

Aus der Klinik für Kardiologie und Pneumologie
(Prof. Dr. med. G. Hasenfuß)
der Medizinischen Fakultät der Universität Göttingen

**Wie entwickelt sich der studentische
Leistungsstand hinsichtlich des
Examensniveaus während der Teilnahme
an semesterbegleitenden formativen
Prüfungen?**

INAUGURAL-DISSERTATION

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

vorgelegt von

Julia Maria Sigrid Henninger

aus

Kassel

Göttingen 2021

Dekan: Prof. Dr. med. W. Brück

Betreuungsausschuss

Betreuer/in Prof. Dr. med T. Raupach, MME

Ko-Betreuer/in: Prof. Dr. med A. Schaper

Prüfungskommission

Referent/in Prof. Dr. med T. Raupach, MME

Ko-Referent/in:

Drittreferent/in:

Datum der mündlichen Prüfung:

Hiermit erkläre ich, die Dissertation mit dem Titel "Wie entwickelt sich der studentische Leistungsstand hinsichtlich des Examensniveaus während der Teilnahme an semesterbegleitenden formativen Prüfungen?" eigenständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Göttingen, den

(Unterschrift)

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
Abkürzungsverzeichnis	V
1 Einleitung	1
1.1 Der Aufbau des Medizinstudiums	1
1.1.1 Lehrplanentwicklung.....	2
1.1.2 Geplante Veränderungen im Medizinstudium.....	3
1.2 Moderne Lehr- und Lernangebote	6
1.2.1 Möglichkeiten an der Universität.....	6
1.2.2 Möglichkeiten des Selbststudiums	7
1.2.3 Lernerfolg durch Digitalisierung.....	8
1.2.4 Aktueller Bedarf an Digitalisierung.....	8
1.3 Motivation	8
1.4 Prüfungen	10
1.4.1 Prüfungsformate.....	10
1.4.2 Fragenformate.....	11
1.4.3 Lernfortschritt durch Prüfungen.....	13
1.5 Zielstellungen und Hypothesen der Arbeit	15
2 Material und Methoden	17
2.1 Studiendesign	17
2.2 Studiendurchführung	17
2.2.1 Übersicht des zeitlichen Ablaufs und des Aufbaus der Studie.....	17
2.2.2 Rekrutierung der Teilnehmer und Datensammlung	21
2.3 Erhebungsinstrumente.....	21
2.3.1 Quantitative Datenanalyse der Fragenpakete von ILIAS mit Excel und SPSS®.....	21
2.3.2 Qualitative Datenanalyse der EvaSys®-Fragebögen am Ende des Semesters.....	22
2.3.3 Qualitative Datenanalyse der Fokusgruppeninterviews	25
2.4 Datenauswertung.....	26
2.4.1 Statistische Analyse	26
2.4.2 Qualitative Analyse.....	30
3 Ergebnisse	31
3.1 Charakterisierung der Studiengruppe	31
3.2 Entwicklung der Teilnehmerzahlen.....	31
3.3 Vergleich des Lernerfolgs zwischen mehrfach wiederholten und neuen Inhalten.....	32
3.4 Entwicklung des Leistungsstandes bei mehrfacher Wiederholung.....	33

3.4.1	Allgemein	34
3.4.2	Leistung in den Interventionsitems zu den Zeitpunkten T1, T2 und T3	36
3.5	Sensitivitätsanalyse.....	39
3.6	Beurteilung des Projekts <i>#clue</i> durch die Studierenden	41
3.6.1	Ergebnisse der Fragebogen-Erhebung.....	41
3.6.2	Ergebnisse der Fokusgruppeninterviews	44
4	Diskussion	55
4.1	Die wesentlichen Ergebnisse	55
4.2	Vergleich des Lernerfolgs zwischen mehrfach wiederholten und neuen Inhalten (Studienfrage eins)	56
4.3	Entwicklung des Leistungsstandes bei mehrfacher Wiederholung (Studienfrage zwei)	58
4.4	Entwicklung der Teilnehmerzahl (Studienfrage drei)	59
4.4.1	Objektive Gründe für die Abnahme der Teilnehmerzahl.....	59
4.4.2	Subjektive Gründe für die Abnahme der Teilnehmerzahl.....	60
4.5	Beurteilung des Projektes <i>#clue</i> (Studienfrage vier).....	62
4.5.1	Qualität des <i>E-Learning</i> -Angebots	63
4.5.2	Emotionen und Lernmotivation	63
4.5.3	Lernverhalten	64
4.5.4	Subjektiver Lernerfolg	64
4.5.5	Zufriedenheit.....	64
4.6	Limitationen der Studie	65
4.6.1	Studiendesign	65
4.6.2	Limitationen durch die Erhebungsinstrumente	67
4.7	Weiterentwicklung von <i>#clue</i> und Ausblick.....	68
5	Zusammenfassung.....	70
6	Anhang.....	72
7	Literaturverzeichnis	88

Abbildungsverzeichnis

Abbildung 1: Übersicht über den klinischen Abschnitt des Medizinstudiums an der UMG.....	2
Abbildung 2: Vorgeschlagenes Strukturmodell des Wissenschaftsrats für das künftige Medizinstudium.....	4
Abbildung 3: Projektverlauf.....	18
Abbildung 4: Verteilung der Interventionsitems.....	19
Abbildung 5: Fragebogen – Persönliche Angaben.....	22
Abbildung 6: Fragebogen – Teilnahme am Projekt.....	22
Abbildung 7: Fragebogen – Wöchentliche Fragenpakete.....	23
Abbildung 8: Fragebogen – Interdisziplinäre Lehrveranstaltungen.....	24
Abbildung 9: Fragebogen – Gesamturteil.....	24
Abbildung 10: Berechnung der Summenscores eines Beispiel-Studienteilnehmers.....	27
Abbildung 11: Teilnehmerzahlen.....	32
Abbildung 12: Durchschnittlich erreichter Wochenscore der Studienteilnehmer (semestergetrennt).....	34
Abbildung 13: Durchschnittlich erreichter Wochenscore der Studienteilnehmer (über alle Semester hinweg).....	35
Abbildung 14: Leistungsunterschiede der primären Analyse (unter Einbeziehung aller Schwierigkeitsgrade).....	36
Abbildung 15: Leistungsunterschiede der Interventions- und Kontrollitems in der primären Analyse (aufgeteilt nach Schwierigkeitsgrad).....	38
Abbildung 16: Leistungsunterschiede in der Sensitivitätsanalyse (unter Einbeziehung aller Schwierigkeitsgrade).....	40
Abbildung 17: Leistungsunterschiede der Interventions- und Kontrollitems in der Sensitivitätsanalyse (aufgeteilt nach Schwierigkeitsgrad).....	40

Tabellenverzeichnis

Tabelle 1: Organisation der Fokusgruppen	26
Tabelle 2: Ergebnisse der Fragebögen hinsichtlich der wöchentlichen Fragenpakete.....	42
Tabelle 3: Ergebnisse der Fragebögen hinsichtlich der interdisziplinären Lehrveranstaltungen	43
Tabelle 4: Ergebnisse der Fragebögen hinsichtlich des Gesamturteils	43
Tabelle 5: Auswertung der Fokusgruppengespräche – Allgemeines (Teil eins)	45
Tabelle 6: Auswertung der Fokusgruppengespräche – Allgemeines (Teil zwei)	47
Tabelle 7: Auswertung der Fokusgruppengespräche – Allgemeines (Teil drei).....	48
Tabelle 8: Auswertung der Fokusgruppengespräche – Fragenpakete (Teil eins)	50
Tabelle 9: Auswertung der Fokusgruppengespräche – Fragenpakete (Teil zwei)	51
Tabelle 10: Auswertung der Fokusgruppengespräche – Interdisziplinäre Lehrveranstaltungen	52
Tabelle 11: Auswertung der Fokusgruppengespräche – Verbesserungsvorschläge	54

Abkürzungsverzeichnis

# <i>clue</i>	<i>cross-linked undergraduate education</i>
ÄApprO	Approbationsordnung für Ärzte
DSGVO	Datenschutz-Grundverordnung
<i>E-Learning</i>	<i>electronical learning</i> = elektronisch unterstütztes Lernen
ILIAS	integriertes Lern-, Informations- und Arbeitskooperations-System
IMPP	Institut für Medizinische und Pharmazeutische Prüfungsfragen
MC.....	<i>multiple choice</i> = Mehrfachauswahl
NKLM.....	Nationaler Kompetenzbasierter Lernzielkatalog Medizin
PJ	Praktisches Jahr
T1	Zeitpunkt eins
T2	Zeitpunkt zwei
T3	Zeitpunkt drei
UMG.....	Universitätsmedizin Göttingen

1 Einleitung

1.1 Der Aufbau des Medizinstudiums

Die Zusammensetzung der ärztlichen Ausbildung ist in der Approbationsordnung für Ärzte¹ (ÄApprO) vom 27. Juni 2002 bundeseinheitlich geregelt. Güntert et al. (2003) beschreiben, dass für das Medizinstudium in Deutschland eine Dauer von sechs Jahren und drei Monaten vorgesehen ist. Das Medizinstudium an der Universitätsmedizin Göttingen (UMG) gliedert sich in einen aus vier Semestern bestehenden vorklinischen und einen aus sechs Semestern bestehenden klinischen Abschnitt, gefolgt vom praktischen Jahr. Die ärztliche Prüfung gliedert sich in drei Prüfungsabschnitte (M1, M2 und M3), welche an unterschiedlichen Zeitpunkten im Studium absolviert werden. Die M1-Prüfung wird nach zwei Jahren im Anschluss an den sog. vorklinischen Studienteil schriftlich und mündlich-praktisch abgelegt. In dem vorklinischen Studienteil werden den Studierenden vor allem naturwissenschaftliche Grundlagen z. B. in den Fächern Anatomie, Physiologie oder Biochemie vermittelt. Zusätzlich ist ein Wahlfach zu belegen, sowie ein Pflegepraktikum und eine Erste-Hilfe-Ausbildung zu absolvieren. Das sich anschließende klinische Curriculum umfasst 21 Module mit einer Dauer zwischen zwei und sieben Wochen, ein Wahlfach, fünf Blockpraktika und 14 Querschnittsfächer (siehe Abbildung 1), in denen den Studierenden klinisch-praktische Kenntnisse und klinisch-theoretisches Wissen vermittelt werden sollen. Der Unterricht wird in Form von Vorlesungen, Seminaren und praktischen Übungen abgehalten. Hierbei behandelt jedes Modul einen thematischen Schwerpunkt. In den ersten beiden klinischen Semestern werden den Studierenden klinisch-praktische und klinisch-theoretische Basisfertigkeiten und Grundlagen vermittelt. Das dritte bis fünfte klinische Semester beinhaltet jeweils Module, in denen die spezifische Krankheitslehre besprochen wird. Am Ende jedes Moduls werden die erlernten Kenntnisse durch eine Klausur, die meistens aus *Multiple-Choice*(MC)-Fragen besteht, abgefragt. Um die in der ÄApprO geforderten Leistungsnachweise zu erhalten, werden in jedem Modul eine bestimmte Anzahl an Leistungspunkten vergeben. Es können pro Fachbereich maximal 100 Leistungspunkte erzielt werden, welche auf die verschiedenen Module eines Fachbereichs verteilt sind und beispielsweise durch das Ablegen einer MC-Prüfung vergeben werden. In den Semesterferien müssen die Studierenden Blockpraktika und eine viermonatige Famulatur ableisten. Im sechsten und letzten Semester vor dem schriftlichen M2-Staatsexamen werden die zuvor vermittelten Fachgebiete vertieft. Hiermit soll den Studierenden differenzialdiagnostisches

¹ Aus Gründen der besseren Lesbarkeit beschränkt sich diese Arbeit auf die Verwendung männlicher oder neutraler Substantive bzw. Pronomina. Diese beziehen sich damit jedoch grundsätzlich in gleichem Maße auch auf alle Geschlechter und implizieren keine Benachteiligungen.

Denken vermittelt und anschließend gefördert werden. Im Anschluss daran absolvieren die Studierenden ein Praktisches Jahr (PJ), das sich aus drei Tertialen mit einer Dauer von je 16 Wochen zusammensetzt. Nach dessen Ende folgt die dritte mündlich-praktische M3-Prüfung. In der Ausbildung sollen unter anderem „praktische Erfahrungen im Umgang mit Patienten, einschließlich der fächerübergreifenden Betrachtungsweise von Krankheiten und der Fähigkeit, die Behandlung zu koordinieren“ (Approbationsordnung für Ärzte 2002), vermittelt werden (Güntert et al. 2003; Deutscher Wissenschaftsrat 2014).

Klinisches Curriculum SoSe 2018 (09.04.2018 - 13.07.2018) - unter Vorbehalt, Änderungen möglich!

Wo 1	Wo 2	Wo 3	Wo 4	Wo 5	Wo 6	Wo 7	Wo 8	Wo 9	Wo 10	Wo 11	Wo 12	Wo 13	Wo 14
09.04.	16.04.	23.04.	30.04.	07.05.	14.05.	21.05.	28.05.	04.06.	11.06.	18.06.	25.06.	02.07.	09.07.
13.04.	20.04.	27.04.	04.05.	11.05.	18.05.	28.05.	01.06.	08.06.	15.06.	22.06.	29.06.	06.07.	13.07.

Grundlagen zu Krankheitslehre und Diagnostik

1. klinisches Semester

M1.1 Ärztliche Basisfertigkeiten I		M1.2 Krankheitslehre/Diagnostik		M1.3 Bildgebende Verfahren		M1.4 Gesundheitssystem/-verfahren							
1	2	3	4	5	6	7	8	9	10	11	12	13	14

2. klinisches Semester

M 2.1 Pharmakologie/Toxikologie (A)		M 2.2 Infektion/Abwehr (A)		M 2.3 UaK		M 2.4 Evidenzbasierte Medizin (A)		OSCE I Ärztl. Basisfertigkeiten					
M 2.2 Infektion/Abwehr (B)		Basisfertigkeiten II (A+B)		M 2.1 Pharmakologie/Toxikologie (B)		M 2.3 U-Kurs							
M 2.4 Evidenzbasierte Medizin (B)		M 2.3 Ärztliche Basisfertigkeiten II (A+B)		M 2.3 UaK		M 2.3 U-Kurs							
1	2	3	4	5	6	7	8		9	10	11	12	13

Module zur Spezifischen Krankheitslehre

3. klinisches Semester

M 3.1 Herz - Kreislauf und Lunge		M 3.2 Urologie/Niere		M 3.3 Tumorerkrankungen/Blut									
1	2	3	4	5	6	7	8	9	10	11	12	13	14

Blockpraktika finden über die gesamten Semesterferien statt

4. klinisches Semester

M 4.1 Haut/Systemerkr.		M 4.2 Bewegungsorgane und Trauma		M 4.3 Verdauungsorgane und Endokrinologie		M 4.4 Auge/HNO/ZMK							
1	2	3	4	5	6	7	8	9	10	11	12	13	14

Blockpraktikum in der UMG/Student
2 Wochen Chirurgie
2 Wochen Innere Medizin
2 Wochen Allgemeinmedizin

5. klinisches Semester

M 5.1 Nervensystem und Psyche		M 5.2 Kinder/Jugend		M 5.3 Fortpflanzung		M 5.4 Allgemeinmedizin und Infektiologie							
1	2	3	4	5	6	7	8	9	10	11	12	13	14

Blockpraktikum in der UMG/Student
1 Woche Frauenheilkunde
1 Woche Kinderheilkunde

Vertiefungsmodule zur Differentialdiagnose und -therapie

6. klinisches Semester

M 6.x Repetitoriumsmodule		OSCE II		M 6.2 Notfall- u. Intensivmedizin									
1	2	3	4	5	6	7	8	9	10	11	12	13	14

Eintritt in das PJ
nach dem SoSe Mitte November,
nach dem WiSe Mitte Mai

Vorlesungsfreie Tage:

Tag der Arbeit: 01.05.2018
Christi Himmelfahrt: 10.05.2018
Pfingstmontag: 21.05.2018
Sport Dies (ab 14 Uhr): 30.05.2018

Abbildung 1: Übersicht über den klinischen Abschnitt des Medizinstudiums an der UMG. Dargestellt ist das klinische Curriculum des Sommersemesters 2018. Die Lehre der UMG ist in 21 Module gegliedert, die jeweils einen klinischen Schwerpunkt behandeln. Die im Rahmen dieser Studie vorgestellte Lehrinnovation wurde dem dritten bis sechsten klinischen Semester zur Verfügung gestellt. Die Verwendung der Abbildung erfolgt mit freundlicher Genehmigung der Universitätsmedizin der Georg-August-Universität Göttingen, vertreten durch Prof. Dr. med. Martin Oppermann (stellv. Studiendekan).

1.1.1 Lehrplanentwicklung

„Ziel der ärztlichen Ausbildung ist der wissenschaftlich und praktisch in der Medizin ausgebildete Arzt, der zur eigenverantwortlichen und selbstständigen ärztlichen Berufsausübung, zur Weiterbildung und zu ständiger Fortbildung befähigt ist. Die Ausbildung soll grundlegende Kenntnisse, Fähigkeiten und Fertigkeiten in allen Fächern vermitteln, die für eine umfassende Gesundheitsversorgung der Bevölkerung erforderlich sind“ (Approbationsordnung für Ärzte 2002).

Hierfür ist ein systematisches Vorgehen notwendig, das sich auf das Gesamtsystem beziehen muss (Biggs 1996). Thomas et al. (2016) beschreiben in ihrem Buch „*Curriculum development for medical education*“ mit sechs Stufen die Möglichkeit für ein systematisches Vorgehen bei der Entwicklung eines Lehrplans. Die erste Stufe beschreibt die Ausgangslage. So müssen zunächst allgemeine Probleme z. B. in der Gesundheitsversorgung erkannt und der optimale Zustand benannt werden. Anschließend muss das Problem genauer definiert werden (zweite Stufe), um daran anknüpfend in der dritten Stufe Ziele und Aufgaben in dem Lehrplan benennen zu können (Thomas et al. 2016). Nach Biggs (1996) sollen Ziele in Form von verschiedenen Kompetenzebenen formuliert werden. Kompetenzen sind

„[...] die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösung in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert 2002).

In der vierten Stufe sollen nun die konkreten Bildungsmethoden festgelegt werden, um die genannten Ziele zu erreichen (Thomas et al. 2016). Hierbei ist es wichtig, dass sich die Prüfungsaufgaben auf die zu vermittelnden Kompetenzen beziehen, welche vorab durch geeignete Lehrmethoden erlernt werden konnten, sog. *constructive alignment* (Biggs 1996). Letztlich kommt es zur Durchführung (Stufe fünf) und Evaluation des Lehrplans (Stufe sechs). Nicht zu vergessen ist, dass die Lehrplanentwicklung nie endet, sondern abhängig von Evaluationsergebnissen und vorhandenen Ressourcen einer stetigen Weiterentwicklung unterliegt (Thomas et al. 2016).

Durch die ÄApprO wird den Universitäten eine Grundlage für die Umsetzung des Curriculums für das Medizinstudium vorgegeben (Deutscher Wissenschaftsrat 2018). 2015 wurde der Nationale Kompetenzbasierte Lernzielkatalog Medizin (NKLM) von Delegierten des Medizinischen Fakultätentags der Bundesrepublik Deutschland e. V. und der Gesellschaft für Medizinische Ausbildung e. V. verabschiedet. Dieser definiert fächerübergreifend, jedoch nicht verbindlich, Kompetenzen, Teilkompetenzen und Lernziele für das Medizinstudium. Seit dem Wintersemester 2017/18 ist die Lehre an der UMG nach dem NKLM ausgerichtet (Göttinger Lernzielkatalog 2017).

1.1.2 Geplante Veränderungen im Medizinstudium

Seit 2012 beschäftigt sich der Wissenschaftsrat mit bestehenden Modellstudiengängen in Deutschland und evaluiert jene. Auf dieser Grundlage und weiteren internationalen Erfahrungen sind im Hinblick auf die Weiterentwicklung des Medizinstudiums in Deutschland am 11. Juli 2014 Empfehlungen verabschiedet worden. Der Beruf als Arzt steht unter einer stetigen Weiterentwicklung hinsichtlich der komplexer werdenden Versorgungssituationen, des demografischen Wandels und des medizinischen Fortschritts (Deutscher Wissenschaftsrat 2014). Nach Kuhn et al. (2018) wird durch die Digitalisierung des

Gesundheitssystem das Berufsbild des Arztes im Hinblick auf Information, Kommunikation und Zusammenarbeit verändert. Für diese Entwicklungen sind neue Kompetenzen notwendig, die in Curricula festgehalten sein müssen (The Digital Turn 2016). Somit wird eine Anpassung des Medizinstudiums als unvermeidlich betrachtet (siehe Abbildung 2).

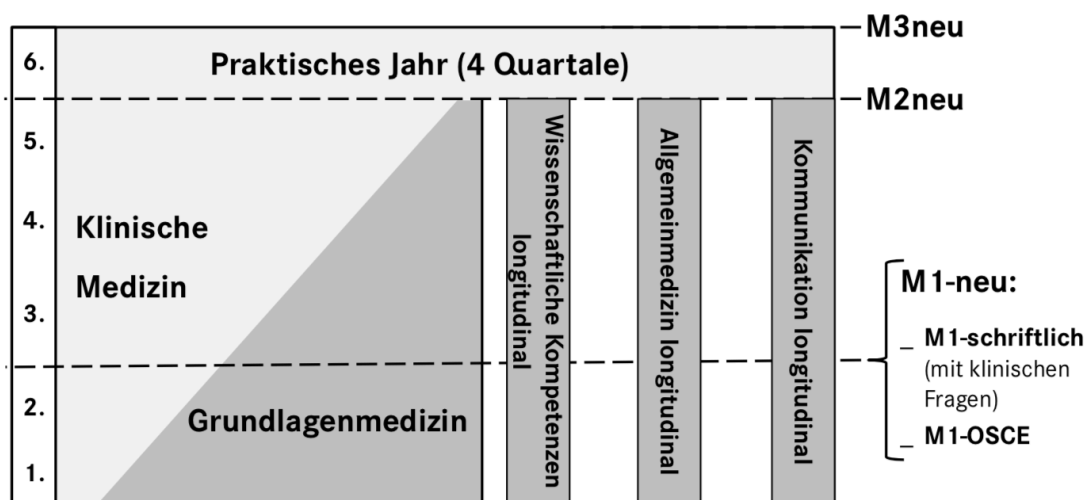


Abbildung 2: Vorgeschlagenes Strukturmodell des Wissenschaftsrats für das künftige Medizinstudium. Dargestellt ist die vorgeschlagene neue Strukturierung des Medizinstudiums. Wissenschaftliche Kompetenzen sollen über mehrere Semester verteilt vermittelt werden. Die Vermittlung der wissenschaftlichen Kompetenzen und der parallel stattfindenden vertikalen Integration (in der Abbildung in unterschiedlichen Grautönen verdeutlicht) verursachen eine neue Strukturierung der M1-Prüfung in zwei Phasen (nach vier Semestern soll eine schriftliche Prüfung stattfinden und eine mündlich-praktische Prüfung). Das PJ soll in vier Ausbildungsabschnitte zu je 12 Wochen aufgeteilt werden. Die Verwendung der Abbildung erfolgt mit freundlicher Genehmigung vom Wissenschaftsrat (Neustrukturierung des Medizinstudiums und Änderung der Approbationsordnung für Ärzte – Empfehlungen der Expertenkommission zum Masterplan Medizinstudium 2020, 2018, S. 60).

