

Aus der Klinik für Mund-, Kiefer- und Gesichtschirurgie
(Prof. Dr. med. Dr. med. dent. H. Schliephake)
im Zentrum Zahn-, Mund- und Kieferheilkunde
der Medizinischen Fakultät der Universität Göttingen

**Evaluation der Auswirkungen eines
strukturierten Peerteachings
zahntechnischer Einzelaspekte auf die
Leistung von Studienanfängern – ein
Pilotprojekt**

INAUGURAL-DISSERTATION

zur Erlangung des Doktorgrades
[für Zahnmedizin]
der Medizinischen Fakultät der
Georg-August-Universität zu Göttingen

vorgelegt von

Ladan Darweshi-Pecnik

aus

Kandahar

Göttingen 2021

Dekan: Prof. Dr. med. W. Brück

Betreuungsausschuss

Betreuer/in: PD Dr. med. dent. S. Sennhenn-Kirchner

Ko-Betreuer/in: Prof. Dr. med. dent. M. Rödiger

Prüfungskommission

Referent/in: PD Dr. med. dent. S. Sennhenn-Kirchner

Ko-Referent/in: Prof. Dr. med. dent. M. Rödiger

Drittreferent/in:

Datum der mündlichen Prüfung: 02.06.2022

Hiermit erkläre ich, die Dissertation mit dem Titel „Evaluation der Auswirkungen eines strukturierten Peerteachings zahntechnischer Einzelaspekte auf die Leistung von Studienanfängern – ein Pilotprojekt“ eigenständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Kassel, den

.....

(Unterschrift)

Inhaltsverzeichnis

Tabellen- und Abbildungsverzeichnis.....	IV
Abkürzungsverzeichnis	V
1 Einleitung	1
1.1 Besonderheiten des Studiengangs Zahnmedizin	1
1.2 Die Lehrformen.....	3
1.2.1 Vorlesung.....	4
1.2.2 Seminar	5
1.2.3 Problemorientiertes Lernen/Problem Based Learning (POL/PBL)	5
1.2.4 Electronic-Learning	6
1.2.5 Praktische Übungen.....	6
1.3 Die Lehrmethoden.....	7
1.3.1 <i>See one, do one</i>	7
1.3.2 4-Schritt-Methode nach Peyton	7
1.3.3 Peerteaching/PAL.....	9
1.3.4 Die Lehrmethoden an der UMG am Beispiel des TPK.....	9
1.4 OSCE/OSPE in der Literatur.....	10
1.4.1 Die OSCE an der UMG.....	11
1.5 Testgütekriterien.....	12
1.5.1 Validität, Reliabilität, Objektivität.....	12
1.6 Überblick über das Thema.....	14
1.7 Formulierung der Hypothese	15
2 Material und Methoden	16
2.1 Materialliste	16
2.2 Die statistischen Methoden	17
2.2.1 Die Itemanalyse	19
2.2.2 Deskriptive Statistik und Verteilungseigenschaften.....	21
2.2.3 Planung der Studie	23
2.2.4 Gruppenbildung	24
2.2.5 Peerteacher – Auswahl	25
2.2.6 Vorbereitung der Lehre	26
2.2.7 Durchführung der Lehre.....	27
2.2.8 Vorbereitung und Durchführung der ersten OSPE	28
2.2.9 Follow-Up nach drei Monaten/zweite OSPE	30
2.2.10 Das Bewertungsformat.....	30
2.2.11 Die Datenauswertung.....	30
3 Ergebnisse	32
3.1 Itemanalyse der Hauptuntersuchung.....	32
3.1.1 Deskriptive Statistik und Verteilungseigenschaften.....	37

3.1.2	Vergleich der Mittelwerte mit dem t-Test für unabhängige Stichproben.....	38
3.2	Itemanalyse der Follow-Up-Untersuchung.....	39
3.2.1	Deskriptive Statistik und Verteilungseigenschaften (Follow-Up).....	41
3.2.2	Vergleich der Mittelwerte mit dem t-Test für unabhängige Stichproben (Follow-Up).....	42
3.2.3	Vergleich der Gruppenmittelwert-Differenz innerhalb einer Gruppe von der ersten zur zweiten OSPE mit dem t-Test für abhängige Stichproben	43
3.2.4	Die lineare Regression	44
3.2.5	Studentische Evaluation des Pilotprojekts	46
4	Diskussion	49
4.1	Diskussion der Methoden.....	49
4.1.1	Studiendesign	49
4.1.2	Zeitlicher Ablauf des Pilotprojekts	50
4.1.3	Die Lehrmethoden in der Literatur.....	51
4.1.4	Retention von Wissen und Fertigkeiten in der Literatur.....	53
4.1.5	Retention erworbener Fertigkeiten.....	53
4.1.6	Peerteaching in der Literatur	55
4.1.7	Bewertung durch studentische Rater.....	56
4.1.8	Effizienz der Raterschulung	56
4.1.9	Bewertung der OSPE und der Checklisten.....	57
4.1.10	Bewertung der studentischen Evaluation	58
4.2	Diskussion der Ergebnisse.....	59
4.2.1	Testgüte der Stationen.....	59
4.2.2	Die Testgüte der Station „Dreiecksklammer biegen“	59
4.2.3	Die Testgüte der Station „Modell eckig trimmen“	60
4.2.4	Die Testgüte der Station „Politur“	61
4.3	Beurteilung der Hypothesen.....	62
4.4	Limitation der Studie	62
4.5	Weiterführende Fragestellung	63
4.6	Implementierung vorklinischer OSPE-Prüfungen an der UMG.....	64
5	Zusammenfassung.....	65
6	Anhang.....	66
6.1	Bögen	66
6.1.1	Einverständniserklärung der Studierenden zur Teilnahme am Pilotprojekt	66
6.1.2	Einverständniserklärung der Studierenden zur Teilnahme an der Follow-Up-Untersuchung.....	67
6.1.3	Evaluationsbogen.....	68
6.2	Item-Checklisten.....	69
6.2.1	Station „Dreiecksklammer biegen“	69
6.2.2	Station „Modell eckig trimmen“	69
6.2.3	Station „Politur“	70
6.3	Lineare Regression	71
7	Literaturverzeichnis	72

Tabellen- und Abbildungsverzeichnis

Abbildung 1: Miller-Pyramide der klinischen Kompetenzen	11
Abbildung 2: Zeitlicher Verlauf	23
Abbildung 3: Die Lehrintervention	28
Tabelle 1: Gliederung des Studiums nach alter und neuer ZApprO.....	2
Tabelle 2: Maastrichter 7-Schritt-Modell	5
Tabelle 3: Materialliste	16
Tabelle 4: Cronbachs Alpha.....	20
Tabelle 5: Itemanalyse Station „Dreiecksklammer biegen“	32
Tabelle 6: Itemanalyse der Station „Modell eckig trimmen"	33
Tabelle 7: Itemanalyse der Station „Modell eckig trimmen" nach Korrektur von Item 6.....	34
Tabelle 8: Itemanalyse der Station „Politur"	35
Tabelle 9: Itemanalyse der Station „Politur" nach Korrektur der Items 3-7.....	36
Tabelle 10: Deskriptive Statistik der Interventionsgruppe (Hauptuntersuchung).....	37
Tabelle 11: Deskriptive Statistik der Kontrollgruppe (Hauptuntersuchung).....	37
Tabelle 12: Vergleich der Gruppenmittelwerte „Dreiecksklammer biegen"	38
Tabelle 13: Vergleich der Gruppenmittelwerte „Modell eckig trimmen"	38
Tabelle 14: Vergleich der Gruppenmittelwerte „Politur"	39
Tabelle 15: Itemanalyse der Station „Dreiecksklammer biegen" (Follow-Up)	39
Tabelle 16: Itemanalyse der Station „Modell eckig trimmen" (Follow-Up)	40
Tabelle 17: Entwicklung der Itemparameter.....	41
Tabelle 18: Deskriptive Statistik der Interventionsgruppe (Follow-Up-Untersuchung)	41
Tabelle 19: Deskriptive Statistik der Kontrollgruppe (Follow-Up-Untersuchung)	42
Tabelle 20: Vergleich der Gruppenmittelwerte „Dreiecksklammer biegen" (Follow-Up).....	42
Tabelle 21: Vergleich der Gruppenmittelwerte „Modell eckig trimmen" (Follow-Up)	43
Tabelle 22: t-Test für abhängige Stichproben (Interventionsgruppe).....	43
Tabelle 23: t-Test für abhängige Stichproben (Kontrollgruppe).....	44
Tabelle 24: Studentische Evaluation zu den Übungen	46
Tabelle 25: Studentische Evaluation zum Peerteaching.....	47

Abkürzungsverzeichnis

AV	Abhängige Variable
E-Learning	<i>Electronic Learning</i>
fMRT	Funktionelle Magnetresonanztomographie
FVDZ	Freier Verband Deutscher Zahnärzte
IT	Informationstechnik
KaPVO	Kapazitätsverordnung
MME	<i>Master of Medical Education</i>
OSCE	<i>Objective Structured Clinical Examination</i>
OSPE	<i>Objective Structured Practical Examination</i>
PAL	<i>Peer Assisted Learning</i>
PBL	<i>Problem Based Learning</i>
POL	Problemorientiertes Lernen
RCT	<i>Randomized Controlled Trial</i>
SINUZ	Studentisches Innovations- und Trainingszentrum Zahnmedizin
STÄPS	Studentisches Trainingszentrum Ärztlicher Praxis und Simulation
TPK	Technisch-Propädeutischer Kurs
UMG	Universitätsmedizin Göttingen
UV	Unabhängige Variable
ZAprO	Approbationsordnung für Zahnärzte und Zahnärztinnen
ZMK	Zahn-Mund-Kieferklinik

1 Einleitung

1.1 Besonderheiten des Studiengangs Zahnmedizin

Der Studiengang der Zahnheilkunde besteht aus elf Semestern, unterteilt in einen vorklinischen und klinischen Abschnitt. Der vorklinische Abschnitt des Studiengangs bezieht sich auf das erste bis fünfte Fachsemester. In der Vorklinik erwerben die Studierenden komplexe Kenntnisse aus dem Feld der Naturwissenschaften und erlernen praktische Fertigkeiten aus dem zahntechnischen Bereich. Nach den ersten beiden Semestern steht die erste staatliche Prüfung an: die naturwissenschaftliche Vorprüfung (Vorphysikum), eine mündliche Prüfung basierend auf den Inhalten aus der Chemie, Physik und Biologie. Die Vorklinik hat für die Studierenden der Zahnmedizin eine weitere Besonderheit im Vergleich zum Studiengang der Humanmedizin. Diese besteht darin, dass es sogenannte Phantomkurse gibt, in denen die Studierenden praktische Fertigkeiten im Bereich des Zahnersatzes (technisch-propädeutischer Kurs/TPK und die Phantomkurse I/II) an einem Phantommodell erlernen. Im Anschluss an das fünfte Semester findet die zweite staatliche Prüfung statt, die zahnärztliche Vorprüfung (Physikum). Diese Prüfung besteht aus einer mündlichen Prüfung in den Fächern Anatomie, Physiologie und Biochemie, sowie einer einwöchigen praktischen Prüfung. Die Anforderungen an die praktische Prüfung ist in der Approbationsordnung (1955) der Zahnmediziner festgehalten und erfordert den Erweis von vier verschiedenen Prüfungsarbeiten. Ein bestandenes Physikum ist Voraussetzung für das Fortführen des zweiten Studienabschnitts. Der klinische Teil umfasst insgesamt fünf Semester (sechs bis zehn). Das sechste Semester beginnt zunächst erneut mit einem Phantomkurs in der Zahnerhaltung.

Die Semester sieben bis zehn beinhalten Patientenbehandlungen in den Abteilungen der zahnärztlichen Prothetik, Zahnerhaltung, zahnärztlichen Chirurgie und Kieferchirurgie. Im elften Semester findet das Staatsexamen statt, bestehend aus praktischen Prüfungen (Prothetik, Zahnerhaltung, Kieferorthopädie und Mund-, Kiefer- und Gesichtschirurgie), sowie zahlreichen zahnärztlichen und allgemeinärztlichen mündlichen Prüfungen. Im gesamten Studium der Zahnmedizin gehen Theorie und Praxis Hand in Hand. Vom ersten Semester an steht das Üben der manuellen Fertigkeiten im Vordergrund. Neben einem großen theoretischen Wissen müssen die Studierenden über handwerkliches Geschick verfügen, um das erlernte Wissen auch umzusetzen.

Zudem fordert die Approbationsordnung in den Prüfungen (Physikum und Staatsexamen) das eigenständige Anfertigen von zahntechnischen Arbeiten an (§ 28, Abs. 4a und § 50). Im Physikum erfolgt die Fertigung vier unterschiedlicher Arbeiten am Phantomkopf. Im Staatsexamen müssen ein festsitzender und ein herausnehmbarer Zahnersatz am Patienten vom Studierenden selbst angefertigt und eingesetzt werden (ZApprO, 1955). Seit einigen Jahren ist eine Debatte um die Anpassung der Approbationsordnung von 1955 entfacht, die seither inhaltlich unverändert blieb.

„Der Verordnungsentwurf zielt auf eine grundlegende Reform des Studiums der Zahnmedizin. Diese ist angesichts der fachlichen Weiterentwicklung der Zahnmedizin und der veränderten Anforderungen einer modernen und interdisziplinären Lehre dringend erforderlich, um auch künftig die Qualität der zahnärztlichen Ausbildung als Voraussetzung für die zahnmedizinische Versorgung der Patientinnen und Patienten in einer älter werdenden Gesellschaft sicherzustellen“ (Verordnung zur Neuregelung der zahnärztlichen Ausbildung, S.1).

In Tabelle 1 wird die Gliederung des Studiums nach alter und novellierter ZApprO dargestellt.

Tabelle 1: Gliederung des Studiums nach alter und neuer ZApprO. Gliederung des Studiums der Zahnmedizin nach alter (links) und neuer (rechts) ZApprO. Die grün hinterlegten Felder weisen auf die jeweiligen staatlichen Prüfungen im Verlauf des Studiums hin.

ZApprO 1955		novellierte ZApprO	
Vorklinischer Abschnitt		Vorklinischer Abschnitt	
1.-2. Semester	TPK, Biologie, Chemie, Physik, Phantom I	1.-4. Semester	Physik, Physiologie, Chemie, Biochemie, Molekularbiologie, Biologie, Anatomie, Psychologie, Soziologie
Naturwissenschaftliche Vorprüfung (Vorphysikum)		Ärztlich-Zahnärztliche Prüfung	
3.-5.Semester	Phantom II, Anatomie, Physiologie, Biochemie	Klinischer Abschnitt, Teil 1:	
Zahnärztliche Vorprüfung (Physikum)		5.-6. Semester	Phantomsemester
Klinischer Abschnitt		2. Abschnitt der zahnärztlichen Prüfung	
6. Semester	Phantomkurs der Zahnerhaltung	Klinischer Abschnitt, Teil 2:	
7.-10. Semester	Behandlung am Patienten	7.-10. Semester	Behandlung am Patienten
11. Semester	Staatsexamen	11. Semester	Staatsexamen

Das Ziel der neuen ZApprO ist es die Studiengänge Zahnmedizin und Humanmedizin in der Vorklinik anzugleichen. Nach den ersten vier Semestern erfolgt eine gemeinsame Ärztlich-Zahnärztliche Prüfung. Die zahnärztliche Ausbildung wird demnach neu strukturiert und in drei Abschnitte eingeteilt. Nach jedem Abschnitt erfolgt eine staatliche Prüfung (in Tabelle 1 grün hinterlegt). Zusätzlich findet eine Neugewichtung der Ausbildungsinhalte zugunsten „der besseren Abbildung von Allgemeinerkrankungen im Zahnmedizinstudium“ statt (ZApprO, S. 2). Für das Erlernen von praktischen Fertigkeiten im zahntechnischen Bereich ist folgende Aussage nachzulesen: „Die zahntechnischen Lehrinhalte werden, auf die für den Zahnarzt und die Zahnärztin erforderlichen zahntechnischen Arbeitsweisen konzentriert“. Des Weiteren sind vor dem Ablegen der Ärztlich-Zahnärztlichen Prüfung ein abgeleiteter Krankenpflegedienst mit der Dauer von einem Monat nachzuweisen. Zusätzlich ist nach bestandener Prüfung eine vierwöchige Famulatur in einer zahnärztlichen Einrichtung abzuleisten. Die neue Approbationsordnung ist ab dem 01. Oktober 2020 offiziell in Kraft getreten. Den Hochschulen wird ein Jahr mithilfe von Übergangsregelungen die volle Umsetzung der neuen ZApprO ermöglicht. Demnach erfolgt die Immatrikulation nach neuer umgesetzter ZApprO ab dem 1. Oktober 2021.

1.2 Die Lehrformen

Von der zahnmedizinischen Lehre im vorklinischen Abschnitt des Studiengangs ist bekannt, dass die Vermittlung von Wissen hauptsächlich über Vorlesungen stattfindet. Das Erlernen praktischer Fertigkeiten findet in den praktischen Kursen (TPK und Phantom I/II) statt. Anhand curricularer Vorgaben, die die Grundstruktur des Unterrichts darstellen, werden die Lernziele klar definiert. Für die Umsetzung einer effizienten und didaktisch sinnvollen Lehre sind neben den angewandten Lehrmethoden die Qualifizierung des Personals, die materielle und personelle Ausstattung, sowie die Organisationsstruktur der Fakultät von großer Bedeutung (Wissenschaftsrat, 2005). Die Voraussetzung für einen erfolgreichen Unterricht besteht unter anderem in der Rekrutierung qualifizierten Personals und der Bereitstellung geeigneter Räumlichkeiten. Qualifiziertes Personal an Universitätskliniken sind in der Regel Forscher und Ärzte, die ihre Lehrbefähigung oft lediglich durch ihre eigenen Erfahrungen während des Studiums erlangt haben. Dadurch ist eine medizindidaktische Qualifizierung des Lehrpersonals unabdingbar (Lammerding-Köppel et al. 2006). Eine Weiterbildung des lehrenden Personals ist nach Zierer (2015) erforderlich, da der Lernerfolg der Studierenden maßgebend durch das Lehrpersonal beeinflusst wird.

Die Weiterbildungen können von einer didaktischen Basisqualifizierung bis hin zu einem Masterprogramm (*Master of Medical Education/MME*) reichen.

In den medizinischen/zahnmedizinischen Fakultäten finden verschiedene Lehrformen ihre Anwendung. Das sind unter anderem klassischerweise Vorlesungen, Seminare, praktische Übungen, problemorientiertes Lernen (POL) und der Einsatz digitaler Medien (E-Learning). Die aktuelle Approbationsordnung für Zahnärzte und Zahnärztinnen sieht an Universitäten folgende Unterrichtsveranstaltungen vor: Vorlesungen, praktische Übungen und Seminare (§ 5, Abs. 2).

Im Folgenden werden die Lehrformen näher erläutert, die an medizinischen Hochschulen Anwendung finden.

1.2.1 Vorlesung

„Die Vorlesung ist eine zusammenhängende Darstellung und Vermittlung von wissenschaftlichen und methodischen Kenntnissen durch den Vortrag von Lehrkräften“ (ZApprO § 6). In der Literatur wird das Konzept Vorlesung als veraltet angesehen, bei dem der Zuhörer nur passiv anwesend ist und ein Informationsbehalten von unter 20% besteht (Widulle 2009). Bligh (2000) zeigte, dass Vorlesungen lediglich zur Vermittlung von Informationen eingesetzt werden sollten und nicht geeignet sind, um zum Nachdenken anzuregen. Um die Aufmerksamkeit der Lernenden aufrechtzuerhalten empfiehlt es sich aktivierende Methoden, wie Gruppendiskussionen, Quiz/Test, Kurzvorträge und regelmäßige Pausen einzubringen (Broadbent 1970). Zusätzlich soll das Mitschreiben einen positiven Effekt auf das Behalten der Informationen haben, insbesondere wenn die Notizen während oder unmittelbar nach der Vorlesung noch einmal gelesen werden (McQueen 1994). Hierbei spielen Handouts eine wichtige Rolle. Diese können den Lernenden helfen, die Vorlesungsinhalte leichter zu verfolgen und zu wiederholen. Dabei haben sich unvollständige Handouts, die vor der Vorlesung verteilt und von den Lernenden eigenständig vervollständigt wurden für die Wissensretention als vorteilig bewährt (Prabhu et al. 2015).

Vorlesungen sind nach wie vor gut geeignet einer großen Anzahl an Studierenden die Einführung einer Thematik zu erleichtern und zeichnen sich durch ein günstiges Kosten-Nutzen-Verhältnis hinsichtlich der Schonung personeller Ressourcen aus.

1.2.2 Seminar

Laut der ZApprO sind Seminare wie folgt definiert: „In den Seminaren wird der durch praktische Übungen und Vorlesungen vermittelte Lehrstoff vertiefend, anwendungs- und gegenstandsbezogen erörtert. Die Seminare sind darauf gerichtet, den Lernenden wichtige medizinische Zusammenhänge zu vermitteln [...] die Teilnehmerzahl darf 20 nicht überschreiten“ (§ 8). Im Vergleich zu den Vorlesungen sind in den Seminaren kleinere Gruppen von Lernenden anwesend, die interaktiv am Unterricht teilnehmen sollen.

Der Aktivierungsgrad der Lernenden macht schlussendlich den Unterschied zu dem Vorlesungsformat und nicht die Teilnehmerzahl. Vorteil der Seminare ist, dass auf individuelle Bedürfnisse wie Rückfragen oder das Lerntempo Rücksicht genommen werden kann.

1.2.3 Problemorientiertes Lernen / *Problem Based Learning (POL/PBL)*

Dieses Prinzip entspricht inhaltlich den curricularen Vorgaben. Dabei werden die Inhalte als Themenblöcke in einem Zeitraum von mehreren Wochen bearbeitet. Die Themenblöcke werden in Kleingruppen behandelt, die von einem Tutor geleitet werden können.

Es handelt sich um ein fallbasiertes Lernen, dessen Ablauf einem klaren Schema folgt, wie z. B. dem Maastrichter 7-Schritt-Modell. Es dient dazu der POL-Gruppe eine Strukturhilfe zu geben, um das Arbeiten in einer Gruppe möglichst effektiv zu gestalten.

Tabelle 2: Maastrichter 7-Schritt-Modell. Das Maastrichter 7-Schritt-Modell zeigt einen schematischen Ablauf für das Strukturieren eines problemorientierten Lernens.

1. Unbekannte Begriffe klären	Fallvorstellung; Klärung von Verständnisfragen
2. Problemdefinition	Klärung der Thematik
3. Brainstorming	Sammlung von Hypothesen und Ideen
4. Strukturieren	Diskussion, Prüfung und Ordnung der Hypothesen
5. Lernzielformulierung	Formulierung spezieller Fragen
6. Informationssuche, Recherche	Individuelles Studium
7. Synthese, Präsentation	Erarbeitetes zusammentragen

Im Vergleich zum klassischen Frontalunterricht ist POL sehr arbeits-, personal- und zeitintensiv. Diese Tatsache stellt ein wichtiges Argument hinsichtlich der engen Ressourcen dar. Studien haben gezeigt, dass das problemorientierte, eigenständig erarbeitete Wissen deutlich länger im Gedächtnis bleibt. Allerdings wird POL von den Studierenden auch kritisch betrachtet, da die klare Strukturierung von Wissen seitens der Dozierenden fehlt (Albanese und Mitchell 1993).

1.2.4 *Electronic-Learning*

Im Rahmen der Digitalisierung seit den 90er Jahren ist auch die Nachfrage nach digitalen Medien an den Hochschulen enorm gestiegen. Informationen zu jeder Zeit von jedem Gerät aus zu erhalten und einzusehen wird heute als selbstverständlich angesehen. Der Begriff *Electronic Learning (E-Learning)* ist breit gefächert. Damit können ganz einfach hochgeladene Vorlesungsfolien, Vorlesungsaufzeichnungen oder Videos gemeint sein. Weitere kombinierte Formen sind das *Flipped-Classroom* Modell, bei dem sich Studierende ein neues Thema anhand von Lernvideos zu Hause aneignen und dieses anschließend in der Vorlesung vertiefen. Ein weiteres *E-Learning* Format ist das *Blended Learning* in dem computerunterstütztes Lernen mit dem klassischen Unterricht kombiniert werden. Vorteile bei dem Modell, durch die Kombination aus Präsenz- und Onlineveranstaltung bestehen darin, dass die Studierenden zum einen den direkten Austausch zum Lehrenden haben und sich zum anderen zu Hause ihr eigenes Lehrtempo einrichten können.

