts-Universität zu Kiel, externer Promovend (berufsbegleitend) am Lehrstuhl für Arbeits-, Organisations- und Marktpsychologie
06/2010 - 02/2014	Universitätsklinikum Hamburg-Eppendorf, Forschungsarbeitsgruppe AdH, wissenschaftlicher Angestellter
06/2008 - 06/2011	Universitätsklinikum Hamburg-Eppendorf, Prodekanat für Lehre, Qualitätssicherung, wissenschaftlicher Angestellter
03/2007 - 02/2008	Weinmann GmbH + Co.KG, Junior-Personalreferent

### **Praktika**

01/2006 - 07/2006	Ameco Beijing, Peking (China), Personalentwicklung
08/2003 - 03/2004	Lufthansa Technik AG, Personalmarketing

### **Akademische Ausbildung**

---

10/1999 - 10/2005	Studium der Psychologie an den Universitäten Konstanz und Christian-Albrechts-Universität zu Kiel
-------------------	---------------------------------------------------------------------------------------------------

### **Berufsausbildung**

---

08/1997 - 06/1999	Ausbildung zum Groß- und Außenhandelskaufmann im Lebensmittelindustriebetrieb Pickenpack Tiefkühlgesellschaft mbH, Lüneburg
-------------------	-----------------------------------------------------------------------------------------------------------------------------

### **Schulbildung, Wehrdienst**

---

09/1996 - 06/1997	Grundwehrdienst
08/1991 - 06/1996	Gymnasium Hagenow

Hamburg, den 30. Januar 2014

Christian Kothe