Die verfassten Grundsätze beschäftigen sich vor allem mit der Vermittlung von wissenschaftlichen Kompetenzen, der patientenorientierten Ausbildung und fächerübergreifenden Curricula. Gefordert wird zum einen ein Verfahrenswechsel von reiner Wissensvermittlung der Fächer hin zu einer kompetenzvermittelten Ausbildung, die den Ansprüchen an das Berufsbild eines Arztes gerecht wird (Deutscher Wissenschaftsrat 2014). Kompetenzen sollen mit Blick auf Wissenschaft, Kommunikation und Interprofessionalität vermittelt werden (Deutscher Wissenschaftsrat 2018). Zum anderen wird eine horizontale und eine bidirektionale vertikale Integration von vorklinischen und klinischen Lehrinhalten verlangt. Mit einer bidirektionalen vertikalen Integration ist die Verknüpfung von Lehrinhalten zwischen dem vorklinischen und klinischen Studienabschnitt gemeint und mit horizontaler Integration wird ein fächerübergreifender klinischer Unterricht gefordert (Deutscher Wissenschaftsrat 2014). So wurde beispielsweise in Schweden beobachtet, dass eine vertikale Integration ein besseres Verständnis biomedizinischer Probleme und dadurch auch ein tiefergreifenderes Lernen anregt. Dadurch sind Studierende besser auf neue

Situationen, Fragen und Herausforderungen vorbereitet. Es wird vermutet, dass durch die vertikale Integration eine bessere Retention von Wissen entsteht und Ärzte hervorgebracht werden, die gut auf die Veränderungen im Gesundheitswesen vorbereitet sind (Dahle et al. 2002). In einer anderen Studie aus Indien wiederum wurde bei der Befragung von Studierenden deutlich, dass im Gegensatz zur regulären, didaktischen Methodik horizontal ausgerichteter Unterricht mehr geschätzt und stärker nachgefragt wird. Vor allem führt die horizontale Integration zu einem besseren Verständnis der behandelten Themen (Kolhe et al. 2018). Die Inhalte des Kerncurriculums sollen nun in Deutschland zentral verbindlich durch den NKLM vorgegeben werden (Deutscher Wissenschaftsrat 2018). Zusätzlich sind Prüfungen im MC-Format ein reines Abfragen von Faktenwissen, weshalb zukünftige M1- und M2-Prüfungen auf die Überprüfung von Wissen, Fertigkeiten und Haltungen ausgerichtet sein sollen. Auch die neue Ausrichtung der Lehrinhalte hinsichtlich der vertikalen und horizontalen Integration soll in den M1- und M2-Prüfungen widerspiegelt werden (Deutscher Wissenschaftsrat 2014). Die schriftlichen Prüfungsfragen werden vom Institut für Medizinische und Pharmazeutische Prüfungsfragen (IMPP) erstellt (Deutscher Wissenschaftsrat 2018). Lehrinhalte sollen nicht einzeln geprüft werden, vielmehr soll es zu einer stärkeren Anwendung der erlernten Fähigkeiten in der Prüfung kommen (Deutscher Wissenschaftsrat 2014). Es sind Veränderungen hinsichtlich der Zulassung für das Medizinstudium festgehalten worden, z. B. sollen unter den Bewerbern diejenigen zum Studium im Auswahlverfahren zugelassen werden, die besondere soziale und kommunikative Kompetenzen vorweisen können oder eine medizinische Ausbildung absolviert haben. Zusätzlich sind Maßnahmen entwickelt worden, um in ländlichen Regionen, die ärztliche Versorgung sicherstellen zu können. Darüberhinausgehend sind Erneuerungen zur Struktur des Studiums, den Inhalten der Ausbildung und der Interdisziplinarität während des Studiums vorgesehen. Im „Masterplan Medizinstudium 2020“ wird ebenfalls deutlich gemacht, dass eine reine Wissensvermittlung nicht ausreichend ist, um Studierenden eine angemessene ärztliche Qualifikation zu vermitteln. So wurde auch hier das fächerübergreifende Denken als Voraussetzung betrachtet, um das Zusammenspiel von Gesundheit und Krankheit begreifen zu können. Die Neuerungen in der Lehre sollen sich auch durch den Einsatz von sog. *Key-Feature-Fällen* (siehe Kapitel 1.4.2) in den Prüfungen widerspiegeln (Bundesministerium für Gesundheit et al. 2017). Zusätzlich wurde eine Expertenkommission zusammengestellt, die in ihrer Publikation „Neustrukturierung des Medizinstudiums und Änderung der Approbationsordnung für Ärzte – Empfehlungen der Expertenkommission zum Masterplan Medizinstudium 2020“ Anpassungen für die Approbation erarbeitet hat (Deutscher Wissenschaftsrat 2018).

Am 15. August 2019 wurde die ÄApprO von 2002 unter Berücksichtigung des Masterplan Beschlusstextes 2020 und den oben aufgeführten Empfehlungen vom Wissenschaftsrat teilweise angepasst. So sind Universitäten nun dazu angehalten, fächerübergreifende Lehre anzubieten (Approbationsordnung für Ärzte 2002).

1.2 Moderne Lehr- und Lernangebote

1.2.1 Möglichkeiten an der Universität

Schon in den 1940er Jahren wurde die Idee entwickelt, das Lernen mittels „Lehrmaschinen“ effizienter zu gestalten. Durch den Fortschritt der Computerentwicklung in den 50er Jahren bot sich diese digitale Technologie zur Erstellung von Lernprogrammen an (Niegemann et al. 2004). Seit den 1990er Jahren gibt es Ansätze, digitale Medien in Bildungsprozesse verschiedener Fachbereiche zu etablieren. Der Computer und das Internet nehmen gegenwärtig einen zentralen Bestandteil im Alltag aller Menschen ein (Arnold et al. 2013). Zum aktuellen Zeitpunkt werden die Möglichkeiten, die digitale Medien mit sich bringen, allerdings noch nicht regulär in der medizinischen Lehre eingesetzt (Kuhn et al. 2018). Es werden verschiedene Kategorien unterschieden, durch die digitale Medien eingesetzt werden können. Dies sind „klassische digitale Medien“ wie digitale Präsentationen, „soziale Kommunikationstools“ wie soziale Netzwerke, „elektronische Prüfungssysteme“ wie *E-Assessments* oder *E-Klausuren*, „audio-/videobasierte Medien“ wie Tutorials oder „interaktive Tools“ wie *Serious Gaming* (Persike und Friedrich 2016). Bis auf *E-Klausuren* werden aktuell die Möglichkeiten der digitalen Medien, Lehr- und individuelle Lernangebote in z. B. kleine Apps zu integrieren, nur an wenigen Hochschulen ausgetestet. Die Vorlesungen werden mit digitalen Elementen wie z. B. dem Programm Microsoft® Power-Point unterstützt. Hierbei werden allerdings nicht die Stärken, die digitale Medien aufweisen, genutzt und zu wenig in den Lernprozess der Studierenden aktiv mit einbezogen (Persike und Friedrich 2016; The Digital Turn 2016). Die Expertenkommission des Wissenschaftsrats sieht die Digitalisierung in der Lehre als eine wertvolle Ergänzung an. So wird explizit der Einsatz von digitalen Medien und Technologien erwünscht. Hierdurch kann die vertikale Integration vereinfacht werden (Deutscher Wissenschaftsrat 2018). Unter *E-Learning*-Angeboten (englisch: *electronical learning* = elektronisch unterstütztes Lernen) werden digitale Lehr- und Lernmethoden verstanden, die den Lernenden durch interaktive Technologien herausfordert und selbstgesteuertes Lernen fördern soll (Hodson et al. 2001; Kerres 2013; Tolks 2016). Hierbei ist es wichtig, dass *E-Learning*-Angebote nicht nur als freiwillige Zusatzangebote vorgestellt werden und dadurch für Studierende eine zusätzliche Belastung bedeuten, sondern gemeinsam mit der Präsenzlehre einen interessanten Bestandteil im Curriculum bilden (Boeker und Klar 2006). Somit kann sich die digitale Form des Lehrens und Lernens zu einem zentralen Element im Alltag der Hochschulen entwickeln. Mittels der Digitalisierung können (z. B. mit einem *Audience-Response*-System, das Umfragen parallel zur Vorlesung einbindet) auch Veranstaltungen mit einer großen Anzahl an Studierenden interaktiv gestaltet werden (The Digital Turn 2016). Mayer et al. (2009) beobachteten, dass durch die regelmäßige Präsentation von Fragen während der Vorlesung in Abschlussprüfungen bessere Ergebnisse erzielt werden. Eine wesentliche Limitation dieser Studie stellten unbekannte Störfaktoren dar, welche allerdings ein grundsätzliches Problem bei Studien dieser Art zu sein scheinen. Nelson et al. (2012) untersuchten in einer Meta-Analyse 21 Studien, die sich mit den

Auswirkungen eines *Audience-Response*-Systems auf die Lernergebnisse beschäftigten. Die beobachteten Auswirkungen waren u. a. vom Studiendesign (randomisiert vs. nicht-randomisiert) abhängig. So kann es sein, dass in nicht-randomisierten Studien die Effekte von *Audience-Response*-Systemen überschätzt werden. Kritisch zu betrachten ist hauptsächlich, dass keine Studie den Einfluss von vorhandenen Störfaktoren wie z. B. vorhandene Fertigkeiten und Wissen der Studienteilnehmer betrachtete. Somit sind eventuelle positive Auswirkungen von *Audience-Response*-Systemen noch nicht ausreichend belegt.

Zusätzlich besteht die Möglichkeit, die Lehre nach dem Ansatz des *Blended Learning* auszurichten. *Blended Learning* beschreibt ein aufeinander abgestimmtes Verfahren mit wechselnden Elementen der Präsenzveranstaltungen und der selbstständigen Vorbereitung theoretischer Inhalte mit digitalen Medien. Dadurch soll eine erhöhte Qualität in der Lehre erreicht werden und durch die *E-Learning*-Angebote ein didaktischer Mehrwert geschaffen werden (Arnold et al. 2013; Kerres 2013). Digitale Lernmaterialien und -medien können den Nutzern durch sog. *Learning Management Systeme* wie z. B. der *Open Source* Lernplattform „integriertes Lern-, Informations- und Arbeitskooperations-System“ (ILIAS; www.ilias.de) zur Verfügung gestellt werden. Diese Systeme bilden die Grundlage für eine webbasierte *E-Learning*-Infrastruktur (Tolks 2016).

1.2.2 Möglichkeiten des Selbststudiums

Studierende haben auch die Möglichkeit im Selbststudium ihren Lernprozess durch digitale und kostenpflichtige Medien wie z. B. „Thieme examen online“ (Georg Thieme Verlag KG, Stuttgart, Deutschland) oder „AMBOSS®“ (AMBOSS GmbH, Berlin, Deutschland, <https://www.amboss.com/de>) zu fördern. AMBOSS® ist eine Lernsoftware und ein Nachschlagewerk für Studierende sowie Ärzte. Mittlerweile nutzen laut eigenen Angaben rund 500.000 Mediziner AMBOSS® weltweit (AMBOSS: Über uns). Für Nutzer sind hiermit die Vorbereitungen auf Prüfungen während des Studiums und Staatsexamina möglich. So können individuelle Fragensitzungen bestehend aus originalen IMPP-Fragen konzipiert werden und die Nutzer erhalten ein ausführliches Feedback über ihren persönlichen Lernstand (Kuhn et al. 2018). Bei AMBOSS® werden den Nutzern die Schwierigkeitsgrade in Form von einem „Hammerindex“ präsentiert. Hierbei werden die Fragen in fünf verschiedene Stufen, nach der Richtig- bzw. Falschbeantwortungsquote der Nutzer eingeteilt. Darüber hinaus erscheint zu jeder Antwortmöglichkeit nach der Beantwortung ein Kommentar mit einer Erklärung (AMBOSS: Handbuch – Kreuzmodus). Einige Universitäten u. a. auch die UMG, bieten sog. Campuslizenzen an, durch die die Studierenden einer Hochschule vergünstigte Konditionen wahrnehmen oder sogar kostenlos auf die Lernsoftware zugreifen können (AMBOSS: Campuslizenzen). Des Weiteren bietet AMBOSS® eine Kooperation mit einer weiteren Lernsoftware namens „Meditricks“ (Meditricks GmbH, Freiburg im Breisgau, Deutschland, <https://www.meditricks.de>) an. Die Funktionsweise von Meditricks basiert auf dem Konzept des Picmonic®-Lernsystems

(Picmonic, Phoenix, AZ, USA). Hier werden Studierenden auf einer multimedialen Website audiovisuelle Mnemotechniken zur Verfügung gestellt, um die Gedächtnisretention und Prüfungsleistung von medizinischen Fakten zu erhöhen, wodurch den Studierenden eine effektivere Prüfungsvorbereitung ermöglicht wird. Unklar ist allerdings noch, inwiefern dieses Lernsystem zu einer längeren Gedächtnisretention bzw. zu einer besseren Anwendbarkeit von Wissen beitragen kann (Yang et al. 2014).

1.2.3 Lernerfolg durch Digitalisierung

Computerbasiertes Lernen ermöglicht es, eine große Bandbreite an unterschiedlichen Lernmöglichkeiten anzubieten (Hodson et al. 2001; Tolks 2016). Durch die Digitalisierung kann die lernende Person herausgefordert werden. Zusätzlich kann das Lernen intensiviert und vor allem individualisiert werden. Diese Effekte können zu einer besseren Lehrqualität führen. Der Lernerfolg beim alleinigen *E-Learning* ist nach Boeker und Klar (2006) vergleichbar mit der herkömmlichen Lehre und dieser noch nicht überlegen. Diese Auffassung bestätigen auch andere Autoren (Cook et al. 2008; Kerres 2013), in Kombination führen die beiden Lehrformate allerdings zu einem größeren Lernerfolg (Boeker und Klar 2006; Kerres 2013). Vor allem in Hinblick auf die bereits erwähnten multimedialen Lernsoftwares gibt es Studien, die bei Verwendung von multimedialen Lernsoftwares einen größeren Lernfortschritt im Vergleich zu traditionellen Lernmethoden feststellen konnten (Grundman et al. 2000). Letztendlich ist der Lernerfolg aber relativ unabhängig von der eingesetzten Technologie und ist v. a. abhängig vom didaktischen Konzept (Kerres 2013).

1.2.4 Aktueller Bedarf an Digitalisierung

Von der Deutschen Hochschulmedizin e.V. wurde am 25. März 2020 die dritte Empfehlung für die Lehre während der COVID-19 Pandemie veröffentlicht. Dadurch, dass an den Hochschulen keine Präsenzlehre stattfinden kann, ist ein Fortschritt hinsichtlich digitaler Lehre unverzichtbar geworden. Empfohlen wird eine gemeinschaftliche Nutzung über das Projekt DigiPal (Nationale Lernplattformen für digitales Patienten-bezogenes Lernen im Medizinstudium, existierend seit dem 1. Juni 2020) bereits bestehender digitaler Angebote sowie eine Entwicklung von neuen und innovativen *E-Learning*-Angeboten. Insgesamt soll hierbei eine Ausrichtung nach dem NKLM erfolgen (Deutsche Hochschulmedizin e.V. 2020).

1.3 Motivation

„*Motivation is relevant to learning because learning is an active process requiring conscious and deliberate effort*“ (Stipek 1988). Die Motivation etwas zu lernen, wird von dem Studenten selbst und seiner Umwelt beeinflusst (Stipek 1988; Pelaccia und Viau 2017). In der Selbstbestimmungstheorie von Deci und Ryan (1985) wird eine allgemeine Motivationstheorie vorgestellt, die die qualitative Motivation eines Individuums in verschiedenen Lebenssituationen betrachtet (Deci und Ryan 1985; Ryan und Deci 2000; Kusurkar et al. 2011a; 2013). Deci und

Ryan (1985) beschreiben, dass der Mensch das Ziel verfolgt, sein Leben selbstbestimmt gestalten zu können. Die Autoren unterscheiden zwischen verschiedenen Motivationsqualitäten, die zu unterschiedlichen Ergebnissen führen können. Sie definieren die „autonome Motivation“, die intrinsisch von einer Person gebildet wird und die „kontrollierte Motivation“, die durch eine externe Quelle hervorgerufen wird (Deci und Ryan 1985; Ryan und Deci 2000; Kusrkar et al. 2011a; Kusrkar et al. 2013).

Intrinsisch motivierte Studierende haben ein aufrichtiges bzw. persönliches Interesse an den zu bewältigenden Aufgaben (Deci und Ryan 1985; Ryan und Deci 2000; Kusrkar et al. 2011a). Die intrinsische Motivation ist nach Deci und Ryan (1985) die selbstbestimmteste Form der Motivation. Auf die intrinsischen Motivationsquellen, und damit auf das Lernen, haben verschiedenen Wahrnehmungen („subjektiver Wert“, „Kompetenzwahrnehmung“, „wahrgenommene Kontrolle“, „Verbundenheit“) Einfluss (Kusrkar et al. 2011a; Pelaccia und Viau 2017). So wird den zu belegenden Kursen oder zu bearbeiteten Aufgaben von den Studierenden ein „subjektiver Wert“ zugeordnet, mit dem sie bewerten, inwiefern die Bearbeitung der Aufgabe sie beim Erreichen ihrer Ziele voranbringt (Pelaccia und Viau 2017). Zusätzlich wird von den Studierenden überprüft, inwiefern die Aufgaben oder Kurse sie darin unterstützen, ihre Kompetenzziele zu erreichen und ob sie selbst die Durchführung kontrollieren können (Ryan und Deci 2000; Kusrkar et al. 2011a; Pelaccia und Viau 2017). Kusrkar et al. (2011a) beschreiben des Weiteren noch ein Bedürfnis nach Verbundenheit zu anderen Menschen. Diese anderen Menschen können Eltern, Dozenten, Kommilitonen oder auch Patienten sein. Durch dieses insgesamt aufrichtige Interesse bei der intrinsischen Motivation findet ein tiefergreifendes Lernen statt, wodurch gelernte Inhalte auch auf neue Situationen angewendet werden können (Entwistle et al. 1974; Seale et al. 2000).

Im Gegensatz dazu steht die extrinsische Motivation, bei der u. a. eine externe Belohnung den hauptsächlichsten Anreiz zum Lernen darstellt. Dabei rückt die Relevanz des Inhalts für den Studenten in den Hintergrund (Fransson 1977; Deci und Ryan 1985; ten Cate et al. 2011). So gelten Studierende dann als extrinsisch motivierte Lernende, wenn sie sich vornehmlich auf die geprüften Lehrinhalte aus dem Curriculum konzentrieren, um eine gute Note zu erreichen (Seale et al. 2000). Auch hohe Testanforderungen oder Prüfungsangst können dazu führen, dass Studierende vermehrt oberflächlich lernen (Fransson 1977). Die extrinsische Motivation selbst setzt sich aus vier verschiedenen Ebenen zusammen, die jeweils einen unterschiedlichen Grad an Selbstbestimmung aufweisen. Bei der „externen Regulierung“ besteht kein Eigeninteresse an dem Studium, bei der „introjizierten Regulierung“ wird zwar der Wert des Studiums anerkannt, der Grund für das Studium ist allerdings dennoch extern zu suchen. Diese Regulierungsformen bilden die kontrollierte und wenig selbstbestimmte Motivation. Bei der „identifizierten Regulierung“ wird darüberhinausgehend noch akzeptiert, dass die Kontrolle außerhalb von einem selbst liegt. Die „integrierte Regulation“ kommt der autonomen und selbstbestimmten Motivation noch am nächsten, da sich hier die Regulation von extern nach intern verlagert hat (Ryan und Deci 2000; Kusrkar et al. 2011a; 2013).

Abhängig von der Qualität der Motivation können unterschiedliche Effekte auf die Lernqualität, Lernstrategien und Leistungen von Studierenden beobachtet werden (Kusurkar et al. 2013; Pelaccia und Viau 2017). Es ist eine Steigerung der autonomen und selbstbestimmten Motivation durch die Anpassung von Lerninhalten an die Bedürfnisse der Studierenden erwünscht (Kusurkar et al. 2013; The Digital Turn 2016). Es sollte das Ziel sein, ein Lernen auf einer tieferen Ebene im Sinne der intrinsischen Motivation bei Studierenden hervorzurufen (Fransson 1977). Um auf die beschriebenen Wahrnehmungen der intrinsischen Motivation Einfluss zu nehmen, empfehlen Pelaccia und Viau (2017), den Studierenden die Lernziele sowie die Möglichkeiten zum Bestehen dieser zu erklären. Dadurch sind sie in der Lage, den Kurs hinsichtlich des Aufgabenwerts einzusortieren (Hopstock 2008; Kusurkar et al. 2011b; Pelaccia und Viau 2017). Zusätzlich soll das Anforderungsniveau an die Studierenden angemessen sein und dabei weder zu leicht noch zu schwer sein. Damit kann verhindert werden, dass Studierende durch Misserfolg demotiviert und durch zu einfachen Erfolg gelangweilt werden (Kusurkar et al. 2011b; Pelaccia und Viau 2017). Zusätzlich werden Studierende vor allem in ihrem Lernprozess unterstützt, wenn die Bewertung für sie nachvollziehbar ist und Fehler mit Kommentaren erläutert werden (Kusurkar et al. 2011b; Pelaccia und Viau 2017). Darüber hinaus empfehlen Kusurkar et al. (2011a; 2011b) den Studierenden ein selbstständiges Planen zu ermöglichen, damit die Autonomie und damit die Selbstbestimmtheit des Lernens gefördert wird. So lässt sich beispielsweise durch *Blended-Learning*-Angebote die Autonomie und damit intrinsische Motivation steigern (Woltering et al. 2009; Kusurkar et al. 2011a). Demnach entstehen nachhaltigere Lerneffekte durch den Einsatz von digitalen Lehrformaten (The Digital Turn 2016). Das Umfeld der Lernenden, wie zum Beispiel auch motivierte Dozierende, kann also einen Einfluss auf die Motivation haben und die Lernenden in der Autonomie der intrinsischen Motivation unterstützen (Williams et al. 1999; Seale et al. 2000).

1.4 Prüfungen

1.4.1 Prüfungsformate

Die Ärztliche Prüfung gliedert sich in mündlich-praktische und schriftliche Teile (Approbationsordnung für Ärzte 2002). Bei der Auswahl von Prüfungsformaten muss beurteilt werden, inwiefern diese sich dazu eignen, die Leistungen der Studierenden in unterschiedlichen Fachbereichen abzufragen (Biggs 1996). Gegenwärtig bestehen nach wie vor Diskrepanzen zwischen den Lernzielen und den verwendeten Prüfungsmethoden (Newble 2016), weshalb es in der kompetenzbasierten medizinischen Ausbildung nötig ist, ein geeignetes Prüfungssystem einzuführen (Holmboe et al. 2010).

Im klinischen Abschnitt des Medizinstudiums werden verschiedene Prüfungsformate wie summative oder formative Prüfungen (Ben-David 2000) sowie der sog. *OSCE* (= *objective structured clinical examination*) zur Überprüfung von klinisch-praktischen Kompetenzen

eingesetzt (Nikendei und Jünger 2006). Formative Prüfungen bewerten die Lernstrategie und bieten Studierenden die Möglichkeit, im Verlauf des Semesters Bewertungen und Rückmeldungen zu ihrem Lernstatus und Lernfortschritt zu bekommen. Ein zentraler Bestandteil von formativen Prüfungen ist ein ausführliches Feedback (Black und Wiliam 1998; Krasne et al. 2006; Roediger und Karpicke 2006b; Cliff et al. 2008). Das Ziel hierbei ist, das Wissen oder die Fähigkeiten der Studierenden zu verbessern (Burgess und Mellis 2015). Formative Prüfungen kommen beispielsweise an der UMG seit 2013 in Form von elektronischen Fallseminaren in bestimmten Modulen zum Einsatz. Diese setzen sich v. a. aus *Key-Feature-Fällen* zusammen (s. Kapitel 1.4.2). Nach Beantwortung jeder Frage bekommen die Teilnehmer ein automatisches Feedback, das Informationen zu den richtigen Antwortoptionen enthält. Darüberhinausgehend können sich die Teilnehmer bei Interesse ihre erreichte Punktzahl sowie die durchschnittliche Punktzahl des Semesters nach jedem Fallseminar per E-Mail mitteilen lassen. Dadurch können Studierende und auch Lehrende ggf. bestehende Wissenslücken feststellen und Studierende können sich des Weiteren an die Prüfungssituation für eine summative Prüfung gewöhnen (Roediger und Karpicke 2006b; Cliff et al. 2008; McDaniel und Little 2019). Summative Prüfungen sollen die Leistungen der Studierenden am Ende eines Lehrabschnitts bewerten (Cliff et al. 2008), den Lernfortschritt messen (Burgess und Mellis 2015) oder auch zu einer Notenvergabe führen (Roediger und Karpicke 2006b).

1.4.2 Fragenformate

Bei der Auswahl des Fragenformats, sollte der sog. *Cueing*-Effekt berücksichtigt werden. Der *Cueing*-Effekt beschreibt ein Szenario, indem ein Student eine MC-Frage durch das Erkennen der Lösung richtig beantworten kann, ohne dass er selbstständig dazu in der Lage wäre, die richtige Antwort zu generieren (Schuwirth und van der Vleuten 2004). Hierdurch kann es im klinischen Alltag vorkommen, dass eine übereilige Diagnose gestellt wird, ohne zuvor weitere Alternativen in Betracht zu ziehen (Graber et al. 2005).

MC-Fragen sind eines der ältesten und am häufigsten verwendeten Fragenformate in der medizinischen Ausbildung (Larsen et al. 2008). Dieses Fragenformat bietet die Möglichkeit Grundlagenwissen abzufragen (Appel et al. 2002) und die Beurteilung des Wissensstands von Studierenden bei formativen Prüfungen zu testen (Anderson 2004). Da MC-Fragen in der Regel eine kurze Bearbeitungsdauer erfordern, besteht die Möglichkeit viele Fragen zu stellen und damit viele Themengebiete abzudecken (Schuwirth und van der Vleuten 2004). MC-Prüfungen können allerdings einen negativen Einfluss auf das Lernverhalten von Studierenden haben, die sich durch das alleinige Abfragen von Faktenwissen zu oberflächlich Lernenden entwickeln (Newble und Entwistle 1986). Bei MC-Prüfungen entsteht das Problem, dass gewöhnlich jede Antwortmöglichkeit als binäre Einheit betrachtet wird („richtig“ oder „falsch“). Eine Skalierung der Wissensinhalte findet nicht statt, da die binären Einheiten lediglich aufsummiert werden. Zusätzlich ist hier die Abfrage des Wissensstands am wenigsten anspruchsvoll, weil die Studierenden die Antwort nur wiedererkennen

müssen (Biggs 1996). Auswendig gelernte Inhalte können in einer neuen Problemsituation nicht unmittelbar anwendbar sein (Dahle et al. 2002), weshalb der Einsatz von MC-Fragen in summativen Prüfungen nicht empfohlen wird (Anderson 2004). Schuwirth et al. (1996a) untersuchten darüber hinaus das Auftreten von *Cueing*-Effekten bei MC-Prüfungen. Hierbei wurden zwei verschiedene Effekte festgestellt: Durch die vorgegebenen Antworten können die Studierenden entweder falsche Antworten ausschließen (negatives *Cueing*) oder werden auf die richtige Antwort hingewiesen (positives *Cueing*) (Schuwirth et al. 1996a). Wenn die Antwortmöglichkeiten so ausgewählt werden, dass sie alle als richtige Antwort plausibel eingestuft werden könnten, müssen Studierende bei der Beantwortung viel überlegen und abwägen. Hierbei kann das Risiko entstehen, dass diese falschen, aber dennoch plausiblen Antwortmöglichkeiten als falsches Wissen verinnerlicht werden und somit bei späteren Tests als falsche Antworten ausgewählt werden. Little und Bjork (2012) beobachteten allerdings, dass regelmäßige MC-Tests kurz nach dem Unterricht während eines langen Retentionsintervalls diese negativen Effekte vermeiden können. Zusätzlich können die MC-Tests zu einer längeren Retention der Lerninhalte führen (Roediger und Marsh 2005; Little et al. 2012; Little und Bjork 2012; Little et al. 2019). Entscheidend dafür, wie tiefgreifend die Verarbeitung und der anschließende Lernprozess einer MC-Frage stattfinden kann, ist also die Konstruktion dieser (McDaniel und Little 2019).