1.2.5 **Praktische Übungen**

Nach der Approbationsordnung umfassen praktische Übungen, Praktika, Unterricht und Behandlung am Patienten. „In den praktischen Übungen bearbeiten die Studierenden eigenständig praktische Aufgaben unter Anleitung, Aufsicht und Verantwortung der ausbildenden Lehrkraft [...]. Sofern es der Lehrstoff erfordert, ist in kleinen Gruppen zu unterrichten“ (§ 7, Abs. 2). Für den Unterricht am Patienten ist nach novellierter ZApprO ein Betreuungsschlüssel von 1:6 und für die Behandlung am Patienten von 1:3 vorgesehen.

Dieser Betreuungsschlüssel wurde dem der humanmedizinischen Ausbildung angepasst, der bereits 2002 definiert wurde. Das gewährleistet am Krankenbett bzw. am Zahnarztstuhl eine hochwertige Qualität und Sicherheit der Behandlung.

1.3 Die Lehrmethoden

Im Folgenden werden die für das Pilotprojekt relevanten Lehrmethoden erläutert.

Diese sind die Standardinstruktion *see one, do one* nach William Stewart Halsted aus dem Jahre 1890, sowie die 4-Schritt-Methode nach J.W. Rodney Peyton (Peyton 1988).

1.3.1 *See one, do one*

Die Lehrmethode *see one, do one* ist dem amerikanischen Chirurgen Dr. William Stewart Halsted (*1852 - †1922) zuzuschreiben. Im späten 19. Jahrhundert gab es in den USA keine effektive Lehrmethode, die dafür geeignet war junge Chirurgen praktisch auszubilden. Der Begriff impliziert, dass eine Tätigkeit lediglich beobachtet und anschließend ausgeführt wird. Es gibt in der Literatur eine Vielzahl an Publikationen, die diese Standardinstruktion auf ihre didaktische Wertigkeit untersucht haben. Für das Erlernen von praktischen Fertigkeiten sei das Konzept überholt und veraltet, zusätzlich fehle nach Rodriguez-Paz und Kollegen (2009) die Betreuung, Überprüfung der Ausführung und das Feedback durch den Lehrenden.

Es gibt aber auch Befürworter, die besagen, dass es im Arbeitsalltag nicht genügend Zeit gibt alle erforderlichen Kompetenzen zu lehren. Nach Kotsis und Chung (2013) ist die Standardinstruktion als Basisinstruktion immer noch anwendbar, sie müsse sich allerdings aufgrund der derzeitigen Entwicklung in der medizinischen Lehre weiterentwickeln, um die Sicherheit der Patienten mithilfe einer evidenzbasierten Behandlung zu gewährleisten. In der Literatur wird das Erlernen praktischer Fertigkeiten durch die Standardinstruktion *see one, do one* kontrovers diskutiert. In der Geschichte der medizinischen Ausbildung galt sie als Meilenstein, ist aber inzwischen durch weitere pädagogische Grundsätze ergänzt worden.

1.3.2 4-Schritt-Methode nach Peyton

Durch die wachsenden Anforderungen an die ärztlich/zahnärztliche Ausbildung im Bereich klinisch-praktischer Fertigkeiten wurden in den Curricula medizinischer Fakultäten weltweit *Skills-Lab*-Trainingseinheiten etabliert. Hierbei haben die Studierenden die Möglichkeit praktische Fertigkeiten an Phantomen oder mit standardisierten Patienten zu erlernen. Eine führende Methode bei der Instruktion praktischer Fertigkeiten ist die Lehrmethode nach Peyton (*Peyton's Four-Step-Approach*), welche aus vier Schritten besteht:

1. Demonstration: Der Lehrende führt die Tätigkeit in gewohnter Schnelligkeit durch.
2. Dekonstruktion: Der Lehrende verbalisiert jeden einzelnen Schritt und führt die Tätigkeit langsam durch.
3. Verständnis: Der Lehrende wird nun die Tätigkeit nach Anleitung des Lernenden durchführen. Dabei ist auf eine reine verbale Instruktion zu achten.
4. Durchführung: Der Lernende führt die Tätigkeit eigenständig unter Verbalisierung der Schritte aus.

Die Unterweisung in praktische Fertigkeiten sollte in kleinen Gruppen von vier bis sechs Personen stattfinden (Peyton 1988, S.139). Das Lernen in Kleingruppen wird in der Literatur unabhängig von der Lehrmethode als äußerst effektiv beschrieben (Kalaian und Kasim 2017). Krautter et al. (2015) konnten zeigen, dass offensichtlich der dritte Schritt entscheidend für den Erfolg der Methode ist. Die Aktivierung entsprechender Hirnareale durch das Vorstellen sowie das wiederholte Abrufen der Abläufe soll zu einer tieferen Prozessierung und damit Verminderung der kognitiven Belastung führen. Zusätzlich beschreiben Krautter et al. (2011) einen positiven Effekt auf die Retentionsfähigkeit durch die Lehrmethode: „Die genaue Verbalisierung der Einzelschritte und die Bewegungsvorstellung führen zu einer besseren objektiven Performance sowie zu einer besseren Verinnerlichung der Einzelschritte und Speicherung im Langzeitgedächtnis“.

In der Literatur wurde die Wirksamkeit der 4-Schritt-Methode mit unterschiedlichen Fertigkeiten ausführlich überprüft. Vom Legen einer Magensonde bis hin zur laparoskopischen Technik konnte diese Lehrmethode überzeugen. Für das Erlernen einfacher Abläufe wie der Herzdruckmassage konnte im Vergleich zu der Standardinstruktion nach einem Beobachtungszeitraum von fünf bis sechs Monaten kein Vorteil gefunden werden (Münster et al. 2016). Hierbei betont Münster, dass die Herzdruckmassage als Fertigkeit nicht komplex genug sei, um von der Lehrmethode nach Peyton zu profitieren. In einer aktuellen Studie wurde die Effizienz der Standardinstruktion *see one, do one* nach Halsted und Peytons 4-Schritt-Methode videobasiert verglichen (Seifert et al. 2020). Auch nach dieser Studie ist die Lehrmethode nach Peyton in Hinblick auf komplexe praktische Fertigkeiten zu bevorzugen.

1.3.3 Peerteaching/PAL

Der Begriff *Peer* bedeutet so viel wie gleichrangig oder gleichgestellt und wird im Bildungsbereich seit dem frühen 19. Jahrhundert eingesetzt. Unter Peerteaching oder auch PAL (*Peer Assisted Learning*) versteht man den Unterricht durch gleichaltrige oder gleichgestellte Studierende. Nach Miller und MacGilchrist (1996) wird das Peerteaching aufgrund der potentiell qualitativen Verbesserung des Lernens eingesetzt: „It suggests that there are positive psychological benefits for both the peer tutor and the tutee“. Ein Peerteaching wird demnach von einem Tutor übernommen, der curriculare Inhalte linear weitervermittelt. Tutorien sind etablierter Bestandteil des universitären Bildungssystems, da sie neben den bereits genannten Vorteilen auch ein günstiges Kosten-Nutzen-Verhältnis aufweisen (Topping 2007). Peerteaching ist kein Ersatz der klassischen Lehre durch Dozenten, es ist eher als ein effektiver Zusatz zu verstehen, welches das Repertoire von Lehren und Lernen erweitert und somit die Qualität der Lehre steigert (Boud et al. 2001).

1.3.4 Die Lehrmethoden an der UMG am Beispiel des TPK

Praktische Fertigkeiten werden anhand von Livedemonstrationen durch den Kursleiter gelehrt. Die studentischen Laborräume verfügen über Bildschirme, über die die Übungen in Echtzeit zu sehen sind. Die Livedemonstrationen finden in Anwesenheit des gesamten Semesters während der Kurszeit statt. Die Anzahl der Studierenden des ersten Fachsemesters an der UMG variiert zwischen ca. 50-55. Die Kurszeit beträgt täglich drei Stunden (wöchentlich 15 Stunden). Die Lernziele des Kurses sind klar formuliert, wobei der Zeitplan einen großen Stellenwert einnimmt. Die Studierenden müssen ihre praktischen Arbeiten anhand eines strukturierten Zeitplans abgeben. Einer verspäteten Abgabe folgt eine ungenügende Benotung. Die Lehrmethoden, die angewandt werden, sind ein Zusammenspiel aus Vorlesungen und der Standardinstruktion *see one, do one*.

Für das Vermitteln praktischer Fertigkeiten wird die Übung gefilmt und zeitgleich auf die Bildschirme des Studentenlabors übertragen. Die Demonstrationen finden während der effektiven Arbeits- bzw. Übungszeit der Studierenden statt. Somit sind die 15 Stunden pro Woche nicht als reine Übungs- bzw. Arbeitszeit anzusehen.

1.4 OSCE/OSPE in der Literatur

Eine OSCE (*Objective Structured Clinical Examination*) ist eine Prüfungsform, die praktische Fertigkeiten und kommunikative Fähigkeiten prüft. Die Notwendigkeit, dass praktische Fertigkeiten auch praktisch geprüft werden müssen, wurde in Dundee (Schottland) bereits in den 70er Jahren wahrgenommen und gelöst (Harden und Gleeson 1979). Hier wurden die ersten Prüfungen dieser Art entwickelt. Dabei handelt es sich um einen Parcours, bestehend aus Stationen, die in einer definierten Zeit durchlaufen werden. Die Aufgabenstellung ist vor Antritt jeder Station präzise formuliert. Pro Station stehen in der Regel fünf bis sechs Minuten zur Verfügung. Nach der entsprechenden Zeit ertönt ein Signal und die Prüflinge wechseln die Station, wofür erneut eine definierte Zeit zur Verfügung steht. An jeder Station befinden sich Rater (Prüfer), die anhand von zuvor erstellten Checklisten die Leistung der Prüflinge beurteilen. Die Prüfungsinhalte basieren in der Regel auf curricularen Vorgaben. Die Aufgabenstellungen umfassen Stationen zur Anamnese, körperlichen Untersuchung, Aufklärungsgespräche, manuelle Tätigkeiten oder auch Notfallsituationen (Ruessler et al. 2010). Das Prüfungsformat OSCE ist in Deutschland weit verbreitet und wurde an medizinischen Fakultäten etabliert, allerdings mit unterschiedlichen Inhalten und Zielen (Müller et al. 2016). Die Planung und Durchführung von OSCE-Prüfungen sind mit einem hohen zeitlichen Aufwand und personellen Kosten verbunden. Nach Schrauth und Kollegen (2005) betragen die Kosten pro Prüfling ca. 160 € bzw. 320 € pro Station.

OSCE-Prüfungen werden in zwei unterschiedliche Prüfungsformen unterteilt und zwar in eine summative und formative Prüfung. Die summative Prüfung findet am Ende eines Semesters oder Kurses statt und hat den Zweck zu entscheiden, ob die Studierenden die erlernten klinischen Kompetenzen ausreichend beherrschen, um das darauffolgende Semester anzutreten. Eine formative Prüfung hat hingegen das Ziel den Leistungsstand der Studierenden in Form eines Feedbacks zu bewerten. Eine formative OSCE ist primär eine Lehr-OSCE und keine Prüfungs-OSCE (Chenot und Ehrhardt 2003). Raupach und Kollegen (2013) konnten zeigen, dass summative Prüfungsformen, bei denen Studierende durchfallen können, den entscheidenden Anreiz für das Lernverhalten darstellen und zu größerem Lernerfolg als formative Prüfungen führen.

Das Pendant zu einer OSCE, die praktische Fertigkeiten prüft, wird OSPE (*Objective Structured Practical Examination*) genannt und hauptsächlich für naturwissenschaftliche Prüfungen eingesetzt. Damit eine OSCE/OSPE-Prüfung eine valide Aussagekraft besitzt, wird in der Literatur die Anzahl der Prüfungsstationen als wichtigstes Kriterium dargestellt. So sollte nach Nikendei und Jünger (2006) eine OSCE-Prüfung aus 10-14 Stationen bestehen.

Die universitären Prüfungen im Studiengang Human- und Zahnmedizin konzentrieren sich vorwiegend auf mündliche und schriftliche Prüfungsformate. Diese Prüfungsformate beurteilen reines Faktenwissen und lassen keine Beurteilung der klinischen Kompetenzen zu. Das Erlangen klinischer Kompetenzen hat George E. Miller im Jahre 1990 entwickelt und ist in Abbildung 1 dargestellt. Die Pyramide nach Miller stellt die vier Ebenen der Kompetenzen dar. Die erste Ebene (*knows*) stellt reines Faktenwissen dar und ist für das Erlangen einer praktischen Kompetenz zwar essentiell, aber allein nicht ausreichend: „there is nothing more useless than a merely well informed man“ (Miller 1990, S. 63). Die nächst höhere Stufe beschreibt, dass derjenige in der Lage ist sein Wissen anzuwenden (*knows how*), dieses Wissen praktisch demonstrieren (*shows how*) und letztlich umsetzen kann (*does*). Die Miller-Pyramide der klinischen Kompetenzen beschreibt in ihrer Gesamtheit die Anforderungen, die an klinisch-praktischen Prüfungen gestellt werden sollten.

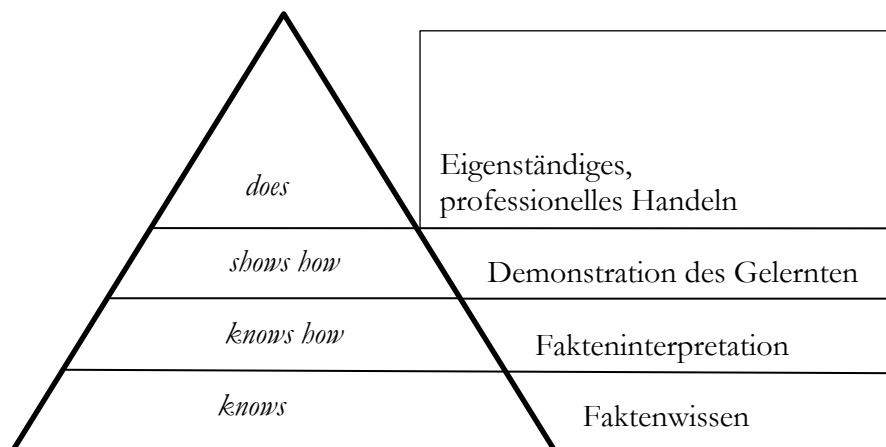


Abbildung 1: Miller-Pyramide der klinischen Kompetenzen basierend auf (Miller 1990).

1.4.1 Die OSCE an der Universitätsmedizin Göttingen

OSCE Prüfungen stellen seit Jahren in der Humanmedizin an der Universitätsmedizin Göttingen (UMG) ein etabliertes Prüfungsverfahren dar. Im Studentischen Trainingszentrum Ärztlicher Praxis und Simulation (STÄPS) werden diese mit hoher Akzeptanz der Studierenden durchgeführt (Simmenroth-Nayda et al. 2014). In der Zahn-Mund-Kieferklinik (ZMK) der UMG sind seit dem Sommersemester 2014 OSCE-Prüfungen ein fester Bestandteil der Lehre (Med. Diss. Göttingen 2017).

Hierbei werden zahnärztlich-chirurgische Inhalte aus den klinischen Semestern überprüft. Diese OSCE beschränkt sich demnach auf Studierende, die sich im klinischen Teil des

Studiums befinden. Eine OSCE im vorklinischen Teil des Studiums findet derzeit an der UMG keine Anwendung, wird jedoch in der Literatur empfohlen (Ratzmann et al. 2012). Das Studium der Zahnmedizin hat bereits einen ersten praktischen Schwerpunkt im vorklinischen Teil. Diese praktischen Fertigkeiten in der Vorklinik beziehen sich auf zahntechnische Arbeiten, die sich im TPK und den Phantomkursen des Zahnersatzes I und II wiederfinden.

1.5 Testgütekriterien

Anhand der Testgütekriterien bestehend aus Validität, Reliabilität und Objektivität können die Inhalte einer OSCE/OSPE-Prüfung gemessen werden. Im Folgenden werden die Kriterien näher erläutert und im Hinblick auf die Testgüte der OSPE des Pilotprojekts im Diskussionsteil (4.2.1) nochmals aufgegriffen.

1.5.1 Validität, Reliabilität, Objektivität

Die Validität oder auch Gültigkeit genannt beschreibt eine Messmethode, die das Konstrukt misst, das gemessen werden soll (Moosbrugger u. Kelava 2012). Sie teilt sich in drei Unterkategorien auf, der Inhalts-, Kriteriums- und Konstruktvalidität.

Die Inhaltsvalidität gibt Aufschluss darüber, ob das Konstrukt/der Test die curricularen Inhalte auch tatsächlich prüft (Schelten 1980; Chenot und Ehrhardt 2003). Sie ist nicht numerisch bestimmbar und wird auch als Augenscheinvalidität bezeichnet, da die subjektive Wahrnehmung miteinfließt. Im Gegensatz dazu wird die Kriteriumsvalidität mithilfe eines Koeffizienten empirisch ermittelt. Dabei wird ein bereits valides Außenkriterium mit dem zu untersuchten Kriterium bewertet. Das Außenkriterium sollte inhaltlich auf das untersuchte Kriterium angepasst sein. Aus diesem Zusammenhang lässt sich somit der Validitäts-Koeffizient ermitteln. Die Konstruktvalidität gibt an, wie genau das zu messende Konstrukt mit anderen ähnlichen Konstrukten korreliert. Eine hohe Korrelation besteht bei zwei Konstrukten, die das Ähnliche messen und demnach eine niedrige Korrelation, wenn sie unterschiedliche Konstrukte messen. Hinsichtlich der Validitätsformen wird der Konstruktvalidität die größte Bedeutsamkeit zugeschrieben, da sie aussagekräftiger als die Inhaltsvalidität und öfter anwendbar als die Kriteriumsvalidität ist (Hodges et al. 1998).

Die Reliabilität ist ein Maß für die Genauigkeit einer Messung, unabhängig davon, ob der Test das misst, was er messen soll (Schelten 1980). Eine hohe Reliabilität ist eine Voraussetzung für eine hohe Validität. Der Zusammenhang zwischen der Reliabilität und Validität eines Testergebnisses ist allerdings nicht gleichzusetzen. So muss ein reliables Testergebnis nicht zwingend heißen, dass das Testergebnis auch valide ist. Hingegen ist ein Testergebnis, das nicht reliabel ist demnach auch nicht valide (Möltner et al. 2006).

Für die Steigerung der Reliabilität eines Tests können die Anzahl der Stationen, eine homogene Aufgabenstellung und die Trennschärfe gesteuert werden. Nach Nikendei und Jünger (2006) sollte eine OSCE-Prüfung mit einer hohen Reliabilität aus 10-14 Stationen bestehen. Durch das zusätzliche Generieren von Daten kann der Einfluss eines zufälligen Ergebnisses verringert werden. Die Items einer Station sollten inhaltlich aufeinander abgestimmt sein, dadurch kann eine höhere Reliabilität erreicht werden. Zuletzt ist die Trennschärfe von großer Bedeutung. Sie zeigt an, wie zuverlässig ein Item den Inhalt misst, den auch die anderen Testaufgaben messen (Schelten 1980). Die genaue statistische Bedeutung der Trennschärfe wird in Abschnitt 2.2.1 erläutert.

Die Objektivität als drittes Testgütekriterium ist durch eine objektive Haltung des Testleiters gekennzeichnet. Somit soll der Test unabhängig von subjektiven Eindrücken fair und gerecht ablaufen. Die Objektivität wird in drei Kategorien, der Durchführungs-, Auswertungs-, und Interpretationsobjektivität unterteilt (Schelten 1980).

Die Durchführungsobjektivität beschreibt das Ausmaß der Unabhängigkeit der Prüfer und gewährleistet eine maximale Standardisierung der Testsituation. Dazu gehören unter anderem eine eindeutige und detaillierte Aufgabenstellung und die Festlegung eines zeitlichen Rahmens zur Absolvierung des Tests.

Die Auswertungsobjektivität ist erfüllt, wenn jede beliebige Person den Test auswerten kann. Ausschlaggebend dafür ist, dass das Auswertungsergebnis nicht durch Interpretationsmöglichkeiten beeinflusst werden darf. Spielraum für Interpretationen kann durch eine dichotome Auswahl (richtig/falsch) vermieden werden. Eine differenziertere Bewertung mittels mehrstufigen Skalen erlaubt mehr Interpretationsmöglichkeit, welches die Auswertungsobjektivität mindern kann. Checklisten haben sich in den Prüfungsformaten bewährt, da sie durch die präzise definierten Auswertungskriterien eine Auswertungsobjektivität verwirklichen.

Die Interpretationsobjektivität ist gegeben, „wenn zwei Beurteiler unabhängig voneinander aus dem gleichen Auswertungsergebnis den gleichen Schluss ziehen“ (Schelten 1980). Die Interpretationsobjektivität ist die am schwierigsten zu gewährleistende Art der Objektivität. Hierbei geht es um die Forderung, dass verschiedene Personen auf Basis der Messergebnisse

dieselbe Interpretation vornehmen. Eindeutiger ist die Interpretationsobjektivität bei normierten Testverfahren gegeben, als bei Verfahren, bei denen die Auswertung der Testergebnisse sehr stark von der Deutung und Interpretation des Testleiters abhängig ist (Kubinger 2010). Strukturierte Prüferschulungen tragen zur Einhaltung der Objektivität bei. Nikendei und Jünger (2006) empfehlen eine 30-minütige Einführung der Prüfer vor Beginn der Prüfung. Der Ablauf der Prüfung, das *Procedere* an den Prüfungsstationen und die erwarteten Lösungen der Aufgabenstellungen sind Bestandteile einer Prüferschulung.

1.6 Überblick über das Thema

„Erzähl mir etwas und ich werde es vergessen; zeig mir etwas und ich werde mich daran erinnern; lass mich etwas tun und ich werde es behalten“ (Konfuzius).

Die Erkenntnis des chinesischen Philosophen Konfuzius ist auch für die heutige universitäre Lehre relevant. Praktische Fertigkeiten sollten demnach auch praktisch vermittelt werden. In diesem Pilotprojekt erlernten Studienanfänger praktische Fertigkeiten durch Tutoren. Der Vergleich zwischen einer einfachen Instruktion (*see one, do one*), wie sie derzeit in den vorklinischen Kursen der UMG ihre Anwendung findet, wurde mit einer zeitintensiveren Lehrmethode (Peytons 4-Schritt-Methode), basierend auf einem strukturierten Peerteaching verglichen. Die Leistungen der Studierenden wurden in einer OSPE-Prüfung zu zwei verschiedenen Zeitpunkten bewertet, sodass die Retention der Lehrformate nach einer dreimonatigen Pause beurteilt werden konnte. Die Untersuchung der Lehrformate mithilfe eines Peerteachings und der rein studentisch betreuten OSPE-Prüfung sind in dieser Form neuartig und wurden bislang in dieser Kombination noch nicht untersucht. Die Arbeit untersucht neben der Effizienz und Retention der Lehrformate, die Akzeptanz bzw. Beurteilung der Studierenden bezüglich eines strukturierten Peerteachings.

1.7 Formulierung der Hypothese

Ziel des Pilotprojekts war es zu ermitteln, inwiefern sich die verschiedenen Lehrmethoden auf die Leistung der Studierenden auswirkte. Zusätzlich wurde die Retention der Lehrmethoden nach einer interventionsfreien Phase von drei Monaten untersucht. Mithilfe einer objektiven Prüfung (OSPE) wurden die Leistungen der studentischen Performance erfasst und anschließend ausgewertet.

Es wurde vorgesehen, mithilfe der hier vorgestellten Studie folgende Hypothesen zu beantworten. Die Alternativhypothese H_{A1} , der im Rahmen dieser Arbeit nachgegangen wird, geht davon aus, dass es signifikante Mittelwertunterschiede zwischen den Ergebnissen der Studierenden in Abhängigkeit von den beiden Lehrformaten (Peyton und E-Learning) gibt.

Die Nullhypothese H_1 besagt hingegen, dass es keine signifikanten Mittelwertunterschiede gibt.

Mathematisch formuliert ergeben sich folgende hypothetische Überlegungen: $H_{A1}: P \neq E$;
 $H_1: P=E$

Für die Untersuchung der Retention der Fertigkeiten in Abhängigkeit von den Lehrformaten geht die Alternativhypothese davon aus, dass nach einer erneuten Untersuchung nach drei Monaten (Follow-Up) immer noch ein signifikanter Mittelwertunterschied besteht:

$H_{A2}: P \neq E$

Die Nullhypothese für die Follow-Up-Mittelwertunterschiede besagt, dass es keinen signifikanten Mittelwertunterschied gibt: $H_2: P = E$

2 Material und Methoden

2.1 Materialliste

Tabelle 3: Materialliste

Schilder mit Aufgabenstellung
Drei iPads (iPad Pro) pro Station, Fa. Apple, Cupertino, Kalifornien, USA
Trillerpfeife
QR-Codes für jeden Prüfling, papierbasiert
OK und UK Gipsmodelle (Gips Typ 3, Fa. SHERA Werkstoff-Technologie GmbH & Co. KG, Lemförde, Deutschland)
Gipstrimmer (Modell HSS AZ, Fa. Wassermann Dental-Maschinen GmbH, Hamburg, Deutschland)
Edelstahl-Draht (remanium®, d: 0,7 mm und 1,00 mm federhart, Fa. Dentaforum, Ispringen, Deutschland)
Zangen: Seitenschneider, Flachspitzzange, Adererzange klein und groß, Kramponzange, Hohlkehlzange, Fa. Aesculap AG, Tuttlingen, Deutschland)
Lineal (16 cm)
Fineliner schwarz (Fa. Stabilo International GmbH, Heroldsberg, Deutschland)
Parodontalsonde (XP23/UNC15, Fa. Hu-Friedy, Chicago, USA)
Metall (remanium®star, Fa. Dentaforum, Ispringen, Deutschland)
Fräsen und Gummierer (Prothetik Set der Universitätsmedizin Göttingen, Fa. Hager & Meisinger GmbH, Neuss, Deutschland)
Universal Polierpaste (Fa. Ivoclar Vivadent, Ellwangen, Deutschland)

2.2 Die statistischen Methoden

Mithilfe des t-Tests für unabhängige Stichproben können Gruppenmittelwerte statistisch miteinander verglichen werden. Liegt ein Unterschied zwischen den Mittelwerten vor, der signifikant geworden ist, so kann die Alternativhypothese bestätigt werden. Ist das Ergebnis hingegen rein zufällig, also nicht signifikant, so wird die Nullhypothese bestätigt.

Der t-Test für unabhängige Stichproben bestimmt die Wahrscheinlichkeit der Differenz der Gruppenmittelwerte unter der Annahme der Nullhypothese (Rasch et al. 2014).

Der t-Wert (T) wird mathematisch aus der Differenz zwischen empirischer und theoretischer Mittelwertdifferenz, dividiert durch den geschätzten Standardfehler errechnet.

„Die Anzahl der Freiheitsgraden gibt an, wie viele Werte in einer Berechnungsformel frei variieren dürfen, damit es zu genau einem bestimmten Ergebnis kommt“ (Rasch et al. 2014, S. 54). Die Freiheitsgrade (df) errechnen sich aus der Gesamtheit, die als normalverteilt gelten sollte. Allerdings kann der t-Test auch bei Verletzung der Normalverteilung eingesetzt werden (Pagano 2010, Wilcox 2012). Die Anzahl der Freiheitsgrade ergeben sich aus der Gesamtheit der Stichprobengröße: $n_1 + n_2 - 2$. Für die Hauptuntersuchung (erste OSPE) resultiert aus der Stichprobengröße von $n = 39$ eine Anzahl an Freiheitsgraden von 37 (df = 37). Der Standardfehler gibt die Streuung der Stichprobenwerte vom arithmetischen Mittelwert an.

Mithilfe des p-Werts wird die Signifikanz bestimmt, die von der Stichprobengröße abhängig ist. Je größer die Stichprobe und geringer die Streuung der Daten, desto signifikanter wird das Ergebnis. Der p-Wert, welches einen Wert zwischen 0 und 1 annehmen kann, gibt die Wahrscheinlichkeit an, mit der die Mittelwertdifferenz als Zufall angesehen wird. Für die Aussage eines zufälligen oder signifikanten Ergebnisses ist das Signifikanzniveau ausschlaggebend. Die Grenze des Signifikanzniveaus (Irrtumswahrscheinlichkeit) ist in der Regel bei 5% ($p = 0,05$) anzusetzen. Ist der p-Wert größer als 0,05 ($p > 0,05$) ist der Mittelwertunterschied rein zufällig entstanden. Ist der p-Wert kleiner als 0,05 ($p < 0,05$) so gilt das Ergebnis als signifikant. Für einen t-Test mit einem Signifikanzniveau von 0,05 bedeutet es folglich, dass die Wahrscheinlichkeit der Mittelwertdifferenz bei einem zufälligen Ergebnis unter 5% liegt. Für einen sehr kleinen p-Wert (z.B. 0,001) ist ein Zufall nahezu ausgeschlossen. Allerdings ist dieses Ergebnis nie zu 100% sicher, da jede empirische Mittelwertdifferenz auch unter der Nullhypothese möglich ist (Rasch et al. 2014, S.56). P-Werte im Bereich $0,2 > p > 0,05$ liegen im Indifferenzbereich und können demnach in beide Richtungen interpretiert werden.

Mithilfe des t-Test für abhängige Stichproben wird der Zusammenhang der Leistung innerhalb einer Gruppe zu zwei verschiedenen Zeitpunkten ermittelt (erste und zweite OSPE). Die Stichprobengröße nahm von der Hauptuntersuchung (Interventionsgruppe: $n = 20$, Kontrollgruppe: $n = 19$) zur Follow-Up-Untersuchung (Interventionsgruppe: $n = 15$, Kontrollgruppe: $n = 12$) ab. Ein Test wird als abhängig bezeichnet, wenn wie im Falle des Pilotprojekts der Leistungsverlauf der Prüflinge zwischen erster und zweiter OSPE-Prüfung untersucht werden soll.

Zusätzlich wurde die Effektstärke (Cohen's d) errechnet, welche Rückschlüsse auf die praktische Bedeutsamkeit eines statistisch signifikanten Ergebnisses zulässt (3.1.2). Hierbei kann die Effektstärke der Mittelwertunterschiede zweier Gruppen ermittelt werden (Cohen 1988). Die Effektstärke verdeutlicht die praktische Relevanz der gefundenen Effekte und ist somit von der Stichprobengröße unabhängig. Cohen's d ist für alle reellen Zahlen definiert und wird wie folgt interpretiert: kleiner Effekt $d=0.2$, mittlerer Effekt $d=0.5$, starker Effekt $d=0.8$. Die Daten der deskriptiven Statistik, die Verteilungseigenschaften Schiefe, Kurtosis, Minimum und Maximum wurden für die einzelnen Gruppen und Stationen erhoben (siehe 3.1.1 und 3.2.1).

Als weiteres Analysetool wurde die lineare Regression angewandt (siehe 3.2.4). Diese statistische Methode ermöglicht es durch Zusammenhänge bzw. Merkmale eine mathematische Vorhersage zu treffen. In Hinblick auf das Pilotprojekt sollte überprüft werden, inwieweit die Lehrmethode (Einflussvariable) Einfluss auf die Leistung der Prüflinge (abhängige Variable) zu beiden Zeitpunkten nahm. Als unabhängige Variable, auch Prädiktor genannt, gelten die beiden Lehrformate (Peyton/E-Learning). Als abhängige Variable, auch Kriterium genannt, gelten die Leistungen der Prüflinge zu den beiden Zeitpunkten. Sind abhängige Variablen durch eine Regressionsgleichung miteinander verknüpft, kann die eine Variable zur Vorhersage der anderen eingesetzt werden. Hierzu werden zwischen Prädiktorvariablen, die zur Vorhersage eingesetzt werden und Kriteriumsvariablen, die vorgesagt werden sollen, unterschieden (Bortz und Schuster 2010).

Der Ausdruck „R-Quadrat“ (R^2) oder auch Bestimmtheitsmaß beschreibt den Anteil der aufgeklärten Varianz in der abhängigen Variable (AV; die Leistung der Prüflinge) der durch die unabhängige Variable (UV; das Lehrformat) erklärt werden kann. Dabei kann das R^2 Werte von 0 bis 1 annehmen. Je höher das R^2 , desto besser passt das Regressionsmodell auf die Daten. Da das R^2 von der Anzahl der UVs beeinflusst wird (Der Anteil der aufgeklärten Varianz nimmt zu, unabhängig davon, ob die UVs einen begründeten Bezug zur AV aufweisen), wird es nach unten korrigiert („Korrigiertes R^2 “).

Der Regressionskoeffizient B gibt an, um wie viele Einheiten sich die AV verändert, wenn sich die UV um eine Einheit verändert (Bortz und Schuster 2010). Die Signifikanz von $p < 0,05$ gibt an, dass mit einer sehr hohen Wahrscheinlichkeit ein linearer Zusammenhang zwischen der Lehrmethode und der Leistung besteht. Ein signifikantes Ergebnis sagt aber nur aus, dass ein Effekt besteht, jedoch nichts über die Bedeutsamkeit des Effekts. Um die Bedeutsamkeit eines Ergebnisses zu beurteilen, wird die Effektstärke einer Regression nach Cohen's f^2 berechnet. Das R-Quadrat, das bei Regressionsanalysen ausgegeben wird, kann in eine Effektstärke f^2 nach Cohen (1992) umgerechnet werden. In diesem Fall ist der Wertebereich der Effektstärke zwischen 0 und unendlich. Für das Maß der Effektstärke gilt Cohen's f^2 und wird wie folgt interpretiert: $f^2 = 0,02$ (kleiner Effekt), $f^2 = 0,15$ (mittlerer Effekt), $f^2 = 0,35$ (großer Effekt). Für die Bezeichnung der Effektstärken gibt es in der Literatur keinen einheitlichen Konsens (Ellis 2010).

2.2.1 Die Itemanalyse

Um die Checklisten-Items statistisch auszuwerten wurden sie in Punktwerte codiert. Dies erfolgte nach einer Drei-Punkt-Codierung. Jedes Item konnte mit einem „trifft zu“ (1 Punkt), „trifft teilweise zu“ (0,5 Punkte) oder „trifft nicht zu“ (0 Punkte) bewertet werden. Durch die Itemanalyse konnte Erkenntnis über die Eignung der Items (Schwierigkeit und Trennschärfe) und die interne Konsistenz der Checklisten (Cronbachs Alpha) erlangt werden. Zur statistischen Analyse wurde für jedes Item und für die gesamte Station die Schwierigkeit (M), die Standardabweichung (SD), die Trennschärfe und Cronbachs Alpha ermittelt. Durch „Cronbachs Alpha, wenn Item weggelassen“ wird das Gesamt-Cronbachs Alpha für die Station verändert, wenn das Item weggelassen wird. Dabei kann sich Cronbachs Alpha für die Station entweder verbessern oder verschlechtern. Verbessert sich der Wert, wenn ein oder mehrere Items weggelassen werden, so sollten diese Items näher betrachtet und neu konzipiert werden.

Die Schwierigkeit (M) gibt an wie viele Prüflinge das Item richtig gelöst haben. Eine Schwierigkeit von $M = 1$ bedeutet, dass 100% der Prüflinge die Aufgabe richtig gelöst haben. Die Werte für die Schwierigkeit können demnach in Prozentzahlen interpretiert werden.

Eine Schwierigkeit von 0 bedeutet, dass niemand in der Lage war die Aufgabe richtig zu lösen und es somit eine schwere Aufgabenstellung war. Liegt der Wert bei 1 war sie demnach zu leicht. Ein Wert von 0,5 beschreibt eine mittelschwere Aufgabenstellung (Schelten 1980).

Mithilfe der Itemanalyse kann die Schwierigkeit der Items und somit ihre Aussagekraft analysiert werden. Items, die entweder zu schwierig oder zu leicht sind, tragen nicht zur Differenzierung zwischen leistungsstarken und leistungsschwachen Prüflingen bei. Demnach sollten diese Items eliminiert werden (Schelten 1980). In der Literatur werden empfohlene Richtwerte zwischen 0,4 bis 0,8 angegeben (Möltner et al. 2006).

Als weiterer Parameter der Itemanalyse ist die Standardabweichung (SD) zu erwähnen. Sie gibt die durchschnittliche Entfernung der Messpunkte zum Mittelwert an. Der Trennschärfe eines Items ist zu entnehmen, wie gut das gesamte Testergebnis aufgrund der Beantwortung eines einzelnen Items vorhersagbar ist (Bortz 2005). Sie ist an der Schwierigkeit (M) gebunden und erhält für große und kleine Schwierigkeiten einen niedrigen Wert. Trennschärfen von $> 0,3$ tragen unter anderem zur Steigerung der Reliabilität bei. Negative Trennschärfen bedeuten, dass Prüflinge das Item entweder gut gelöst haben aber insgesamt schlecht waren oder das Item schlecht gelöst haben und insgesamt gut waren. Items mit negativen Trennschärfen können unter anderem ihren Ursprung in einer fehlerhaften Aufgabenstellung haben und sollten im Anschluss an die Itemanalyse überarbeitet werden. Die höchsten Trennschärfen findet man bei Items mit mittlerer Schwierigkeit (Lienert 1989). Zur Steigerung der Reliabilität einer Prüfung tragen zum einen die Trennschärfe, die Anzahl der Items einer Station und der Reliabilitätskoeffizient (Cronbachs Alpha) bei. Cronbachs Alpha kann Werte zwischen 0 und 1 annehmen. In der folgenden Tabelle werden die Werte nach Blanz (2015) dargestellt.

Tabelle 4: Cronbachs Alpha

Cronbachs α	Interpretation
> .9	Exzellent
> .8	Gut / Hoch
> .7	Akzeptabel
> .6	Fragwürdig
> .5	Schlecht / Niedrig
< .5	Inakzeptabel

Wie in der Tabelle (4) zu entnehmen stellt ein Cronbachs-Alpha-Wert von 1 eine absolute Messgenauigkeit dar. Bei einem Wert von über 0,8 kann von einer guten Reliabilität ausgegangen werden. Diese hohe Reliabilität ist erforderlich, um eine Zuverlässigkeit der Reproduzierbarkeit von Prüfungsergebnissen zu erreichen (Möltner et al. 2006) und zwischen leistungsstarken bzw. leistungsschwachen Prüflingen zu unterscheiden (Chenot und Ehrhardt 2003). Neben Cronbachs Alpha für die gesamte Station wird „Cronbachs Alpha, wenn Item weggelassen“ berechnet. Dieser Wert zeigt, inwiefern sich Cronbachs Alpha für die gesamte Station verändert, wenn ein Item weggelassen wird. Das kann dazu führen, dass sich der Wert für Cronbachs Alpha der gesamten Station vermindert oder erhöht. Werte, die auf Cronbachs Alpha vermindert wirken sollten überarbeitet werden (Möltner et al. 2006).

2.2.2 Deskriptive Statistik und Verteilungseigenschaften

Im Rahmen der deskriptiven Statistik und Verteilungseigenschaften werden für die Checklisten der Mittelwert (M), die Standardabweichung (SD), die Schiefe (Sch), die Kurtosis (Ex), sowie das Maximum (Max) und Minimum (Min) erhoben.

Der Mittelwert, auch arithmetisches Mittel genannt, errechnet sich aus der Summe aller Werte und wird durch die Anzahl der Werte dividiert. Er ist das gebräuchlichste Maß der zentralen Tendenz der Verteilung eines metrischen Merkmals (Bortz 2005). Der Mittelwert wird gegenüber Extremwerten (Ausreißern) stark beeinflusst, welches einerseits als Vorteil bezüglich der Sensitivität des Mittels betrachtet wird aber andererseits eine mögliche Ursache eines Fehlers in der Berechnung bedeuten kann. Mithilfe des Mittelwertes werden die einzelnen Items einer Station bezüglich ihrer Verteilungseigenschaften analysiert und mit einer Normalverteilung verglichen. Die Dichtefunktion einer Normalverteilung hat einen glockenförmigen Verlauf und wird auch als unimodal und symmetrisch beschrieben. Die Standardabweichung (SD) ist ein Maß für die Streubreite der Werte um den Mittelwert. Für die Normalverteilung gilt, dass Zweidrittel (68,27%) aller Messwerte innerhalb der Entfernung einer Standardabweichung zum Mittelwert liegen (Bortz 2005). Im Umkreis von zwei Standardabweichungen befinden sich 95% aller Messwerte.

Die im Rahmen dieser Studie angewandte statistische Analyse zur Ermittlung der Mittelwert-Differenz zwischen der studentischen Leistung kann mithilfe des unabhängigen t-Tests unter der Voraussetzung einer Normalverteilung Anwendung finden (Bortz 1985).

Studien zeigen allerdings, dass die Voraussetzung auf Normalverteilung die unwichtigste aller Voraussetzungen der meisten parametrischen Verfahren ist (Salkind 2010).

Zusätzlich finden sich in der Literatur zahlreiche Aussagen darüber, dass der t-Test gegenüber Verletzungen der Normalverteilungsannahme relativ robust ist (Pagano 2010; Wilcox 2012).

Weitere Parameter zur theoretischen Verteilung einer Normalverteilung sind die Schiefe (Sch) und Kurtosis (Ex), auch Exzess bzw. Wölbung genannt. Die Schiefe wird aus der Differenz zwischen dem Mittelwert und dem Modalwert, dividiert durch die Standardabweichung errechnet. Der Modalwert ist der Wert einer Verteilung, der am häufigsten vorkommt (Bortz 2005, S.28). Nimmt die Schiefe einen Wert von 0 an, so ist die Verteilung symmetrisch. Ein positiver Wert bedeutet eine linkssteile oder rechtsschiefe Verteilung, wohingegen ein negativer Wert eine rechtssteile oder linksschiefe Verteilung bedeutet. Die Kurtosis lässt Aussagen über die Wölbung bzw. Steilheit einer Verteilung zu. Bei einem Wert von 0 ist sie moderat gewölbt (mesokurtisch). Nimmt sie positive Werte an, so wird die Wölbung als stark gewölbt/schmalgipflig (leptokurtisch) beschrieben. Bei negativen Werten wird die Wölbung als flach/schmalgipflig (platykurtisch) bezeichnet (Bortz 1985). Liegen Schiefe und Kurtosis nahe dem Wert 0, so gilt die Verteilung als annähernd normalverteilt. Für die Schiefe liegen die Grenzwerte bei $-0,5$ und $+0,5$ und für die Kurtosis bei -1 und $+1$.

2.2.3 Planung der Studie

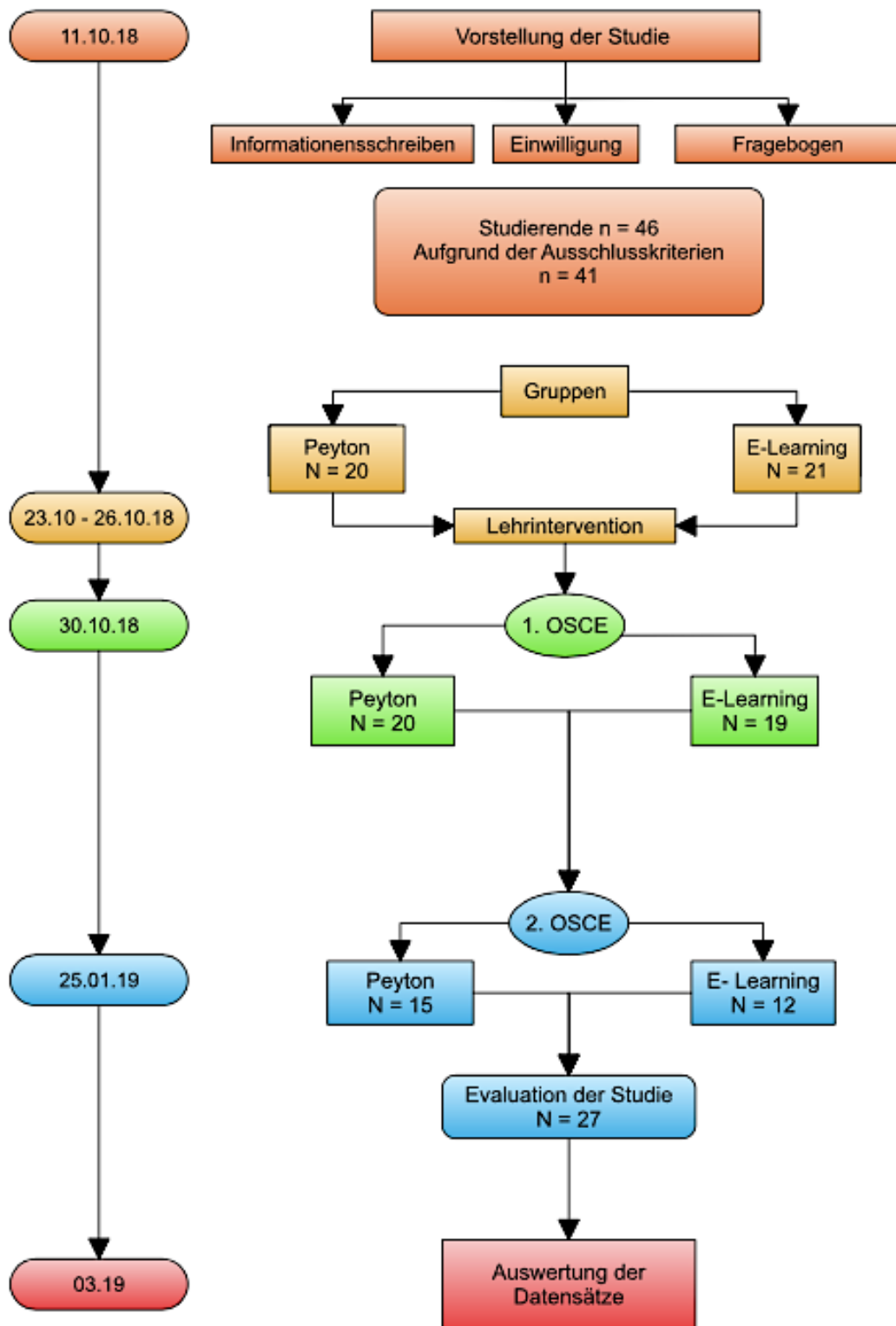


Abbildung 2: Zeitlicher Verlauf

2.2.3.1 Studiendesign

Bei der hier vorliegenden Studie handelt es sich um eine prospektive, monozentrische, randomisierte Interventionsstudie, an der im Wintersemester 2018/2019 Studierende des ersten Fachsemesters Zahnmedizin an der UMG auf freiwilliger Basis teilnehmen konnten. Der Durchführung des im Folgenden beschriebenen Pilotprojekts wurde durch die Ethik-Kommission der Universitätsmedizin Göttingen unter der Antragsnummer 11/8/18 zugestimmt.

2.2.3.2 Vorbereitung und Planung

Die Vorbereitungen zur Umsetzung des Pilotprojekts wurden im Frühjahr 2018 getroffen. Die zahntechnischen Übungen des Pilotprojekts wurden vorab in der Abteilung für Prothetik der UMG ausführlich diskutiert. Eine Überschneidung der Übungen aus dem Pilotprojekt und den prothetischen Kursen konnte vermieden werden, indem die Auswahl auf praktische Übungen aus dem Bereich der Kieferorthopädie fiel. Die Übungen „Dreiecksklammer biegen“ und „Modell eckig trimmen“ sind hinsichtlich ihrer OSPE-Tauglichkeit zuvor im Studentenlabor simuliert und als geeignet erachtet worden. Die Abbildung (2) stellt den zeitlichen Verlauf der Studie dar. Zwischen der ersten und der zweiten OSPE wurde ein Zeitraum von drei Monaten festgelegt.

2.2.3.3 Generierung der Mittel

Für die Umsetzung des Pilotprojekts wurden die Verbrauchsmaterialien (25kg Hartgips und 20m Edelstahl-Draht) von der Studienleitung zur Verfügung gestellt. Die Firma Hager & Meisinger GmbH unterstützte das Pilotprojekt mit zahntechnischen Fräsen und Gummieren ihres Prothetik-Sets der UMG. Alle anderen Mittel stammen aus den eigenen studentischen Laborräumen der Universitätsmedizin Göttingen (siehe Materialliste 2.1).

2.2.4 Gruppenbildung

Die Einführung der Studierenden des ersten Fachsemesters in das Pilotprojekt erfolgte während der Orientierungsphase (Oktober 2018). Mithilfe digitaler Medien, einem Video und einer Power-Point-Präsentation wurde der Inhalt und Zweck der Studie den Studierenden erläutert. Ein Aufklärungsbogen im Hinblick auf die Datenschutz- und Persönlichkeitsrechte und eine Einwilligung zur freiwilligen Teilnahme wurden ausgehändigt und eingeholt (6.1.1).