Wenn die Zusammenhänge eines Sachverhaltes verstanden sind, fällt es Studierenden leichter, Probleme bei ähnlichen aber nicht identischen Situationen zu lösen (Dahle et al. 2002). Für diese Problemlösungen sind meistens nur einige entscheidende „kritische Schritte“ (Page und Bordage 1995) zu beachten (= *Key-Feature*), die in Form von konstruierten Patientenfällen in Prüfungen eingebaut werden können. Dieser Fall wird zunächst kurz erläutert (u. a. Patientendaten, Beschwerden) und im Anschluss durch mehrere Fragen zu Differentialdiagnosen, Diagnostik oder Therapieansätzen ergänzt. Hierbei wird durch jede Frage die klinische Entscheidungsfindung von den Studierenden getestet (Page und Bordage 1995; Hatala und Norman 2002; Farmer und Page 2005; Hrynchak et al. 2014). Die Studierenden können ihre Antworten entweder selbst formulieren (*Write-in-Format*) oder bekommen die Lösungen in Form von *Short-menu-* bzw. *Long-menu-*Listen präsentiert. *Short-menu-*Listen beinhalten 15 bis 25 Antwortoptionen, die die richtigen Antworten oder mögliche Fehlannahmen beinhalten. Hierbei sollten pro richtige Antwort vier bis fünf falsche Optionen präsentiert werden, um die *Cueing*-Effekte zu reduzieren (Page und Bordage 1995; Farmer und Page 2005). *Long-menus* „[...] sind lange alphabetisch geordnete Listen, die alle Antwortmöglichkeiten enthalten. Die Liste soll aus mindestens 500 Einträgen bestehen, um einen *Cueing*-Effekt minimal zu halten“ (Kopp et al. 2006). Studierende können in ein Eingabefeld die ersten drei Buchstaben ihrer selbst generierten Lösung eintippen und erhalten daraufhin eine Auswahl an Begriffen, anhand derer sie sich für eine Antwort entscheiden können (van Bruggen et al. 2012; Cerutti et al. 2016). Durch den speziellen Aufbau der *Long-Menu-*Listen kann der Prozess der Entscheidungsfindung mit der Realität in der Praxis oder Klinik verglichen werden und ist durch die vielen Antwortoptionen mit einem offenen

Fragenformat vergleichbar. Der oben beschriebene *Cueing*-Effekt kann durch *Long-menus* verringert werden, da Studierende aktiv beginnen müssen, eine Antwort einzugeben (Schuwirth et al. 1996b; Kopp et al. 2006; Cerutti et al. 2016). Ein Nachteil des *Long-menus*-Formats ist, dass eine korrekte Benennung des Wortstammes sowie die Rechtschreibung der Lösung mit mindestens drei Buchstaben notwendig ist, wodurch v. a. Studierenden mit einer anderen Muttersprache ein Nachteil entsteht (Cerutti et al. 2016).

Die oben beschriebenen Prüfungs- und Fragenformate werden mittlerweile, wie auch teilweise schon erläutert, vermehrt digital eingesetzt (Appel et al. 2002). Hierbei können zwei verschiedene Formate unterschieden werden. Bei dem ersten Format findet ein computer-gestütztes Lesegerät Anwendung, das handschriftlich ausgefüllte Antwortbögen auslesen kann (Cantillon et al. 2004). Das zweite Format bilden die komplett am Computer durchgeführten *E-Klausuren*, welche zum einen die Möglichkeit bieten, Prüfungsfragen in einem „Fragenpool“ zu sammeln und zum anderen auch mit weniger Aufwand eine große Anzahl an Studierenden zu prüfen (Michel et al. 2015). Weitere Vorteile bestehen darin, dass in den Tests andere Medien (wie z. B. Röntgenaufnahmen, Bilder oder Videos) zum Einsatz kommen können und Dozierende und Studierende eine sofortige Rückmeldung über die Leistung erhalten (Cantillon et al. 2004). Zusätzlich wird die Umsetzung der *Key-Feature*-Fälle ermöglicht, da in papierbasierter Form ein Zurückblättern möglich ist, wodurch es zu sog. *backward cueing*, also der Vermittlung von Lösungshinweisen in nachfolgenden Informationen, kommen kann. Durch die Bearbeitung der Fälle am Computer können Lösungshinweise zu bereits beantworteten Fragen gegeben werden, ohne dass die gegebenen Antworten vorheriger Fragen rückwirkend angepasst werden können (Kopp et al. 2006).

1.4.3 Lernfortschritt durch Prüfungen

Die Aussage „*Assessment drives learning*“ ist mittlerweile zu einer allgemein anerkannten Überzeugung geworden (Joughin 2010). So bilden Prüfungen einen dominierenden, wenn nicht sogar den wichtigsten Einfluss auf das Lernverhalten und die Lernergebnisse von Studierenden (Newble und Jaeger 1983; McLean 2001; Cilliers et al. 2010). Hauptsächlich kommen summative Prüfungen als Beurteilungsinstrumente der Studierenden zum Einsatz (Larsen et al. 2008). Raupach et al. (2013) zeigten in einer Studie, dass summative Prüfungen einen starken Einfluss auf den Lernerfolg haben können. So wird durch summative Prüfungen nicht nur das Prüfungsergebnis verbessert, sondern auch die extrinsische Motivation gesteigert, sich im Selbststudium mit weiterem Lernmaterial zu beschäftigen (Raupach et al. 2013).

Formative Prüfungen wiederum können zu Lerngewinnen führen, da sie noch während des Lernprozesses Rückmeldungen über den aktuellen Lernstand der Studierenden geben und dadurch ein selbstgesteuertes Lernen ermöglicht wird (Hudson und Bristow 2006). Diese Retention von Lerninhalten und der Einfluss auf die Leistungsverbesserung werden auch *testing effect* genannt. Hierbei kann das Testen entweder einen direkten Einfluss (= direkter

testing effect) auf den Lernprozess haben oder einen indirekten Einfluss (= indirekter *testing effect*) (Roediger und Karpicke 2006a; Larsen et al. 2008). Hinsichtlich der Funktionsweise des direkten *testing effects* auf die bessere Retention von Lerninhalten bestehen zwei Hypothesen. Durch formative Prüfungen kann zum einen der Effekt des sogenannten *retrieval efforts*, also dem Wiederaufrufen von Informationen, gefördert werden, da dies eine aktive Verarbeitung beinhaltet und zu einer besseren Merkfähigkeit führt (Wheeler et al. 2003; Roediger und Karpicke 2006b; Larsen et al. 2008; McDaniel und Little 2019). Verschiedene Studien zeigten, dass das Abrufen und die Rekonstruktion von Wissen das Lernen fördern können (Wheeler et al. 2003; Karpicke und Blunt 2011). Hierbei gilt, dass je aufwendiger sich das Wiederaufrufen gestaltet, der direkte *testing effect* umso besser wirken kann und dadurch ein langfristiges Lernen gefördert wird (Bjork und Bjork 1992; Roediger und Karpicke 2006b). Zum anderen kann sich durch das wiederholte Prüfen auch ein gewisser Trainingseffekt bemerkbar machen, da sich die Testformate ähneln und die gleichen kognitiven Prozesse zur Beantwortung der Fragen genutzt werden (Morris et al. 1977; Raupach et al. 2016). Das erneute Lesen des Lernmaterials dagegen trainiert nicht das aktive Abrufen von Informationen (Roediger und Karpicke 2006b). Das regelmäßige Testen im Anschluss an eine Vorlesung oder nach dem Lesen von Material führt zusätzlich noch zu einer besseren Beibehaltung und Verarbeitung von Wissen im Vergleich zum erneuten Lesen des Materials (Larsen et al. 2008; McDaniel et al. 2012).

Beim indirekten Einfluss der formativen Prüfungen wird den Studierenden durch das Feedback bewusst, wie er oder sie den Lernprozess selbst effektiver gestalten und eine Verbesserung der eigenen Lernstrategien vornehmen kann (Roediger und Karpicke 2006b; Larsen et al. 2008). Diese indirekten Einflüsse scheinen in der Unterrichtspraxis einen größeren Effekt zu haben, als es die Labor-Experimente zur kognitiven Psychologie bisher gezeigt haben. So wird u. a. durch regelmäßige Prüfungen ein kontinuierliches Lernen bei den Studierenden hervorgerufen und nicht erst kurz vor der Abschlussprüfung praktiziert. Dadurch können sie dem Unterricht selbst besser folgen und mehr davon profitieren. Darüber hinaus lernen Studierende durch regelmäßiges Testen sich selbst bzgl. ihres Wissensstandes besser einzuschätzen (sog. *metacognitive monitoring*), wodurch sie ihr anschließendes Lernen gezielter ausrichten können (McDaniel und Little 2019).

Es scheint, dass das Fragenformat einen gewissen Einfluss auf die Retention der Lerninhalte hat. So haben Fragenformate in Labor-Experimenten der kognitiven Psychologie, bei denen die Antwort durch die Studierenden selbst generiert werden muss zu einer besseren Retention als MC-Prüfungen geführt, da bei Letzteren die Abfrage weniger aufwendig ist. Hier muss nur die Antwort widererkannt werden. (Foos und Fisher 1988; Bjork und Bjork 1992; Roediger und Karpicke 2006b; McDaniel et al. 2007). In jüngsten Ergebnissen zeigt sich allerdings eine Dissoziation bei der Anwendung des *testing effects* im Unterrichtskontext (McDaniel und Little 2019). McDaniel et al. (2012) haben in einer Studie im medizinischen Unterrichtskontext nachgewiesen, dass der *testing effect* bei MC-Fragen, genauso groß ist wie bei Kurzantwortfragen, die den gleichen Fragenstamm behandeln. Wie bereits in

Kapitel 1.4.2 beschrieben können auch gut konstruierte MC-Fragen zu einem tieferen Lerngewinn führen. Entscheidend für den Lernfortschritt ist also nicht primär das Fragenformat selbst, sondern welche indirekten Wirkungen durch die Frage bei den Studierenden hervorgerufen werden (McDaniel und Little 2019).

Die Vorteile von formativen Prüfungen auf den Lernfortschritt lassen sich noch ausweiten, indem die Tests über einen längeren Zeitraum verteilt werden (Karpicke und Roediger 2007). Hierbei ist es irrelevant, wie gut die Leistung während der formativen Prüfung ist oder ob eine ausführliche Rückmeldung über den Lernstand gegeben wird (Roediger und Karpicke 2006b). Dennoch kann ein Feedback den Nutzen der Prüfung erhöhen, indem der Student eine Rückmeldung über die richtige Antwort bekommt (Butler und Roediger 2008). Darüber hinaus sorgt das regelmäßige Testen zu einer signifikant längeren Speicherung der Lerninhalte (Larsen et al. 2009). Nicht nur auf die theoretischen, sondern auch auf die praktischen Fähigkeiten der Studierenden oder klinisches Denken können sich formative Prüfungen positiv auswirken (Kromann et al. 2009; Raupach et al. 2016). Studierende sind sich allerdings dieser positiven Effekte nicht unbedingt bewusst, weshalb sie formative Prüfungen teilweise schlecht evaluieren (Karpicke und Roediger 2008; Schmidmaier et al. 2011). Bei einem selbstregulierten Lernen, wie es bei der intrinsischen Motivation gefordert wird, kann es dadurch zu einer falschen Auswahl an Lernstrategien kommen (Karpicke 2009).

Durch den Aufbau des medizinischen Curriculums und das fächergetrennte Unterrichten finden sich in den MC-Prüfungen wenige horizontale oder vertikale Verknüpfungen zu anderen Disziplinen. Dadurch werden Studierende dazu angehalten, Fakten auswendig zu lernen, ohne dass diese anwendungsbezogen auf den späteren klinischen Alltag übertragbar sind (McLean 2001). Die M2-Prüfung setzt sich allerdings aus fächerübergreifenden Fragen zusammen und wird zukünftig durch die geforderten Neuerungen in der ÄApprO hinsichtlich einer stärkeren horizontalen und vertikalen Verknüpfung in der Lehre neue Schwerpunkte setzen (Bundesministerium für Gesundheit et al. 2017). Prüfungen sind und bleiben ein untrennbarer Bestandteil in der medizinischen Ausbildung, aber sie sollten einen positiven Einfluss auf das Lernen der Studierenden haben (McLean 2001).

1.5 Zielstellungen und Hypothesen der Arbeit

Wie in Kapitel 1.1 beschrieben ist das Medizinstudium an der UMG modular strukturiert. In den summativen Modulabschlussprüfungen wird durch MC-Prüfungen vorwiegend fächer-spezifisches Faktenwissen abgefragt. In Kapitel 1.1.2 sind die vom Wissenschaftsrat und im Masterplan 2020 (2017) vorgesehenen, weitreichenden Änderungen beschrieben, die das Medizinstudium ab 2025 betreffen werden. Die aufgeführten Änderungen werden sich sowohl in der Lehre als auch in den Prüfungen widerspiegeln. Der Schwerpunkt liegt hierbei auf einer fächerübergreifenden und longitudinalen Verknüpfung der klinischen Module und einer ausgedehnteren Kompetenzvermittlung. Des Weiteren soll die Lehre vermehrt

digitalisiert werden (Deutscher Wissenschaftsrat 2018). Welche Möglichkeiten sich hierfür anbieten wurde in Kapitel 1.2 erläutert. Zusätzlich wurde in Kapitel 1.3 die Theorie der extrinsischen und intrinsischen Motivation vorgestellt und wie es geschafft werden kann, Studierende zu einem tiefgreifenden Lernen zu motivieren (Seale et al. 2000). Wie in Kapitel 1.4.3 beschrieben, können formative Prüfungen einen direkten positiven Effekt auf das Lernen selbst haben. Darüberhinausgehend motivieren sie die Studierenden noch, sich kontinuierlich mit der Materie zu beschäftigen (Roediger und Karpicke 2006b), verbessern die Lernstrategien und die Wissensbewahrung der Studierenden und lassen sie ihre eigenen Schwächen durch das *metacognitive monitoring* erkennen (Larsen et al. 2008; McDaniel und Little 2019). Die beschriebenen Auswirkungen von formativen Prüfungen wurden bisher vor allem in kognitionspsychologischen Labor-Studien gefunden. Die Wirksamkeit von Prüfungen und der indirekte *testing effect* in der medizinischen Ausbildung sind noch nicht ausreichend erwiesen (Larsen et al. 2008; McDaniel und Little 2019). Durch den Einsatz von formativen Prüfungen während des Studiums lassen sich Prüfungen nicht mehr nur zur Bewertung von Studierenden einsetzen, sondern als eine Lernstrategie zum Erwerb und zur Aufbewahrung von Wissen (Larsen et al. 2008).

Das Projekt „*cross-linked undergraduate education*“ (*#clue*), das in dieser Studie vorgestellt wird, richtet sich an Studierende des dritten bis sechsten klinischen Semesters der UMG und bietet ihnen die Möglichkeit durch digitale freiwillige formative Prüfungen, ihren Lernstand über mehrere Semester longitudinal und v. a. fächerübergreifend zu kontrollieren. Zusätzlich können sie sich semesterbegleitend auf die M2-Prüfung vorbereiten. Hierbei werden überwiegend MC-Fragen präsentiert, die im Hinblick auf die anstehenden neuen Änderungen in der ÄApprO einen Schwerpunkt in der horizontalen Integration setzen und in interdisziplinären Lehrveranstaltungen von Lehrpersonen aus unterschiedlichen Fachrichtungen diskutiert werden. Bereits gelernte Inhalte sollen durch den direkten *testing effect* gefestigt werden (siehe Forschungsfrage zwei) und der indirekte *testing effect* wird beobachtet (siehe Forschungsfrage eins). Ziel dieser Arbeit besteht in der Vorstellung der Entwicklung von *#clue* und der Untersuchung der Wirkung dieser Lerninnovation auf die Studierenden (siehe Forschungsfragen drei und vier).

Für die Studie stellten sich folgende Forschungsfragen:

1. Führt die mehrfache Wiederholung von Inhalten während des Semesters bezüglich dieser Inhalte zu einem größeren Lernerfolg als bei Inhalten, die nicht mit Fragen wiederholt wurden?
2. Wie entwickelt sich der studentische Leistungsstand im Laufe des Semesters, wenn Inhalte anhand von Fragen mehrfach wiederholt werden?
3. Wie entwickelt sich die Teilnehmerzahl bei einem freiwilligen Projekt, das zum Verbessern des eigenen Lernerfolgs angeboten wird?
4. Wie beurteilen die Studierenden dieses innovative Projekt?

2 Material und Methoden

2.1 Studiendesign

In der hier vorliegenden Arbeit wurde eine prospektive, nicht-randomisierte monozentrische Interventionsstudie durchgeführt, an der im Sommersemester 2018 alle Studierenden des dritten, vierten, fünften und sechsten klinischen Semesters der UMG teilnehmen konnten. Im Folgenden werden zunächst der Aufbau und der Ablauf der Studie erläutert.

Zur Beantwortung der in Kapitel 1.5 aufgeworfenen Studienfragen kamen qualitative und quantitative Erhebungsinstrumente zum Einsatz, die anschließend näher beschrieben werden sollen.

Gemäß dem Votum der Ethik-Kommission der medizinischen Fakultät der Universität Göttingen vom 13.03.2018 bestanden keine ethischen und rechtlichen Bedenken gegen die Durchführung des vorliegenden Forschungsvorhabens (Antragsnummer 17/3/18). Diese Arbeit wurde im Rahmen einer Projektförderung des Göttinger Qualitätspakt-Projekts Campus Q^{PLUS} finanziell unterstützt.

2.2 Studiendurchführung

2.2.1 Übersicht des zeitlichen Ablaufs und des Aufbaus der Studie

Der Ablauf des Projektes gliederte sich in drei Phasen (siehe Abbildung 3). Zunächst wurden in Phase eins die vom IMPP bereitgestellten Staatsexamens-Fragen der M2-Prüfung kategorisiert und nach den Modulplänen der UMG sortiert. Hierfür wurden die Fragen der Sommerstaatsexamina von 2015, 2016 und 2017 sowie die Herbstexamina 2016 und 2017 verwendet. Es fand eine thematische Einsortierung aller Fragen in die jeweiligen Semester, Module und passenden Wochen statt. Parallel dazu wurden Querschnittsmodule notiert, deren Inhalte von den IMPP-Fragen ebenfalls abgedeckt wurden. In der allen Studierenden der UMG zur Verfügung stehenden Lern-Software AMBOSS[®] sind zudem fünf Schwierigkeitsgrade für alle Fragen in Form eines „Hammerindexes“ angegeben. Diese Kennwerte wurden für die ausgewählten Fragen dokumentiert, wobei die zweiten und dritten, sowie die vierten und fünften Schwierigkeitsgrade zusammengefasst worden sind, woraus sich für *#clue* drei Schwierigkeitsgrade ergaben (leicht, mittelschwer, schwer). Prinzipiell wurden die originalen Fragen verwendet. Vereinzelt wurden lange Fragen auf das Wesentliche gekürzt. Zusätzlich wurden ausgewählten Fragen relevante Informationen aus der Fallvignette hinzugefügt, welche einigen Fragen des IMPP vorausgeht, um eine Beantwortung der Fragen auch ohne das Lesen der Fallvignette möglich zu machen. Im Anschluss daran wurde für jede Woche eine Auswahl an Fragen getroffen, die thematisch mit den Vorlesungsinhalten der jeweiligen Woche identisch waren und an die jeweiligen Organisatoren der verschiedenen

Module versendet. Bei der Auswahl wurde eine ausgeglichene Präsentation von Fragen verschiedener Schwierigkeitsgrade berücksichtigt.

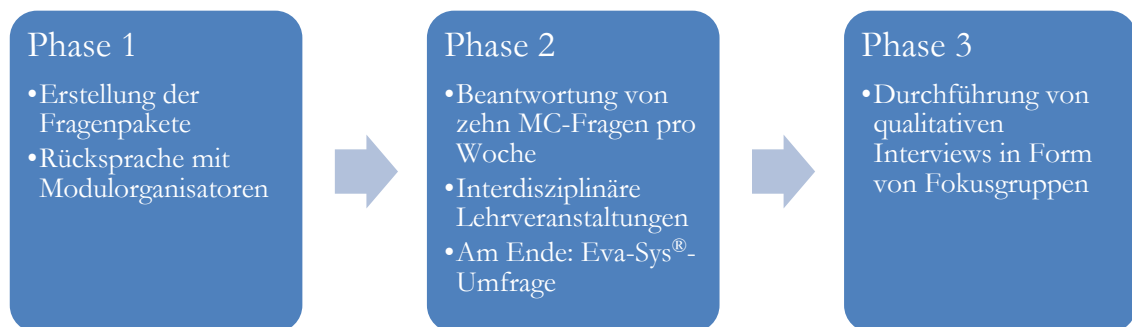


Abbildung 3: Projektverlauf. Dargestellt ist die Aufteilung des Forschungsprojektes in drei verschiedene Phasen. Die Vorbereitungen der Studie (Phase eins) fanden vor dem Sommersemester 2018 statt, während der Vorlesungszeit (Phase zwei) wurden die Daten für den quantitativen (Fragenpakete) und auch teilweise für den qualitativen Teil (Eva-Sys®-Fragebögen) dieser Arbeit erhoben. In der sich anschließenden vorlesungsfreien Zeit (Phase drei) erfolgte durch Fokusgruppengespräche die Erhebung der restlichen Daten für den qualitativen Teil dieser Arbeit.

Zudem wurden möglichst keine zu umfangreichen Fragen gewählt, um die Darstellung für mobile Endgeräte zu optimieren. Des Weiteren wurde darauf geachtet, dass Querschnittsfragen mit aktuellen Inhalten und Inhalten aus den vorangegangenen Semestern Anwendung fanden. Nach Rücksprache mit den Modulorganisatoren wurde für jede Semesterwoche ein individuelles Fragenpaket – bestehend aus zehn Fragen – für das dritte bis sechste klinische Semester erstellt. Hierbei wurde versucht, möglichst viele Themen der jeweiligen Woche durch die Fragen zu behandeln. Durch dieses Vorgehen konnte eine optimale zeitliche und inhaltliche Passung zu den Vorlesungen gewährleistet werden. Dem dritten und fünften klinischen Semester wurden insgesamt 14, dem vierten 13 und dem sechsten klinischen Semester zehn Fragenpakete angeboten. Die ungleiche Anzahl an Fragenpaketen ist zum einen auf die abweichende Semesterdauer zurück zu führen und zum anderen auf die unterschiedliche Teilnahmebereitschaft der Organisatoren der verschiedenen Module. Hinsichtlich der oben beschriebenen Kategorien wurden pro Wochenpaket fünf bis sechs leichte, zwei bis drei mittelschwere und ein bis zwei schwere Fragen ausgewählt (Schwierigkeiten in Anlehnung an AMBOSS®). Eine nachträgliche Analyse der Anforderungsniveaus der Schwierigkeitsgrade wird in Kapitel 2.4.1 erläutert. Die Fragenpakete beinhalteten drei verschiedene Arten von Fragen. Die erste Art setzte sich aus Fragen zusammen, die nur einmalig vorkamen. Die zweite Art bestand aus sogenannten Interventionsitems (wurden insgesamt den Studierenden insgesamt dreimal präsentiert) und die dritte Art aus Kontrollitems (wurden den Studierenden lediglich einmal am Ende des Semesters präsentiert).

Im Verlauf der zehn bis 14 Wochen (abhängig vom Semester) wurden insgesamt pro Semester fünf Fragen mit unterschiedlichen Schwierigkeitsgraden als sogenannte

Interventionsitems präsentiert. Hierbei wurden den Studierenden die gleichen Fragestellungen zweimal zu den Zeitpunkten eins und zwei (T1 und T2) präsentiert (siehe Abbildung 4). Das Auftreten der Interventionsitems fand an keinem bestimmten Datum statt und war auf die verschiedenen Wochen verteilt. Das letzte Fragenpaket (= *Exit Exam*) zum dritten Zeitpunkt (T3) bestand aus fünf Interventionsitems und fünf Kontrollitems. Bei den Kontrollitems wurde darauf Wert gelegt, dass diese den gleichen Schwierigkeitsgrad aufwiesen, wie die Interventionsitems, aber die gleiche Fragestellung in den vorherigen Wochen noch nicht behandelt wurde. Außerdem mussten zwischen dem *Exit Exam* und dem letzten Wiederholen des Interventionsitems mindestens zwei Wochen liegen, um die Möglichkeit, dass sich die Studierenden speziell an diese Frage erinnerten, zu minimieren. Die absolute Anzahl an schweren, mittelschweren und leichten Fragen war in den vier *Exit Exams* der Semester nicht identisch.

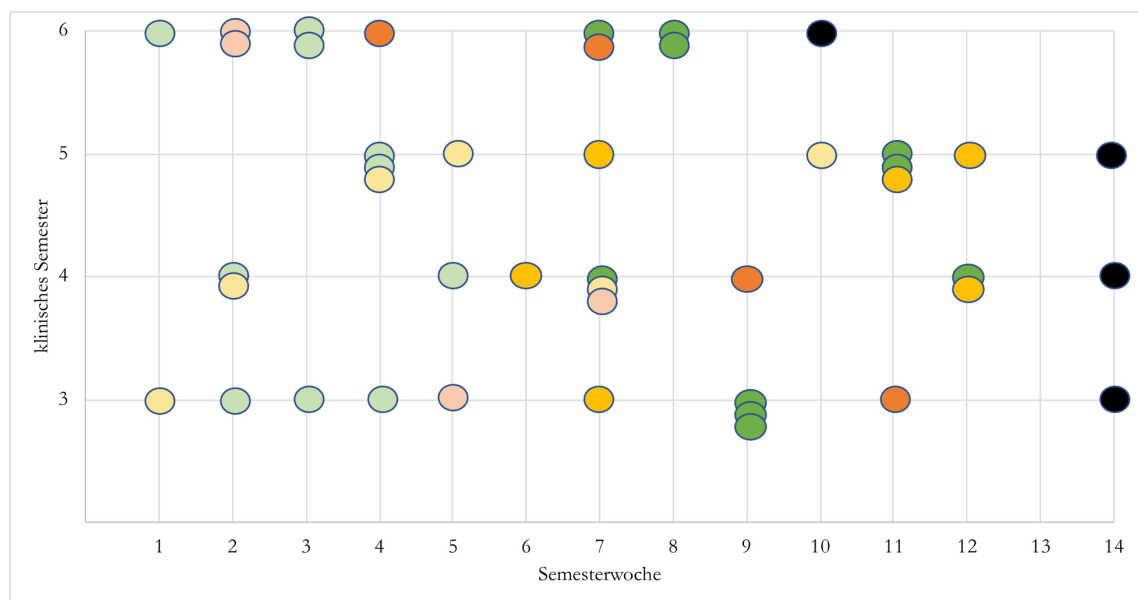


Abbildung 4: Verteilung der Interventionsitems. Dargestellt ist die Verteilung der Interventionsitems über das Sommersemester 2018. Jedem Semester wurden im Verlauf der 14 Wochen zweimal fünf Interventionsitems präsentiert (T1: matte Farben; T2: kräftige Farben). Grün sind die leichten Interventionsitems, gelb die mittelschweren und rot die schweren. Das *Exit Exam* fand zum Zeitpunkt T3 statt und ist durch die schwarze Farbe gekennzeichnet.

Die Fragenpakete wurden mit Hilfe von ILIAS erstellt und mit dem universitätseigenen Studierendenportal „e-campus“ auf „Stud.IP“ verknüpft. Jedes Fragenpaket wurde im Vorfeld von Mitarbeitern aus der Arbeitsgruppe Medizindidaktik der UMG, die einen Stud.IP-Zugang haben und nicht in diese Studie einbezogen wurden, pilotiert, um technische Probleme minimieren zu können.

Zusätzlich wurde eine studentische Gruppe mit Mitgliedern aus verschiedenen vorklinischen und klinischen Semestern gegründet. Hierfür wurden die Semestersprecher vorab per E-Mail kontaktiert, über das Projekt informiert und im Anschluss gebeten, in ihrem Semester Freiwillige zu rekrutieren. Es fanden regelmäßige Treffen in Phase eins und zwei statt, um

aktuelle Meinungen, Ideen und eventuell auftretende Probleme direkt von den Studierenden mitgeteilt zu bekommen. Mitglieder der studentischen Gruppe konnten an der Studie teilnehmen.

Während der zweiten Phase (siehe Abbildung 3) wurde den Studierenden immer sieben Tage lang Zeit gegeben, ein Fragenpaket zu bearbeiten. Es wurde eigens für dieses Forschungsprojekt auf dem „e-campus“ der UMG eine Veranstaltung erstellt. Die Studierenden hatten die Möglichkeit auf zwei Dokumente zuzugreifen, die je eine Anleitung für den Zugriff auf die Fragenpakete mithilfe des Computers bzw. mithilfe mobiler Endgeräte beinhalteten. In dieser Veranstaltung wurden die Links zu den aktuellen Fragenpaketen online gestellt, wodurch die Studierenden auf die Lernplattform ILIAS gelangten. Das entsprechende Fragenpaket wurde jeweils am Mittwochabend für sieben Tage freigeschaltet. Die Studierenden durften das Fragenpaket einmal bearbeiten. Die Fragen erschienen in einer festgelegten Reihenfolge von einem leichten, über einen mittelschweren bis hin zum schweren Schwierigkeitsgrad. Nach Beantwortung einer Frage war es dem Studenten nicht möglich, die Antwort zu ändern. Nach der Bearbeitung des Fragenpakets erhielt der Student eine Rückmeldung darüber, ob er „bestanden“ (mindestens 60% der Antworten richtig) oder „nicht bestanden“ (weniger als 60% richtig) hat und konnte noch einmal die zehn Fragen des Fragenpakets mit den richtigen Lösungen durchsehen. Da *#clue* ein neues und freiwilliges Projekt war, wurden die Studierenden auf verschiedenen Wegen an die Bearbeitung der Fragenpakete erinnert. Zum einen wurde zu Beginn jeder Woche eine E-Mail an alle Studierenden versandt, mit der Information, bis wann das aktuelle Fragenpaket noch bearbeitet werden konnte. Zum anderen wurde eine *Facebook*-Gruppe gegründet, in der ebenfalls wöchentliche Erinnerungen verschickt wurden. Durch die Datenschutz-Grundverordnung (DSGVO) vom 25.05.2018 musste allerdings die *Facebook*-Gruppe nach der siebten Woche während Phase zwei geschlossen werden. Auch die automatischen wöchentlichen Erinnerungs-E-Mails wurden ab der neunten Woche eingestellt. Die Studierenden hatten jedoch die Möglichkeit sich für die wöchentlichen Erinnerungs-E-Mail schriftlich aktiv anzumelden.

In jedem Semester wurde den Studierenden zudem die Möglichkeit angeboten, an interdisziplinären Lehrveranstaltungen teilzunehmen. Das dritte klinische Semester hatte die Möglichkeit an drei Veranstaltungen teilzunehmen. Dem vierten, fünften und sechsten klinischem Semester wurde jeweils eine Veranstaltung angeboten. Diese Veranstaltungen sollten von verschiedenen Dozierenden fächerübergreifend gehalten werden, mit dem Ziel besonders schwere oder interessante IMPP-Fragen im Diskurs interaktiv zu erörtern. Hierfür fanden zum einen solche Fragen Anwendung, die von besonders wenig Studierenden richtig beantwortet worden sind. Das prozentuale Antwortverhalten des Semesters wurde den Studierenden in Form einer Präsentation bei Microsoft® Professional Plus PowerPoint 2010 für Windows, Version 14.0 vorgestellt. Zum anderen wurden den Studierenden noch weitere modulübergreifende IMPP-Fragen präsentiert.

Unter den Studierenden, die regelmäßig die Fragenpakete beantwortet hatten, wurden leistungsunabhängig 20 Büchergutscheine einer akademischen Buchhandlung im Wert von je 20 Euro verlost. Zusätzlich wurde Studierenden, die regelmäßig teilnahmen, ein entsprechendes Zertifikat ausgestellt.

Am Ende des Semesters wurden EvaSys®Fragebögen zur Evaluation des Projektes verteilt (siehe Kapitel 2.3.2). In der anschließenden vorlesungsfreien Zeit (Phase drei, siehe Abbildung 3) wurden vier qualitative Interviews in Form von Fokusgruppengesprächen durchgeführt und analysiert (siehe Kapitel 2.3.3).

2.2.2 Rekrutierung der Teilnehmer und Datensammlung

Zwei Wochen vor Beginn des Sommersemesters 2018 wurden alle Studierenden des dritten bis sechsten klinischen Semesters per E-Mail zur Teilnahme an der Studie eingeladen. Die E-Mail beinhaltete allgemeine Informationen zum Projekt in Form eines Anschreibens. Des Weiteren wurden die Studierenden über das Studiendesign und die Studienziele informiert, die Probandeninformation wurde der E-Mail als PDF-Format beigelegt. In den Einführungsveranstaltungen zu Semesterbeginn der ersten Module wurde das Projekt *#clue* in den betreffenden vier Semestern erneut vorgestellt. Außerdem wurden die Probandeninformation und Einverständniserklärung ausgeteilt und Letztere nach dem Ausfüllen durch die Studierenden wieder eingesammelt. Zusätzlich wurden die Studierenden über die Freiwilligkeit der Teilnahme an der Studie und über die Möglichkeit des Widerrufs ihrer Einwilligung ohne Angabe von Gründen hingewiesen. Während des ganzen Semesters hatten die Studierenden die Möglichkeit, den Modulkoordinator per E-Mail zu kontaktieren und weitere Informationen einzuholen. Studierenden, die ihre Daten nicht zur Verfügung stellen wollten, entstand dadurch kein Nachteil und sie konnten uneingeschränkt das Lehrangebot nutzen.

2.3 Erhebungsinstrumente

2.3.1 Quantitative Datenanalyse der Fragenpakete von ILIAS mit Excel und SPSS®

Die Datensätze der Personen, die eine Einverständniserklärung abgegeben hatten, wurden von der Lernplattform ILIAS heruntergeladen und anschließend zunächst pseudonymisiert. Hierbei wurde mittels eines Schlüssels jeder Name durch eine Nummer ersetzt, um eine longitudinale Analyse zu ermöglichen. Diese Nummer enthielt keine identifizierenden Informationen mehr. Nach Abschluss der Datenerhebung wurde der Schlüssel für die Zuordnung zwischen Namen und Nummern gelöscht. Dadurch entstand ein vollständig anonymisierter Datensatz. In diesem Datensatz war nicht mehr ersichtlich, welche Daten zu welchem individuellen Studienteilnehmer gehörten. Anschließend wurden die Daten sowohl mit dem Programm Microsoft® Excel für Mac, Version 16.23, als auch mit dem Programm

IBM® SPSS® Statistics, Version 24 (IBM Corp., Armonk, New York, USA) quantitativ analysiert.

2.3.2 Qualitative Datenanalyse der EvaSys®-Fragebögen am Ende des Semesters

Das jeweils dritte, vierte und fünfte klinische Semester füllte am Semesterende einen EvaSys®-Fragebogen (Electric Paper Evaluationssysteme GmbH, Lüneburg) aus. Die Beantwortung erfolgte anonym und papierbasiert.

Insgesamt setzte sich der Bogen aus fünf verschiedenen Abschnitten zusammen und beinhaltete 24 Fragen. Zunächst wurden die Studienteilnehmer um Angaben zu ihrem Alter, dem Geschlecht und ihrem klinischen Semester gebeten (siehe Abbildung 5).

1. Persönliche Angaben			
Wie alt sind Sie?			
<input type="text"/>			
Geschlecht	<input type="checkbox"/> männlich	<input type="checkbox"/> weiblich	<input type="checkbox"/> anderes Geschlecht
klinisches Semester	<input type="checkbox"/> 3.	<input type="checkbox"/> 4.	<input type="checkbox"/> 5.
	<input type="checkbox"/> 6.		

Abbildung 5: Fragebogen – Persönliche Angaben. Die Studienteilnehmer wurden gebeten ihr Alter in Ziffern einzutragen und die zutreffende Antwort zu ihrem Geschlecht und Semester anzukreuzen.

Anschließend wurde im zweiten Abschnitt nach der individuelle Teilnahmehäufigkeit an der Beantwortung der wöchentlichen Fragenpakete und den interdisziplinären Lehrveranstaltungen gefragt. In diesen beiden Abschnitten fanden vor allem geschlossene Fragen Anwendung. Die Studienteilnehmer hatten die Möglichkeit aus drei vorgegebenen Antworten die Zutreffendste auszuwählen (siehe Abbildung 6).

2. Teilnahme am Projekt			
Haben Sie an der Beantwortung der wöchentliche Fragenpakete teilgenommen?	<input type="checkbox"/> ja, immer	<input type="checkbox"/> ja, manchmal	<input type="checkbox"/> nein
Haben Sie an der interdisziplinären Lehrveranstaltung teilgenommen?	<input type="checkbox"/> ja, immer	<input type="checkbox"/> ja, manchmal	<input type="checkbox"/> nein

Abbildung 6: Fragebogen – Teilnahme am Projekt. Die Studienteilnehmer wurden gebeten hier ihre persönliche Einschätzung hinsichtlich ihrer eigenen Teilnahmehäufigkeit einzutragen.

Der dritte und vierte Abschnitt bezog sich auf den Inhalt und die konkrete Gestaltung der „Wöchentlichen Fragenpakete“ und der „Interdisziplinären Veranstaltungen“ (siehe Abbildung 7 und Abbildung 8). Die Studienteilnehmer konnten hier ihre Meinung auf einer sechsstufigen Likert-Skala äußern (eins = „trifft voll zu“, sechs = „trifft überhaupt nicht zu“). Voraussetzung für die Beantwortung dieser beiden Abschnitte war, dass der Studienteilnehmer mindestens einmal das Lehrangebot wahrgenommen hatte.

3. Wöchentliche Fragenpakete (nur wenn Sie an der Beantwortung mindestens einmal teilgenommen haben)						
Der Stoffumfang der wöchentlichen Fragenpakete war angemessen	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> trifft überhaupt nicht zu
Die wöchentlichen Mails waren als Erinnerungstütze hilfreich	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> trifft überhaupt nicht zu
Die semesterbegleitende Wiederholung von Lerninhalten aus vorangegangenen Semestern ist sinnvoll	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> trifft überhaupt nicht zu
Die wöchentlichen Fragenpakete waren dem zeitlichen und inhaltlichen Rahmen der Module angemessen	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> trifft überhaupt nicht zu
Die Bearbeitung der wöchentlichen Fragenpakete hat mir geholfen, Inhalte aus verschiedenen Modulen interdisziplinär miteinander zu verknüpfen	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> trifft überhaupt nicht zu
Die Beantwortung der Fragenpakete hat wesentlich zu meinem Lernerfolg beigetragen	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> trifft überhaupt nicht zu
Die Beantwortung der wöchentlichen Fragenpakete hat mich dazu angeregt, mich intensiver mit den abgefragten Themen auseinanderzusetzen	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> trifft überhaupt nicht zu
Die Bearbeitung der wöchentlichen Fragenpakete hat mich bei der Wiederholung der Modulinhalte unterstützt	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> trifft überhaupt nicht zu
Das Anforderungsniveau der wöchentlichen Fragenpakete war	viel zu hoch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> viel zu gering

Abbildung 7: Fragebogen – Wöchentliche Fragenpakete. Die Studienteilnehmer wurden gebeten auf einer sechsstufigen Likert-Skala die wöchentlichen Fragenpakete hinsichtlich ihrer inhaltlichen Gestaltung und der organisatorischen Durchführung zu bewerten. Das Anforderungsniveau konnten die Studienteilnehmer durch eine fünfstufige Likert-Skala bewerten.

4. Interdisziplinäre Lehrveranstaltungen (nur wenn Sie an den Lehrveranstaltungen mindestens einmal teilgenommen haben)	
Der Stoffumfang der interdisziplinären Lehrveranstaltungen war angemessen	trifft voll zu <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> trifft überhaupt nicht zu
Die interdisziplinären Lehrveranstaltungen haben bei mir zu einem tieferen Verständnis der geprüften Themen geführt	trifft voll zu <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> trifft überhaupt nicht zu
Der Besuch der interdisziplinären Lehrveranstaltungen hat mich dazu angeregt, mich intensiver mit den abgefragten Themen auseinanderzusetzen	trifft voll zu <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> trifft überhaupt nicht zu
Die interdisziplinären Lehrveranstaltungen haben wesentlich zu meinem Lernerfolg beigetragen	trifft voll zu <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> trifft überhaupt nicht zu
Das Anforderungsniveau der interdisziplinären Lehrveranstaltungen war	viel zu hoch <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> viel zu gering
In der interdisziplinären Lehrveranstaltung hätte mehr ... stattfinden sollen. (Mehrfachnennungen sind möglich)	
<input type="checkbox"/> Diskussion der geprüften Fragen <input type="checkbox"/> Vorstellung der Fragelösungen <input type="checkbox"/> Interaktion mit den Studierenden	

Abbildung 8: Fragebogen – Interdisziplinäre Lehrveranstaltungen. Auf dieser sechsstufigen Likert-Skala konnten die Studienteilnehmer ihre Meinung zur Gestaltung der interdisziplinären Lehrveranstaltungen äußern.

5. Gesamturteil	
Das Projekt hat mein Interesse an Interdisziplinarität gefördert	trifft voll zu <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> trifft überhaupt nicht zu
Ich würde im nächsten Semester wieder am Projekt teilnehmen	trifft voll zu <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> trifft überhaupt nicht zu
Ich würde das Projekt weiterempfehlen	trifft voll zu <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> trifft überhaupt nicht zu
In einem Gesamturteil bewerte ich das Projekt als	sehr gut <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> sehr schlecht
Was haben Sie als Stärken und Schwächen des Projekts #clue empfunden? Was war gut, was kann man besser machen? Warum haben Sie ggf. nicht teilgenommen? Bitte schreiben Sie innerhalb des umrandeten Bereichs.	
Vielen Dank für Ihre Teilnahme an der Umfrage!	

Abbildung 9: Fragebogen – Gesamturteil. An dieser Stelle konnten die Studienteilnehmer auf einer sechsstufigen Likert-Skala z. B. angeben, ob #clue ihr Interesse an Interdisziplinarität beeinflusst hat. Zusätzlich konnten sie noch eigene Gedanken in einem Freitextfeld formulieren.

Im fünften Abschnitt konnten die Studienteilnehmer das Projekt #clue in einem „Gesamturteil“ erneut auf einer sechsstufigen Likert-Skala bewerten und hatten anschließend die Möglichkeit, in einem Freitextfeld weitere Kritik und Anregungen anzubringen (siehe Abbildung 9).

2.3.3 Qualitative Datenanalyse der Fokusgruppeninterviews

Die Fokusgruppengespräche wurden in Form eines halb-standardisierten Interviews durchgeführt. Hierzu wurde vorab ein Fragenkatalog (siehe Anhang 1) erstellt. Als Anhaltspunkt für mutmaßliche Interessenschwerpunkte der Studienteilnehmer dienten die Freitextkommentare der Fragebögen. Der Fragenkatalog diente der Moderatorin als Leitfaden während des Interviews. Beim Erstellen der Fragen wurde darauf geachtet, dass diese möglichst offengehalten wurden. Die exakte Reihenfolge wurde nicht streng festgelegt und konnte im Verlauf des Interviews variiert werden, sofern alle relevanten Punkte erhoben wurden. Im Vorfeld fand eine Pilotierung des Fragenkatalogs mit anschließender Anpassung statt. Der Fragenkatalog selbst befasste sich mit vier Hauptthemen: „Allgemeines“, „Wöchentliche Fragenpakete“, „Interdisziplinäre Lehrveranstaltungen“ und „Verbesserungsvorschläge“.

Die Interviewteilnehmer wurden durch Beschäftigte der Medizindidaktik direkt angesprochen und des Weiteren durch „Mund-zu-Mund-Propaganda“ gewonnen. Aufgrund der DSGVO vom 25. Mai 2018 war es nicht möglich Studierende per E-Mail zur Teilnahme einzuladen. Die Teilnahme selbst war freiwillig. Organisatorisch war es nicht möglich, die Teilnehmergruppen nach Semestern getrennt zu gestalten. Es wurde darauf geachtet, dass aus jedem Semester im Verlauf aller Interviews mindestens zwei Teilnehmer vertreten waren. Außerdem wurden Teilnehmer berücksichtigt, welche *#clue* unterschiedlich intensiv genutzt hatten, um eine möglichst breit gefächerte Meinung der Studienteilnehmer zu erhalten. Die Terminabsprache fand immer im Einzelnen direkt mit den Teilnehmern statt (siehe Tabelle 1). Es wurden so viele Gespräche durchgeführt, bis sich die Meinungen der Teilnehmer stark zu wiederholen begannen und keine neuen Faktoren mehr zur Sprache gekommen sind. Die Fokusgruppengespräche wurden im Büro des Bereichs Medizindidaktik und Ausbildungsforschung der UMG geführt. Vor Beginn des eigentlichen Interviews gab es eine Einleitung, in der sich die Moderatorin, eine Mitarbeiterin des Bereichs Medizindidaktik und Ausbildungsforschung der UMG, kurz vorstellte. Die Moderatorin arbeitete unabhängig von *#clue* und war im Vorfeld nicht in das Projekt involviert. Die Teilnehmer des Interviews wurden darüber aufgeklärt, dass das Interview einer qualitativen Analyse des Projektes diene und elektronisch-akustisch aufgezeichnet wurde. Die Transkription und Auswertung wurde ausschließlich durch eine in der Medizindidaktik beschäftigte Person vorgenommen. Zu Beginn des Gesprächs wurden Teilnehmerinformationen und Einverständniserklärungen ausgeteilt und darauf hingewiesen, dass die Teilnahme freiwillig war. Außerdem wurden die Teilnehmer über den groben Ablauf und den Umgang mit den Daten aufgeklärt. Nach der Unterzeichnung der Einverständniserklärungen wurden die Gesprächsregeln erläutert und die Diskussion eingeleitet, indem die Moderatorin die Teilnehmer fragte, was ihrer Meinung nach das Ziel von *#clue* sei. Im Verlauf des Gesprächs regte die Moderatorin die Studienteilnehmer unter anderem dazu an, offen Kritik zu äußern und die Vor- und Nachteile von *#clue* gegenüber anderen Lehrformaten zu diskutieren.

Tabelle 1: Organisation der Fokusgruppen

Gruppe	Datum des Gesprächs	Uhrzeit	Dauer in min.	Teilnehmerzahl	Semester
1	06.08.2018	18.00 Uhr	40:47	5 Personen	1x 3. klin. Semester 1x 4. klin. Semester 3x 6. klin. Semester
2	13.08.2018	19.00 Uhr	35:29	6 Personen	4x 3. klin. Semester 2x 4. klin. Semester
3	21.08.2018	16.30 Uhr	37:24	4 Personen	2x 3. klin. Semester 1x 4. klin. Semester 1x 5. klin. Semester
4	29.08.2018	16.30 Uhr	35:14	5 Personen	4x 4. klin. Semester 1x 5. klin. Semester

Die Fokusgruppeninterviews fanden in vier Gruppen an vier verschiedenen Terminen statt. Insgesamt haben 20 Personen aus vier verschiedenen Semestern teilgenommen.

2.4 Datenauswertung

2.4.1 Statistische Analyse

Im Folgenden wird beschrieben, wie die statistische Auswertung erfolgte. Das Signifikanzniveau wurde auf $0,05 = 5\%$ festgelegt. Die Studienfragen eins und zwei kohärieren miteinander, weshalb die statistischen Analysen und Ergebnisse teilweise in gemeinsamen Abbildungen präsentiert werden. Es wurde eine gesonderte Excel-Tabelle der Teilnehmer des *Exit Exams* des dritten bis sechsten klinischen Semesters erstellt (siehe Abbildung 10). Die Ergebnisse der Interventionsitems zu den Zeitpunkten T1, T2 und T3 und der Kontrollitems zum Zeitpunkt T3 wurden aufgelistet und dichotomisiert. In der primären Analyse wurden die Datensätze der Studienteilnehmer für jede richtig beantwortete Frage mit jeweils einem Punkt versehen. Wenn ein Studienteilnehmer ein Item nicht beantwortet hat, wurde dieses Item mit null Punkten bewertet. Für jeden Studienteilnehmer wurden verschiedene Arten von sog. Summenscores berechnet. Bei der ersten Summscore-Art wurden die drei Schwierigkeitsgrade zusammengefasst (*Overall*), während die zweite Summscore-Art für die jeweiligen Schwierigkeitsgrade (leicht, mittelschwer, schwer) getrennt berechnet wurde.

Für die erste Art, den *Overall*-Summscore, wurde die Anzahl richtig beantworteter Items ins Verhältnis zur Gesamtzahl der Items gesetzt, ausgedrückt in Prozent. Die *Overall*-Summenscores wurden für die Interventionsitems zu den Zeitpunkten T1, T2 und T3 und

für die Kontrollitems zum Zeitpunkt T3 unter Einbeziehung der drei Schwierigkeitsgrade berechnet. Die *Overall*-Summenscores setzten sich jeweils aus fünf Items zusammen.

Darüberhinausgehend wurde die zweite Art von Summenscores abhängig vom Schwierigkeitsgrad (leicht, mittelschwer, schwer) bestimmt. Bei den unterschiedlichen Semestern variierte das Angebot der Interventionsitems hinsichtlich der angebotenen Schwierigkeitsgrade. Somit setzen sich diese Summenscores aus einer unterschiedlichen Anzahl an Items pro Schwierigkeitsgrad (je nach Semester, siehe Abbildung 4) zusammen. In Abbildung 10 ist die Berechnung für das dritte klinische Semester beispielhaft dargestellt. Die Berechnung der leichten Summenscores setzte sich aus drei Items zusammen, während sich die Berechnung der mittelschweren und schweren Summenscores aus jeweils einem Item zusammensetzte.