Anhand des ausgewerteten Eingangsfragebogens erfolgte zunächst der Ausschluss der Studierenden, deren Teilnahme aufgrund der unter 2.2.4.1 angeführten Ausschlusskriterien nicht möglich war. Die übrigen Teilnehmer wurden randomisiert in eine Interventions- und Kontrollgruppe zugeordnet. Die Gruppenzugehörigkeit wurde über Stud.IP kommuniziert. Stud.IP ist eine Online-Plattform der Universität Göttingen, die es den Lehrenden und Lernenden ermöglicht sich auszutauschen und stellt die zentrale Komponente der E-Learning-Infrastruktur dar.

2.2.4.1 Ein- und Ausschlusskriterien

Von der Studie waren Studierende mit einer abgeschlossenen zahntechnischen Berufsausbildung und Wiederholer des ersten Fachsemesters ausgeschlossen. Studierende, die im Ausland ein Studium der Zahnmedizin begonnen, aber keinen praktischen Kurs absolviert hatten, wurden nicht von der Studie ausgeschlossen.

2.2.5 Peerteacher – Auswahl

Die Auswahl der Peerteacher fiel auf zwei Studierende des neunten Fachsemesters, die eine abgeschlossene Berufsausbildung als Zahntechnikerin und zahnmedizinische Fachassistentin nachwiesen. Prinzipiell gilt, dass jeder Studierende eines höheren Fachsemesters als Tutor geeignet ist. In diesem Fall waren die Peerteacher den Studierenden des Pilotprojekts im Studium weit voraus. Da zahntechnische Fertigkeiten im Bereich der Kieferorthopädie Gegenstand der Lehrmethode waren, fiel die Wahl auf Peerteacher, die den Kurs der kieferorthopädischen Technik (achtes Semester) bereits absolviert hatten.

2.2.5.1 Schulung und Kalibrierung der Peerteacher

Für die beiden praktischen Übungen „Modell eckig trimmen“ und „Dreiecksklammer biegen“ nahmen die Peerteacher zunächst an einer internen Schulung bezüglich der Peyton-Lehrmethode (4-Schritt-Methode) teil. Dabei wurde zeitgleich zu dem praktischen Vorgehen jeder Handgriff verbalisiert und die Übung anschließend zu einem Text zusammengefasst. Eine einheitliche Instruktion der Übungen hat den Vorteil, dass alle Studierende auf Basis derselben „Sprache“ eingewiesen werden können. Jede Übung wurde von einem Peer vermittelt und geleitet.

2.2.5.2 Raterauswahl/Raterschulung

Die Auswahl der Rater für die Prüfungen fiel hauptsächlich auf Studierende des neunten Fachsemesters, die bezüglich der Übungen (Dreiecksklammer biegen, Modell eckig trimmen) wie auch die Peerteacher mit einem zahntechnischen Basiswissen ausgestattet waren. Die Peerteacher, die die Lehrintervention durchgeführt haben, konnten aus Gründen der Befangenheit nicht als Rater fungieren, um somit die Objektivität der OSCE-Prüfung nicht zu gefährden. Zwischen den beiden Prüfungen gab es keinen Raterwechsel um eine mögliche Interrater-Inkongruenz zu vermeiden. Die Auswahl der Rater einer OSCE/OSPE beschränkt sich nicht nur auf Ärzte und Zahnärzte aus der Klinik. Studien zeigen, dass Studierende aus höheren Fachsemestern (Peers/Tutoren) sowohl im Bezug der Reliabilität und Anerkennung durch die Prüflinge als Rater geeignet sind (Chenot et al. 2007). Die Rater wurden den Stationen im Vorfeld zugeteilt, um eine gezielte Vorbereitung zu ermöglichen. Ein weiterer Rater mit einem handwerklichen Background konnte für die Station „Dreiecksklammer biegen“ rekrutiert werden. Die Stationen werden im Folgenden noch genauer beschrieben. Die Rater erhielten Einblick in die Lehrvideos, welche die zentralen Punkte der Checkliste widerspiegelten. Eine Übersicht über die Item-Checklisten ist im Anhang zu finden (6.2). Die Handhabung der iPads und der tOSCE App wurde zusammen mit den Ratern trainiert. Jeder Rater wurde für den Fall eines technischen Ausfalls mit einem weiteren iPad ausgestattet. Die Rater erhielten die Anweisung sich während der Prüfung neutral zu verhalten und wurden darauf hingewiesen keine Hilfestellung zu geben und keine Kommentierung der studentischen Ausführung vorzunehmen. Ziel einer Raterschulung ist die Standardisierung der Rater und der Prüfungsbedingungen für die Studierenden (Nikendei und Jünger 2006).

2.2.6 Vorbereitung der Lehre

Das Lehrvideo für die Lehrmethode „E-Learning“/Standardinstruktion (*see one, do one*) wurde von den Peerteachern erstellt. Dabei dienten die Lehrvideos aus den klinischen Kursen der Abteilung für Prothetik als Inspiration und Vorlage. Das Lehrvideo wurde den Ratern vorgespielt und auf Verständlichkeit geprüft. In Bezug auf das Lehrformat nach Peyton (4-Schritt-Methode) erfolgte eine interne Schulung und Unterweisung der beiden Peerteacher in die Lehrmethode durch die Studienleitung mit einer entsprechenden Qualifizierung als MME.

Zur Übung und Festigung der Instruktionen durch die Peerteacher wurden vor Beginn der ersten OSPE mehrere Testläufe mit Studierenden aus dem dritten und vierten Fachsemester vorgenommen, die sich freiwillig zu Verfügung gestellt hatten.

2.2.7 Durchführung der Lehre

Die Unterweisung der Studierenden der Interventionsgruppe erfolgte durch die Peerteacher, die sich zuvor einer intensiven internen Schulung unterzogen hatten. Die Kontrollgruppe wurde mit Hilfe der erstellten Lehrvideos nach der Standardinstruktion (*see one, do one*) in die zahntechnischen Übungen eingewiesen. Die Kontrollgruppe bestand aus 21 und die Interventionsgruppe aus 20 Studierenden, mit einer weiteren Unterteilung der Interventionsgruppe in vier Untergruppen von jeweils fünf Studierenden. Alle Teilnehmer wurden in derselben Woche in die beiden Übungen im Studentenlabor der UMG von Peerteachern eingewiesen. Der zeitliche Ablauf der Unterweisung ist in Abbildung 3 dargestellt. Aus organisatorischen Gründen war es nicht möglich alle Gruppen an demselben Tag einzuweisen. Die Übungszeit betrug für jede Gruppe bzw. Untergruppe exakt drei Stunden. Alle Studierenden, die ihre freiwillige Teilnahme bestätigt hatten, nahmen an der Lehrintervention teil. Die Studierenden der Kontrollgruppe schauten sich das Lehrvideo in einem Seminarraum des SINUZ´ (Studentisches Innovations- und Trainingszentrum Zahnmedizin der Universitätsmedizin Göttingen) an. Das hatte den Vorteil, dass sie sich ausschließlich auf das Video konzentrieren und Notizen erstellen konnten. Zu Beginn des Lehrvideos stand die Schutzkleidung und die Sicherheitsbelehrung im Vordergrund. Die praktischen Arbeiten wurden zwei Mal gefilmt, zunächst Schritt für Schritt aus verschiedenen Perspektiven und in Echtzeit. Anschließend wurden Fragen anhand der Videopassagen beantwortet. Im Anschluss an die Lehrintervention verblieb allen Teilnehmern ein Zeitraum von ca. zwei Stunden im Labor zu üben und ihre Skills zu festigen.

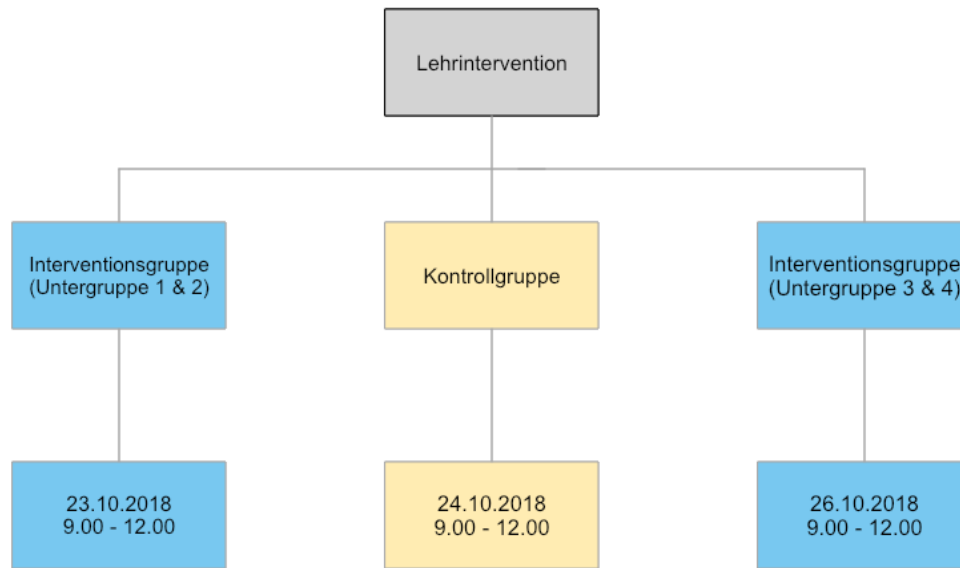


Abbildung 3: Die Lehrintervention. Dargestellt ist der zeitliche Verlauf der Lehreinheiten, beginnend mit den Untergruppen eins und zwei der Interventionsgruppe (Untergruppe eins, $n = 5$; Untergruppe zwei, $n = 5$). Am Folgetag erfolgte die Lehrintervention der gesamten Kontrollgruppe ($n = 21$). Am letzten Tag fand die Lehrintervention der Untergruppen drei und vier der Interventionsgruppe statt (Untergruppe drei, $n = 5$; Untergruppe vier, $n = 5$).

2.2.8 Vorbereitung und Durchführung der ersten OSPE

Die Vorbereitungen zur Durchführung der ersten OSPE fanden bereits am Vortag der Prüfung statt. Unter anderem wurden die QR-Codes zur Zuordnung der Prüflinge und die iPads mit der angelegten Prüfung aus der IT der UMG entgegengenommen. Der OSPE-Parcours konnte hingegen erst am Prüfungstag aufgebaut werden. Den Prüflingen wurde vorab mitgeteilt ihre gewohnte Laborkleidung und Schutzausrüstung, demnach einen weißen Kittel und eine Schutzbrille zur Prüfung mitzubringen.

Die erste OSPE fand am 30.10.2018 von 8.15-12.45 Uhr in den studentischen Laborräumen der UMG statt. Von den 41 Studierenden, die bei den Lehreinheiten anwesend waren, erschienen insgesamt 39 Prüflinge. Zwei Teilnehmer aus der Kontrollgruppe hatten sich zuvor abgemeldet. Die Kontrollgruppe bestand aus 19, die Interventionsgruppe aus 20 Teilnehmern. Jeder Prüfling erhielt vor Antritt der Prüfung seinen eigenen QR-Code papierbasiert. Die OSPE-Prüfung war mit drei Stationen besetzt. Die Teilnehmer wurden in Kleingruppen von jeweils drei Studierenden einbestellt, sodass an jeder Station zeitgleich ein Prüfling die

Aufgabenstellung absolvieren konnte. Vor der jeweiligen Station hatten die Prüflinge eine Minute Zeit die Aufgabenstellung zu lesen. Nach dieser Zeit ertönte ein Pfiff als Signal, welches eine Prüfungszeit von fünf Minuten definierte. Der nächste Pfiff signalisierte einen Stationswechsel und die anschließende Lesezeit von einer Minute für die nächste Station. Die Prüfung war nach dem Durchlauf der drei Stationen beendet. Eine komplette Runde dauerte folglich 18 Minuten. In einem zeitlichen Takt von 20 Minuten wurden die Prüflinge in Dreiergruppen einbestellt. Jeder Rater hatte an seiner Station ein iPad, mit dem er die QR-Codes der Prüflinge einscannte. So konnte jeder Prüfling sicher registriert und zugeordnet werden.

2.2.8.1 Station 1 – Modell eckig trimmen

Diese Station fand in der Gipsküche des Studentenlabors der UMG statt. Die Prüflinge erhielten entweder ein Oberkiefer- oder Unterkiefermodell. Die unterschiedliche Trimmung der Kiefer wurde vorab in den Lehreinheiten gelehrt und geübt. Die Prüflinge sollten unter anderem die Handhabung des Geräts, die akkurate Umsetzung der Übung und ein sicheres Arbeiten demonstrieren. Die richtige Trimm-Geometrie des Modells unter Berücksichtigung der anatomischen Strukturen wurde zusätzlich bewertet.

2.2.8.2 Station 2 – Dreiecksklammer biegen

Diese Station fand an den zahntechnischen Arbeitsplätzen des studentischen Labors statt. Hierzu gab es die genaue Anweisung über die Lage der Klammer am Modell. Die Prüflinge sollten eine Dreiecksklammer im 3. Quadranten (Unterkiefer links) zwischen 34 und 35 (Prämolaren unten links) biegen. Es wurden unter anderem die Einzelschritte der Biegungen, die fertige Dreiecksklammer sowie die Handhabung der Zangen und das sichere Arbeiten bewertet.

2.2.8.3 Station 3 – Politur

Diese Station war nicht Bestandteil der Lehrintervention. Die Prüflinge wurden vorab darüber informiert, dass sie während der OSPE zusätzlich zu den beiden bekannten Stationen eine weitere unbekannt Station durchlaufen. Der Inhalt der Station wurde nicht kommuniziert, damit kein Studierender die Möglichkeit hatte zu üben oder sich vorab zu informieren. Die Station diente dem Vergleich der Gruppen untereinander und sollte Aufschluss darüber geben ob eine Gruppe statistisch signifikant besser oder schlechter abschneidet.

2.2.9 Follow-Up nach drei Monaten / zweite OSPE

Die Teilnehmer des Pilotprojekts hatten eingewilligt an einer Follow-Up Untersuchung drei Monate nach der ersten OSPE-Prüfung teilzunehmen (siehe 6.1.2). Die zweite OSPE, die sich hinsichtlich der Rahmenbedingungen nicht von der ersten OSPE unterschied, fand am 25.01.2019 statt. Die Prüfungsstationen waren den Teilnehmern aus der ersten Prüfung bekannt. Lediglich die dritte Station „Politur“ war nicht mehr Bestandteil der Untersuchung. An der zweiten OSPE nahmen in der Interventionsgruppe noch 15 Studierende teil, in der Kontrollgruppe waren es noch 12. Somit reduzierte sich die Gesamtteilnehmerzahl auf 27 Prüflinge. In der Zeit zwischen den beiden OSPE-Prüfungen wurden die Teilnehmer angewiesen sich nicht gezielt auf die Prüfungsinhalte vorzubereiten. Im Anschluss an die Prüfung erfolgte eine Evaluation des Projekts. Der Evaluationsbogen (EvaSys®, Electronic Paper, Lüneburg Germany) ist im Anhang 6.1.3 zu finden.

2.2.10 Das Bewertungsformat

Da es sich bei den Übungen um eine reine Beurteilung von Fertigkeiten handelt, wurde eine Bewertung mit Hilfe einer Item-Checkliste ohne zusätzliche Vergabe einer Globalnote verwendet. Checklisten werden bei technischen Fertigkeiten als adäquates Evaluationsinstrument empfohlen (Newble 2004). Auf eine zusätzliche Vergabe von Globalnoten wurde bewusst verzichtet, da sie Fähigkeiten wie Empathie miteinbezieht (Nikendei und Jünger 2006).

2.2.11 Die Datenauswertung

Die Datenerhebung erfolgte mit Hilfe der tOSCE App des Umbrella Consortium for Assessment Network (UCAN). Die App wurde auf die iPads installiert, welche das Institut für Medizinische Informatik (IT) der UMG für die Datenerhebung zur Verfügung stellte. Die App erfasste die Datensätze aus den beiden OSPE-Prüfungen. Zusätzlich erstellte die IT der UMG QR-Codes individuell für jeden Prüfling. Die Erstellung der Item-Checklisten erfolgte durch die Studienleitung, die anschließend in der IT auf die App überspielt wurden. Die Auswertung der Daten fand in der Abteilung Medizindidaktik und Ausbildungsforschung der UMG statt via IBM SPSS Statistics (Version 26).

Die Datenauswertung beinhaltet die deskriptive Statistik, die Item-Analyse jeder Station und die lineare Regression. Der t-Test für unabhängige Stichproben kam zum Einsatz um die

Gruppenmittelwert-Differenzen für die Hauptuntersuchung (3.1.2) und die Follow-Up-Untersuchung (3.2.2) zu ermitteln. Der t-Test für abhängige Stichproben wurde verwendet, um die Leistungsunterschiede der Teilnehmer innerhalb einer Gruppe zwischen der ersten OSPE und der zweiten OSPE zu untersuchen (3.2.3).

3 Ergebnisse

3.1 Itemanalyse der Hauptuntersuchung

In den folgenden Tabellen (5-8) sind die Mittelwerte (M) mit der Standardabweichung (SD), sowie der Trennschärfe und „Alpha, wenn Item weggelassen“ für jedes einzelne Item dargestellt. Zusätzlich wurden die bereits erwähnten Parameter und Cronbachs Alpha für die gesamte Station ermittelt. Die gelb markierten Items werden im Diskussionsteil näher erläutert. Die Item-Checklisten sind im Anhang 6.2 zu finden.

Tabelle 5: Itemanalyse Station „Dreiecksklammer biegen“. Die in der Tabelle gelb markierten Felder werden im Diskussionsteil aufgrund ihrer Schwierigkeit und der dazugehörigen Trennschärfe erklärt.

Item	M	SD	Trennschärfe	Alpha, wenn Item weggelassen
Item 1	1,00	0,00	0,00	0,76
Item 2	0,74	0,44	0,27	0,76
Item 3	0,88	0,24	0,35	0,75
Item 4	0,64	0,39	0,32	0,75
Item 5	0,92	0,27	0,40	0,74
Item 6	0,96	0,13	0,51	0,74
Item 7	0,88	0,24	0,49	0,74
Item 8	0,51	0,45	0,55	0,72
Item 9	0,73	0,41	0,44	0,73
Item 10	0,56	0,20	0,13	0,76
Item 11	0,79	0,31	0,68	0,71
Item 12	0,76	0,42	0,12	0,77
Item 13	0,41	0,49	0,43	0,74
Item 14	0,25	0,44	0,62	0,71
M	0,71	0,33	0,37	0,74
Cronbachs Alpha	0,76			

Die Item-Mittelwerte liegen zwischen 0,25 (SD 0,44) und 1,00 (SD 0,00). Der Mittelwert der gesamten Station liegt bei 0,71 (SD 0,37) und besagt, dass 71% der Prüflinge die Station richtig gelöst haben. Die Trennschärfen variieren zwischen 0,00 und 0,68. Cronbachs Alpha ist für die gesamte Station mit einem Wert von 0,76 in einem akzeptablen Bereich.

Tabelle 6: Itemanalyse der Station „Modell eckig trimmen“. Die gelb markierte Zeile (Item 6) wurde einer Korrektur unterzogen und wird im Diskussionsteil ausführlich erklärt.

Item	M	SD	Trennschärfe	Alpha, wenn Item weggelassen
Item 1	1,00	0,00	0,00	0,61
Item 2	1,00	0,00	0,00	0,61
Item 3	0,92	0,27	0,13	0,62
Item 4	0,97	0,16	0,13	0,61
Item 5	0,97	0,16	0,56	0,55
Item 6	0,61	0,49	0,003	0,72
Item 7	0,91	0,22	0,56	0,53
Item 8	0,56	0,30	0,48	0,53
Item 9	0,83	0,28	0,29	0,58
Item 10	0,92	0,21	0,58	0,53
Item 11	0,76	0,27	0,58	0,51
M	0,85	0,21	0,30	0,58
Cronbachs Alpha	0,61			

Die Item-Mittelwerte schwanken zwischen 0,56 und 1,00. Der Mittelwert mit 0,85 besagt, dass 85% der Prüflinge diese Station richtig gelöst haben. Cronbachs Alpha liegt bei 0,61 für die gesamte Station (siehe: Tabelle 4). Item 6 hat eine moderate Schwierigkeit (0,61) aber eine geringe Trennschärfe (0,003). Dieses Item wurde einer Korrektur unterzogen und wird im Diskussionsteil näher erläutert. In der folgenden Tabelle (7) ist die Itemanalyse mit Korrektur von Item 6 erfolgt.

Tabelle 7: Itemanalyse der Station „Modell eckig trimmen“ nach Korrektur von Item 6

Item	M	SD	Trennschärfe	Alpha, wenn Item weggelassen
Item 1	1,00	0,00	0,00	0,69
Item 2	1,00	0,00	0,00	0,69
Item 3	0,92	0,27	0,09	0,70
Item 4	0,97	0,16	0,15	0,68
Item 5	0,97	0,16	0,57	0,63
Item 6	0,79	0,40	0,16	0,72
Item 7	0,91	0,22	0,68	0,60
Item 8	0,56	0,30	0,51	0,62
Item 9	0,83	0,28	0,42	0,64
Item 10	0,92	0,21	0,57	0,62
Item 11	0,76	0,27	0,52	0,62
M	0,87	0,20	0,33	0,65
Cronbachs Alpha	0,68			

Durch die Korrektur von Item 6 erhöht sich der Schwierigkeits-Mittelwert der gesamten Station auf 0,87 (vorher 0,85), dadurch sinkt die Schwierigkeit der Station. Auch die Gesamt-Trennschärfe weist einen höheren Wert auf und ist von 0,30 auf 0,33 gestiegen. Die Diskrepanz zwischen Cronbachs Alpha und „Alpha, wenn weggelassen“ ist nach der Korrektur geringer. Das Herausnehmen von Item 6 würde einen Anstieg auf 0,72 für Cronbachs Alpha bedeuten und somit als akzeptabel gelten.

Tabelle 8: Itemanalyse der Station „Politur“. Die in der Tabelle gelb markierten Zeile (Item 5) wird aufgrund der negativen Trennschärfe im Diskussionsteil erläutert.

Item	M	SD	Trennschärfe	Alpha, wenn Item weggelassen
Item 1	1,00	0,00	0,00	0,29
Item 2	0,94	0,22	0,08	0,30
Item 3	0,15	0,36	0,07	0,29
Item 4	0,12	0,33	0,17	0,23
Item 5	0,26	0,16	-0,21	0,35
Item 6	0,10	0,30	0,12	0,26
Item 7	0,07	0,27	0,27	0,19
Item 8	0,64	0,37	0,07	0,29
Item 9	0,17	0,24	0,16	0,24
Item 10	0,23	0,42	0,21	0,19
M	0,45	0,26	0,09	0,26
Cronbachs Alpha	0,29			

Die Trennschärfen der zehn Items liegen unter 0,27. Im Durchschnitt wird ein Wert von 0,09 erreicht und ein negativer Wert von Item 5 ist zu sehen. Die Schwierigkeit der Items liegt zwischen 0,07 und 1,00. Cronbachs Alpha hat einen Wert von 0,29 für die gesamte Station. Item 5 hat bezüglich der Trennschärfe einen negativen Wert (-0,21).

Die Items 3-7 haben bezüglich ihrer Trennschärfe geringe Werte. Diese Items wurden einer Korrektur unterzogen und neu codiert. Die Itemanalyse mit der Korrektur ist in der folgenden Tabelle (9) dargestellt.

Tabelle 9: Itemanalyse der Station „Politur“ nach Korrektur der Items 3-7. Die Items 3-7 wurden nach der Neucodierung zusammengefasst.

Item	M	SD	Trennschärfe	Alpha, wenn Item weggelassen
Item 1	1,00	0,00	0,00	0,43
Item 2	0,94	0,22	0,78	0,47
Item 3-7	0,31	0,22	0,72	0,43
Item 8	0,64	0,37	0,31	0,29
Item 9	0,17	0,24	0,37	0,28
Item 10	0,23	0,42	0,32	0,28
M	0,54	0,24	0,41	0,36
Cronbachs Alpha	0,42			

Das gelb markierte Feld (Item 3-7) sind die Items, die nachträglich korrigiert und neucodiert wurden. Insgesamt hat die Station nach Neucodierung nur noch sechs Items.