Für die deskriptive Analyse wurden, wenn nicht anders angegeben, Summscore-Mittelwerte und entsprechende Standardfehler aus den Summenscores aller Studienteilnehmer bestimmt. Hierfür wurden für die unterschiedlichen Zeitpunkte T1, T2 und T3 jeweils die Mittelwerte und Standardfehler der *Overall*-Summenscores und der schwierigkeitsgradgetrennten Summenscores für alle Studienteilnehmer bestimmt. In die Analyse wurden alle Teilnehmer des *Exit Exams* unabhängig vom Semester mit einbezogen.

	I T1	I T2	I T3	K T3	I-T1	I-T2	I-T3	K T3	I-T1	I-T2	I-T3	K T3	I-T1	I-T2	I-T3	K T3
erhobene Daten	1	1	1	1	1	1	1	1								
	0	1	1	1	1	1	1	1								
	0	0	0	0	33%	66%	66%	100%								
	0	1	1	1	0				0%	100%	100%	0%				
	0	1	1	1									0%	100%	100%	100%
Summscore	20%	80%	80%	80%	33%	66%	66%	100%	0%	100%	100%	0%	0%	100%	100%	100%
Bezeichnung	I-T1 overall	I-T2 overall	I-T3 overall	K-T3 overall	I-T1 (l)	I-T2 (l)	I-T3 (l)	K-T3 (l)	I-T1 (m)	I-T2 (m)	I-T3 (m)	K-T3 (m)	I-T1 (s)	I-T2 (s)	I-T3 (s)	K-T3 (s)

Abbildung 10: Berechnung der Summenscores eines Beispiel-Studienteilnehmers. Dargestellt ist die beispielhafte Berechnung der Summenscores (= hell-gelb markierte Prozentzahlen). Im Verlauf des Semesters wurden fünf Interventionsitems zu den Zeitpunkten T1, T2 und T3 und fünf Kontrollitems zum Zeitpunkt T3 präsentiert. Die erhobenen Daten wurden dichotomisiert (Item richtig beantwortet = eins oder falsch beantwortet = null). Der schwarze Füllbalken allegorisiert die Nichtbeantwortung des dritten leichten Interventionsitems zum Zeitpunkt T1 und geht in die Berechnung der Summenscores I-T1 mit null Punkten ein. Zur Berechnung der Summenscores wurde die Anzahl richtig beantworteter Items ins Verhältnis zur Gesamtzahl der Items gesetzt, ausgedrückt in Prozent. Es wurden Summenscores unter Einbeziehung der drei Schwierigkeitsgrade (*Overall*) und für die unterschiedlichen Schwierigkeitsgrade (leicht (l) = grün, mittelschwer (m) = gelb, schwer (s) = rot) zu den Zeitpunkten T1, T2 und T3 für die Interventionsitems und zu dem Zeitpunkt T3 für die Kontrollitems bestimmt.

Zur Beantwortung der ersten Studienfrage wurden die Daten deskriptiv mit Hilfe des Programms Microsoft® Excel für Windows, Version 16 ausgewertet. Das Ziel dieser Studienfrage ist die Gegenüberstellung der Interventionsitems zum Zeitpunkt T3 (Items wurden das dritte Mal abgefragt) mit den Kontrollitems zum Zeitpunkt T3 (Items wurden das erste Mal abgefragt). Für die Beantwortung dieser Frage wurden die Summscore-Mittelwerte und Standardfehler zum Zeitpunkt T3 verwendet. Die Nullhypothese („Es existiert kein signifikanter Unterschied zwischen den *Overall*-Summenscores der

Kontrollitems und den *Overall*-Summenscores der Interventionsitems.“) wurde mit Hilfe des Statistikprogramms IBM® SPSS® Statistics, Version 24 (IBM Corp., Armonk, New York, USA) überprüft. In einer explorativen Analyse kam zunächst der Kolmogorov-Smirnov-Anpassungstest zur Anwendung. Im Anschluss daran wurden, in Abhängigkeit vom Ergebnis, die Daten mittels eines gepaarten T-Tests oder eines Wilcoxon-Tests weiter ausgewertet. Zum einen wurde untersucht, ob sich ein Unterschied zwischen den Summenscores der Kontrollitems zum Zeitpunkt T3 und den Summenscores der Interventionsitems zum Zeitpunkt T3 unter Einbeziehung der drei Schwierigkeitsgrade (*Overall*) feststellen ließ (siehe Abbildung 10). Zum anderen wurde geprüft, ob sich signifikante Unterschiede zwischen den Summenscores abhängig vom Schwierigkeitsgrad zwischen Interventionsitems und Kontrollitems zum Zeitpunkt T3 (leicht, mittelschwer, schwer) feststellen ließen. Darüberhinausgehend wurde untersucht, ob das Anforderungsniveau der fünf Kontroll- und fünf Interventionsitems allgemein vergleichbar war. Hierfür wurden die *Overall*-Summenscores von den Kontrollitems zum Zeitpunkt T3 und den Interventionsitems zum Zeitpunkt T1 auf signifikante Unterschiede untersucht. Hierbei wurde der Wilcoxon-Test verwendet. Des Weiteren wurden auch die Summenscores der unterschiedlichen Schwierigkeitsgrade der Interventionsitems zum Zeitpunkt T1 und der Kontrollitems zum Zeitpunkt T3 jeweils im Wilcoxon-Test miteinander verglichen, um zu untersuchen, ob den Studienteilnehmern zum Zeitpunkt T3 die Beantwortung von neuen Fragen genauso leicht bzw. schwer fällt wie zum Zeitpunkt T1.

Um die zweite Studienfrage zu beantworten, wurden die wöchentlichen Ergebnisse der Studienteilnehmer, die regelmäßig teilgenommen hatten (d. h. ein Studienteilnehmer musste mindestens 75% der ihm zur Verfügung stehenden Fragenpakete bearbeitet haben), semestergetrennt aufgelistet. An dieser Stelle wurden sogenannte Wochenscores mit dem Programm Microsoft® Excel für Mac, Version 16.23 bestimmt. Ein Wochenscore ist der Mittelwert aus der erreichten Punktzahl der Studienteilnehmer eines Semesters pro Woche. Anschließend wurden die Wochenscores der verschiedenen Semester über die Zeit graphisch aufgetragen. Zusätzlich wurden jeweils für die regelmäßigen und die unregelmäßigen Teilnehmer über alle Semester hinweg gemittelte Wochenscores bestimmt, graphisch dargestellt und anschließend deskriptiv beschrieben.

In einer primären Analyse (Subskript PA) wurden deskriptiv die Summscore-Mittelwerte und Standardfehler (*Overall*) der Interventionsitems zu den Zeitpunkten T1, T2 und T3 von den Teilnehmern am *Exit Exam* verwendet (siehe Abbildung 10). Durch den Friedman-Test (Subskript F) und dem sich ggf. anschließenden Wilcoxon-Test (Subskript W) wurde untersucht, ob sich die signifikanten Unterschiede der Summenscores zwischen T1 und T2, zwischen T2 und T3 oder zwischen T1 und T3 feststellen ließen.

Zusätzlich wurden für die Beantwortung der Studienfrage außerdem die Summscore-Mittelwerte im zeitlichen Verlauf (T1, T2, T3) für leichte, mittelschwere oder schwere Interventionsitems der *Exit Exam* Teilnehmer verwendet. Durch den Friedmann-Test und dem

sich ggf. anschließenden Wilcoxon-Test wurde untersucht, ob sich die signifikanten Unterschiede der Summenscores zwischen T1 und T2, zwischen T2 und T3 oder zwischen T1 und T3 feststellen ließen.

In einer nachfolgenden Sensitivitätsanalyse (Subskript SA) wurden komplette Interventionsitems zu den Zeitpunkten T1, T2 und T3 von Studienteilnehmern aus den Summenscores genommen, die zu mindestens zu einem Zeitpunkt (T1, T2 oder T3) ein Item nicht beantwortet hatten und somit unvollständige Datensätze aufwiesen. In dem oben aufgeführten Beispiel (siehe Abbildung 10) wird ein nicht beantwortetes Interventionsitem durch einen schwarzen Füllbalken dargestellt. An dieser Stelle würden sich somit die Summenscores der leichten Interventionsitems nur aus den ersten beiden und nicht aus den ursprünglichen drei Items und die *Overall*-Summenscores der Interventionsitems nur aus vier Items (zwei leichte Items, ein mittelschweres Item, ein schweres Item) zusammensetzen. Die Summenscores der Kontrollitems setzten sich nach wie vor aus allen Items des *Exit Exams* zusammen. Um einen Wilcoxon-Test durchführen zu können, müssen immer zwei entsprechende Summenscores von einer Person miteinander gepaart werden. Bei komplett fehlenden Summenscores eines Schwierigkeitsgrades der Interventionsitems zu T1 oder T2 bei einer Person wäre es somit nicht möglich ein Testpaar für diese Person zu bilden. Aus diesem Grund wurden die Summenscores der Kontrollitems für einen Schwierigkeitsgrad nicht weiter in der statistischen Auswertung betrachtet, wenn zu T1 oder T2 kein Summenscore für die Interventionsitems des entsprechenden Schwierigkeitsgrades gebildet werden konnte. Mit diesem überarbeiteten Datensatz wurde die primäre Analyse überprüft.

Um die dritte Studienfrage zu beantworten wurde die wöchentliche Teilnehmerzahl von den Studienteilnehmern an den Fragenpaketen untersucht. Hierfür wurden mit Hilfe einer Excel-Tabelle die Teilnehmerzahlen graphisch über die Zeit aufgetragen und deskriptiv beschrieben. Für jedes Semester wurde pro Woche die prozentuale Teilnahmerate der Studienteilnehmer semestergetrennt berechnet. Im Anschluss daran wurde der Mittelwert der prozentualen Teilnahmerate über den kompletten Projektverlauf für jedes Semester berechnet.

Zur Beantwortung der vierten Studienfrage wurden die EvaSys[®]-Fragebögen des dritten, vierten und fünften klinischen Semesters zusammengefasst, deskriptiv ausgewertet und anschließend tabellarisch dargestellt. Es wurden die Skalenoptionen eins und zwei der Likert-Skala zusammengefasst und als positive Rückmeldung gewertet. Bei den Aussagen, die die Anforderungsniveaus betrafen, wurden die mittleren Skalenoptionen als ausgewogenes Anforderungsniveau gewertet. Da im fünften klinischen Semester die interdisziplinäre Veranstaltung von den Studienteilnehmern nicht besucht wurde, wurden Bewertungen aus diesem Semester bzgl. der interdisziplinären Lehrveranstaltung nicht berücksichtigt.

2.4.2 Qualitative Analyse

Zusätzlich wurden zur Beantwortung der vierten Studienfrage die Fokusgruppeninterviews ausgewertet.

Die Interviews wurden elektronisch-akustisch aufgezeichnet und im Anschluss in Anlehnung an die Regeln von Dresing und Pehl (2013) mit Hilfe der Software MaxQDA (VERBI Software GmbH, Berlin) zunächst transkribiert. Im Anschluss wurden die Transkripte ebenfalls mit MaxQDA ausgewertet. Bei der Verschriftlichung wurden folgende Regeln angewandt: Wortwiederholungen wurden mit aufgenommen und auch abgebrochene Sätze wurden erfasst. Wortverschleifungen wurden geglättet und in korrektes Hochdeutsch umgeschrieben. Längere Pausen wurden ebenfalls durch Punkte in Klammern transkribiert (1 Sekunde = (.), 2 Sekunden = (..), 3 Sekunden = (...)). Besonders betonte Worte wurden in Großbuchstaben festgehalten. Einwürfe von anderen Gesprächspartnern wurden an entsprechenden Stellen mit zwei Querstrichen gekennzeichnet.

Während des Interviews war eine weitere Mitarbeiterin der Medizindidaktik anwesend, um Beobachtungsdaten wie Lautäußerungen und Mimik festzuhalten. Diese Beobachtungsdaten wurden in Kursiv-Schrift in die Transkription eingearbeitet. In einem zweiten Durchgang wurde das Transkript noch einmal überprüft und korrekte Satzstellungen gebildet. Im Anschluss wurde die Audio-Datei gelöscht.

Die Fokusgruppeninterviews wurden mittels einer strukturierten qualitativen Inhaltsanalyse in Anlehnung nach Mayring (2010) mit dem Programm MaxQDA von zwei voneinander unabhängigen Personen ausgewertet. Hierbei fand ein Kodierungssystem nach Kategorien Anwendung. Es wurden zunächst deduktive Ober- und Unterkategorien (sog. „Codes“) aus dem Leitfaden und der Pilotierung erstellt und diese anschließend während der Analyse mit induktiven *Codes* ergänzt. Insgesamt wurden deduktiv fünf Oberkategorien gebildet („Allgemeines“, „Fragenpakete“, „Interdisziplinäre Lehrveranstaltungen“, „Verbesserungsvorschläge“, „Thematischer Exkurs“). Die Inhalte des „Thematischen Exkurs“ hatten in der Analyse keine weitere Bedeutung. Die induktiven *Codes* reichten bis zu fünf Ebenen tief. Die Textstellen des Interviews wurden den *Codes* zugeordnet (sog. „Coding“). Parallel dazu wurde ein Kodierleitfaden mit allen Oberkategorien und Unterkategorien angelegt. In diesem wurden Regeln aufgestellt, um eine Textstelle einem *Code* zuordnen zu können (siehe Anhang 2). Außerdem wurden Beispiele mit besonders passenden Zitaten festgehalten. Nach dem Vergleich der Ergebnisse beider Personen und der Zusammenfassung des Kategoriensystems wurden die Themenschwerpunkte der Fokusgruppengespräche identifiziert. Die Ergebnisse wurden in einer MindMap (XMind, Ltd, Hongkong) zusammengefasst (siehe Anhang 3 bis Anhang 7).

3 Ergebnisse

3.1 Charakterisierung der Studiengruppe

Im Sommersemester 2018 nahmen 149 Studierende am dritten klinischen Semester, 145 Studierende am vierten klinischen Semester, 147 Studierende am fünften klinischen Semester und 132 Studierende am sechsten klinischen Semester teil. Insgesamt ergab sich somit eine potenzielle Gesamtteilnehmerzahl von 573 Studierenden. Hiervon willigten 141 des dritten klinischen Semesters, 75 des vierten klinischen Semesters, 99 des fünften klinischen Semesters und 90 Studierende des sechsten klinischen Semesters ein, an der Studie teilzunehmen. Daraus resultierte eine Studienteilnehmerzahl von 405 Personen (Rücklaufquote 70,7%). Die Rücklaufquote hinsichtlich der Nutzung der Fragenpakete zu T3 betrug 13,1%.

3.2 Entwicklung der Teilnehmerzahlen

Zur Beantwortung der dritten Studienfrage (Wie entwickeln sich die Teilnehmerzahlen bei einem freiwilligen Projekt, das zum Verbessern des eigenen Lernerfolgs angeboten wird?) wurden die wöchentlichen Teilnehmerzahlen graphisch dargestellt (siehe Abbildung 11).

Das dritte klinische Semester zeigte fast durchgehend eine höhere Teilnahmerate im Vergleich zu den anderen Semestern. Im Verlauf der 14 Wochen nahmen durchschnittlich 53% der 141 Studienteilnehmer an der Bearbeitung der Fragenpakete teil.

Da im vierten klinischen Semester erst ab der zweiten Semesterwoche Fragenpakete zur Verfügung gestellt wurden, existierten für die erste Semesterwoche keine Teilnehmerzahlen. 40% der 75 Studienteilnehmer haben durchschnittlich die 13 Fragenpakete bearbeitet.

28% der 99 Studienteilnehmer aus dem fünften klinischen Semester haben durchschnittlich die 14 Fragenpakete bearbeitet.

Das sechste klinische Semester bestand lediglich aus zehn Wochen, weshalb hier keine Daten mehr ab der elften Woche erhoben wurden. Im Verlauf dieser zehn Wochen haben durchschnittlich 47% der Studienteilnehmer die Fragenpakete bearbeitet.

Im Verlauf der 14 Wochen ließ sich ein annähernd kontinuierlicher Verlust der Teilnehmerzahlen in den teilnehmenden Semestern dokumentieren. In Woche zehn konnte ein Minimum an Teilnehmern (aus den vier Semestern bearbeiteten insgesamt 61 Personen die Fragenpakete) festgestellt werden.

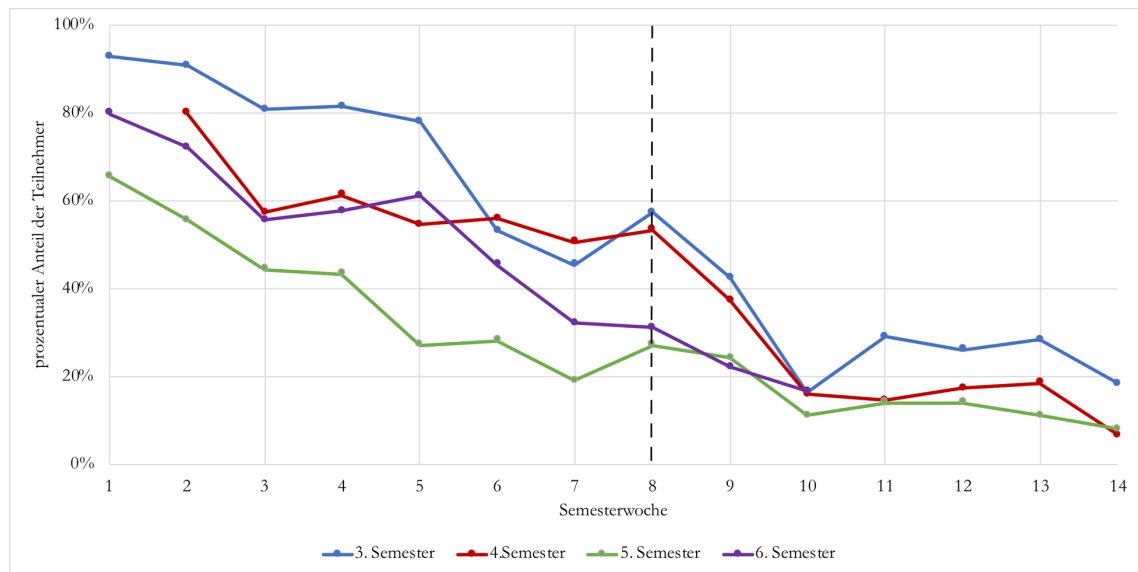


Abbildung 11: Teilnehmerzahlen. Dargestellt sind die prozentualen Teilnehmerzahlen vom dritten bis zum sechsten klinischen Semester während der Datenerhebung in der zweiten Projektphase. Bis einschließlich zur achten Woche wurden die Erinnerungs-E-Mails automatisch versendet. Durch die DSGVO mussten sich anschließend die Studierenden, die auch weiterhin per E-Mail erinnert werden wollten, aktiv für die Erinnerungs-E-Mail anmelden.

3.3 Vergleich des Lernerfolgs zwischen mehrfach wiederholten und neuen Inhalten

Die erste Studienfrage (Führt die mehrfache Wiederholung von Inhalten während des Semesters bezüglich dieser Inhalte zu einem größeren Lernerfolg als bei Inhalten, die nicht mit Fragen wiederholt wurden?) wurde mit Hilfe der Summenscore-Mittelwerte des *Exit Exams* beantwortet.

Am *Exit Exam* nahmen insgesamt 53 Studierende teil, deren Summenscores zusammengefasst wurden. Jedem Studienteilnehmer wurden fünf Interventionsitems jeweils zu den Zeitpunkten T1, T2 und T3 präsentiert, sowie fünf Kontrollitems zum Zeitpunkt T3. Die *Overall*-Summenscore-Mittelwerte (mit Standardfehler) betragen für die Teilnehmer aller vier Semester zusammen zum Zeitpunkt T3 bei den Interventionsitems $78,9\% \pm 2,7\%$ und bei den Kontrollitems $74,7\% \pm 2,7\%$ (siehe Abbildung 14). Der Kolmogorov-Smirnov-Test (ohne Lilliefors-Korrektur) zeigte zum Zeitpunkt T3, dass keine Normalverteilung für die *Overall*-Summenscores der Interventions- und Kontrollitems zum Zeitpunkt T3 vorliegt ($p_{I-T3 \text{ overall}} = 0,002$; $p_{K-T3 \text{ overall}} = 0,001$; $n = 53$).

Der Unterschied zwischen den *Overall*-Summenscores der Kontrollitems und Interventionsitems zum Zeitpunkt T3 im Wilcoxon-Test war nicht signifikant ($p = 0,281$). Darüberhinausgehend wurden die Summenscores der Interventionsitems und Kontrollitems der verschiedenen Schwierigkeitsgrade zum Zeitpunkt T3 jeweils verglichen (siehe Abbildung 15). Die Summenscores der leichten Kontrollitems zum Zeitpunkt T3 (Summenscore-

Mittelwert: $91,4\% \pm 2,8\%$) waren nicht signifikant ($p_{PA-W} = 0,059$) höher im Wilcoxon-Test als die Summenscores der leichten Interventionsitems zum Zeitpunkt T3 (Summenscore-Mittelwert: $83,5\% \pm 3,5\%$). Die mittelschweren Kontrollitems wurden im *Exit Exam* mit einem Summenscore-Mittelwert von $57,4\% \pm 7,4\%$ im Wilcoxon-Test nicht signifikant schlechter beantwortet ($p_{PA-W} = 0,616$) als die Interventionsitems zum Zeitpunkt T3 ($63,5\% \pm 7,1\%$). Zum Zeitpunkt T3 konnte ein signifikanter Unterschied der schweren Summenscores zwischen den Interventionsitems (Summenscore-Mittelwert: $78,9\% \pm 5,4\%$) im Vergleich zu den Summenscores der Kontrollitems (Summenscore-Mittelwert: $53,3\% \pm 7,0\%$) im Wilcoxon-Test festgestellt werden ($p_{PA-W} = 0,014$).

Bezogen auf die Untersuchung des Anforderungsniveaus der Interventionsitems ließen sich folgende Beobachtungen machen. Hierfür wurden jeweils die *Overall*-Summenscores und die Summenscores der unterschiedlichen Schwierigkeitsgrade (leicht, mittelschwer, schwer) der Interventionsitems zum Zeitpunkt T1 mit den entsprechenden Summenscores der Kontrollitems zum Zeitpunkt T3 im Wilcoxon-Test miteinander gepaart. Es konnten in den *Overall*-Summenscores signifikante Unterschiede zwischen den jeweiligen Summenscores ($p_{PA-W} = 0,021$) sowie bei den leichten Items zwischen den jeweiligen Summenscores festgestellt werden ($p_{PA-W} = 0,001$). Bei den mittelschweren ($p_{PA-W} = 0,187$) und schweren ($p_{PA-W} = 0,839$) Items bestanden keine signifikanten Unterschiede zwischen den Summenscores.

Allgemein lässt sich bei der Beantwortung der Interventionsitems (die den Studienteilnehmern im Laufe des Semesters mehrfach präsentiert worden waren) und der Kontrollitems (die den Studienteilnehmern neu waren) im *Exit Exam* bei der Betrachtung der jeweiligen *Overall*-Summenscores somit kein signifikanter Unterschied im Vergleich feststellen. Durch die Aufschlüsselung nach Schwierigkeitsgraden lässt sich ein signifikanter Unterschied zwischen den schweren Summenscores der Interventionsitems und schweren Summenscores der Kontrollitems zum Zeitpunkt T3 feststellen.

3.4 Entwicklung des Leistungsstandes bei mehrfacher Wiederholung

Die zweite Studienfrage (*Wie entwickelt sich der studentische Leistungsstand im Laufe des Semesters, wenn Inhalte anhand von Fragen mehrfach wiederholt werden?*) wurde mit Hilfe der berechneten Wochenscores und Summenscores der vorhandenen Fragenpakete von Woche eins bis 14 beantwortet.

3.4.1 Allgemein

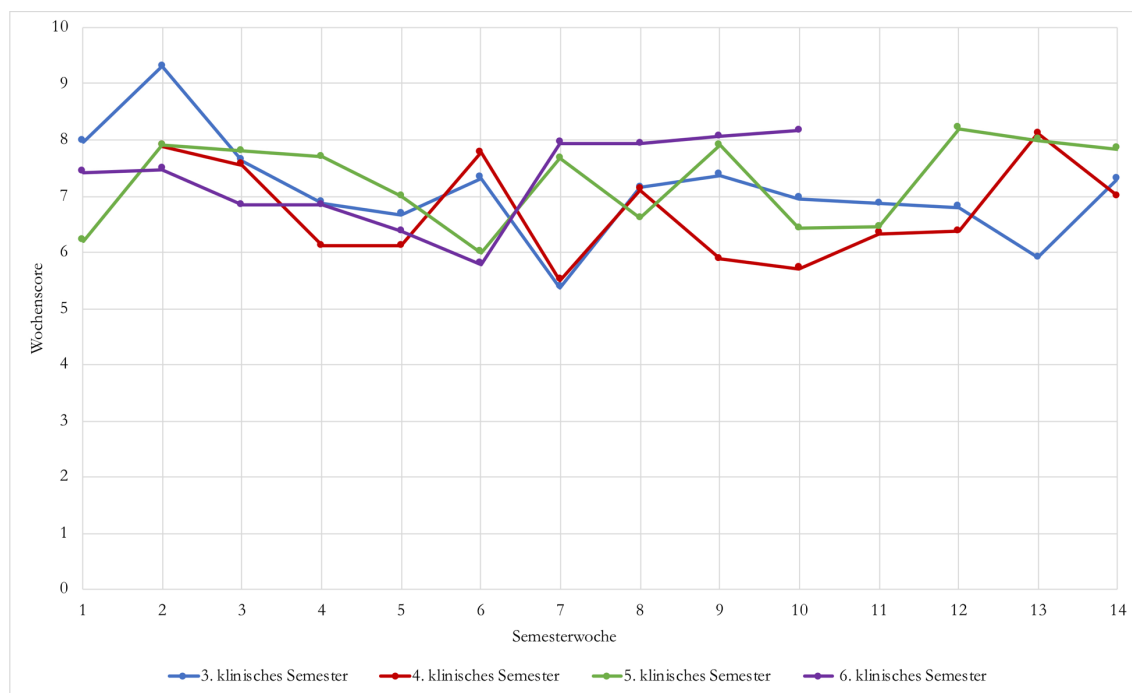


Abbildung 12: Durchschnittlich erreichter Wochenscore der Studienteilnehmer (semestergetrennt).