Der Schwierigkeits-Mittelwert steigt nach Korrektur von 0,45 auf 0,54. Cronbachs Alpha verbessert sich von 0,29 auf 0,42. Die Trennschärfe für die gesamte Station verbessert sich von 0,09 auf 0,41.

3.1.1 Deskriptive Statistik und Verteilungseigenschaften

In den folgenden Tabellen 10 und 11 sind für die Interventions- und Kontrollgruppe folgende Parameter dargestellt: Stichprobengröße (N), der Mittelwert (M) für die jeweilige Station, Maximum (Max) und Minimum (Min), die Standardabweichung (SD) und die Verteilungseigenschaften Schiefe (Sch) und Kurtosis (Ex). Als annähernd normalverteilt gelten Werte für die Schiefe zwischen -0,5 und +0,5 und Kurtosis zwischen -1 und +1. Als nicht mehr normalverteilt gelten die gelb markierten Werte.

Tabelle 10: Deskriptive Statistik der Interventionsgruppe (Hauptuntersuchung). Die gelb markierten Werte liegen außerhalb der Grenzen einer Normalverteilung.

Interventionsgruppe	N	M	Max	Min	SD	Sch	Ex
„Dreiecksklammer biegen“	20	0,82	0,88	0,75	0,14	-0,70	-0,31
„Modell eckig trimmen“	20	0,85	0,87	0,82	0,05	-1,04	0,46
„Politur“	20	0,55	0,62	0,47	0,15	0,69	-0,93

Tabelle 11: Deskriptive Statistik der Kontrollgruppe (Hauptuntersuchung). Die gelb markierten Werte liegen außerhalb der Grenzen einer Normalverteilung.

Kontrollgruppe	N	M	Max	Min	SD	Sch	Ex
„Dreiecksklammer biegen“	19	0,60	0,66	0,54	0,12	-0,84	0,95
„Modell eckig trimmen“	19	0,76	0,82	0,69	0,13	-1,14	1,52
„Politur“	19	0,55	0,61	0,49	0,13	0,22	-0,42

Die Verteilung der Station „Dreiecksklammer biegen“ der Kontrollgruppe zeigt eine Schiefe von -0,84 und eine Kurtosis von 0,95. Die Schiefe beschreibt hier somit linksschiefe Daten. Die Kurtosis von 0,95 beschreibt eine annähernd mesokurtische Verteilung. Die Verteilung der Station „Modell eckig trimmen“ zeigt eine Schiefe von -1,14 welche eine starke Linksverschiebung darstellt. Die Kurtosis von 1,52 zeigt eine leptokurtische Verteilung. Eine Normalverteilung ist im Hinblick der geringen Stichprobengröße nicht zu realisieren, da sie durch Ausreißer stark beeinflusst wird.

3.1.2 Vergleich der Mittelwerte mit dem t-Test für unabhängige Stichproben

In den folgenden Tabellen (12-14) werden für beide Gruppen jeweils die Stichprobengröße (N), der Mittelwert (M), die Standardabweichung (SD), der T-Wert (T), die Anzahl an Freiheitsgraden (df) und der p-Wert dargestellt. P (2-seitig) stellt die Signifikanz der Mittelwert-Differenz dar. Cohen's d ist ein Maß für die Effektstärke der Mittelwert-Differenz.

Signifikante Ergebnisse $p < 0.05$ sind grün, während nicht signifikante Ergebnisse $p > 0.2$ rot unterlegt sind.

Tabelle 12: Vergleich der Gruppenmittelwerte „Dreiecksklammer biegen“

Interventionsgruppe			Kontrollgruppe			t-Test für unabhängige Stichproben			Cohen's d
N	M	SD	N	M	SD	T	df	p (2-seitig)	
20	0,82	0,14	19	0,60	0,12	5,00	37	0,000	1,61

Für die Station „Dreiecksklammer biegen“ ergibt sich ein statistisch signifikanter Unterschied zwischen der Interventions- und der Kontrollgruppe: $t(37) = 5.00$, $p < 0.001$

Cohen's d mit 1,61 weist einen sehr großen Effekt der untersuchten Parameter (Lehrmethode) auf.

Tabelle 13: Vergleich der Gruppenmittelwerte „Modell eckig trimmen“

Interventionsgruppe			Kontrollgruppe			t-Test für unabhängige Stichproben			Cohen's d
N	M	SD	N	M	SD	T	df	p (2-seitig)	
20	0,85	0,05	19	0,76	0,13	2,23	37	0,031	0,71

Für die Station „Modell eckig trimmen“ ergibt sich ein statistisch signifikanter Unterschied zwischen der Interventions- und der Kontrollgruppe: $t(37) = 2.23$, $p = 0.031$. Cohen's d mit 0,71 weist einen großen Effekt der untersuchten Parameter (Lehrmethode) auf.

Tabelle 14: Vergleich der Gruppenmittelwerte „Politur“

Interventionsgruppe			Kontrollgruppe			t-Test für unabhängige Stichproben			Cohen´s d
N	M	SD	N	M	SD	T	df	p (2-seitig)	
20	0,55	0,15	19	0,55	0,13	-0,05	37	0,954	-0,012

Für die Station „Politur“ ergibt sich statistisch kein signifikanter Unterschied zwischen der Interventions- und der Kontrollgruppe: $t(37) = -0,05$, $p = 0,954$. Cohen´s d ist in einem negativen Bereich mit $-0,012$ und stellt keinen Effekt dar.

3.2 Itemanalyse der Follow-Up-Untersuchung

Tabelle 15: Itemanalyse der Station „Dreiecksklammer biegen“ (Follow-Up)

Item	M	SD	Trennschärfe	Alpha, wenn Item weggelassen
Item 1	1,00	0,00	0,00	0,50
Item 2	1,00	0,00	0,00	0,50
Item 3	0,85	0,30	0,16	0,48
Item 4	0,68	0,41	-0,05	0,55
Item 5	0,88	0,32	-0,02	0,57
Item 6	1,00	0,00	0,00	0,50
Item 7	0,96	0,13	0,26	0,48
Item 8	0,90	0,24	0,16	0,48
Item 9	0,66	0,33	0,56	0,38
Item 10	0,88	0,25	0,06	0,50
Item 11	0,74	0,32	0,24	0,47
Item 12	0,59	0,50	0,20	0,48
Item 13	0,63	0,49	0,52	0,35
Item 14	0,37	0,49	0,47	0,37
M	0,79	0,27	0,18	0,47
Cronbachs Alpha	0,50			

Die Schwierigkeiten der 14 Items schwanken zwischen 0,37 und 1,00. Für die gesamte Station ergibt sich im Durchschnitt eine Schwierigkeit von 0,79. Zwei Trennschärfen sind in

einem negativen Bereich (Item 4 und 5). Der Mittelwert der Trennschärfen liegt bei 0,18. Für Cronbachs Alpha ergibt sich ein Wert von 0,50. Die gelb markierten Felder stellen negative Trennschärfen dar.

Tabelle 16: Itemanalyse der Station „Modell eckig trimmen“ (Follow-Up)

Item	M	SD	Trennschärfe	Alpha, wenn Item weggelassen
Item 1	1,00	0,00	0,00	0,41
Item 2	1,00	0,00	0,00	0,41
Item 3	0,96	0,19	-0,15	0,50
Item 4	0,92	0,26	0,42	0,21
Item 5	1,00	0,00	0,00	0,41
Item 6	0,96	0,19	-0,006	0,44
Item 7	0,88	0,21	0,28	0,32
Item 8	0,79	0,25	0,21	0,35
Item 9	0,92	0,18	0,27	0,33
Item 10	1,00	0,00	0,00	0,41
Item 11	0,98	0,09	0,63	0,28
M	0,94	0,12	0,15	0,37
Cronbachs Alpha	0,40			

Die Item-Schwierigkeiten liegen zwischen 0,79 und 1,00. Der Schwierigkeits-Mittelwert für die gesamte Station hat einen Wert von 0,94. Die Items 3 und 6 haben negative Trennschärfen. Der Mittelwert für die Trennschärfe liegt bei 0,15. Cronbachs Alpha für die gesamte Station hat einen Wert von 0,40. Die gelb markierten Felder stellen negative Trennschärfen dar.

Tabelle 17: Entwicklung der Itemparameter. Dargestellt ist die Entwicklung der Parameter Schwierigkeit (M), Trennschärfe und Cronbachs Alpha der Stationen „Dreiecksklammer biegen“ und „Modell eckig trimmen“ von der ersten zur zweiten OSPE-Prüfung.

	Dreiecksklammer biegen	Modell eckig trimmen
M ₁	0,71	0,87
M ₂	0,79	0,94
Trennschärfe ₁	0,37	0,33
Trennschärfe ₂	0,18	0,15
Cronbachs Alpha ₁	0,76	0,68
Cronbachs Alpha ₂	0,50	0,40

3.2.1 Deskriptive Statistik und Verteilungseigenschaften (Follow-Up)

In den folgenden Tabellen (18 und 19) sind für die Interventions- und Kontrollgruppe folgende Parameter dargestellt: Strichprobengröße (N), der Mittelwert (M), Maximum (Max) und Minimum (Min), die Standardabweichung (SD) und die Verteilungseigenschaften Schiefe und Kurtosis. Als annähernd normalverteilt gelten Werte für die Schiefe zwischen -0,5 und +0,5 und Kurtosis zwischen -1 und +1. Als nicht mehr normalverteilt gelten die gelb markierten Werte.

Tabelle 18: Deskriptive Statistik der Interventionsgruppe (Follow-Up-Untersuchung)

Interventionsgruppe	N	M	Max	Min	SD	Sch	Ex
„Dreiecksklammer biegen“	15	0,84	0,90	0,78	0,10	-0,47	0,27
„Modell eckig trimmen“	15	0,97	0,99	0,95	0,02	-0,80	-0,12

Die Station „Dreiecksklammer biegen“ kann mit einer Schiefe von -0,47 und einer Kurtosis von 0,27 als annähernd normalverteilt bezeichnet werden. Die Station „Modell eckig trimmen“ zeigt eine linksschiefe und mesokurtische Verteilung.

Tabelle 19: Deskriptive Statistik der Kontrollgruppe (Follow-Up-Untersuchung)

Kontrollgruppe	N	M	Max	Min	SD	Sch	Ex
„Dreiecksklammer biegen“	12	0,74	0,81	0,67	0,11	0,20	-0,02
„Modell eckig trimmen“	12	0,91	0,96	0,86	0,07	-1,24	2,43

Wie auch in der Interventionsgruppe zeigt hier die Station „Dreiecksklammer biegen“ eine annähernd normale Verteilung an. Die Station „Modell eckig trimmen“ zeigt eine stark links-schiefe und platykurtische Verteilung.

3.2.2 Vergleich der Mittelwerte mit dem t-Test für unabhängige Stichproben (Follow-Up)

In den folgenden Tabellen (20 und 21) werden für beide Gruppen jeweils die Stichprobengröße (N), den Mittelwert (M), die Standardabweichung (SD), der T-Wert (T), die Anzahl an Freiheitsgraden (df) und der p-Wert dargestellt. P (2-seitig) stellt die Signifikanz der Mittelwert-Differenz dar. Cohen's d ist ein Maß für die Effektstärke der Mittelwert-Differenz. Signifikante Ergebnisse $p < 0.05$ sind grün unterlegt, während nicht signifikante Ergebnisse $p > 0.2$ rot unterlegt sind.

Tabelle 20: Vergleich der Gruppenmittelwerte „Dreiecksklammer biegen“ (Follow-Up)

Interventionsgruppe			Kontrollgruppe			t-Test für unabhängige Stichproben			Cohen's d
N	M	SD	N	M	SD	T	df	p (2-seitig)	
15	0,84	0,10	12	0,74	0,11	2,32	25	0,028	0,935

Für die Station „Dreiecksklammer biegen“ ergibt sich nach drei Monaten ein statistisch signifikanter Unterschied zwischen der Interventions- und der Kontrollgruppe: $t(25) = 2.32$, $p = 0.028$. Cohen's d weist mit 0,935 einen großen Effekt auf.

Tabelle 21: Vergleich der Gruppenmittelwerte „Modell eckig trimmen“ (Follow-Up)

Interventionsgruppe			Kontrollgruppe			t-Test für unabhängige Stichproben			Cohen´s d
N	M	SD	N	M	SD	T	df	p (2-seitig)	
15	0,97	0,02	12	0,91	0,07	2,74	25	0,011	1,098

Für die Station „Modell eckig trimmen“ ergibt sich nach drei Monaten ein statistisch signifikanter Unterschied zwischen der Intervention- und der Kontrollgruppe: $t(25) = 2.74$, $p = 0.011$. Cohen´s d weist mit 1,098 einen großen Effekt auf.

3.2.3 Vergleich der Gruppenmittelwert-Differenz innerhalb einer Gruppe von der ersten zur zweiten OSPE mit dem t-Test für abhängige Stichproben

In den nachfolgenden Tabellen (22 und 23) wurden die Gruppenmittelwerte von der Hauptuntersuchung (M1) und der Follow-Up-Untersuchung (M2) untersucht, sowie die Standardabweichung (SD), der T-Wert (T), die Anzahl an Freiheitsgraden (df) und der p-Wert. Der Zusatz „1“ bezieht sich auf die Hauptuntersuchung und der Zusatz „2“ auf die Follow-Up-Untersuchung. P (2-seitig) stellt die Signifikanz der Mittelwert-Differenz dar. Signifikante Ergebnisse $p < 0.05$ sind grün unterlegt, während nicht signifikante Ergebnisse $p > 0.2$ rot unterlegt sind.

Tabelle 22: t-Test für abhängige Stichproben (Interventionsgruppe)

Station	N1	M1	SD1	N2	M2	SD2	T	df	p (2-seitig)
„Dreiecksklammer biegen“	20	0,82	0,14	15	0,84	0,10	-0,11	14	0,911
„Modell eckig trimmen“	20	0,85	0,05	15	0,97	0,02	-2,28	14	0,038

Für die Interventionsgruppe ergibt sich in der Station „Dreiecksklammer biegen“ eine leichte Verbesserung des Mittelwerts, allerdings ergibt dieses Ergebnis keine statistische Signifikanz: $t(14) = -0.11$, $p = 0.911$.

Die Prüflinge der Interventionsgruppe konnten sich in der Station „Modell eckig trimmen“ statistisch signifikant verbessern: $t(14) = -2.28$, $p = 0.038$

Tabelle 23: t-Test für abhängige Stichproben (Kontrollgruppe)

Station	N1	M1	SD1	N2	M2	SD2	T	df	p (2-seitig)
„Dreiecksklammer biegen“	19	0,60	0,12	12	0,74	0,11	-2,76	11	0,018
„Modell eckig trimmen“	19	0,76	0,13	12	0,91	0,07	-0,85	11	0,409

Bei der Kontrollgruppe verhält es sich genau umgekehrt zur Interventionsgruppe.

Sie konnte sich in der Station „Dreiecksklammer biegen“ signifikant verbessern:

$$t(11) = -2.76, p = 0.018$$

In der Station „Modell eckig trimmen“ ist anhand des Mittelwerts eine leichte Verbesserung zu sehen, die jedoch statistisch nicht signifikant ist: $t(11) = -0.85, p = 0.409$

3.2.4 Die lineare Regression

Im Folgenden werden die Ergebnisse einer einfachen linearen Regression mit dichotomen Prädiktor für die Stationen „Dreiecksklammer biegen“ und „Modell eckig trimmen“ zu zwei Messzeitpunkten dargestellt. Die Bezeichnungen T_1 und T_2 stehen für die Zeitpunkte der ersten bzw. zweiten OSPE-Prüfung. Für das Maß der Effektstärke wird Cohen's f^2 angegeben. Im Anhang 6.3 sind die Tabellen zur linearen Regression dargestellt.

3.2.4.1 Ergebnisse der linearen Regression für die Station „Dreiecksklammer biegen“

Die Lehrmethode hat zum Zeitpunkt T_1 einen signifikanten Einfluss auf die Leistung der Studierenden ($F(1,37) = 25,02, p < 0.001$). Sie erklärt zu 38,7% die Leistung der Studierenden in der Station. Im Durchschnitt erreichen die Prüflinge der Interventionsgruppe 3,024 Punkte mehr in der Stationsleistung, als die Prüflinge der Kontrollgruppe, was nach Cohen (1992) einen großen Effekt ($f^2 = 1,85$) aufweist.

Auch nach drei Monaten hat die Lehrmethode einen signifikanten Einfluss auf die Leistung ($F(1,25) = 5,42, p = 0,028$). Sie erklärt zu 14,5% die Leistung der Studierenden, welches nach Cohen einen großen Effekt ($f^2 = 0,46$) aufweist. Die Prüflinge der Interventionsgruppe erreichten zum Zeitpunkt T_2 im Schnitt 1,3 Punkte mehr als die der Kontrollgruppe.


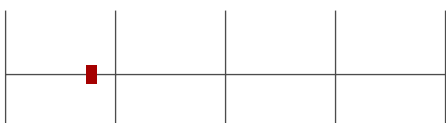
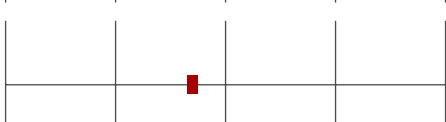
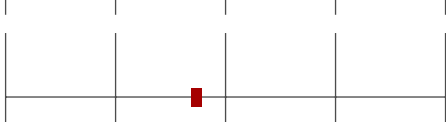
3.2.4.2 Ergebnisse der linearen Regression für die Station „Modell eckig trimmen“

Die Lehrmethode hat zum Zeitpunkt T1 einen signifikanten Einfluss auf die Leistung der Studierenden ($F(1,37) = 5,01, p = 0.031$). Sie erklärt zu 9,5% die Leistung der Studierenden in der Station. Im Durchschnitt erreichen die Prüflinge der Interventionsgruppe 0,8 Punkte mehr in der Stationsleistung, als die Prüflinge der Kontrollgruppe, welches nach Cohen einen großen Effekt ($f^2 = 0,38$) aufweist. Auch nach drei Monaten hat die Lehrmethode einen signifikanten Einfluss auf die Leistung ($F(1,25) = 7,53, p = 0,011$). Die Leistung der Studierenden kann zu 20,1% vorhergesagt werden, welches einem großen Effekt nach Cohen ($f^2 = 0,55$) entspricht. Die Prüflinge der Interventionsgruppe erreichten zum Zeitpunkt T₂ im Schnitt 0,6 Punkte mehr als die der Kontrollgruppe.

3.2.5 Studentische Evaluation des Pilotprojekts

Die Teilnehmer des Pilotprojekts evaluierten im Anschluss an die Follow-Up-Untersuchung das Pilotprojekt (6.1.3). In den folgenden Tabellen (23 und 24) ist die Auswertung dargestellt mit den Parametern: Anzahl (N), Mittelwert (M), Median (MD) und Standardabweichung (SD). An der Evaluation nahmen 27 Studierende teil. Die Fünf-Punkt-Skala ist von links „trifft voll zu“ nach rechts „trifft nicht zu“ angeordnet.

Tabelle 24: Studentische Evaluation zu den Übungen

	trifft voll zu	trifft nicht zu	N = 27
Ich kann wesentliche Anteile der eckigen Modell-Trim-mung in der Theorie benennen			M = 1,8 MD = 2,0 SD = 0,6
Ich kann die eckige Modell-Trim-mung eigenständig durchführen			M = 1,8 MD = 2,0 SD = 0,6
Ich kann wesentliche Anteile der Dreiecksklammer-Bie-gung in der Theorie benennen			M = 2,7 MD = 3,0 SD = 0,8
Ich kann wesentliche Anteile der Dreiecksklammer-Bie-gung in der Theorie eigenständig durchführen			M = 2,7 MD = 3,0 SD = 0,8

Aus den Fragen hinsichtlich beider Übungen ist zu sehen, dass die Teilnehmer sich in der Übung „Modell eckig trimmen“ bezüglich der theoretischen und praktischen Kenntnisse besser als in der Übung „Dreiecksklammer biegen“ einschätzten. Zur Bewertung der Station „Modell eckig trimmen“ sind für den praktischen und theoretischen Teil ein Schnitt von 1,8 ermittelt worden. Die theoretischen und praktischen Kenntnisse der Dreiecksklammer-Bie-gung liegen auf der Skala im Schnitt bei 2,7.

Tabelle 25: Studentische Evaluation zum Peerteaching

	trifft voll zu	trifft nicht zu	N = 27
Ich erachte ein Peerteaching (Student lehrt Student) als sinnvoll			M = 1,4 MD = 1,0 SD = 0,6
Die Atmosphäre während des Peerteachings ist im Vergleich zum Kurs (TPK) angenehmer			M = 1,1 MD = 1,0 SD = 0,4
Die Möglichkeit der Eigenreflektion und Diskussion mit dem Peerteacher bezüglich der Übungen führt zu einer verbesserten Wahrnehmung meiner Fähigkeiten			M = 1,6 MD = 2,0 SD = 0,6
Ich habe in der regulären Kurszeit ausreichend Zeit über meine Arbeit zu diskutieren und reflektieren			M = 3,9 MD = 4,0 SD = 0,8
Durch das Feedback des Peerteachers konnte ich die Übungen korrekt umsetzen			MW = 1,7 MD = 2,0 SD = 0,6
Die betreute Übungszeit im Rahmen des Projekts ist sinnvoller als die unbetreute freie Übungszeit (Di/Do 17-20 Uhr)			M = 1,7 MD = 2,0 SD = 0,9
Ein Peerteaching in Bezug auf zahntechnische Fertigkeiten sollte in der Vorklinik angeboten werden			M = 1,2 MD = 1,0 SD = 0,4
Die OSCE Erfahrung im Rahmen des Projekts bewerte ich als positiv			M = 1,3 MD = 1,3 SD = 0,5
Die OSCE-Erfahrung ist eine gute Vorbereitung auf die chirurgische OSCE am Ende des 8. Semesters			M = 2,1 MD = 2,0 SD = 0,9

Die Bewertung des im Pilotprojekt erfolgten Peerteachings wurden von den Teilnehmern im Schnitt von 1,1 bis 1,7 gewertet. Hinsichtlich der Sinnhaftigkeit eines Peerteachings, der Atmosphäre, des Feedbacks und der Möglichkeit der Eigenreflexion und Diskussion ist die Evaluation sehr positiv ausgefallen. Die Studierenden erachten ein Peerteaching als äußerst sinnvoll ($M = 1,4$) und würden sich ein zahntechnisches Peerteaching in der Vorklinik wünschen ($M = 1,2$). Die angenehme Atmosphäre und die Möglichkeit der Diskussion und Eigenreflexion aufgrund einer stressfreien Umgebung wurde als sehr gut empfunden. Dies fehlt während der regulären Kurszeit des technisch propädeutischen Kurses (TPK) und wurde dementsprechend geringwertig bewertet ($M = 3,9$). Zusätzlich wurde die OSCE-Erfahrung als positiv bewertet ($M = 1,3$) und als eine gute Vorbereitung für klinische OSCEs angesehen ($M = 2,1$).

4 Diskussion

4.1 Diskussion der Methoden

Im Rahmen der vorliegenden Arbeit wurde die Vermittlung praktischer Fertigkeiten anhand von zwei in der Literatur ausführlich untersuchten Lehrformaten (Standardinstruktion *see one, do one* und Peytons 4-Schritt-Methode 1988) mithilfe eines strukturierten Peerteachings im Wintersemester 2018/2019 an der UMG durchgeführt. Im Folgenden werden die Methoden kapitelweise diskutiert und mit der aktuellen Literatur in Zusammenhang gestellt.

4.1.1 Studiendesign

Bei der hier vorliegenden Studie handelt es sich um eine prospektive, monozentrische, randomisierte Interventionsstudie, die im Wintersemester 2018/2019 an der UMG in Form eines Pilotprojektes stattfand. Die Teilnahme an diesem Pilotprojekt erfolgte auf freiwilliger Basis der Studierenden des 1. Fachsemesters Zahnmedizin. Die Lehrinhalte des Pilotprojektes wurden nicht in den curricularen Lehrplan der UMG implementiert. Hieraus kann bereits ein Nachteil der Studie hinsichtlich der Motivation der Teilnehmer resultieren, womit sich möglicherweise die Rücklaufquote erklären lässt. Die Rücklaufquote zur ersten formativen Prüfung betrug 95,12% (39 von 41 möglichen Teilnehmern). Zur zweiten formativen Prüfung hin betrug die effektive Rücklaufquote 65,85% (27 von 41 möglichen Teilnehmern). Im Vergleich zu ähnlichen Studien ist die Rücklaufquote des Pilotprojektes vermindert (z.B. 97,01% in einer Studie von Münster et al. 2016 bzw. 93,13% in einer Studie von Seifert et al. 2020). Die Einteilung der Studiengruppen erfolgte randomisiert, dadurch entfiel eine mögliche Präferenz hinsichtlich des Lehrformates, zu der sich gegebenenfalls motivierte Studierende entschieden hätten. Randomisierte Studien gelten in der evidenzbasierten Medizin als Goldstandard zum empirischen Nachweis von Behandlungen. Dies lässt sich auch auf die Lehrforschung übertragen. Zusätzlich zur Randomisierung ist eine kontrollierte Studie (*randomized controlled trial/RCT*) als bestes Studiendesign anzusehen.