Während der zweiten Phase des Projekts hatten die Studierenden jede Woche die Möglichkeit, Fragenpakete bestehend aus zehn Fragen zu beantworten. Dem vierten klinischen Semester wurde ab der zweiten Semesterwoche Fragenpakete zur Verfügung gestellt. Das sechste klinische Semester bestand lediglich aus zehn Wochen. In dieser Abbildung sind die Wochenscores der Studienteilnehmer abgebildet, die regelmäßig (mindestens 75% der Fragenpakete mussten bearbeitet werden) teilgenommen haben. Ein Wochenscore setzt sich aus der durchschnittlich erreichten Punktzahl der Studienteilnehmer eines Semesters pro Woche zusammen. Im Fragenpaket der ersten Woche haben beispielsweise die Studienteilnehmer des dritten klinischen Semesters durchschnittlich 7,9 Punkte erreicht.

In Abbildung 12 sind die erreichten Durchschnittspunktzahlen (in Form von Wochenscores) der Studienteilnehmer semestergetrennt dargestellt, die regelmäßig die Fragenpakete beantwortet hatten. Im dritten klinischen Semester bearbeiteten 33 Studienteilnehmer, im vierten klinischen Semester bearbeiteten neun Studienteilnehmer, im fünften klinischen Semester bearbeiteten zehn Studienteilnehmer und im sechsten klinischen Semester bearbeiteten 19 Studienteilnehmer regelmäßig (mindestens 75% der zur Verfügung stehenden Fragenpakete mussten bearbeitet werden) die Fragenpakete. Bei deskriptiver Betrachtung der Abbildung 12 zeigt sich, dass im Verlauf der 14 Wochen bei den Studienteilnehmern im Hinblick auf den erreichten Wochenscore kein klarer Trend in eine Richtung erkennbar ist – insbesondere kein Aufwärtstrend zu besseren Wochenscores. In vereinzelten Wochen betrug der Wochenscore und damit der Durchschnittswert aller Studienteilnehmer eines Semesters weniger als 60%, was einem Durchschnitt unterhalb der Bestehensgrenze entsprechen würde (siehe z. B. Woche sieben: drittes und viertes klinisches Semester).

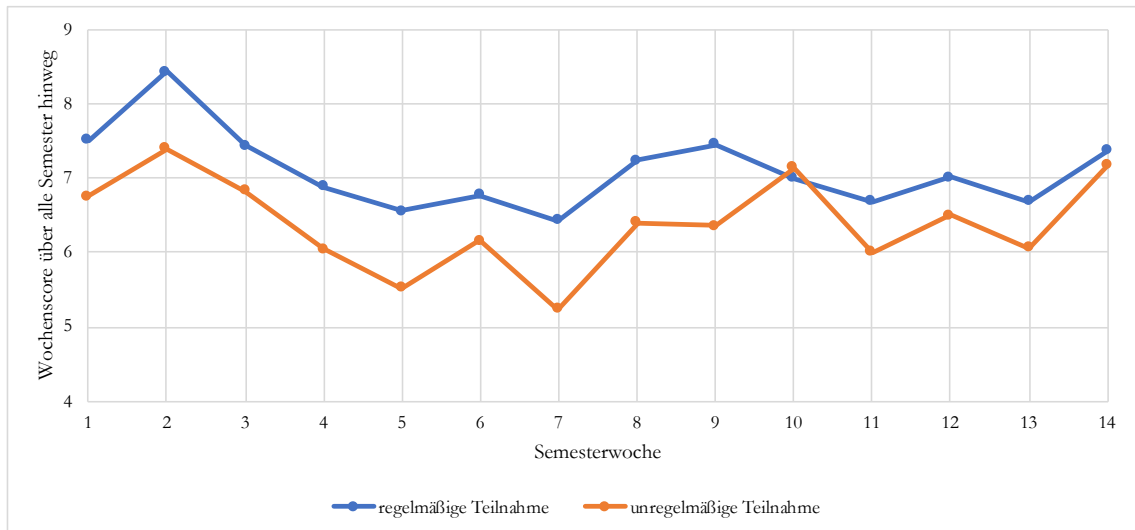


Abbildung 13: Durchschnittlich erreichter Wochenscore der Studienteilnehmer (über alle Semester hinweg). Es wurden jeweils für die regelmäßigen und die unregelmäßigen Teilnehmer über alle Semester hinweg gemittelte Wochenscores bestimmt, graphisch dargestellt und anschließend deskriptiv beschrieben. Die Studierenden, die eine regelmäßige Teilnahme vorweisen konnten, erreichten von Beginn an höhere Wochenscores als ihre Mitstudierenden (mit Ausnahme der zehnten Woche). Die Ordinate wurde für eine optimale Darstellung angepasst.

In Abbildung 13 sind die Wochenscores der Studienteilnehmer über alle Semester hinweg dargestellt. Deskriptiv lässt sich ein höherer Kurvenverlauf der regelmäßigen Teilnehmer beobachten als der ihrer unregelmäßig teilnehmenden Mitstudierenden (mit Ausnahme der zehnten Woche).

3.4.2 Leistung in den Interventionsitems zu den Zeitpunkten T1, T2 und T3

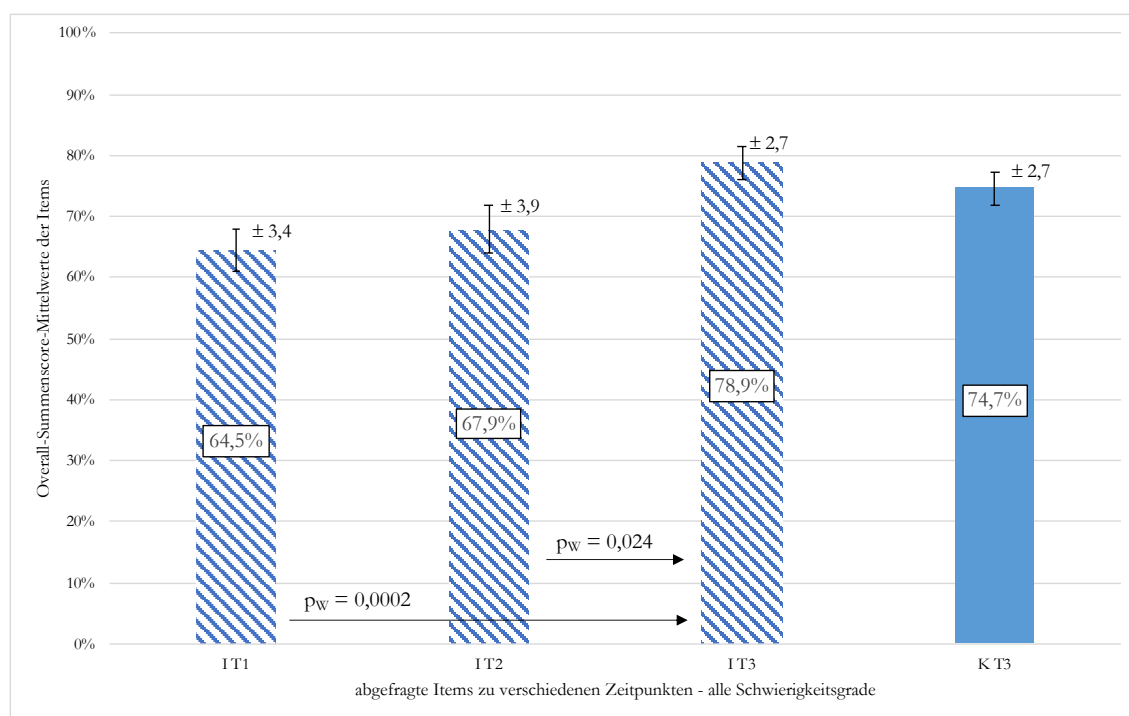


Abbildung 14: Leistungsunterschiede der primären Analyse (unter Einbeziehung aller Schwierigkeitsgrade). Dargestellt sind zum einen die *Overall-Summscore-Mittelwerte* nach Beantwortung der gleichen Interventionsitems (I) zu unterschiedlichen Zeitpunkten (T1, T2 und T3) und zum anderen die *Overall-Summscore-Mittelwerte* nach erstmaliger Beantwortung der Kontrollitems (K) zum Zeitpunkt T3. Jedem *Overall-Summscore-Mittelwert* ist auch ein Fehlerbalken in Form des Standardfehlers zugeordnet. Jeder signifikante Unterschied zwischen den *Overall-Summscores* zu verschiedenen Zeitpunkten, die im Wilcoxon-Test in der primären Analyse ermittelt wurden, ist mit einem Pfeil markiert.

Zwischen den *Overall-Summscores* der Interventionsitems zu den drei Zeitpunkten (T1, T2, T3) bestanden signifikante Unterschiede ($p_{PA-F} = 0,00005$ im Friedman-Test der primären Analyse). Wie in Abbildung 14 verdeutlicht, zeigte sich, dass zum Zeitpunkt T1 die Studienteilnehmer einen *Overall-Summscore-Mittelwert* von $64,5\% \pm 3,4\%$ erreicht haben. Zum Zeitpunkt T2 betrug der *Overall-Summscore-Mittelwert* der Studienteilnehmer $67,9\% \pm 3,9\%$ und zum Zeitpunkt T3 $78,9\% \pm 2,7\%$. Der Unterschied zwischen den *Overall-Summscores* ließ sich zu T2 und T3 ($p_{PA-W} = 0,024$) sowie zu T1 und T3 ($p_{PA-W} = 0,0002$) im Wilcoxon-Test signifikant nachweisen. Der Unterschied zwischen den *Overall-Summscores* zu T1 und T2 war nicht signifikant ($p_{PA-W} = 0,389$). Werden die Interventionsitems der verschiedenen Schwierigkeitsgrade als *Overall-Summscore* zusammengefasst, lässt sich bei mehrfachem Präsentieren bestimmter Fragestellungen ein signifikanter Unterschied zwischen den *Overall-Summscores* feststellen.

Zusätzlich wurde der Unterschied zwischen den *Summscores* der Interventionsitems zu den Zeitpunkten T1, T2 und T3 abhängig vom Schwierigkeitsgrad untersucht (siehe Abbildung 15).

Bei den leichten Interventionsitems zeigte sich im Friedman-Test ein signifikanter Unterschied ($p_{PA-F} = 0,016$) zwischen den Summenscores zu T1 (Summenscore-Mittelwert: $73,0\% \pm 3,8\%$), T2 (Summenscore-Mittelwert: $67,4\% \pm 5,0\%$) und T3 (Summenscore-Mittelwert: $83,5\% \pm 3,5\%$). Zu T1 und T2 ließ sich im Wilcoxon-Test kein signifikanter Unterschied zwischen den Summenscores feststellen ($p_{PA-W} = 0,454$). Der Unterschied zwischen den Summenscores war zu T2 und T3 ($p_{PA-W} = 0,007$) sowie zu T1 und T3 ($p_{PA-W} = 0,028$) im Wilcoxon-Test signifikant.

Bei den Interventionsitems, die einen mittelschweren Schwierigkeitsgrad aufwiesen, konnte ein Unterschied im Friedman-Test zwischen den Summenscores zu T1 (Summenscore-Mittelwert: $42,9\% \pm 7,4\%$), T2 (Summenscore-Mittelwert: $67,5\% \pm 6,9\%$) und T3 (Summenscore-Mittelwert: $63,5\% \pm 7,1\%$) festgestellt werden ($p_{PA-F} = 0,028$). Zwischen den Summenscores zu den Zeitpunkten T1 und T2 sowie zu T1 und T3 konnte ein signifikanter Unterschied mit dem Wilcoxon-Test nachgewiesen werden ($p_{PA-W T1-T2} = 0,003$ bzw. $p_{PA-W T1-T3} = 0,028$). Der Unterschied zwischen den Summenscores zu T2 und T3 war im Wilcoxon-Test nicht signifikant ($p_{PA-W} = 0,526$).

Bei den Interventionsitems mit dem höchsten Schwierigkeitsgrad konnte ein signifikanter Unterschied im Friedman-Test zwischen den Summenscores der Interventionsitems zu den Zeitpunkten T1 (Summenscore-Mittelwert: $53,3\% \pm 6,8\%$), T2 (Summenscore-Mittelwert: $68,9\% \pm 6,0\%$) und T3 (Summenscore-Mittelwert: $78,9\% \pm 5,4\%$) festgestellt werden ($p_{PA-F} = 0,004$). Der Unterschied zwischen den Summenscores zu T1 und T2 ($p_{PA-W} = 0,027$) sowie zu T1 und T3 ($p_{PA-W} = 0,005$) war im Wilcoxon-Test signifikant. Der Unterschied zwischen den Summenscores zu T2 und T3 war im Wilcoxon-Test nicht signifikant ($p_{PA-W} = 0,140$).

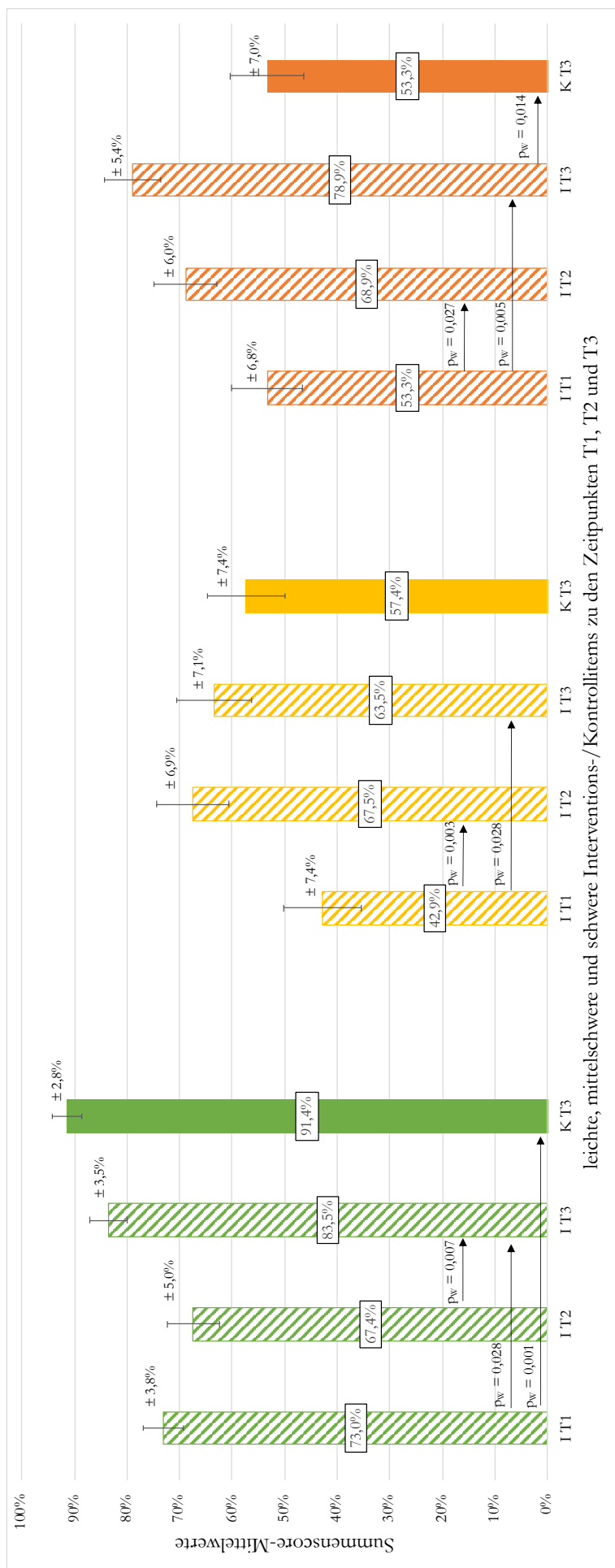


Abbildung 15: Leistungsunterschiede der Interventions- und Kontrollitems in der primären Analyse (aufgeteilt nach Schwierigkeitsgrad). In dieser Abbildung sind die Summscore-Mittelwerte zu den Zeitpunkten T1 (erstes Abfragen eines Items), T2 (zweites Abfragen eines Items) und T3 (drittes Abfragen eines Items im *Exit Exam*) der Interventionsitems I (gestreifter Farbverlauf des Balkens) und dem erstmaligen Abfragen der Kontrollitems K (einfarbige Füllung des Balkens) zum Zeitpunkt T3 (im *Exit Exam*) dargestellt. Gesondert betrachtet werden hier die drei verschiedenen Schwierigkeitsgrade, wobei grün die leichten Items darstellt, gelb die mittleren darstellt und rot die schweren darstellt. Jedem Summscore-Mittelwert ist auch ein Fehlerbalken in Form eines Standardfehlers zugeordnet. Jeder signifikante Unterschied zwischen den Summscores in der primären Analyse der Wilcoxon-Tests ist durch einen Pfeil markiert.

3.5 Sensitivitätsanalyse

Zusätzlich zur primären Analyse wurde eine Sensitivitätsanalyse (SA) durchgeführt (siehe Abbildung 16 und Abbildung 17). In der Sensitivitätsanalyse wurde ein Item eines Studienteilnehmers aus den Summenscores genommen, sobald ein unvollständiger Datensatz vorlag. Dies betraf zum Zeitpunkt T1 14 Items (= 5% der insgesamt präsentierten Interventionsitems zu T1) und zum Zeitpunkt T2 wurden 33 Items (= 13% der insgesamt präsentierten Interventionsitems zu T2) im Zuge der Sensitivitätsanalyse herausgenommen. 70% der Items, die in der primären Analyse mit null Punkten bewertet worden sind, betreffen Zeitpunkt T2. Wie in Kapitel 2.4.1 beschrieben, wurde für manche Teilnehmer bei bestimmten Schwierigkeitsgraden kein Summenscore mehr gebildet, wenn es keine beantworteten Items mehr für diesen Schwierigkeitsgrad gab. Aus diesem Grund gab es insgesamt fünf leichte, vier mittelschwere und drei schwere Summenscores weniger, die in die jeweilige Berechnung der Summenscore-Mittelwerte eingehen konnten. Im Folgenden wird auf die wesentlichen Unterschiede im Vergleich zu den oben beschriebenen Ergebnissen aus der primären Analyse eingegangen.

In Bezug zur ersten Studienfrage (vergleiche Kapitel 3.3) ließen sich keine Änderungen im Hinblick auf die berechneten signifikanten Unterschiede der primären Analyse feststellen.

In Bezug zur zweiten Studienfrage (vergleiche Kapitel 3.4) waren die Unterschiede zwischen den *Overall*-Summenscores der Interventionsitems (unter Einbeziehung aller Schwierigkeitsgrade) zu T1, T2 und T3 nach wie vor signifikant in der Sensitivitätsanalyse ($p_{SA-F} = 0,002$ im Friedman-Test). Allerdings fanden sich die signifikanten Unterschiede im Wilcoxon-Test zwischen den *Overall*-Summenscores zu den Zeitpunkten T1 und T2 ($p_{SA-W} = 0,009$) und nicht mehr zu T2 und T3 ($p_{SA-W} = 0,970$), wie in der primären Analyse. Der signifikante Unterschied zwischen den *Overall*-Summenscores zu T1 und T3 ließ sich nach wie vor im Wilcoxon-Test nachweisen ($p_{SA-W} = 0,008$). Des Weiteren fanden sich keine signifikanten Unterschiede mehr zwischen den Summenscores der leichten Interventionsitems zu T1, T2 und T3 ($p_{SA-F} = 0,607$ im Friedman-Test). Der Unterschied zwischen den mittelschweren Summenscores zu T1 und T3 war im Wilcoxon-Test nicht mehr signifikant ($p_{SA-W} = 0,62$). Im Übrigen ließen sich keine Änderungen im Hinblick auf die berechneten signifikanten Unterschiede der primären Analyse feststellen.

Bezogen auf die Untersuchung des Anforderungsniveaus der Interventionsitems ließen sich in den *Overall*-Summenscores keine signifikanten Unterschiede mehr zwischen den jeweiligen Summenscores finden ($p_{SA-W} = 0,481$).

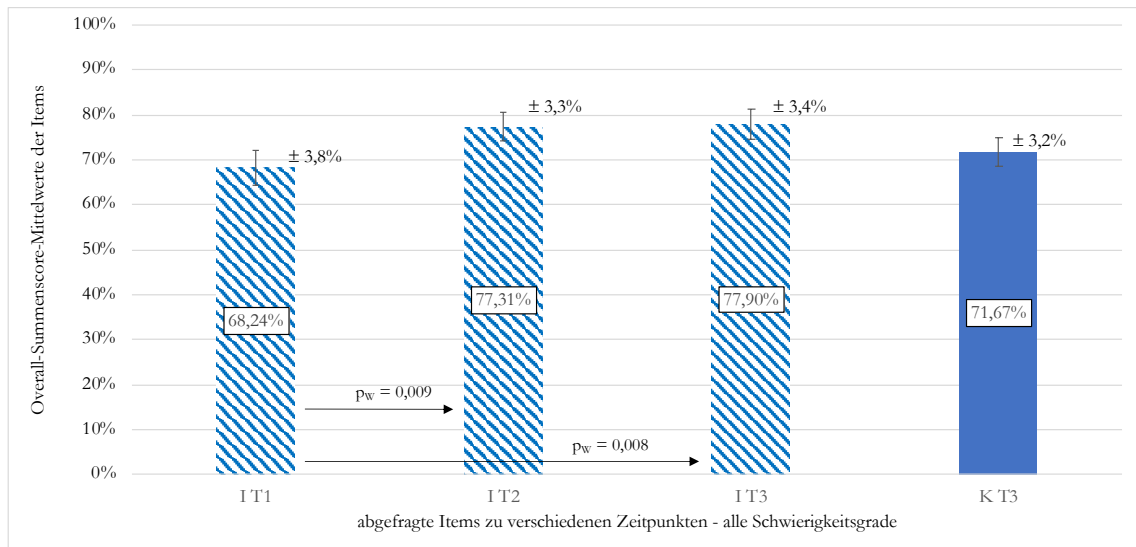


Abbildung 16: Leistungsunterschiede in der Sensitivitätsanalyse (unter Einbeziehung aller Schwierigkeitsgrade). Dargestellt sind zum einen die *Overall*-Summscore-Mittelwerte nach Beantwortung der gleichen Interventionsitems (I) zu unterschiedlichen Zeitpunkten (T1, T2 und T3) und zum anderen die *Overall*-Summscore-Mittelwerte nach erstmaliger Beantwortung der Kontrollitems (K) zum Zeitpunkt T3. Jedem *Overall*-Summscore-Mittelwert ist auch ein Fehlerbalken in Form des Standardfehlers zugeordnet. Jeder signifikante Unterschied zwischen den *Overall*-Summscores zu verschiedenen Zeitpunkten, die im Wilcoxon-Test in der primären Analyse ermittelt wurden, ist mit einem Pfeil markiert.

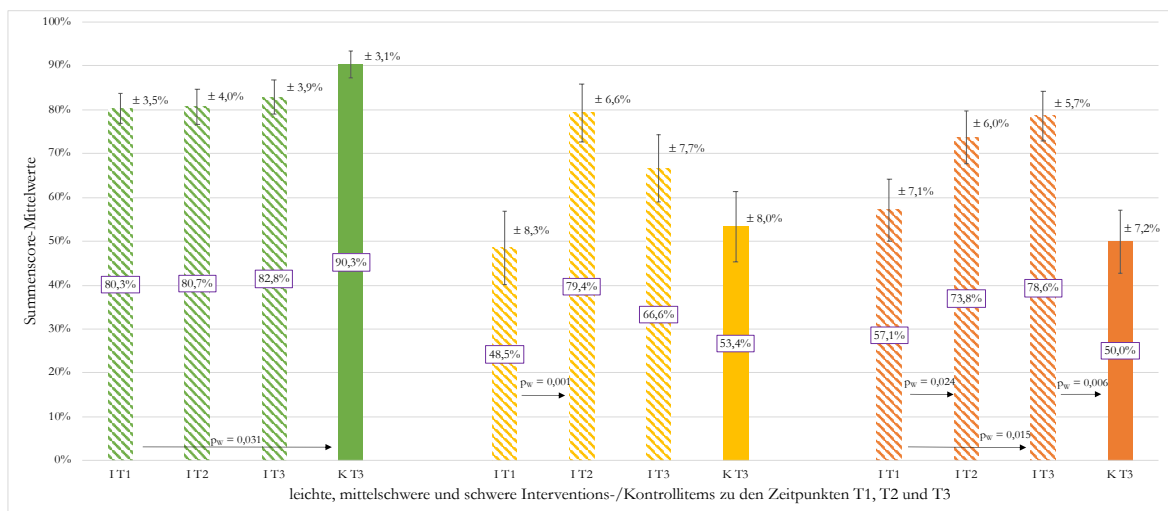


Abbildung 17: Leistungsunterschiede der Interventions- und Kontrollitems in der Sensitivitätsanalyse (aufgeteilt nach Schwierigkeitsgrad). In dieser Abbildung sind die Summscore-Mittelwerte zu den Zeitpunkten T1 (erstes Abfragen eines Items), T2 (zweites Abfragen eines Items) und T3 (drittes Abfragen eines Items im *Exit Exam*) der Interventionsitems I (gestreifter Farbverlauf des Balkens) und dem erstmaligen Abfragen der Kontrollitems K (einfarbige Füllung des Balkens) zum Zeitpunkt T3 (im *Exit Exam*) dargestellt. Gesondert betrachtet werden hier die drei verschiedenen Schwierigkeitsgrade, wobei grün die leichten Items darstellt, gelb die mittleren darstellt und rot die schweren darstellt. Jedem Summscore-Mittelwert ist auch ein Fehlerbalken in Form eines Standardfehlers zugeordnet. Jeder signifikante Unterschied zwischen den Summscores in der primären Analyse der Wilcoxon-Tests ist durch einen Pfeil markiert.

3.6 Beurteilung des Projekts #*clue* durch die Studierenden

Zur Beantwortung der vierten Studienfrage (Wie beurteilen die Studierenden dieses innovative Projekt?) wurden die EvaSys[®]-Fragebögen und die Fokusgruppeninterviews ausgewertet.

3.6.1 Ergebnisse der Fragebogen-Erhebung

Die Ergebnisse der Studierendenbefragung mittels EvaSys[®]-Fragebögen sind in Tabelle 2, Tabelle 3 und Tabelle 4 aufgeführt. Die Fragebögen wurden von 135 Teilnehmern aus drei verschiedenen Semestern ausgefüllt. Es wurde nicht jede Frage von jedem Studienteilnehmer beantwortet, weshalb sich Diskrepanzen hinsichtlich der angegebenen Grundgesamtheiten (n) finden. 69,3% der Teilnehmer waren weiblich (n = 127). Das Durchschnittsalter betrug 24,8 Jahre (n = 111). 24,8% (n = 129) der Teilnehmer gaben an, immer an der Beantwortung der wöchentlichen Fragenpakete teilgenommen zu haben. 64,4% (n = 129) der Studienteilnehmer hatten die Fragenpakete unregelmäßig beantwortet und 10,8% (n = 129) hatten das Angebot gar nicht wahrgenommen.

Insgesamt 24,1% der Teilnehmer besuchten alle interdisziplinären Lehrveranstaltungen, 41,4% besuchten mindestens eine interdisziplinäre Lehrveranstaltung. 34,5% besuchten nie eine solche Lehrveranstaltung (n = 87).

Der zweite Teil der Fragebögen befragte die Studienteilnehmer zu den wöchentlichen Fragenpaketen auf einer sechsstufigen Likert-Skala. Hierbei wurden wie in Kapitel 2.4.1 erläutert, die Skalenoptionen eins und zwei als positive Rückmeldung zusammengefasst. Die positive Rückmeldung ist im Folgenden in Prozent angegeben. 84,3% der Teilnehmer bewerteten den Stoffumfang von zehn Fragen pro Woche als angemessen (n = 108). Die wöchentlichen E-Mails zur Erinnerung an das Kreuzen wurden als hilfreich eingestuft (71,8%, n = 110). 83,8% der Studienteilnehmer fanden es sinnvoll, dass mit Hilfe von Fragen Lehrinhalte aus vergangenen Semestern wiederholt wurden (n = 111). 69,7% der Studienteilnehmer stuften das Anforderungsniveau der Fragen weder als zu leicht noch als zu schwer ein. 43,5% der Teilnehmer gaben im dritten Teil der Fragebögen an, dass die interdisziplinären Lehrveranstaltungen zu einem tieferen Verständnis der behandelten Themen geführt haben (n = 62) und dass sie durch den Besuch motiviert wurden, sich mit den Themen intensiver auseinander zu setzen (30,0%, n = 60). 37,5% der Teilnehmer hätten sich eine vermehrte Vorstellung der Fragelösungen aus den Fragenpaketen gewünscht (n = 64).

67,0% der Studienteilnehmer bewerteten im Gesamturteil das Projekt #*clue* mit einer guten oder sehr guten Gesamtnote (n = 109). 75,2% der Studienteilnehmer würden das Projektangebot auch im folgenden Semester nutzen (n = 113).

Tabelle 2: Ergebnisse der Fragebögen hinsichtlich der wöchentlichen Fragenpakete

ERGEBNISSE DER FRAGEBÖGEN I	
Wöchentliche Fragenpakete	
Der Stoffumfang der wöchentlichen Fragenpakete war angemessen	84,3% (n = 108)
Die wöchentlichen Mails waren als Erinnerungsstütze hilfreich	71,8% (n = 110)
Die semesterbegleitende Wiederholung von Lerninhalten aus vorangegangenen Semestern ist sinnvoll	83,8% (n = 111)
Die wöchentlichen Fragenpakete waren dem zeitlichen und inhaltlichen Rahmen der Module angemessen	73,3% (n = 107)
Die Bearbeitung der wöchentlichen Fragenpakete hat mir geholfen, Inhalte aus verschiedenen Modulen interdisziplinär miteinander zu verknüpfen	43,1% (n = 109)
Die Beantwortung der Fragenpakete hat wesentlich zu meinem Lernerfolg beigetragen	25,7% (n = 109)
Die Beantwortung der wöchentlichen Fragenpakete hat mich dazu angeregt, mich intensiver mit den abgefragten Themen auseinanderzusetzen	32,4% (n = 108)
Die Bearbeitung der wöchentlichen Fragenpakete hat mich bei der Wiederholung der Modulinhalte unterstützt	60,8% (n = 109)
Das Anforderungsniveau der wöchentlichen Fragenpakete war ausgewogen	69,7% (n = 109)

Eventuelle Inkonsistenzen in den Summen sind auf eine jeweils unterschiedliche Anzahl an fehlenden Werten pro Item zurückzuführen. Die Skalenoptionen eins und zwei der ursprünglichen Likert-Skala wurden als positive Wertung zusammengefasst und es wurde der Anteil der Studienteilnehmer berechnet, der eine eins oder zwei gewählt hat.

Tabelle 3: Ergebnisse der Fragebögen hinsichtlich der interdisziplinären Lehrveranstaltungen

ERGEBNISSE DER FRAGEBÖGEN II		
Interdisziplinäre Lehrveranstaltungen		
Der Stoffumfang der interdisziplinären Lehrveranstaltungen war angemessen		67,2% (n = 64)
Die interdisziplinären Lehrveranstaltungen haben bei mir zu einem tieferen Verständnis der geprüften Themen geführt		43,5% (n = 62)
Der Besuch der interdisziplinären Lehrveranstaltungen hat mich dazu angeregt, mich intensiver mit den abgefragten Themen auseinanderzusetzen		30,0% (n = 60)
Die interdisziplinären Lehrveranstaltungen haben wesentlich zu meinem Lernerfolg beigetragen		27,9% (n = 61)
Das Anforderungsniveau der interdisziplinären Lehrveranstaltungen war ausgewogen		75,4% (n = 57)
In der interdisziplinären Lehrveranstaltung hätte mehr ... stattfinden sollen	... Diskussion der geprüften Fragen ...	31,3% (n = 20)
	... Vorstellung der Fragelösungen ...	37,5% (n = 24)
	... Interaktion mit den Studierenden ...	14,0% (n = 9)

Eventuelle Inkonsistenzen in den Summen sind auf eine jeweils unterschiedliche Anzahl an fehlenden Werten pro Item zurückzuführen. Die Skalenoptionen eins und zwei der ursprünglichen Likert-Skala wurden als positive Wertung zusammengefasst und es wurde der Anteil der Studierenden berechnet, der eine eins oder zwei gewählt hat.

Tabelle 4: Ergebnisse der Fragebögen hinsichtlich des Gesamturteils

ERGEBNISSE DER FRAGEBÖGEN III	
Gesamturteil	
Das Projekt hat mein Interesse an Interdisziplinarität gefördert	52,7% (n = 108)
Ich würde im nächsten Semester wieder am Projekt teilnehmen	75,2% (n = 113)
Ich würde das Projekt weiterempfehlen	64,8% (n = 111)
In einem Gesamturteil bewerte ich das Projekt als sehr gut oder gut	67,0% (n = 109)

Eventuelle Inkonsistenzen in den Summen sind auf eine jeweils unterschiedliche Anzahl an fehlenden Werten pro Item zurückzuführen. Die Skalenoptionen eins und zwei der ursprünglichen Likert-Skala wurden als positive Wertung zusammengefasst und es wurde der Anteil der Studierenden berechnet, der eine eins oder zwei gewählt hat.

3.6.2 Ergebnisse der Fokusgruppeninterviews

3.6.2.1 Teilnehmer

An den vier Fokusgruppengesprächen, die im Sommersemester 2018 durchgeführt wurden, nahmen insgesamt 20 Personen teil. Davon waren elf Gesprächsteilnehmer Männer und neun Gesprächsteilnehmerinnen Frauen. Aus dem dritten klinischen Semester nahmen sieben, aus dem vierten klinischen Semester acht, aus dem fünften klinischen Semester zwei und aus dem sechsten klinischen Semester drei Personen teil.

3.6.2.2 Inhaltliche Ergebnisse

Im Rahmen der qualitativen Inhaltsanalyse nach Mayring (2010) entstanden 765 *Codings*, die sich auf fünf Oberkategorien verteilten („Allgemeines“, „Fragenpakete“, „Interdisziplinäre Lehrveranstaltungen“, „Verbesserungsvorschläge“ und „thematischer Exkurs“). Diese Oberkategorien orientierten sich an den Themenschwerpunkten des Gesprächsleitfadens. Durch die Zuteilung von Unterkategorien wurden die *Codings* weiter aufgearbeitet und den Oberkategorien zugeordnet. Im Folgenden werden im Fließtext die o. g. Themenschwerpunkte mit Unterkategorien mit mindestens fünf *Codings* vorgestellt. In Tabelle 5 bis Tabelle 11 finden sich die gesamten Ergebnisse. Die aufgeführten Zitate dienen zur Veranschaulichung der Beiträge der einzelnen Teilnehmer.

Mit 340 von 765 aller *Codings* lag der Schwerpunkt der Gespräche auf der Oberkategorie „Allgemeines“ (siehe Tabelle 5, Tabelle 6, Tabelle 7).

Zunächst äußerten sich die Studienteilnehmer dazu, was ihrer Meinung nach das Ziel des Projektes *#clue* sei (42 *Codings*). Als vorrangig betrachteten sie hier, dass die Lehre mit dem IMPP verknüpft wurde (16 *Codings*) und somit eine Vorbereitung auf das anstehende Staatsexamen erfolgen sollte (15 *Codings*). Zusätzlich diene das Projekt laut den Studienteilnehmern dazu, den Lernstoff zu wiederholen (14 *Codings*), indem das Selbststudium angeregt wurde (sechs *Codings*): „So wie ich das verstanden habe, ging es darum, die Studierenden des klinischen Studienabschnitts mit wöchentlichen Fragen, die die Themen der Vorlesungen widerspiegeln, abzufragen, um die Studierenden besser auf das Staatsexamen vorzubereiten“ (Gruppe04_Absatz3_Person3; männlich). Des Weiteren betrachteten die Teilnehmer das Projekt als Möglichkeit, den Lehrenden ein Feedback über die Vermittlung des Lernstoffes zu geben (neun *Codings*): „Ich denke, es war auch dafür gedacht oder hat auch dazu beigetragen zu gucken, wie die Lehrinhalte rübergebracht wurden.“ (Gruppe02_Absatz06_Person6; weiblich).

Primär sahen die Teilnehmer in dem Projekt eine Lernunterstützung (105 *Codings*), indem sie ein persönliches Feedback (44 *Codings*) zu ihren Wissenslücken (13 *Codings*) erhielten: „Oh, das habe ich jetzt noch nicht gewusst, aber ich weiß zumindestens, dass ich das Thema schon einmal hatte und es wissen hätte können.“ (Gruppe03_Absatz15_Person3; weiblich). Außerdem empfanden die Teilnehmer das Projekt als „eine sehr gute Rückmeldung, [...] ob

man wenigstens so ungefähr das Thema verstanden hat“ (Gruppe02_Absatz14_Person2; männlich) (15 *Codings*). Zusätzlich nutzten die Teilnehmer das Projekt, um Wissen aufzuarbeiten (sieben *Codings*) und sich mit anderen Kommilitonen zu vergleichen (fünf *Codings*). Des Weiteren nutzten die Teilnehmer das Projekt, um ihr Wissen zu verfestigen (15 *Codings*) und zur Examensvorbereitung (zwölf *Codings*), um Schwerpunkte, die vom IMPP gesetzt werden, kennen zu lernen (fünf *Codings*). Bei der Überlegung, ob die Teilnehmer einen Lernerfolg (34 *Codings*) aus dem Projekt ziehen konnten, unterschieden sie zwischen einem „generellen“ Lernerfolg (fünf *Codings*), der vor allem abhängig von der Lehre sei und einem „individuellen“ Lernerfolg (29 *Codings*). Prinzipiell konnten die Studienteilnehmer einen individuellen Lernerfolg feststellen (15 *Codings*), wobei „das [...] extrem davon abhängig war, wie sehr man schon die Gelegenheit hatte, sich selbst mit dem Stoff auseinanderzusetzen“ (Gruppe02_Absatz08_Person4; weiblich) (fünf *Codings*). Manche Teilnehmer konnten auch keinen individuellen Lernerfolg für sich erkennen (5 *Codings*).

Tabelle 5: Auswertung der Fokusgruppengespräche – Allgemeines (Teil eins)

Code	Codings	Interview- gruppe	Anzahl Interview- gruppen
I. Allgemeines – Teil eins	340	1,2,3,4	4
I.1 Ziel des Projektes	42	1,2,3,4	4
I.1.1 Lehre mit IMPP verknüpfen	16	1,2,3,4	4
I.1.1.1 Vorbereitung auf das Staatsexamen	15	1,2,3,4	4
I.1.2 Wiederholung von Lernstoff	14	1,2,3,4	4
I.1.2.1 Selbststudium fördern	6	1,2,3	3
I.1.2.2 aktueller Lernstoff	6	2,4	2
I.1.2.3 Lernstoff aus vergangenen Semestern	2	2	1
I.1.3 Feedback an Lehre	9	1,2,3	3
I.1.3.1 Anpassung der Lehre	2	1,2	2
I.1.4 persönlichen Wissensstand überprüfen	2	2	1
I.1.5 Klausurvorbereitung	1	2	1
I.2 Lernunterstützung	105	1,2,3,4	4
I.2.1 persönliches Feedback	44	1,2,3,4	4
I.2.1.1 Wissenslücken finden	13	1,2,3	3
I.2.1.2 zum aktuellen Wissens-/Leistungsstand	15	1,2,3,4	4
I.2.1.3 Vergleich zu Mitstudierenden	5	2,3	2
I.2.1.4 Wissen aufarbeiten	7	1,2,4	3
I.2.1.5 zu älteren Modulen	3	1,2	2
I.2.2 Lernerfolg	34	1,2,3,4	4
I.2.2.1 generell	5	2,3	2
I.2.2.1.1 abhängig von der Lehre	5	2,3	2
I.2.2.2 individuell	29	1,2,3,4	4
I.2.2.2.1 ja	15	1,2,4	3
I.2.2.2.2 teilweise	9	2,3	2
I.2.2.2.2a abhängig vom persönlichen Lernstatus	5	2,3	2
I.2.2.2.2b abhängig vom Zeitpunkt des Kreuzens	3	2,3	2
I.2.2.2.3 nein	5	3	1
I.2.2.2.3a nicht klausurrelevant	4	3	1
I.2.3 Wissen verfestigen	15	1,2,3,4	4
I.2.4 Examensvorbereitung	12	1,3,4	3
I.2.4.1 Schwerpunktsetzung im Staatsexamen	5	3,4	2

Dargestellt sind tabellarisch die Ergebnisse der Fokusgruppengespräche. Den einzelnen Codes sind die Anzahlen der jeweiligen Codings zugeordnet. Zusätzlich finden sich die Interviewgruppen, in denen die Codes genannt worden sind und die Anzahl der Interviewgruppen, in denen das Code thematisiert worden ist.

Die Teilnehmer erläuterten unter der Oberkategorie „Allgemeines“ ihre Motivation am Projekt teilzunehmen (42 *Codings*). Als größte Einflussfaktoren sahen sie hier primär eine motivierende Lehrperson (11 *Codings*), gefolgt vom Staatsexamen (sechs *Codings*): „Ich fand das ganz interessant zu sehen, was mich im Examen erwartet, weil ich jetzt über AMBOSS® gar nicht lerne und da auch keine IMPP-Fragen kreuze.“ (Gruppe03_Absatz47_Person1; weiblich). Des Weiteren bildete die Verlosung (fünf *Codings*) „[...] einen Anreiz, daran teilzunehmen.“ (Gruppe04_Absatz17_Person3; männlich). Prinzipiell sahen die Teilnehmer das Projekt als sinnvolle Zeitinvestition (fünf *Codings*): „Also das ist einfach wie gesagt schon ein guter kleiner Ansporn noch einmal etwas zu tun. Und wenn es nur ein paar Minuten zu Hause sind, diese paar Minuten hätte man wahrscheinlich sonst nicht investiert.“ (Gruppe01_Absatz29_Person2; männlich).

Der *Social-Media*-Aspekt (50 *Codings*) bestand aus den wöchentlichen Erinnerungs-E-Mails und der *Facebook*-Gruppe und wurde als „[...] richtig wichtig“ (Gruppe02_Absatz74_Person5; männlich) eingestuft (16 *Codings*). Einige Teilnehmer hatten nur wegen der Erinnerungs-E-Mail noch einmal das wöchentliche Fragenpaket bearbeitet (zehn *Codings*): „Also ich habe oft dann noch einmal die Fragen Dienstagnachmittag gemacht, weil dann die Erinnerungs-E-Mail kam.“ (Gruppe01_Absatz40_Person4; männlich). Das Ausbleiben der E-Mails ab der neunten Woche wurde nicht positiv aufgenommen (elf *Codings*): „Was mir sehr negativ aufgefallen ist (...) – vermutlich außerhalb der Handhabe der Studienleitung [...], dass die datenschutzrechtlichen Bestimmungen sich geändert hatten und dass es dann ziemlich abrupt zum Versiegen der Informationen der wöchentlichen Erinnerungen kam. Und dass sich das eben – ich glaube, da kann man auch ziemlich genau den Punkt festmachen, das behaupte ich jetzt einfach – wo die Teilnahme der einzelnen Studierenden zurückgegangen ist. Das ist mir negativ aufgefallen.“ (Gruppe04_Absatz31_Person3; männlich).

Tabelle 6: Auswertung der Fokusgruppengespräche – Allgemeines (Teil zwei)

Code	Codings	Interview- gruppe	Anzahl Interview- gruppen
I. Allgemeines – Teil zwei			
I.3 Motivation teilzunehmen	42	1,2,3,4	4
I.3.1 Lehrperson	11	1,2,4	3
I.3.2 Einblick ins Staatsexamen	6	3,4	2
I.3.3 Verlosung	7	1,2,3,4	4
I.3.3.1 Einfluss	5	1,2,3,4	4
I.3.3.2 kein Einfluss	2	1,4	2
I.3.4 sinnvolle Zeitinvestition	5	1,2,3	3
I.3.5 sich verbessern	2	1,4	2
I.3.6 lernen, sich selbst einschätzen zu können	2	2	1
I.3.7 Studie	2	2	1
I.3.8 Regelmäßigkeit	2	2,3	2
I.3.9 weil es alle gemacht haben	2	2	1
I.3.10 Vergleichbarkeit	2	2	1
I.3.11 Wiederholung alter Module	1	2	1
I.4 Social-Media	50	1,2,3,4	4
I.4.1 ist ein wichtiger Aspekt	16	1,2,3,4	4
I.4.2 Erinnerungs-E-Mails	10	1,2,3,4	4
I.4.2.1 wegen E-Mail gekreuzt	10	1,2,3,4	4
I.4.3 facebook	13	1,2,3,4	4
I.4.3.1 genutzt	8	1,2,3,4	4
I.4.3.2 nicht genutzt	5	1,2,3,4	4
I.4.4 wegen Datenschutzgrundverordnung weggefallen	11	1,2,3,4	4
I.4.4.1 Erinnerungen zu unregelmäßig	4	1,3,4	3
I.4.4.2 als Projektende gewertet	3	2,3,4	3

Dargestellt sind tabellarisch die Ergebnisse der Fokusgruppengespräche. Den einzelnen Codes sind die Anzahlen der jeweiligen Codings zugeordnet. Zusätzlich finden sich die Interviewgruppen, in denen die Codes genannt worden sind und die Anzahl der Interviewgruppen, in denen das Code thematisiert worden ist.

Fast ein Drittel der Aussagen in der Kategorie „Allgemeines“ beschäftigte sich mit der Bewertung von *#clue* (101 *Codings*). Vor allem Vorteile (44 *Codings*) wurden von den Teilnehmern genannt. So befürworteten sie die Anpassung der Fragenpakete an den Göttinger Modulplan (elf *Codings*): „Ich fand es irgendwie cool, dass man sich so die Mühe gemacht hat, das auf die Woche anzupassen.“ (Gruppe03_Absatz48_Person2; weiblich) oder auch die zeitliche Flexibilität, mit der die Fragenpakete beantwortet werden konnten (zehn *Codings*). Die digitale Umsetzung (acht *Codings*) und die damit einhergehende Möglichkeit, der Beantwortung am „Handy“ (Gruppe02_Absatz120_Person5; männlich) fanden ebenfalls Zuspruch. Nicht zu unterschätzen war der Aspekt der Freiwilligkeit an der Teilnahme des Projektes (sieben *Codings*). „Abgesehen davon, finde ich das Konzept ziemlich gut, weil es nicht vorgeschrieben ist. Man kann freiwillig teilnehmen, man wird also nicht bestraft, wenn man es nicht macht.“ (Gruppe04_Absatz17_Person3; männlich). Als Nachteil (16 *Codings*) empfanden die Teilnehmer vor allem, dass die möglichen sieben Tage zum Bearbeiten der Fragenpakete zu kurz seien (sieben *Codings*). Zusätzlich vertraten einige die folgende Auffassung (sieben *Codings*): „Ich kenne mein Altbewährtes, ich nutze AMBOSS®. Es funktioniert. Wieso soll ich noch was Neues probieren?“ (Gruppe02_Absatz77_Person6, weiblich). Im Vergleich zu anderen Lehrformaten, sahen sie *#clue* dennoch als gute Ergänzung (neun *Codings*).

Ein Großteil der Studienteilnehmer würde erneut am Projekt teilnehmen (17 *Codings*) und es auch weiterempfehlen (zwölf *Codings*).

Tabelle 7: Auswertung der Fokusgruppengespräche – Allgemeines (Teil drei)

Code	Codings	Interview- gruppe	Anzahl Interview- gruppen
I. Allgemeines – Teil drei			
I.5 Bewertung von #clue	101	1,2,3,4	4
I.5.1 Vorteile	44	1,2,3,4	4
I.5.1.1 an Göttinger Modulplan angepasst	11	1,2,3,4	4
I.5.1.2 zeitliche Flexibilität	10	1,2,3	3
I.5.1.3 digitale Umsetzung	8	1,2	2
I.5.1.4 freiwillig	7	1,2,4	3
I.5.1.5 ähnlich der Prüfungssituation	3	1,2,4	3
I.5.1.6 räumliche Flexibilität	2	1,2	2
I.5.1.7 neuer Blickwinkel	2	4	1
I.5.1.8 unverfälschtes Ergebnis	1	1	1
I.5.2 Nachteile	16	1,2,3,4	4
I.5.2.1 Fragenpakete zu kurz freigeschaltet	2	2,3,4	3
I.5.2.2 vorhandene Medien reichen aus	7	1,2,3	3
I.5.2.3 Fragen können nicht wiederholt werden	1	2	1
I.5.3 Lerninnovation	12	1,2,3,4	4
I.5.3.1 Vergleich zu anderen Lehrformaten	9	1,2,3,4	4
I.5.3.1.1 gute Ergänzung	9	1,2,3,4	4
I.5.3.2 keine Innovation	3	2,3	2
I.5.4 wieder teilnehmen	17	1,2,3,4	4
I.5.5 weiter empfehlen	12	1,2,3,4	4

Dargestellt sind tabellarisch die Ergebnisse der Fokusgruppengespräche. Den einzelnen Codes sind die Anzahlen der jeweiligen Codings zugeordnet. Zusätzlich finden sich die Interviewgruppen, in denen die Codes genannt worden sind und die Anzahl der Interviewgruppen, in denen das Code thematisiert worden ist.

Etwa ein Viertel aller *Codings* beschäftigte sich mit der Oberkategorie „Fragenpakete“ (193 *Codings*). Es wurde die Zusammenstellung der Fragenpakete (65 *Codings*) diskutiert, der auffallende Rückgang der Teilnehmerzahl (60 *Codings*), wie ernsthaft die Fragenpakete beantwortet wurden (28 *Codings*) und welche Wirkung das Beantworten der Fragenpakete auf die Teilnehmer hatte (40 *Codings*). Diese Auswertungen sind in Tabelle 8 und Tabelle 9 aufgeführt.