In diesem Pilotprojekt wurde auf eine klassische Kontrollgruppe, der keinerlei Lehrformat zuzuordnen wäre, bewusst verzichtet. Hieraus würde man keine Erkenntnis generieren, da etwas, was zuvor nicht gelehrt wurde, auch nicht geprüft werden kann.

Stattdessen wurde das Leistungsniveau der beiden Gruppen hinsichtlich praktischer

Vorkenntnisse in Form eines Prätestes mit Hilfe der unbekanntesten Station (Politur) aus der ersten formativen Prüfung überprüft. Da sich die Teilnahme an dem Pilotprojekt ausschließlich auf das erste Fachsemester beschränkte, handelt es sich um eine monozentrische Studie. Die praktischen Übungen, die im Rahmen des Pilotprojektes gelehrt wurden beziehen sich speziell auf zahntechnische Fertigkeiten aus dem Bereich der Kieferorthopädie. Aus diesem Grund können die Ergebnisse dieser Studie nicht auf das gesamte Gebiet der Zahnmedizin übertragen werden.

In Folgestudien könnte untersucht werden, inwiefern die Ergebnisse auf andere Semester und andere zahnmedizinische Fachbereiche zutreffen.

4.1.2 Zeitlicher Ablauf des Pilotprojekts

Die Gruppen wurden an unterschiedlichen Tagen in die Übungen eingewiesen (Abbildung 3). Daraus ergaben sich bis zur ersten formativen Prüfung unterschiedliche Zeitabstände. Demnach hatte die erste Gruppe den längsten zeitlichen Abstand bis zur ersten formativen Prüfung, was als Nachteil angesehen werden kann. Die Gruppe, die als letzte unterwiesen wurde, hatte den Vorteil, dass die Unterrichteinheit mit einem geringeren Zeitabstand zur Prüfung stattfand. Für optimale Prüfungsbedingungen wäre es von Vorteil gewesen alle Gruppen am selben Tag zu unterweisen, welches allerdings im Rahmen des Pilotprojekts organisatorisch nicht möglich war. Im Vergleich zu ähnlichen Studien, wie z.B. von Münster et al. (2016) und Seifert et al. (2020) geht nicht konkret hervor, ob die Lehrunterweisung in allen Gruppen am selben Tag stattfand. Eine Ausnahme bietet die Studie von Krautter et al. (2015), bei der alle Gruppen an einem Tag unterwiesen wurden.

Zur Überprüfung der Retention durch eine zweite formative Prüfung wurde ein Zeitraum von drei Monaten definiert, welcher sich aus der Dauer des Wintersemesters 2018/2019 ergab. Hinsichtlich der Untersuchung des Retentionszeitraums gibt es in der Literatur starke Abweichung, welche in Kapitel 4.1.4 dargestellt werden.

4.1.3 Die Lehrmethoden in der Literatur

Zu den Lehrmethoden der Standardinstruktion (*see one, do one*) und der 4-Schritt-Methode (Peyton 1988), die in der medizinischen Lehre fest etabliert sind gibt es zahlreiche Studien, die sich unter anderem mit der Effizienz, der Retention und dem Kosten-Nutzen-Verhältnis beschäftigen. Im Folgenden werden beide Lehrmethoden kritisch betrachtet und ihre Relevanz für das Pilotprojekt erläutert.

Die Lehrmethode durch eine Standardinstruktion (*see one, do one*) wird in der Literatur kontrovers diskutiert. Nach Khodaverdi (2018) hat die Standardinstruktion unter bestimmten Voraussetzungen weiterhin ihre Daseinsberechtigung in der medizinischen Lehre. Er betont, dass in der Vergangenheit die Auffassung von *see one, do one* mit einer inadäquaten Betreuung assoziiert worden sei. Durch eine adäquate Betreuung hingegen und das Feedback einer geeigneten Aufsicht kann eine Standardinstruktion durchaus ausreichend sein. „*There is clearly a place for observing a procedure before being expected to perform it – ‘do one’ does not necessarily imply being left unsupervised to perform a procedure no matter the historical expectations that this may have been so*” (Khodaverdi 2018).

Demgegenüber argumentieren Speirs und Brazil (2018), dass *see one, do one* eine historische Lehrmethode sei und aus der Zeit nicht-evidenzbasierter Medizin stamme. Dabei beziehen sie sich auf Studien von Barsuk et al. (2009, 2018), in denen es durch *simulated-based learning*, also durch Simulationen, zu weniger Komplikationen beim Legen eines zentralen Venenkatheters und zu besseren Leistungen hinsichtlich der Durchführung einer Lumbalpunktion geführt hat. „*The principle of ‘see one, do one, teach one’ is one that has been a cornerstone of medical education in previous times. However, with increasing evidence around medical education, it is no longer a defensible method of achieving procedural competency...*”

Hinsichtlich aktueller Studien ergibt sich eine Tendenz, die sich für einen Ausbau der Standardinstruktion aussprechen, um einer evidenzbasierten Medizin gerecht zu werden (Albert und Burns 2018, Kotsis und Chung 2013).

Die 4-Schritt-Methode nach Peyton (1988) ist im Vergleich zur Standardinstruktion eine modernere und jüngere Lehrmethode, welche Bestandteil aktuellerer Studien ist. Im Folgenden werden einige Studien vorgestellt, die Parallelen zum Pilotprojekt aufweisen.

In einer Studie von Krautter et al. (2011) wurde das Legen einer Magensonde unter Studierenden des zweiten und dritten Semesters der Humanmedizin hinsichtlich der studentischen Leistung in einer OSCE-Prüfung beobachtet. Untersucht wurden die Leistungsunterschiede einer Interventionsgruppe (n = 17), die nach der Peyton Lehrmethode und einer Kontrollgruppe (n = 17), die nach einer Standardinstruktion unterwiesen wurde.

Hinsichtlich der Leistungsunterschiede, die nach einer binarischen Checkliste beurteilt wurde, ergab sich kein signifikanter Unterschied ($p < .802$). Bezüglich der globalen Bewertung zur „Professionalität“ und „Arzt-Patienten-Kommunikation“ konnte hingegen ein signifikanter Unterschied zu Gunsten der Peyton-Lehrmethode ermittelt werden ($p < .001$). Zusätzlich war die Interventionsgruppe im Vergleich zur Kontrollgruppe beim Legen der Magensonde signifikant schneller ($168 \pm 30s$ vs. $242 \pm 53s$; $p < .001$). Die Autoren geben an, dass die Studierenden, vor allem bei der erstmaligen Umsetzung einer Fertigkeit von der Peyton-Lehrmethode hinsichtlich Schnelligkeit und professionellem Auftreten profitieren können.

In einer weiteren Studie untersuchten Krautter et al. (2015) die einzelnen der vier Peyton-Schritte. Hierzu wurden in einer randomisierten und kontrollierten Studie vier Gruppen untersucht, gebildet aus 97 Studierenden der Humanmedizin, die in einer Kombination der Einzelschritte das Legen eines zentralen Venenkatheters erlernen sollten. Mithilfe einer binarischen Checkliste erfolgte die Beurteilung der Leistung in einer OSCE-Prüfung einen Tag nach der Interventionsphase. Anschließend erfolgte am zweiten Tag eine telefonische Befragung der Teilnehmer hinsichtlich der Einzelschritte der Fertigkeit. Gruppe drei, die als einzige Gruppe den dritten Schritt von Peytons 4-Schritt-Methode durchlief erzielte sowohl in der Prüfung als auch bei der telefonischen Befragung das höchste Ergebnis im Vergleich zu den restlichen Gruppen. Die Autoren geben an, dass der dritte Schritt „der Lehrende führt die Tätigkeit nach Anleitung des Lernenden durch“, der wichtigste Schritt hinsichtlich des Erwerbes prozeduraler Fertigkeiten sei. Das Visualisieren bzw. die Vorstellung der auszuübenden Fertigkeit führe zu einem tieferen Kodierungsprozess im Gegensatz zur alleinigen Observierung (Krautter et al. 2015).

Der Benefit eines mentalen Trainings bzw. eines verbalisierten Wiederholens einer Fertigkeit konnte auch in einer Studie von Immenroth et al. (2007) beobachtet werden. In dieser Studie schnitt die Gruppe der Chirurgen mit einem mentalen Training hinsichtlich einer laparoskopischen Untersuchung signifikant besser als die Kontrollgruppe und die Gruppe mit einem praktischen Training ab. Die mentale Vorstellung einer auszuübenden Tätigkeit *motor imagery* ist eine etablierte Methode das Gehirn hinsichtlich motorischer Abläufe zu trainieren. Mithilfe der funktionellen Magnetresonanztomographie (fMRT) konnte bereits in Studien die neurale Aktivität des primär motorischen Cortex' und des Kleinhirns durch *motor imagery* gezeigt werden (Roth et al. 1996).

4.1.4 Retention von Wissen und Fertigkeiten in der Literatur

In der Lehrforschung gehört die Retention von Wissen und Fertigkeiten zu einem der Kernpunkte. Unter Retention wird in der Lehrforschung die Fähigkeit verstanden, Wissen langanhaltend zu speichern, abzurufen und somit das Wissen vom Kurzzeitgedächtnis in das Langzeitgedächtnis zu übertragen. Um Wissen in das Langzeitgedächtnis einzubauen bedarf es an speziellen Lehrmethoden, wie z.B. das wiederholte Üben oder Testen, in der Literatur als *testing effect* bekannt. Im Bereich praktischer Fertigkeiten wird das Wissen in das prozedurale Gedächtnis gespeichert, welches Fertigkeiten, wie motorische Abläufe umfasst, die ohne Nachdenken umsetzbar sind. Der *testing effect* besagt, dass Lernprozesse effektiver sind, wenn sie aktiv geübt werden, indem nach der initialen Lernphase ein Wissensabruf in Form eines Tests stattfindet. Seit 2006 wird der *testing effect* bezogen aufs Medizinstudium untersucht. In einigen Studien konnte die Effizienz des *testing effects* bewiesen werden (Roediger et al. 2006, Karpicke und Roediger 2008, Larsen et al. 2013, Rowland 2014). Für die Langzeitretention von Wissen fanden Karpicke und Roediger (2007) heraus, dass der entscheidende Faktor hierzu in der verzögerten Wissensabfrage, unabhängig der anschließenden Testintervalle liegt. Demnach sollte ein Wissensabruf nicht unmittelbar nach der Interventionsphase stattfinden, sondern mit einer Verzögerung einhergehen.

An der UMG wurde 2013 erstmalig der *testing effect* anhand klinischer Fallgeschichten und anschließenden Tests untersucht (Raupach et al 2016). Hierbei konnte gezeigt werden, dass das wiederholte Testen im Gegensatz zum wiederholten Lesen effektiver war.

Sennhenn-Kirchner und Kollegen (2017) konnten zeigen, dass das wiederholte Prüfen von Fertigkeiten auch in der zahnmedizinischen Ausbildung im Vergleich zu einem wiederholten Üben mit einer besseren Langzeitretention der Fertigkeit verbunden ist, „*repeated testing produces more favourable skills retention than repeated practice*“.

4.1.5 Retention erworbener Fertigkeiten

In der vorliegenden Studie fand die erste formative Prüfung wenige Tage nach der letzten Unterrichtseinheit statt. Auf Basis der Ergebnisse der ersten formativen Prüfung kann dementsprechend der kurzfristige Effekt der Lehrmethode im Hinblick auf die Retention von Fertigkeiten beurteilt werden. Die zweite formative Prüfung, die nach einer interventionsfreien Phase von drei Monaten stattfand, lässt Rückschlüsse auf den mittelfristigen Erwerb praktischer Fertigkeiten ziehen.

In publizierten Studien erfolgten die Posttests zur Evaluation der Retention der Lehrmethoden unmittelbar im Anschluss an die Interventionsphase (Krautter et al. 2011; 2015), zusätzlich im Abstand von acht Wochen (Seifert et al. 2020) und nach fünf bzw. sechs Monaten (Münster et al. 2016).

In diesem Zusammenhang stellt sich die Frage, inwiefern die erworbenen Fertigkeiten durch die unterschiedlichen Lehrmethoden ins Langzeitgedächtnis übertragen werden können. In einer ähnlichen Pilotstudie von Münster et al. (2016) wurde der Effekt von Peytons 4-Schritt-Methode und der Standardinstruktion (*see one, do one*) bezogen auf die kardiopulmonale Reanimation untersucht. 134 freiwillige Studierende der Humanmedizin des zweiten und dritten Fachsemesters nahmen an der randomisierten Studie teil. Die Leistung der Prüflinge wurde eine Woche und fünf bzw. sechs Monate nach der Interventionsphase mithilfe einer OSCE-Prüfung untersucht. Hinsichtlich des mittelfristigen Überprüfungszeitraumes von fünf bzw. sechs Monaten zeigte sich kein signifikanter Unterschied zwischen den Gruppen. Hierfür betonen die Autoren, dass die Herzdruckmassage als Fertigkeit nicht komplex genug sei, um von der Peyton-Methode zu profitieren.

In einer Studie von Seifert et al. (2020), zeigten Studierende, die videobasiert entweder mit Peytons 4-Schritt-Methode oder mithilfe der Standardinstruktion gelernt haben, acht Wochen nach Abschluss der Interventionsphase eine gleich gute Retention von Wissen und Fähigkeiten zur Gesichtsuntersuchung und Durchführung einer Bellocq-Tamponade.

Direkt im Anschluss an die Interventionsphase schnitt die Gruppe nach Peytons 4-Schritt-Methode ($n = 45$) signifikant besser ($p < .01$) als die Gruppe der Standardinstruktion ($n = 57$) ab. Nach acht Wochen konnten keine Unterschiede festgestellt werden ($p < .362$). Die Ergebnisse des Pilotprojekts stimmen mit den bereits erwähnten Studien teilweise überein. Im Gegensatz zu Seifert et al. (2020) und Münster et al. (2016) konnte allerdings nach drei Monaten ein signifikanter Unterschied zwischen den beiden Gruppen beobachtet werden (Dreiecksklammer biegen, $p = .028$; Modell eckig trimmen, $p = .011$).

Für zukünftige Studien wird vorgeschlagen die Langzeitretention von Fertigkeiten, die sich mit dem Vergleich dieser Lehrmethoden befassen, durch Postmessungen zu überprüfen. Diese sollten im Abstand von mehreren Monaten bzw. Jahren zur Interventionsphase durchgeführt werden und eine repräsentative Stichprobe einschließen.

4.1.6 Peerteaching in der Literatur

Peerteaching oder PAL (*Peer Assisted Learning*) ist im Hinblick auf den positiven kognitiven, pädagogischen und sozio-ökonomischen Effekt in der Literatur ausführlich untersucht worden (Topping 1996, 1998, 2007; Maheady 1998). Aufgrund der Herausforderungen an universitären Einrichtungen mit stetig steigenden Zahlen der Studierenden bietet PAL eine Möglichkeit personelle Ressourcen zu schonen (Ten Cate und Durning 2007). PAL kann nach Perkins et al. (2002) als eine Ergänzung zum Curriculum angesehen werden.

In der Literatur finden sich auch Nachweise, dass PAL nicht nur für die Lernenden, sondern auch für die Tutoren von Vorteil ist (Bardach et al. 2003, English et al. 2006).

Wie die empirische Evidenz zeigt, ist PAL in der medizinischen Lehre eine effektive und effiziente Methode klinische und prozedurale Fertigkeiten zu unterrichten (Nestel und Kidd 2005, Ross und Cameron 2007, Ten Cate und Durning 2007, Dandavino et al. 2007). PAL bietet Führungsqualitäten, Coaching, Training von Fertigkeiten, verstärkt das Selbstvertrauen und steigert die innere Motivation (Ten Cate und Durning 2007).

Die Akzeptanz und Durchführbarkeit einer OSCE, die ausschließlich von Tutoren betreut wurde haben Lee et al. (2018) untersucht. Die Prüfungsgruppe bestand aus Studierenden des 3. Fachsemesters, Standardpatienten aus dem 1./2. Fachsemester und studentischen Prüfern aus dem 4. Fachsemester der Humanmedizin. Die Bewertung erfolgte durch eine studentisch erstellte Checkliste mit zusätzlicher Vergabe einer Globalnote. Inhaltlich befassten sich die Stationen mit der Anamnese, einer körperlichen Untersuchung, einem Beratungsgespräch und einer gezielten Behandlung. Die Inhalte der Stationen basierten auf curricularen Vorgaben und wurden somit nicht OSCE-spezifisch gelehrt. Anhand einer Evaluation bezüglich der Zufriedenheit der Tutoren und Prüflinge konnte eine hohe Akzeptanz der rein studentischen OSCE ermittelt werden. Die Autoren belegen einen Benefit für alle beteiligten Fachsemester, mit dem positiven Übungseffekt dieser OSCE. *„First- and second-year students observed examinees while reflecting on their own clinical skills, whereas fourth-year students gained important experience in a teaching role before their transition into residency“* (Lee et al. 2018).

Aufgrund der eindeutigen Datenlage zu PAL wurden in dem hier untersuchten Pilotprojekt die praktischen Übungen ausschließlich von Tutoren durchgeführt. In Übereinstimmung mit der Studie von Lee et al. (2018) konnte die hohe Akzeptanz der Tutoren durch die Studierenden des 1. Fachsemesters (4.1.10) evaluiert werden.

4.1.7 Bewertung durch studentische Rater

Hinsichtlich der Bewertung studentischer Leistungen in einem Prüfungsformat durch Tutoren ist die Datenlage unstimmg. Nach Reiter et al. (2004) haben Tutoren im Vergleich zu Dozenten die Prüflinge einer OSCE signifikant besser bewertet. Auch Burgess et al. (2012) konnten herausfinden, dass Tutoren in der Vergabe einer OSCE-Globalnote die Benotung nicht kompetent genug beurteilten. Demgegenüber konnten Bucknall et al. (2008) belegen, dass Tutoren in einer OSCE mit *basic life support* im Gegensatz zu den Dozenten das Bestehen der Station strenger bewerteten.

In einer Studie der UMG (Med. Diss. Göttingen 2017) wurden studentische Rater mit ärztlich/zahnärztlichen Ratern verglichen. Hierbei stellte sich heraus, dass studierende Rater Prüflinge weder signifikant besser noch schlechter beurteilten als ärztliche Rater. Nach Chenot et al. (2007) ist die formative Bewertung durch die studentischen Rater qualitativ besser, wenn sie zusätzlich als Peerteacher tätig sind: „*senior students who are involved in the teaching of the skills can reliably assess their peers at OSCE stations [...]*“.

In Anlehnung an die erwähnten Studien (Med. Diss. Göttingen und Chenot et al. 2007) erfolgte die Bewertung der studentischen Leistung des Pilotprojekts ausschließlich durch studentische Rater.

4.1.8 Effizienz der Raterschulung

Im Rahmen dieses Pilotprojekts wurde eine Raterschulung von ca. 30 Minuten durch die Peerteacher durchgeführt. Die erste Raterschulung fand am Tag vor der ersten OSPE-Prüfung statt. Aus organisatorischen Gründen war es nicht möglich den OSPE-Parcours zuvor aufzubauen und die Stationen zur Probe zu durchlaufen. Empfohlen wird eine 30-minütige Einführung direkt vor der Prüfung (Nikendei und Jünger 2006). Nach Friedmann (2000) ist es sinnvoll den Ratern ein jeweils negatives und positives Beispiel einer vergangenen Prüfung zur Verfügung zu stellen. Eine gute Schulung ist nicht vom Stand der Ausbildung abhängig (Wilkinson et al. 2003), somit ist jeder als Rater geeignet ist, der zuvor an einer intensiven Schulung teilgenommen hat. Dennoch ist es sinnvoll, den Prüfungsablauf zuvor durchzuspielen (Boulet et al. 2003). Allerdings fehlt es an Daten, die belegen wie sich Raterschulungen konkret auf die Bewertung auswirken. Dennoch lässt sich sagen, dass ungeschulte Rater im Vergleich zu geschulten Ratern in ihrer Bewertung unbeständig sind (Holmboe 2004, Pell 2008). Reid et al. 2016 untersuchten in ihrer Studie eine Raterschulung mit insgesamt 391

Ratern, bestehend aus 171 Ärzten und 220 Studierenden in Form einer Probe-OSCE. Die einstündige Raterschulung erfolgte durch einen einzigen Schulungsleiter. Die Autoren erwähnten nicht, ob für die Schulung eine spezielle Schulungsmethode angewandt wurde.

Zwischen den Ratern konnte in Bezug auf die Beurteilungen kein signifikanter Unterschied festgestellt werden. Diese Studie ist hinsichtlich der Effizienz und Schonung personeller Ressourcen ein gelungenes Modell.

Das Missverständnis in der Station „Modell eckig trimmen“ (4.2.3), bei dem die Ergebnisse nachträglich mit dem Rater korrigiert werden mussten, hätte durch eine intensive Schulung vor Prüfungsbeginn vermieden werden können. Ein Probedurchlauf der Stationen zusammen mit den Ratern, sowie eine Schulung unmittelbar vor der Prüfung sind für zukünftige Projekte zu empfehlen.

4.1.9 Bewertung der OSPE und der Checklisten

Eine curricular implementierte OSPE im vorklinischen Abschnitt des Studiengangs Zahnmedizin, welche praktische Fertigkeiten beinhaltet, ist derzeit an der UMG nicht vorgesehen. Im nationalen Vergleich führt die Universität Münster im vierten Semester OSPE-Prüfungen im Rahmen des Phantomkurses der Zahnersatzkunde durch. Hierbei finden OSPE-Prüfungen nach jedem Kursteil statt, welche theoretisches Wissen und praktische Fertigkeiten prüfen. Als Abschlussprüfung findet eine finale OSPE-Prüfung statt, die das Aufwachsen von zwei Zähnen und das Präparieren von einem Front- und Seitenzahn beinhaltet. Studien, die sich mit OSPE-Prüfungen im vorklinischen Abschnitt der Zahnmedizin befassen sind im Vergleich zur Humanmedizin rar. In einer Studie von Ratzmann et al. (2012) wurde die Integration einer OSCE in das zahnmedizinische Physikum hinsichtlich des frühen Patientenkontaktes überprüft. Insgesamt nahmen 72 Studierende, aus zwei Semestern an dem Kurs „früher Patientenkontakt“ teil, welcher sich inhaltlich mit einem Anamnesegespräch und einem Zahn- und Mundhygienestatus befasste. Die Interventionsgruppe unterzog sich einer zusätzlichen Trainingseinheit in Vorbereitung auf die OSCE-Prüfung. Diese Trainingseinheit bestand darin in Zweiergruppen gegenseitig eine Anamnese und einen zahnärztlichen Befund zu erheben. Die Studie fand heraus, dass die Leistung der Interventionsgruppe in der OSCE-Prüfung signifikant besser waren als die der Kontrollgruppe ($p < .01$). Im Vergleich zu praktischen Fertigkeiten sind in der Literatur mehr Studien zu kommunikativen Kompetenzen im Zahnmedizinstudium untersucht worden (Manogue und Brown 1998, Schoonheim-Kleim et al. 2006, Cannick et al. 2007, Larsen und Jeppe-Jensen 2008).

Hinsichtlich der Beurteilung von Checklisten eignen sich zur Evaluation der Testgüte einer OSCE sowohl globale als auch Checklisten-Ratings (Nikendei und Jünger 2006).

In der Literatur gibt es eine Vielzahl an Studien, die ausschließlich Checklisten verwenden (Krautter et al. 2015, Münster et al. 2016, Seifert et al. 2020) und Studien, die zusätzlich Globalnoten anwenden (Krautter et al. 2011, Med. Diss. Göttingen 2017). Da die OSPE-Prüfungen des Pilotprojekts ausschließlich praktische Fertigkeiten enthielten ist ein reines Checklistenformat das adäquate Mittel, während kommunikative Fertigkeiten eher global bewertet werden können (Newble 2004; Nikendei und Jünger 2006). Verhaltensweisen, z.B. Empathie, die durch die reinen Checklisten unberücksichtigt bleiben, könnten in die Globalnote miteinfließen (Chenot et al. 2007), wobei in der hier vorgestellten Studie auf eine Beurteilung von Empathie bewusst verzichtet wurde. Die ausführliche Diskussion der Items und der Stationen erfolgt in Kapitel 4.2.

4.1.10 Bewertung der studentischen Evaluation

An der Evaluation nahmen im Anschluss zur zweiten formativen Prüfung insgesamt 27 Studierende teil. Der Evaluationsbogen (6.1.3) bestand aus Fragen zu den praktischen Übungen, zum Peerteaching und zur OSPE-Prüfung. Die Fragen zu beiden Übungen zeigen, dass die Teilnehmer sich in der Übung „Modell eckig trimmen“ bezüglich der theoretischen und praktischen Kenntnisse besser als in der Übung „Dreiecksklammer biegen“ einschätzten. Aus der Itemanalyse konnte unter anderem eine Aussage über die Schwierigkeit (M) der Stationen generiert werden. Zu beiden Zeitpunkten der OSPE-Prüfung war die Station „Modell eckig trimmen“ vom Schwierigkeitsgrad her die leichtere Station. Diese Tatsache spiegelte sich auch in der Wahrnehmung der Studierenden wider. In Anlehnung zu den Studien zum Peerteaching, die in Kapitel 4.1.6 erwähnt wurden, kann der positive Effekt durch Peerteaching auch durch dieses Pilotprojekt belegt werden. Zusätzlich ist in Übereinstimmung an die Studie von Ratzmann et al. (2012) die OSPE Erfahrung hinsichtlich zukünftiger OSPE-Prüfungen durch die Studierenden sehr gut bis gut bewertet worden.

4.2 Diskussion der Ergebnisse

4.2.1 Testgüte der Stationen

Im Folgenden werden die drei Prüfungsstationen (Dreiecksklammer biegen, Modell eckig trimmen und Politur) in Bezug auf ihre Testgüte diskutiert. Hierfür werden die einzelnen Items der Checklisten, die im Anhang 6.2 zu finden sind, betrachtet.

4.2.2 Die Testgüte der Station „Dreiecksklammer biegen“

Im Folgenden werden die Items erklärt, die in der Itemanalyse (Tabelle 5) gelb markiert sind (Items 1, 5, 10 und 12).

Item 1 stellt mit einem Mittelwert von 1,00 und einer dazugehörigen Trennschärfe von 0,00 eine sehr einfache Aufgabenstellung dar, die keine Unterscheidung in leistungsstarke bzw. leistungsschwache Studierende zulässt. Jeder Prüfling war in der Lage das Item richtig zu lösen. Dieses Item ist in jeder der drei Prüfungsstationen identisch und überprüft, ob der Prüfling während des Durchlaufs der Station die Schutzbrille trägt. Obwohl dieses Item keine praktische Fertigkeit prüft, ist sie essentiell für das sichere Arbeiten im Labor. Dieses Item sollte aufgrund des Arbeitsschutzes keinesfalls gestrichen werden, um dadurch eine Verbesserung von Cronbachs Alpha zu bewirken. Nach Möltner et al. (2006) sollten Items, die basale Fertigkeiten prüfen, nicht unbedingt eliminiert werden, nur weil die dazugehörige Trennschärfe nicht im Optimum liegt. „Im Sinne einer kriteriumsorientierten Prüfung ist es jedoch durchaus auch sinnvoll, wichtige basale Fertigkeiten oder Kenntnisse in einer Prüfung unabhängig von ihrer erwarteten Schwierigkeit abzufragen (Möltner et al. 2006, S.3).

Item 5 befasst sich mit der Wahl des richtigen Instruments. 92% der Prüflinge haben dieses Item richtig gelöst. Die dazugehörige Trennschärfe ist mit 0,40 in einem sehr guten Bereich und besagt, dass zwischen guten und weniger guten Prüflingen unterschieden werden konnte. Die Prüflinge hatten während der Prüfung die Möglichkeit mehrere Zangen auszuwählen. In der Lehrintervention wurde allerdings nur die Handhabung der Flachspitzzange gelehrt, sodass der Prüfling die eine richtige Zange in der Prüfung auswählen sollte. Die Prüflinge, die die Flachspitzzange gewählt haben, erreichten auch in der gesamten Station bessere Ergebnisse. Die Items 6 bis 9 weisen Trennschärfen von über 0,4 auf und liegen somit in einem optimalen Bereich. Diese Items stellen die vier essentiellen, aufeinanderfolgenden Biegungen der Dreiecksklammer dar.

Item 10 konnte zu 56% richtig gelöst werden und stellt somit ein mittelschweres Item dar. Die dazugehörige Trennschärfe von 0,13 ist gering, sodass hier weniger zwischen leistungsstarken und leistungsschwachen Prüflingen unterschieden werden konnte. Dieses Item bezog sich auf die korrekte Handhabung der Flachspitzzange. Fiel die Wahl bereits auf eine falsche Zange (Item 5), hatte dies auch negative Auswirkung auf die Bewertung der Handhabung (Item 10). Zur Optimierung des Items könnte entweder keine Auswahl mehrerer Zangen möglich sein oder die Handhabung aller Zangen bewertet werden. Wird die Handhabung aller Zangen bewertet, so muss im Voraus die entsprechende Lehrintervention diese miteinbeziehen. Item 12 hat eine angemessene Schwierigkeit von 0,76 aber eine geringe Trennschärfe von 0,12. Die Prüflinge sollten die korrekte Position der Klammer am Modell bestimmen. 76% der Prüflinge haben dieses Item korrekt gelöst, davon haben die verbliebenen 24% das Item nicht richtig gelöst, die aber in den restlichen Items gut abschnitten. Die Wahl des richtigen Quadranten und der korrekten Lokalisation der Klammer wurde in der Aufgabenstellung explizit vorgegeben. Die Prüflinge hatten jedoch zum Zeitpunkt der ersten OSPE keine anatomischen Vorkenntnisse. Zur Optimierung dieses Items sollten anatomische Grundlagen Bestandteil der Lehrintervention sein, wenn diese in einer Prüfung abgefragt werden.

4.2.3 Die Testgüte der Station „Modell eckig trimmen“

Die Item-Mittelwerte (Tabelle 6) liegen in einem Bereich zwischen 0,56 und 1,00. Der Mittelwert mit 0,85 besagt, dass 85% der Prüflinge diese Station richtig gelöst haben und stellt somit eine eher leichte Station dar. Die Trennschärfen liegen zwischen 0,00-0,58 und im Durchschnitt bei 0,30. Mit insgesamt elf Items liegt die innere Konsistenz (Cronbachs Alpha) der Checkliste bei 0,61 und somit hinsichtlich der Reliabilität in einem fragwürdigen Bereich (Tabelle 4). Wie bereits im vorherigen Kapitel erwähnt ist Item 1 aufgrund der Arbeitssicherheit trotz der Parameter zu belassen. Gleiches gilt für Item 2, welches sich auf das Einschalten des Gerätes bezieht. Item 6 sticht mit einer sehr geringen Trennschärfe von 0,003 hervor. „Der Studierende/die Studierende schließt die Arbeit am Trimmer in der vorgegebenen Zeitspanne ab“, welches eine moderate Schwierigkeit (0,61) aber eine sehr geringe Trennschärfe (0,003) aufweist. Es hängt nicht mit den anderen Items zusammen, da das Weglassen des Items mit einer Verbesserung von Cronbachs Alpha einhergehen würde (0,61 auf 0,72). Hierzu wurde eine sorgfältige Überprüfung der Gruppenleistung vorgenommen, die zeigte, dass die Studierenden der Interventionsgruppe in allen anderen Items besser als die Kontrollgruppe abschnitt. Nur in Item 6 verhält es sich umgekehrt. Eine Kontrolle der

beschrifteten Modelle aus der Prüfung wurde von den Peerteachern zusammen mit dem Rater der Station durchgeführt. Da die Modelle in der Prüfung beschriftet wurden, konnte im Nachhinein eine exakte Zuordnung der Prüflinge erfolgen. Von insgesamt 15 Prüflingen, die mit der Trimmung innerhalb von fünf Minuten laut Rater nicht fertig geworden sind, wurden die Modelle einer visuellen Kontrolle unterzogen. Von den 15 Modellen sind neun mit vollständiger Trimmung im Nachhinein als gültig bewertet worden. In Rücksprache mit dem Rater hat sich ein Missverständnis herausgestellt, welches mit der zeitlichen Verfügbarkeit der Station zusammenhing. Dieser Fehler in der Raterschulung fiel dementsprechend erst nach der Datenauswertung der ersten OSPE-Prüfung auf und konnte aufgrund der Dokumentation im Nachhinein noch korrigiert werden.

4.2.4 Die Testgüte der Station „Politur“

Die Station „Politur“ war die dritte Station des OSPE-Parcours, dessen Inhalt nicht Bestandteil der Lehrintervention gewesen ist. Anhand der unbekannt Station konnte das vorherige Leistungsniveau der Studierenden evaluiert und die Lernerfolge, die durch die Lehrintervention zustande gekommen sind, verglichen werden. Eine weitere Besonderheit der Station lag in der Konzeption der Items 3-7, die sich neben der Auswahl zusätzlich auf die richtige Reihenfolge der Instrumente bezog. Dies resultierte in eine zu hohe Schwierigkeit von 0,45, niedrigen Trennschärfen von 0,09 und Cronbachs Alpha in einem inakzeptablen Bereich von 0,29. Nachträglich wurden die Items 3-7 neu codiert, sodass die Reihenfolge der Instrumente vernachlässigt werden konnte. Dadurch erfolgte eine leichte Verbesserung von Cronbachs Alpha von 0,29 auf 0,42. Das Überprüfen von zwei Parametern, nämlich „richtiges Instrument“ und „Reihenfolge“ sollte nicht in ein einziges Item hineinfließen und für zukünftige Aufgabenstellungen vermieden werden. Mit sechs verbliebenen Items ist die Reliabilität hinsichtlich der Beurteilung der Aufgabenstellung nicht optimal. Die unbekannt Station führte zu der Erkenntnis, dass von Items, die mehrere Arbeitsschritte beinhalten abzuraten ist. Des weiteren veranschaulicht die Station den Grundsatz, dass etwas nicht Gelehrtes auch nicht geprüft werden sollte. So ist auf Ebene einer kompetenzorientierten Prüfung im Bereich praktischer Fertigkeiten eine Intervention vonnöten, die diese Fertigkeiten lehrt (Euler 2011).

Von einer Diskussion der Testgüte der zweiten OSPE-Prüfung wurde abgesehen, da das Durchlaufen einer bereits bekannten Prüfung die Ergebnisse verfälscht und aufgrund des

testing effects eine Verbesserung der studentischen Leistung eintritt (Roediger und Karpicke 2006). Als Konsequenz hinsichtlich der inhaltlich bereits bekannten Aufgabenstellung war von einer Verschlechterung der Parameter (Schwierigkeit, Trennschärfe, Cronbachs Alpha) auszugehen, die in Tabelle 17 dargestellt ist.

4.3 Beurteilung der Hypothesen

Auf die in Kapitel 1.7 aufgestellten Hypothesen ergaben sich folgende Antworten. Für den Vergleich der Mittelwerte ergab sich zum Zeitpunkt der ersten OSPE ein signifikanter Unterschied, sodass die Nullhypothese H_1 verworfen und die Alternativhypothese H_{A1} bestätigt werden konnte ($H_{A1}: P \neq E$). Die Retentionsfähigkeit der Lehrformate wurde nach einem dreimonatigen Abstand zur Interventionsphase untersucht und die Mittelwertunterschiede erneut miteinander verglichen. Auch hier gab es einen signifikanten Unterschied zwischen den Mittelwerten, sodass die Nullhypothese H_2 verworfen und die Alternativhypothese H_{A2} bestätigt werden konnte ($H_{A2}: P \neq E$).

4.4 Limitation der Studie

Limitierend handelt es sich um ein monozentrisches Pilotprojekt, an dem lediglich das erste vorklinische Fachsemester Zahnmedizin teilnehmen konnte. Die praktischen Übungen des Pilotprojekts beschränkten sich auf zahntechnische Fertigkeiten, die Bestandteil des 6. Semesters sind. Somit lassen sich anhand dieser Studie keine Rückschlüsse auf die praktische Bandbreite des zahnmedizinischen Studiums ziehen.

Die freiwillige Teilnahme der Studierenden und die Möglichkeit die Teilnahme auch bereits nach Antreten zurückzuziehen, führte zu einer niedrigen Rücklaufquote von 65,85% (27 von 41 möglichen Teilnehmern). Des weiteren sollte der geringe Stichprobenumfang kritisch betrachtet werden. Die dadurch verzerrenden Faktoren verringern die Aussagekraft der Studienergebnisse. Als weitere Limitation der Studie ist der formative Charakter der Prüfung zu erwähnen. Eine formative Prüfung stellt für die Studierenden keinen Grund zur gezielten Prüfungsvorbereitung dar (Raupach et al. 2013), wodurch der Effekt der Intervention abgeschwächt werden kann. Wie bereits in Kapitel 4.1.5 erwähnt wurde hinsichtlich der Studie ein Retentionszeitraum von drei Monaten definiert. Hieraus kann dementsprechend nur der mittelfristige Effekt der Lehrmethoden beurteilt werden.

4.5 Weiterführende Fragestellung

In Hinblick auf die neue Approbationsordnung für Zahnärzte und die damit verbundene Umstrukturierung des Studiums können praktische Fertigkeiten nicht in dem Ausmaß wie zuvor gelehrt werden. Wie bereits im Einleitungsteil erwähnt, soll sich bezüglich zahntechnischer Fertigkeiten die Ausbildung auf das für den Behandler Relevante konzentrieren. Durch das Verschieben der Schwerpunkte ist aber mit einem deutlichen zeitlichen Nachlass in der praktischen Ausbildung zu rechnen. In der Empfehlung des zahnmedizinischen Wissenschaftsrates (2005) heißt es: „Entlastung des Curriculums von unnötigen Inhalten [...], sodass der Arzt die zahntechnischen Arbeitsweisen zwar kennen, nicht aber alle selber ausführen können muss“. In einer Pressemeldung des FVDZ-Bundesvorsitzenden vom 7.6.19, betont Schrader, dass die zahntechnische und prothetische Ausbildung im Zahnmedizinstudium erhalten bleibt. „[...]denn ein Zahnarzt muss in der Lage sein, zahntechnische Werkstücke beurteilen, herstellen und bearbeiten zu können“. Die Aussagen darüber, in welchem Ausmaß zahntechnische Fertigkeiten gelehrt werden sollen, bleibt weiterhin offen.

„Zahnärzte und Zahnärztinnen müssen weiterhin befähigt werden qualifizierte Aufträge zur Erstellung von Zahnersatz zu geben, die Qualität des Zahnersatzes zu beurteilen und über die Eingliederbarkeit von Zahnersatz zu entscheiden. Sie müssen hinsichtlich der Arbeitsvorbereitung und Nacharbeiten notwendigen Leistungen sowie die im Praxisumfeld anfallenden zahntechnischen Arbeiten erbringen können“ (Zahnmedizinischer Wissenschaftsrat 2005).

Niedergelassene Zahnärzte/-innen arbeiten eng mit zahntechnischen Laboren zusammen. Diese Zusammenarbeit beruht auf einer beidseitigen Abhängigkeit. Eine britische Studie konnte zeigen, dass aus Sicht der Zahntechniker die Kommunikation zwischen Behandler und Techniker aufgrund von mangelndem Fachwissen leidet. *“It was the view of dental technicians who responded that newly qualified dentists do not have an appropriate understanding of technical techniques. Dental schools are still not preparing new graduates to communicate effectively with dental laboratories“* (Juszczuk et al. 2009).

Inwiefern sich die Umstrukturierung der zahnmedizinischen Lehre auf die Qualität der zahnmedizinischen Ausbildung auswirkt lässt sich aus heutiger Sicht nicht abschätzen. Mögliche Entwicklungen hinsichtlich langfristiger Konsequenzen, wie ein gesteigertes Stressempfinden der Studierenden und der Dozenten sollte in Zukunft eruiert werden.

4.6 Implementierung vorklinischer OSPE-Prüfungen an der UMG

Das Universitätsklinikum Göttingen verfügt über drei studentische Laborräume in denen jeweils ca. 50 Studierende gleichzeitig arbeiten können. Demnach können maximal drei komplette Semester gleichzeitig im Labor arbeiten. Zusätzlich bietet das SINUZ für die Studierenden die Möglichkeit an zwanzig Phantomköpfen sowohl konservierende, prothetische als auch chirurgische Skills zu üben und zu vertiefen. Für die Durchführung eines Peerteachings bedarf es an engagierten und motivierten Studierenden, die zusammen mit der Abteilung für Prothetik ein Lehrkonzept für ein Tutorium erstellen. In diesen Tutorien können in Kleingruppen von vier bis sechs Teilnehmern zahntechnische Fertigkeiten durch zeitintensive Lehrmethoden gelehrt werden. Durch das Einsetzen von Tutoren werden die personellen Ressourcen geschont, da sie nach der Kapazitätsverordnung (KaPVO) nicht zu den personellen Kapazitäten gezählt werden. Die Übungszeit sollte, wie in unserem Pilotprojekt geschehen, von Tutoren begleitet werden, um den Studierenden Hilfestellung zu leisten.

Anschließend kann eine formative OSPE-Prüfung erfolgen, die den Studierenden ihren aktuellen Leistungsstand aufzeigt. Summative OSPE-Prüfungen sind allerdings aufgrund der inneren Motivation der Studierenden zu bevorzugen. Aufgrund der räumlichen und technischen Ausstattung sind die Voraussetzungen für die Umsetzung eines effektiven Peerteachings an der UMG gegeben.

5 Zusammenfassung

Das Studium der Zahnmedizin beinhaltet insbesondere im aktuellen vorklinischen Abschnitt des Studiengangs neben der Vermittlung naturwissenschaftlicher Kenntnisse die Vermittlung praktischer zahntechnischer Fertigkeiten. Ressourcenintensive Lehrformate wie Kleingruppenunterricht wurden bislang zur Vermittlung dieser Fertigkeiten nicht eingesetzt.

Die vorliegende Arbeit stellt ein Pilotprojekt der Zahnklinik der UMG vor, das die Lehrmethode nach Peyton (1988) zur Vermittlung praktischer Fertigkeiten mit strukturiertem Peerteaching kombiniert, somit einerseits personelle Ressourcen schont und zeitgleich ein didaktisch fundiertes Lehrformat nutzt.

Im Rahmen dieses Pilotprojekts wurden Studierenden des ersten Fachsemesters im Wintersemester 2018/2019 zwei verschiedene zahntechnische Fertigkeiten unter Einsatz von Peerteaching und zwei unterschiedlichen Lehrmethoden vermittelt. Der randomisierte, kontrollierte Studienaufbau verglich die Ergebnisse der Prüfungsleistung (formative OSPE) der beiden unterschiedlich unterrichteten Gruppen direkt im Anschluss an die Intervention und nach drei Monaten (Retentionswirkung). Gruppe 1 wurde nach der Standardinstruktion *see one, do one* gelehrt, Gruppe 2 erhielt die Unterweisung nach der 4-Schritt-Methode (Peyton 1988). Zudem evaluierten die Teilnehmenden das Projekt im Rahmen der Datenerhebung. Die erste formative Prüfung zeigte signifikant höhere Punktzahlen für die Teilnehmenden am Peyton-Lehrformat (Fertigkeit 1: $82\% \pm 1,4\%$ vs. $60\% \pm 1,2\%$; $p < 0,001$; Fertigkeit 2: $85\% \pm 0,5$ vs. $76\% \pm 1,3\%$; $p = 0,031$). Die Ergebnisse der zweiten formativen Prüfung bestätigten das Ergebnis der ersten OSPE. Die Teilnehmenden am Peyton-Lehrformat erzielten auch nach drei Monaten eine höhere Punktzahl (Fertigkeit 1: $84\% \pm 0,1\%$ vs. $74\% \pm 0,1\%$; $p = 0,028$ und Fertigkeit 2: $97\% \pm 0,2\%$ vs. $91\% \pm 0,7\%$; $p = 0,011$). Die studentische Evaluation des Pilotprojekts zeigte hohe Zufriedenheit sowohl mit Peerteaching als auch mit Kleingruppenunterricht.

Das vorgestellte Projekt bietet die Grundlage für weitere zahnmedizinisch-didaktische Studien zur Evaluation effektiver Implementierung von Peerteaching in Kleingruppen in der ZMK-Klinik der UMG. Dieser Aspekt der Lehre praktischer Fertigkeiten ist vor dem Hintergrund der neuen Approbationsordnung besonders relevant. Des Weiteren sollte die Retention der Fertigkeiten über längere Zeiträume in weiteren Untersuchungen evaluiert werden.

6 Anhang

6.1 Bögen

6.1.1 Einverständniserklärung der Studierenden zur Teilnahme am Pilotprojekt

Universitätsmedizin Göttingen,
Klinik für Mund-, Kiefer- und Gesichtschirurgie

Einverständniserklärung für Studierende zur Studie

**„Evaluation der Auswirkungen eines strukturierten Peerteachings
zahntechnischer Einzelaspekte auf die Leistung von
Studienanfängern - ein Pilotprojekt“**

Bitte eintragen:*

Teilnehmer/in

Nachname:

Vorname:

Ich wurde von der Projektleiterin vollständig über Wesen, Bedeutung und Tragweite der Studie aufgeklärt. Ich hatte die Möglichkeit, Fragen zu stellen, habe falls erforderlich Antworten erhalten und diese verstanden.

Mir ist bekannt, dass meine persönlichen Daten zunächst mit Bezug zu meiner Matrikelnummer gespeichert werden. Sobald die Datensammlung abgeschlossen ist, werden die Daten vollständig anonymisiert, so dass dann keine persönlichen Studien-Daten mehr gespeichert werden.

Ich weiß, dass die Teilnahme an dieser Studie freiwillig ist. Ich weiß, dass ich jederzeit und ohne Angabe von Gründen diese Zustimmung widerrufen kann, ohne dass sich dieser Entschluss nachteilig auf den Kursverlauf auswirken wird. Mir ist bekannt, dass nach Anonymisierung eine Löschung der Daten nicht mehr möglich ist. Die Verwendung meiner Evaluationsdaten setzt vor der Teilnahme an der Befragung folgende freiwillig abgegebene Einwilligungserklärung voraus, d.h. ohne die nachfolgende Einwilligung kann ich nicht an dem Projekt teilnehmen.

Ich erkläre mich damit einverstanden, dass im Rahmen dieser Studie erhobene Daten in Papierform und auf elektronischen Datenträgern aufgezeichnet werden. Die Speicherung erfolgt in anonymisierter Form. Die Daten sind vor unberechtigtem Zugriff geschützt. Die Einwilligung kann jederzeit ohne Angabe von Gründen widerrufen werden.

Ich habe die Studierenden-Information erhalten. Hiermit erkläre ich mich mit der Teilnahme an der Studie einverstanden.

Ort und Datum

Unterschrift der/des Studierenden

Göttingen, den

Ort und Datum

Unterschrift der Studienverantwortlichen

6.1.2 Einverständniserklärung der Studierenden zur Teilnahme an der Follow-Up-Untersuchung

Universitätsmedizin Göttingen,
Klinik für Mund-, Kiefer- und Gesichtschirurgie

Einverständniserklärung für Studierende zur Follow-Up-Studie

**„Evaluation der Auswirkungen eines strukturierten Peerteachings
zahntechnischer Einzelaspekte auf die Leistung von Studienanfängern –
ein Pilotprojekt.“**

Bitte eintragen:*

Teilnehmer/in

Nachname:

Vorname:

Ich wurde von der Projektleiterin vollständig über Wesen, Bedeutung und Tragweite der Studie aufgeklärt. Ich hatte die Möglichkeit, Fragen zu stellen, habe falls erforderlich Antworten erhalten und diese verstanden.

Mir ist bekannt, dass meine OSCE-Ergebnisse zunächst mit Bezug zu meiner Person gespeichert werden. Sobald die Datensammlung abgeschlossen ist, werden die Daten vollständig anonymisiert, so dass dann keine persönlichen Studien-Daten mehr gespeichert werden.

Ich weiß, dass die Teilnahme an dieser Studie freiwillig ist. Ich weiß, dass ich jederzeit und ohne Angabe von Gründen diese Zustimmung widerrufen kann, ohne dass sich dieser Entschluss nachteilig auf den Kursverlauf auswirken wird. Mir ist bekannt, dass nach Anonymisierung eine Löschung der Daten nicht mehr möglich ist. Ohne die nachfolgende Einwilligung kann ich nicht an der Follow-Up-Studie teilnehmen.

Ich erkläre mich damit einverstanden, dass im Rahmen dieser Studie erhobene Daten in Papierform und auf elektronischen Datenträgern aufgezeichnet werden. Die Speicherung erfolgt in anonymisierter Form. Die Daten sind vor unberechtigtem Zugriff geschützt. Die Einwilligung kann jederzeit ohne Angabe von Gründen widerrufen werden.

Ich habe die Studierenden-Information erhalten. Hiermit erkläre ich mich mit der Teilnahme an der Studie einverstanden.

Ort und Datum

Unterschrift der/des Studierenden

Göttingen, den

Ort und Datum

Unterschrift der Studienverantwortlichen

6.1.3 Evaluationsbogen

MUSTER

EvaSys		Electric Paper <small>elektronisches Papier</small>
Ladan Darweshi		UMG
Evaluation zur Studie		

Bitte so markieren: Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift. Dieser Fragebogen wird maschinell erfasst.
 Korrektur: Bitte beachten Sie im Interesse einer optimalen Datenerfassung die links gegebenen Hinweise beim Ausfüllen.

1.

1.1 Gruppenzugehörigkeit Standardgruppe Interventionsgruppe

2. Übungen

2.1 Ich kann wesentliche Anteile der eckigen Modell-Trimmung in der Theorie benennen trifft voll zu trifft überhaupt nicht zu

2.2 Ich kann die eckige Modell-Trimmung eigenständig durchführen trifft voll zu trifft überhaupt nicht zu

2.3 Ich kann wesentliche Anteile der Dreiecksklammer-Biegung in der Theorie benennen trifft voll zu trifft überhaupt nicht zu

2.4 Ich kann die Dreiecksklammer-Biegung eigenständig durchführen trifft voll zu trifft überhaupt nicht zu

3. Peer Teaching

3.1 Ich erachte ein Peer Teaching (Student lehrt Student) als sinnvoll trifft voll zu trifft überhaupt nicht zu

3.2 Die Atmosphäre während des Peer Teachings ist im Vergleich zum Kurs (TPK) angenehmer trifft voll zu trifft überhaupt nicht zu

3.3 Die Möglichkeit der Eigenreflexion und Diskussion mit dem Peer Teacher bezüglich der Übungen führt zu einer verbesserten Wahrnehmung der eigenen Fähigkeiten trifft voll zu trifft überhaupt nicht zu

3.4 Ich habe in der regulären Kurszeit (TPK) ausreichend Zeit über meine Arbeit zu diskutieren und zu reflektieren. trifft voll zu trifft überhaupt nicht zu

3.5 Durch das Feedback des Peer Teachers konnte ich die Übungen korrekt umsetzen trifft voll zu trifft überhaupt nicht zu

3.6 Die betreute Übungszeit im Rahmen des Projekts im Labor ist sinnvoller als die unbetreute Übungszeit (Di/Do 17-20 Uhr) trifft voll zu trifft überhaupt nicht zu

3.7 Ein Peer Teaching im Bezug auf zahntechnische Fertigkeiten sollte in der Vorklinik angeboten werden trifft voll zu trifft überhaupt nicht zu

4. OSCE

4.1 Die OSCE-Erfahrung im Rahmen des Projekts bewerte ich als positiv trifft voll zu trifft überhaupt nicht zu

4.2 Die OSCE-Erfahrung ist eine gute Vorbereitung auf den chirurgischen OSCE am Ende des 8. Semesters trifft voll zu trifft überhaupt nicht zu

F14566U0P1PL0V0
24.01.2019, Seite 1/2

MUSTER

Abbildung A1: Evaluationsbogen

6.2 Item-Checklisten

6.2.1 Station „Dreiecksklammer biegen“

Zur Herstellung einer Dreiecksklammer geht der/die Studierende entsprechend der Lehrsituation wie folgt vor: Er/Sie

Tabelle A1: Checkliste Dreiecksklammer biegen

Item 1	trägt während des gesamten Vorgangs eine Schutzbrille
Item 2	wählt die korrekte Drahtstärke aus
Item 3	knipst den Draht korrekt (ca. 8 cm) ab
Item 4	gummirt die beiden Drahtenden sorgfältig
Item 5	wählt die Flachspitzzange aus
Item 6	setzt die erste Biegung korrekt um
Item 7	setzt die zweite Biegung korrekt um
Item 8	setzt die dritte Biegung korrekt um
Item 9	setzt die vierte Biegung korrekt um
Item 10	handhabt die Zange korrekt
Item 11	zeigt ein metrisch korrektes Ergebnis
Item 12	erreicht eine korrekte Position der Klammer am Modell
Item 13	beendet die Klammerbiegearbeit in der vorgesehenen Zeit
Item 14	legt eine korrekt gebogene, suffiziente Dreiecksklammer vor

6.2.2 Station „Modell eckig trimmen“

Ein zahntechnisches Gipsmodell wird auf die vorher vermittelte Art und Weise am Trimmer bearbeitet. Der/die Studierende:

Tabelle A2: Checkliste Modell eckig trimmen

Item 1	trägt während des gesamten Vorgangs eine Schutzbrille.
Item 2	schaltet den Trimmer ein.
Item 3	schaltet die Wasserkühlung ein.
Item 4	ordnet das Modell dem Trimmvorgang richtig zu.
Item 5	belässt die Zähne unversehrt.

Item 6	schließt die Arbeit am Trimmer in der vorgegebenen Zeitspanne ab.
Item 7	hält die korrekte Reihenfolge der Trimmung ein.
Item 8	trimmt das Modell symmetrisch
Item 9	lässt eine ausreichende Menge vom vestibulären Gips stehen.
Item 10	zeigt durchgehend eine korrekte Hand- und Körperhaltung während der Arbeit am Gerät.
Item 11	produziert ein akzeptables Ergebnis.

6.2.3 Station „Politur“

Ein metallischer Werkstoff soll nach dem eigenen Wissens- und Kenntnisstand auf Hochglanz poliert werden. Der/die Studierende:

Tabelle A3: Checkliste Politur

Item 1	trägt während des gesamten Vorgangs eine Schutzbrille.
Item 2	Spannt die benötigten rotierenden Instrumente selbständig in das Handstück ein
Item 3	1. Instrument (Nr. 5)
Item 4	2. Instrument (Nr. 3)
Item 5	3. Instrument (Nr. 2)
Item 6	4. Instrument (Nr. 4)
Item 7	5. Instrument (Nr.1)
Item 8	handhabt das Handstück korrekt in der vorgegebenen Weise
Item 9	beendet die Politur auf Hochglanzniveau
Item 10	beendet die Arbeit in der vorgegebenen Zeit

6.3 Lineare Regression

Tabelle A4: Lineare Regression Dreiecksklammer biegen

	R ²	Korrigiertes R ²	Regressionskoeffizient B	Signifikanz	Cohen´s f ²
T ₁	0,404	0,387	3,024	0,000	1,855
T ₂	0,178	0,145	1,383	0,028	0,465

Tabelle A5: Lineare Regression Modell eckig trimmen

	R ²	Korrigiertes R ²	Regressionskoeffizient B	Signifikanz	Cohen´s f ²
T ₁	0,119	0,095	0,889	0,000	0,385
T ₂	0,232	0,201	0,650	0,011	0,549

7 Literaturverzeichnis

Albanese MA, Mitchell S (1993): Problem-based learning: a review of literature on its outcomes and implementation issues. *Acad Med* 68, 52-81

Albert BD, Burns JP (2018): Is “See One, Do One, Teach One” Still Relevant in the 21st Century? *Ped Crit Care Med* 19, 678-679

Bardach NS, Vedanthan R, Haber RJ (2003): ‘Teaching to Teach’: enhancing fourth year medical students’ teaching skills. *Med Educ* 37, 1031–1032

Barsuk JH, Cohen ER, Wayne DB, McGaghie WC, Yudkowsky R (2018): A Comparison of Approaches for Mastery Learning Standard Setting. *Acad Med* 93, 1079-1084

Barsuk JH, McGaghie WC, Cohen ER, O’Leary KJ, Wayne DB (2009): Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit. *Crit Care Med* 37, 2697-2701

Blanz M: *Forschungsmethoden und Statistik für die Soziale Arbeit: Grundlagen und Anwendungen*. 2. Auflage; Kohlhammer, Stuttgart 2015

Bligh DA: *What’s the use of lectures?* Jossey-Bass Publishers, 1. Auflage; San Francisco 2000

Bortz J: *Lehrbuch der Statistik für Sozialwissenschaftler*. 2. Auflage; Springer Medizin Verlag, Berlin 1985

Bortz J: *Statistik für Human- und Sozialwissenschaftler*. 6. Auflage; Springer Medizin Verlag, Heidelberg 2005

Bortz J, Schuster C: *Statistik für Human- und Sozialwissenschaftler*. 7. Auflage; Springer Medizin Verlag, Heidelberg 2010

Boud D, Cohen R, Sampson J: *Peer learning in higher education. Learning from & with each other*; Stylus Publishing Inc, Sterling 2001

Boulet J, De Champlain A, McKinley D (2003): Setting defensible performance standards on OSCEs and standardized patient examinations. *Med Teach* 25, 245-249

Broadbent DE (1970): Review lecture. *Proceedings of the Royal Society*, 1, 333-350

Bucknall V, Sobic EM, Wood HL, Howlett SC, Taylor R, Perkins GD (2008): Peer assessment of resuscitation skills. *Resuscitation* 77, 211–215

Burgess A, Clark T, Chapman R, Mellis C (2012): Senior medical students as examiners in the OSCE. *Med Teach* 35, 1–5

Cannick GF, Horowitz AM, Garr DR, Reed SG, Neville BW, Day TA, Woolson RF, Lackland DTI (2007): Use of the OSCE to evaluate brief communication skills training for dental students. *J Dent Educ* 71, 1203-1209

Chenot JF, Ehrhardt M (2003): Objective structured clinical examination (OSCE) in der medizinischen Ausbildung: Eine Alternative zur Klausur. *Z Allg Med* 79, 437–442

Chenot JF, Simmenroth-Nayda A, Koch A, Fischer T, Scherer M, Emmert B, Stanske B, Kochen MM, Himmel W (2007): Can student tutors act as examiners in an objective structured clinical examination? *Med Educ* 41, 1032–1038

Cohen J: *Statistical Power Analysis for the Behavioral Sciences*. 2. Auflage; Lawrence Erlbaum Associates, Hillsdale 1988

Cohen J (1992): A Power Primer. *Psychol Bull* 112, 155-159

Dandavino M, Snell L, Wiseman J (2007): Why medical students should learn how to teach. *Med Teach* 29, 558–565

Ellis P (2010): Effect sizes and the interpretation of research results in international business. *Int Bus Stud* 41, 1581–1588

English R, Brookes ST, Avery K, Blazeby JM, Ben-Shlomo Y (2006): The effectiveness and reliability of peer marking in first year medical students. *Med Educ* 40, 965–972

Euler D: Kompetenzorientiert prüfen - eine hilfreiche Vision? In: E. Severing & R. Weiß: *Prüfungen und Zertifizierungen in der beruflichen Bildung*. Bertelsmann Verlag, Bielefeld 2011, 55-66

Friedmann BD (2000): Standard setting in student assessment. an extended summary of AMEE Medical Education Guide No 18. *Med Teach* 22, 120-130

Harden RM, Gleeson FA (1979): Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 13, 41-54

Hodges B, Regehr G, Hanson M, McNaughton N (1998): Validation of an objective structured clinical examination in psychiatry. *Acad Med* 73, 910-912

Holmboe ES, Hawkins RE, Huot SJ (2004): Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Ann Intern Med* 140, 874-881

Immenroth M, Burger T, Brenner J, Nagelschmidt M, Eberspacher H, Troidl H. (2007): Mental training in surgical education: a randomized controlled trial. *Ann Surg* 245, 385–391

Juszczuk AS, Clark RKF, Radford DR (2009): UK dental laboratory technicians' views on the efficacy and teaching of clinical-laboratory communication. *Brit dent journ* 206, 21

Kalaian SA, Kasim RM (2017): Effectiveness of various innovative learning methods in health science classrooms: a meta-analysis. *Adv Health Sci Educ* 22, 1151–1167

Karpicke JD, Roediger HL (2007): Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *J Exp Psychol Learn Mem Cogn* 33, 704–719

Karpicke JD, Roediger HL (2008): The critical importance of retrieval for learning. *Science* 319, 966–968

Khodaverdi D. (2018): See one, do one, teach one: Is it enough? Yes. *Emerg Med Australas* 30, 107–108

Krautter M, Weyrich P, Schultz JH, Buss SJ, Maatouk I, Jünger J, Nikendei C (2011): Effects of Peyton's four-step approach on objective performance measures in technical skills training: a controlled trial. *Teach Learn Med* 23, 244–250

Krautter M, Dittrich R, Safi A, Maatouk I, Krautter J, Moeltner A, Nikendei C, Herzog W (2015): Peyton's four-step approach on objective performance measures in technical skills training: a controlled trial. *Adv Med Educ Pract* 6, 399–406

Kotsis SV, Chung KC (2013): Application of the "See One, Do One, Teach One" Concept in Surgical Training. *Plast reconstr Surg* 131, 194–201

Kubinger KD: Gütekriterien. In: Westhoff K, Hagemeister C, Kersting M, Lang F, Moosbrugger H, Reimann G, Stemmler G: *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430*. 3. Auflage; Pabst Science Publishers, Lengerich 2010, 8–10

Lammerding-Köppel M, Fabry G, Hofer M, Ochsendorf F, Schirlo C (2006): Hochschuldidaktische Anforderung in der Medizin: II. Anforderungsprofil der Qualifizierungsangebote. *GMS Z Med Ausbild* 23, 72

Larsen T, Jeppe-Jensen D (2008): The introduction and perception of an OSCE with an element of self- and peer-assessment. *Eur J Dent Educ* 12, 2–7

Larsen DP, Butler AC, Roediger HL (2013): Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Med Educ* 47, 674–682

Lee CB, Madrazo L, Khan U, Thangarasa T, McConnell M, Khamisa K (2018): A student-initiated objective structured clinical examination as a sustainable cost-effective learning experience. *Med Educ Online* 23, 1440111

Lienert GA: Testaufbau und Testanalyse. 4. Auflage; PVU, München 1989

Maheady L (1998): Advantages and disadvantages of peer-assisted learning strategies. In: Topping K. Mahwah NJ: Peer-Assisted Learning. Lawrence Erlbaum Associates, New York 1998, 45–62

Manogue M, Brown G (1998): Developing and implementing an OSCE in dentistry. *Eur J Dent Educ* 2, 51-57

McQueen W (1994): Improving graduate students performance in cognitive assessment. The saga continues. *Prof Psychol Res Pract* 25, 283-287

Miller GE (1990): The assessment of clinical skills/competence/performance. *Acad Med* 65, 63-67

Miller W & MacGilchrist L (1996): A model for peer-led work. *Health Educa* 96, 24-29

Möltner A, Schellberg D, Jünger J (2006): Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Z Med Ausbild* 23, 53-63

Moosbrugger H, Kelava A: Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In: Moosbrugger H, Kelava A: Testtheorie und Fragebogenkonstruktion. Springer, Berlin 2012, 7-26

Müller S, Dahmen U, Settmacher U: Objective Structured Clinical Examination (OSCE) an Medizinischen Fakultäten in Deutschland - eine Bestandsaufnahme. *Das Gesundheitswesen*. Thieme Verlag, Stuttgart 2016

Münster T, Stosch C, Hindrichs N, Franklin J, Matthes J (2016): Peyton's 4-Steps-Approach in comparison: Medium-term effects on learning external chest compression – a pilot study. *GMS J Med Educ* 33, 60

Nestel D, Kidd J (2005): Peer assisted learning in patient-centred interviewing: the impact on student tutors. *Med Teach* 27, 439–444

Newble D (2004): Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ* 38, 199–203

Nikendei C, Jünger J (2006): OSCE - praktische Tipps zur Implementierung einer klinisch-praktischen Prüfung. *GMS Z Med Ausbild* 23, 47-54

Pagano, RR.: Understanding statistics in the behavioral science. 9. Auflage; Thomson Wadsworth, Belmont 2010

Pell G, Homer MS, Roberts TE (2008): Assessor training: its effects on criterion-based assessment in a medical context. *Int J Res Meth Educ* 31, 143-154

Perkins GD, Hulme J, Bion JF (2002): Peer led resuscitation training for healthcare students: a randomised controlled study. *Intensive Care Med* 28, 698–700

Peyton JWR: Teaching and Learning in Medical Practice. Manticore Books, Heronsgate 1988

Prabhu V, Sahoo S, Soe HHK (2015): Effectiveness of skeleton handouts during ophthalmology theory lectures of undergraduate medical students. *Int J App Basic Med Res* 5, 29-31

Rasch B, Malte F, Wilhelm H, Naumann E: Quantitative Methoden 1.Einführung in die Statistik für Psychologen und Sozialwissenschaftler. Band 1, 4 Auflage; Springer Verlag, Berlin 2014

Ratzmann A, Wiesmann U, Kordaß B (2012): Integration einer OSCE in das zahnmedizinische Physikum. *GMS Z Med Ausbild* 29, 9-15

Raupach T, Brown J, Anders S, Hasenfuss G, Harendza S (2013): Summative assessments are more powerful drivers of student learning than resource intensive teaching formats. *BMC Med* 11, 61-70

Raupach T, Andresen JC, Meyer K, Strobel L, Koziolk M, Jung W, Brown J, Anders S (2016): Test-enhanced learning of clinical reasoning: a crossover randomised trial. *Med Educ* 50, 711-720

Reid K, Smallwood D, Collins M, Sutherland R, Dodds A (2016): Taking OSCE examiner training on the road: reaching the masses. *Med Educ Online* 28, 32389

Reiter HI, Rosenfeld J, Nandagopal K, Eva KW (2004): Do clinical clerks provide candidates with adequate formative assessment during objective structured clinical examinations? *Adv Health Sci Educ* 9, 189–199

Rodriguez-Paz JM, Kennedy M, Salas E, Wu AW, Sexton JB, Hunt EA, Pronovost PJ (2009): Beyond „see one, do one, teach one“: toward a different training paradigm. *BMJ Quality & Safety* 18, 63-68

Roediger HL, Karpicke JD (2006): Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci* 17, 249–255

Ross MT, Cameron HS (2007): Peer assisted learning: a planning and implementation framework: AMEE Guide no. 30. *Med Teach* 29, 527–545

Roth M, Decety J, Raybaudi M, Massarelli R, Delon-Martin C, Segebarth C, Morand S, Gemignani A, Décorps M, Jeannerod M (1996): Possible involvement of primary motor cortex in mentally simulated movement: a functional magnetic resonance imaging study. *Neuroreport* 7, 1280-1284

Rowland CA (2014): The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol Bull* 140, 1432-1463

Ruesseler M, Weinlich M, Byhahn C, Müller MP, Jünger J, Marzi I, Walcher F (2010): Increasing authenticity in practical assessment using emergency case OSCE stations. *Adv Health Sci Educ Theory Pract* 15, 81-95

Salkind NJ: *Encyclopedia of Research Design*, 2. Auflage; Sage, Los Angeles 2010

Schelten A: *Grundlagen der Testbeurteilung und Testerstellung. Teststatistik und Testtheorie für Pädagogen und Ausbilder in der Praxis*. Quelle & Meyer, Heidelberg 1980

Schoonheim-Klein ME, Habets LL, Aartman IH, van der Vleuten CP, Hoogstraten J, van der Velden U (2006): Implementing an Objective Structured Clinical Examination (OSCE) in dental education: effects on students' learning strategies. *Eur J Dent Educ* 10, 226-235

Schrauth M, Riessen R, Wirtz TSH (2005): Praktische Prüfungen sind machbar. *Med Ausbild* 22, 20-22

Schwarzer SKE: *Die Etablierung und Evaluation einer zahnärztlich-chirurgischen OSCE-Prüfung mit sechs Stationen in der ZMK-Klinik Göttingen - die Bewertung ärztlicher und studentischer Rater im Vergleich*. Med. Diss. Göttingen 2017

Seifert LB, Schnurr B, Stefanescu MC, Sader R, Ruesseler M, Sterz J (2020): Comparing video-based versions of Halsted's 'see one, do one' and Peyton's '4-step approach' for teaching surgical skills: a randomized controlled trial. *BMC Med Educ* 20, 194

Sennhenn-Kirchner S, Goerlich Y, Kirchner B, Notbohm M, Schiekirka S, Simmenroth A, Raupach T (2017): The effect of repeated testing vs repeated practice on skills learning in undergraduate dental education. *Eur J Dent Educ* 22, 42–47

Simmenroth-Nayda A, Görlich Y, Wagner M, Müther M, Lohse C, Utte L, Leiterholt S, Hoerauf H, Feltgen N (2014): Studentische Lehre in der Augenheilkunde: Sind standardisierte praktische Prüfungen sinnvoll? *Ophthalmol* 111, 235–240

Speirs C, Brazil V. (2018): See one, do one, teach one: Is it enough? No. *Emerg Med Australas.* 30, 109-110

Ten Cate O, Durning S (2007): Dimensions and psychology of peer teaching in medical education. *Med Teach* 26, 546–552

Topping KJ (1996): The effectiveness of peer tutoring in further and higher education: a typology and review of the literature. *High Educ* 32, 321–345

Topping KJ (1998): Peer assessment between students in college and university. *Rev Educ Res* 68, 249–276

Topping KJ (2007): Trends in peer learning. *Educ psych* 25, 631-645

Widulle W: Handlungsorientiert Lernen im Studium. Arbeitsbuch für soziale und pädagogische Berufe. VS, Wiesbaden 2009

Wilcox RR.: Introduction to robust estimation and hypothesis testing. Statistical modeling and decision science. 3. Auflage; Academic Press, Amsterdam 2012

Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T (2003): objectivity in objective structured clinical examinations: checklists are no substitute for examiner commitment. *Acad Med* 78, 219-223

Wissenschaftsrat (2005): Empfehlungen zur Weiterentwicklung der Zahnmedizin an den Universitäten in Deutschland (Zuletzt gesehen: 02. August 2020)

ZApprO 2019: Approbationsordnung für Zahnärzte und Zahnärztinnen vom 8. Juli 2019 in der Fassung der Bekanntmachung vom 11. Juli 2019, zuletzt geändert am 19. Mai 2020

Zierer K: Hattie für gestresste Lehrer; Kernbotschaften und Handlungsempfehlungen aus John Hatties „Visible Learning“ und „Visible Learning for Teachers“. Schneider Verlag Hohengrehren, Baltmannsweiler 2015

Internetquellen:

FVDZ – Pressemeldungen – 07.06.2019; <https://www.fvdz.de/nachrichten-detailansicht/bundesrat-verabschiedet-zahnaerztliche-approbationsordnung-8282>;

Zugriff am 28.09.2020

Zahnmedizin Uni Münster – Service – Studium – 4. Semester <https://zahnmedizin-ms.de/studium/4-semester/phantom-i/>; Zugriff am 01.10.2020

Danksagung

An dieser Stelle möchte ich mich herzlich bei meiner Doktormutter Frau Dr. med. dent. Sabine Sennhenn-Kirchner für die herzliche Betreuung und professionelle Unterstützung während der gesamten Bearbeitungsphase bedanken. Durch ihr Engagement zur Weiterentwicklung der Lehre konnte dieses Pilotprojekt realisiert werden.

Darüber hinaus gilt mein Dank Marcel Nothbom und allen weiteren Mitarbeitern des Instituts für Medizinische Informatik der UMG für die Bereitstellung des technischen Equipments sowie der technischen Unterstützung.

Besonders bedanken möchte ich mich bei Theresa Seifert aus dem Bereich der Medizindidaktik und Ausbildungsforschung der UMG für ihre Hilfsbereitschaft und Unterstützung bei der Datenauswertung.

Insbesondere möchte ich mich bei den studentischen Ratern bedanken, die sich Zeit genommen und somit dieses Projekt unterstützt haben.

Der größte Dank gilt allerdings den Studierenden des Wintersemesters 2018/2019. Ohne die freiwillige Teilnahme der Studierenden wäre dieses Projekt nicht machbar gewesen.