Mit der Zusammenstellung der Fragenpakete waren die Teilnehmer im Großen und Ganzen zufrieden. So fanden sie „[...] den Umfang von zehn Fragen gut.“ (Gruppe_04_Absatz139_Person5, weiblich) (acht *Codings*) und auch der Schwierigkeitsgrad wurde als angemessen empfunden (16 *Codings*): „Sie waren mal sehr einfach und mal sehr schwierig zu beantworten. Aber insgesamt glaube ich waren sie auf einem relativ guten Niveau.“ (Gruppe01_Absatz55_Person2; männlich). Teilweise wurde auch festgestellt, dass die einzelnen Fragen innerhalb eines Fragenpaketes unterschiedliche Schwierigkeitsgrade aufwiesen (fünf *Codings*). Die abgefragten Themen gingen inhaltlich teilweise über die Vorlesungsinhalte hinaus, was von den Teilnehmern positiv bewertet wurde (sieben *Codings*): „Aber wie vorhin schon gesagt wurde, fand ich es ziemlich gut, dass man so auch Aspekte, die so in der universitären Lehre hier nicht schwerpunktmäßig behandelt wurden, sozusagen wieder aufgegriffen wurden.“ (Gruppe04_Absatz60_Person 3;

männlich). Im Gegensatz dazu hatten die Teilnehmer teilweise das Gefühl, „[...]“, dass gerade zu Beginn eines Modules [...] sich die Fragen nicht komplett mit den Lehrinhalten einer Woche gedeckt haben.“ (Gruppe 03_Absatz 64_Person 4; männlich) und somit die zeitliche Passung der Fragen weiter perfektioniert werden müsste (fünf *Codings*). Der Aspekt, dass sich einzelne Fragen wiederholten (24 *Codings*), ist einigen Teilnehmern aufgefallen (neun *Codings*). So entstand bei manchen Fragen das Gefühl, dass sich die Frage durch die Wiederholung festigte (acht *Codings*): „Wenn man die Fragen schon mal gekreuzt hat oder schon mal konnte, hatte das keinen so großen Effekt. Wenn man sie schon einmal nicht konnte, hatte das schon einen bisschen größeren Lerneffekt.“ (Gruppe 01_Absatz 58_Person 1; männlich).

Für den Rückgang der Teilnehmerzahl bezüglich der Beantwortung der Fragenpakete sahen die Teilnehmer zum einen allgemeine (45 *Codings*) und zum anderen semesterbezogene Gründe ausschlaggebend (15 *Codings*). Zu den allgemeinen Gründen gehörte das Inkrafttreten der DSGVO (elf *Codings*), welche durch das Einstellen der Erinnerungen zu einem Vergessen des Bearbeitens der Fragenpakete führte (acht *Codings*). Als weiteres Problem sahen die Studienteilnehmer einen Zeitmangel (zehn *Codings*) und „[...] dass es zum Ende des Semesters dann auch stressiger wird“ (Gruppe 03_Absatz 79_Person 2; weiblich). Des Weiteren sahen die Teilnehmer noch einen Grund in einem mangelnden Durchhaltevermögen (fünf *Codings*) bzw. in einer nachlassenden Motivation, sobald die Beantwortung der Fragenpakete einmal vergessen worden ist (fünf *Codings*): „Einerseits würde ich sagen, dass die Motivation zu Beginn des Semesters am Höchsten ist, mit einem kleinen Durchhänger in der Mitte zum Schluss noch einmal mit einem Anstieg an Motivation.“ (Gruppe 03_Absatz 78_Person 4; männlich). Zusätzlich kam noch hinzu, dass die Fragenpakete keine akute Relevanz für „[...] die Klausuren, die zwischendurch noch kamen [...]“ (Gruppe 04_Absatz 74_Person 5; weiblich) hatten (fünf *Codings*) und damit einhergehend auch die Prioritäten der Teilnehmer anders gesetzt worden sind (fünf *Codings*). Zu den semesterbezogenen Gründen zählte vor allem, dass sich die Teilnehmer wünschten, dass die Lehrenden mehr auf das Projekt hätten aufmerksam machen sollen (neun *Codings*).

Tabelle 8: Auswertung der Fokusgruppengespräche – Fragenpakete (Teil eins)

Code	Codings	Interview- gruppe	Anzahl Interview- gruppen
I. Fragenpakete – Teil eins	193	1,2,3,4	4
I.1 Bewertung der Zusammenstellung der Fragenpakete	65	1,2,3,4	4
I.1.1 Umfang angemessen	8	1,2,3,4	4
I.1.2 inhaltliche Passung	7	2,3,4	3
I.1.2.1 gut	7	2,3,4	3
I.1.3 zeitliche Passung	5	3,4	2
I.1.3.1 könnte teilweise besser sein	5	3,4	2
I.1.4 Schwierigkeitsgrad	21	1,2,3,4	4
I.1.4.1 gut gewählt	16	1,2,4	3
I.1.4.2 unterschiedlichen Schwierigkeitsgrad bemerkt	5	1,2,3,4	4
I.1.5 Wiederholung von einzelnen Fragen	24	1,2,3,4	4
I.1.5.1 bemerkt	11	1,2,3,4	4
I.1.5.1.1 ja	9	1,2,3,4	4
I.1.5.1.2 nein	2	4	1
I.1.5.2 Effekt	13	1,2,3,4	4
I.1.5.2.1 Frage gefestigt	8	1,2,3,4	4
I.1.5.2.1a Lernzuwachs bei schwierigen Fragen	3	1,4	2
I.1.5.2.2 bei leichten Fragen keinen Effekt	1	1	1
I.1.5.2.3 negative Emotionen	4	2,3	2
I.2 Rückgang der Teilnahmezahl	60	1,2,3,4	4
I.2.1 allgemeine Gründe	45	1,2,3,4	4
I.2.1.1 Datenschutzgrundverordnung	11	1,2,3,4	4
I.2.1.1.1 vergessen	8	1,2,3,4	4
I.2.1.2 Zeitmangel	10	1,2,3,4	4
I.2.1.3 mangelndes Durchhaltevermögen	5	1,2,3	3
I.2.1.4 nicht akute Relevanz	5	1,2,4	3
I.2.1.5 andere Prioritäten	5	1,2,3	3
I.2.1.6 einmal vergessen → Motivation verloren	5	1,2,3	3
I.2.1.7 freiwillig	2	1,4	2
I.2.1.8 nicht genügend Mehrwert	2	2,3	2
I.2.2 semesterbezogen	15	1,2,3,4	4
I.2.2.1 Lehrende haben zu wenig Werbung gemacht	9	1,2,3	3
I.2.2.2 Abschlussklausur	4	1,2,4	3
I.2.2.3 Qualität der Lehre	2	2,3	2

Dargestellt sind tabellarisch die Ergebnisse der Fokusgruppengespräche. Den einzelnen Codes sind die Anzahlen der jeweiligen Codings zugeordnet. Zusätzlich finden sich die Interviewgruppen, in denen die Codes genannt worden sind und die Anzahl der Interviewgruppen, in denen das Code thematisiert worden ist.

Viele Teilnehmer hatten die Fragen ernsthaft nach ihrem „besten Wissen und Gewissen“ (Gruppe04_Absatz62_Person2; weiblich) beantwortet (zwölf *Codings*), wobei es auch einige gab, die die Fragen nur teilweise (15 *Codings*) ernsthaft beantwortet hatten. Hierbei gab es Teilnehmer, die dann nur rieten (sechs *Codings*): „Wenn ich jetzt das Gefühl hatte, dass ich überhaupt nicht weiß, wovon geredet wird, habe ich einfach irgendetwas gekreuzt.“ (Gruppe03_Absatz70_Person2; weiblich).

Die Fragenpakete wirkten auf manche Teilnehmer motivierend (16 *Codings*) und förderten, dass sie hinterher noch einmal Inhalte „nachlesen“ (Gruppe03_Abschnitt70_Person3, weiblich) und somit fehlendes Wissen nacharbeiteten (elf *Codings*). An ausgelösten Emotionen (21 *Codings*) durch die Fragenpakete wurde zwischen positiven Emotionen (13 *Codings*) im Sinne von zum Beispiel Ansporn (fünf *Codings*) und negativen Emotionen (acht *Codings*) unterschieden: „ZWIESPÄLTIGE EMOTIONEN. Also wirklich tatsächlich. Manchmal (.“

hat es echt total Spaß gemacht, wenn man gut war und der Medizinstudent ist ja auch immer ehrgeizig, aber gleichzeitig natürlich auch Frustration, wenn man irgendwie gemerkt hat, dass man einfach nicht so erfolgreich ist, wie man vielleicht gehofft hatte in der Woche.“ (Gruppe02_Abschnitt31_Person4; weiblich).

Tabelle 9: Auswertung der Fokusgruppengespräche – Fragenpakete (Teil zwei)

Code	Codings	Interview- gruppe	Anzahl Interview- gruppen
I. Fragenpakete – Teil zwei			
I.3 ernsthaft beantwortet	28	1,2,3,4	4
I.3.1 ja	12	1,2,3,4	4
I.3.2 teilweise	15	1,2,3,4	4
I.3.2.1 mit Hilfestellung	1	1	1
I.3.2.2 ohne Hilfestellung	6	1,3,4	3
I.3.2.3 geraten	6	2,3	2
I.3.3 nein	1	2	1
I.3.3.1 keine Werbung	1	2	1
I.4 Wirkung auf einen selbst	40	1,2,3,4	4
I.4.1 motivierend	19	1,2,3,4	4
I.4.1.1 ja	16	1,2,3,4	4
I.4.1.1.1 nacharbeiten	11	1,2,3,4	4
I.4.1.2 nein	3	1,2,4	3
I.4.1.2.1 erst kurz vor der Klausur kreuzen	1	1	1
I.4.1.2.2 wenn man zu wenig weiß	2	2,4	2
I.4.2 Emotionen	21	1,2,3,4	4
I.4.2.1 positive Emotionen	13	1,2,3,4	4
I.4.2.1.1 Ansporn	5	1,3	2
I.4.2.1.2 Spaß	3	2,4	2
I.4.2.1.3 beruhigend	3	1,4	2
I.4.2.1.4 Freude	2	1,3	2
I.4.2.2 negative Emotionen	8	2,3,4	3
I.4.2.2.1 Frustration	6	2,3,4	3
I.4.2.2.2 Ärger	2	2,3	2

Dargestellt sind tabellarisch die Ergebnisse der Fokusgruppengespräche. Den einzelnen Codes sind die Anzahlen der jeweiligen Codings zugeordnet. Zusätzlich finden sich die Interviewgruppen, in denen die Codes genannt worden sind und die Anzahl der Interviewgruppen, in denen das Code thematisiert worden ist.

Mit der Oberkategorie der „Interdisziplinären Lehrveranstaltungen“ beschäftigten sich 86 *Codings* (siehe Tabelle 10). Es wurden Gründe erörtert, diese zu besuchen (42 *Codings*) bzw. nicht zu besuchen (44 *Codings*).

Der Besuch wurde als „[...] erstens sehr interessant und zweitens oft auch sehr aufschlussreich [...]“ (Gruppe03_Absatz20_Person3; weiblich) eingestuft (neun *Codings*). Wobei sich die Teilnehmer auch „[...] schon Klausurtipps erhofft [...]“ (Gruppe02_Absatz86_Person4; weiblich) hatten (sieben *Codings*). Ebenfalls wurde die Besprechung und Diskussion von schweren Fragen (sieben *Codings*) als Grund zum Besuchen einer interdisziplinären Lehrveranstaltung betrachtet. Positiv wurde auch die Interdisziplinarität an sich und damit einhergehend die Tatsache bewertet, dass den Studienteilnehmern der „Blick auf das komplette Ganze“ (Gruppe01_Abschnitt87_Person3) eröffnet wurde (sieben *Codings*).

Ein Hauptproblem, weshalb die Teilnehmer auch teilweise selbst keine Veranstaltung besucht hatten, sahen sie darin, dass diese nicht genug publik gemacht wurden (19 *Codings*): „Aber ich muss ehrlich gesagt auch dazu sagen, also ich habe jetzt das vierte Klinische hinter mir, dass mir auch nicht bewusst gewesen ist, dass es so interdisziplinäre Veranstaltungen gegeben hätte, wenn es die gab. Also (..) da hätte ich gerne teilgenommen, aber ich wusste auch nicht, dass ich die verpasst habe.“ (Gruppe01_Absatz87_Person3; männlich).

Nach der Meinung der Teilnehmer sei eine Lehrveranstaltung durch die zeitliche und räumliche Festlegung (sechs *Codings*) recht unflexibel (13 *Codings*). Zusätzlich sahen die Teilnehmer teilweise keinen ausreichenden Mehrwert (sechs *Codings*): „Die Lehrveranstaltungen waren meistens kurz vor der Klausur. Da hatte ich zumindest den Kopf oder den Fokus mehr auf die Klausur gesetzt und die Veranstaltung war dann inhaltlich doch nicht so klausurrelevant“ (Gruppe02_Absatz92_Person3; männlich).

Tabelle 10: Auswertung der Fokusgruppengespräche – Interdisziplinäre Lehrveranstaltungen

Code	Codings	Interviewgruppe	Anzahl Interviewgruppen
I. Interdisziplinäre Lehrveranstaltungen	86	1,2,3,4	4
I.1 Gründe zu besuchen	42	1,2,3,4	4
I.1.1 Interesse	9	1,2,3	3
I.1.2 Klausurtipps	7	2,3	2
I.1.3 Besprechung von schwierigen Fragen	7	2,3,4	3
I.1.4 Vermittlung eines allgemeinen Blickes	7	1,2,3	3
I.1.5 realitätsnah	4	1,2	2
I.1.6 Vermittlung neuer Sichtweisen	4	1,4	2
I.1.7 Einblick in ärztlichen Alltag	1	1	1
I.1.8 Anwendung des eigenen Wissens	1	1	1
I.1.9 Wiederholung vergangener Module	1	2	1
I.2 Gründe nicht zu besuchen	44	1,2,3,4	4
I.2.1 nicht genug publik gemacht	19	1,2,3,4	4
I.2.1.1 steht nicht im Stundenplan	4	1,2,3	3
I.2.2 unflexibel	13	1,2,3,4	4
I.2.2.1 ungünstiger Termin	6	1,2,3,4	4
I.2.2.2 vorgegebener Ort	4	1,2,3	3
I.2.3 kein ausreichender Mehrwert	6	1,2,3	3
I.2.4 mangelndes Interesse	2	1,3	2
I.2.5 Prioritäten anders gesetzt	4	1,3	2

Dargestellt sind tabellarisch die Ergebnisse der Fokusgruppengespräche. Den einzelnen Codes sind die Anzahlen der jeweiligen Codings zugeordnet. Zusätzlich finden sich die Interviewgruppen, in denen die Codes genannt worden sind und die Anzahl der Interviewgruppen, in denen das Code thematisiert worden ist.

Zuletzt diskutierten die Teilnehmer noch die Oberkategorie „Verbesserungsvorschläge“ (122 *Codings*) für *#clue* (siehe Tabelle 11). Diese betrafen die Fragenpakete (70 *Codings*), die interdisziplinären Lehrveranstaltungen (34 *Codings*) und mögliche Maßnahmen zur weiteren Steigerung des Lernerfolgs (16 *Codings*).

Die Verbesserungsvorschläge zu den Fragenpaketen gliederten sich in technische (37 *Codings*), organisatorische (24 *Codings*) und inhaltliche Faktoren (sieben *Codings*). Letztere werden im Folgenden nicht weiter aufgegriffen, da die einzelnen inhaltlichen Faktoren mit weniger als fünf *Codings* betroffen sind, welches die Untergrenze für eine nähere Beschreibung bildet. In Tabelle 11 können die einzelnen Faktoren nachgelesen werden. An der technischen Umsetzung kritisierten die Studienteilnehmer vor allem die Lernplattform ILLIAS (23 *Codings*), bei der zum einen das Layout wenig ansprechend erschien (15 *Codings*) und zum anderen das Aufrufen über den e-Campus zu umständlich war (acht *Codings*): „[...] dass dann Leute, die das wirklich nervt, sagen: ‚Ich habe da keine Lust mehr drauf, ich lasse es sein.‘. Das ist ein bisschen schade, (.) wenn das der Grund ist.“ (Gruppe_01_Absatz103_Person 5; weiblich). Um dies zu vermeiden, entstand der Wunsch danach, dass „[...] eine kleine App“ (Gruppe01_Absatz104_Person2; männlich) entwickelt werden sollte (fünf *Codings*). Organisatorisch (24 *Codings*) wurde sich eine längere Verfügbarkeit der Fragenpakete über mehr als eine Woche gewünscht (neun *Codings*). Zusätzlich wurde „[...] eine mögliche gleichmäßige Erinnerung in allen Modulen“ (Gruppe02_Abschnitt112_Person2) gefordert (fünf *Codings*).

Bezüglich der interdisziplinären Lehrveranstaltungen wurden primär Ideen über die Gestaltung ausgetauscht (34 *Codings*). Die Lehrveranstaltungen würden vor allem dann positiv von den Studienteilnehmern angenommen werden, wenn sie abwechslungsreich gestaltet werden würden (elf *Codings*) und den Studienteilnehmern eine längerfristige Planung ermöglicht werden würde (neun *Codings*). Einen weiteren wichtigen Faktor sahen die Teilnehmer auch darin, eine direkte Relevanz (neun *Codings*) zur nächsten Klausur zu schaffen (fünf *Codings*).

Der Lernerfolg der Teilnehmer ließe sich durch ein detaillierteres Feedback (zehn *Codings*) durch z. B. Hintergrundinformationen zu den richtigen Antworten weiter steigern (neun *Codings*).

Die Verbesserungsvorschläge wurden in einem gesonderten Dokument erfasst und teilweise bei der Weiterentwicklung des Projektes im Wintersemester 2018/2019 bereits beachtet (siehe Anhang 8).

Tabelle 11: Auswertung der Fokusgruppengespräche – Verbesserungsvorschläge

Code	Codings	Interview- gruppe	Anzahl Interview- gruppen
I. Verbesserungsvorschläge	122	1,2,3,4	4
I.1 Fragenpakete	70	1,2,3,4	4
I.1.1 technisch	37	1,2,3,4	4
I.1.1.1 ILIAS	23	1,2,3,4	4
I.1.1.1.1 Layout	15	1,2,3,4	4
I.1.1.1.1a "bestätigen und weiter"-Button abschaffen	6	1,2,3,4	4
I.1.1.1.1b runterscrollen zur richtigen Antwort	6	1,3,4	3
I.1.1.1.1c Antwortpunkte ordnen	3	3,4	2
I.1.1.1.2 Aufrufen zu umständlich	8	1,3,4	3
I.1.1.2 eine App entwickeln	5	1,4	2
I.1.1.3 Vergleich zu Mitstudierenden	3	2,3	2
I.1.1.4 Anzeige der Fragenpakete	2	3	1
I.1.1.5 Quick-Evaluation pro Fragenpaket	1	4	1
I.1.1.6 Forum zum gemeinsamen Diskutieren	1	4	1
I.1.1.7 Sitzung bei AMBOSS erstellen	1	1	1
I.1.1.8 Module und Fragen aktiv selbst auswählen	1	2	1
I.1.2 organisatorisch	24	1,2,3,4	4
I.1.2.1 Verfügbarkeit der Fragenpakete	11	1,2,3,4	4
I.1.2.1.1 mehr als eine Woche freigeschaltet lassen	9	1,2,3,4	4
I.1.2.1.2 begrenzte Bearbeitungsdauer von 50 Minuten	2	1,3	2
I.1.2.1.2a kürzer	1	3	1
I.1.2.1.2b abschaffen	1	1	1
I.1.2.2 Erinnerungen	13	1,2,3,4	4
I.1.2.2.1 gleichmäßig	5	1,2,3,4	4
I.1.2.2.2 früher als einen Tag vor Ablauf	4	4	1
I.1.2.2.3 persönliche Hinweise von aktuellen Lehrenden	4	2,3	2
I.1.3 inhaltlich	7	1,3,4	3
I.1.3.1 Wiederholung von älteren Modulen verstärken	4	1,3,4	3
I.1.3.2 mehr Themengebiete pro Fragenpakete abdecken	2	4	1
I.1.3.3 vereinzelt Fragen von Lehrenden	1	3	1
I.1.4 Schwierigkeitsgrad	1	2	1
I.1.4.1 angeben	1	2	1
I.1.5 Projekt auch in anderen Fächern anbieten	1	1	1
I.2 Interdisziplinäre Lehrveranstaltungen	34	1,2,3,4	4
I.2.1 Vorstellungen zur Gestaltung	34	1,2,3,4	4
I.2.1.1 abwechslungsreich	11	1,2,3,4	4
I.2.1.1.1 Bier und Brezel	9	2,3,4	3
I.2.1.1.2 motivierte Lehrende	1	2	1
I.2.1.2 längerfristige Planung	9	2,3,4	3
I.2.1.2.1 im Stundenplan aufnehmen	4	2,3	2
I.2.1.3 Relevanz	9	1,2,4	3
I.2.1.3.1 für aktuelle Klausur	5	1,2	2
I.2.1.4 für mehrere Semester anbieten	2	1,4	2
I.2.1.5 informativ	2	1,2	2
I.2.1.6 live übers Internet übertragen	1	4	1
I.3 Lernerfolg steigern	16	1,2,3,4	4
I.3.1 detaillierteres Feedback	10	1,2,3,4	4
I.3.1.1 Hintergrundinformationen zur Antwort	9	1,2,3,4	4
regelmäßigere Teilnahme	2	1,4	2
noch besser auf den Lehrplan abstimmen	2	3	1
Fächerwiederholungen verstärken	1	3	1
von Beginn an fachübergreifend anbieten	1	1	1
I.4 Motivation steigern	2	1,2	2
unmittelbare Relevanz	1	1	1
noch detaillierteres Feedback der eigenen Leistung	1	2	1

Dargestellt sind tabellarisch die Ergebnisse der Fokusgruppengespräche. Den einzelnen Codes sind die Anzahlen der jeweiligen Codings zugeordnet. Zusätzlich finden sich die Interviewgruppen, in denen die Codes genannt worden sind und die Anzahl der Interviewgruppen, in denen das Code thematisiert worden ist.

4 Diskussion

4.1 Die wesentlichen Ergebnisse

Im Rahmen der vorliegenden Arbeit wurde das Projekt *#clue* als innovatives *E-Learning*-Angebot in die klinische Lehre der UMG eingeführt. Hierdurch sollte Studierenden semesterbegleitend eine langfristige Vorbereitung auf das Staatsexamen geboten und eine gedankliche Vernetzung (horizontale Integration) der einzelnen klinischen Module ermöglicht werden.

Auf die unter Kapitel 1.5 aufgeworfenen Studienfragen ergaben sich folgende Antworten:

(1) Im Verlauf des Semesters wurden wiederholt aufgetretene Fragen bei zusammenfassender Betrachtung der Schwierigkeitsgrade am Ende des Semesters nicht signifikant besser beantwortet, als am Ende des Semesters erstmalig aufgetretene Fragen. Bei differenzierter Betrachtung nach Schwierigkeitsgrad ließen sich bei Items mit einem höheren Schwierigkeitsgrad signifikante Unterschiede zwischen den Summenscores der Interventions- und Kontrollitems finden.

(2) Im Verlauf von 14 Wochen ließ sich bei Studienteilnehmern im Hinblick auf den erreichten Wochenscore weder eine deutliche Verbesserung noch Verschlechterung feststellen. Nach Zusammenfassung aller Summenscores der verschiedenen Schwierigkeitsgrade, ließ sich bei mehrfachem Abfragen im zeitlichen Verlauf ein signifikanter Lernzuwachs bei den Interventionsitems feststellen.

(3) Die Teilnehmerrate nahm bei diesem freiwilligen Projekt im zeitlichen Verlauf ab. Je nach Semester betrug die Teilnehmerrate durchschnittlich zwischen 28% (fünftes klinisches Semester) und 53% (drittes klinisches Semester) der Studienteilnehmer im jeweiligen Semester.

(4) Das Projekt wurde von den Studienteilnehmern insgesamt als gut bewertet und würde auch in weiteren Semestern von diesen genutzt werden.

In den folgenden Kapiteln werden die quantitativen und qualitativen Ergebnisse der hier vorgestellten Studie bezüglich der in Kapitel 1.5 aufgeworfenen Studienfragen mit bestehender Literatur diskutiert.

4.2 Vergleich des Lernerfolgs zwischen mehrfach wiederholten und neuen Inhalten (Studienfrage eins)

In Kapitel 1.4 wurde dargelegt, dass Prüfungen das Lernverhalten beeinflussen (Cilliers et al. 2010). Zudem wurden verschiedene Prüfungsformate vorgestellt. In der hier vorliegenden Arbeit wurde der *testing effect* bei formativen Prüfungen untersucht. Die verschiedenen Theorien, wie ein *testing effect* zustande kommen kann, wurden in Kapitel 1.4.3 vorgestellt. In der hier beschriebenen Studie konnte unter Einbeziehung aller Schwierigkeitsgrade im *Exit Exam* zum Zeitpunkt T3 kein signifikanter Unterschied zwischen der Bearbeitung der Interventionsitems und Kontrollitems festgestellt werden.

Dieser Beobachtung können verschiedene Ursachen zugrunde liegen. Zum einen müssen an dieser Stelle die zugrundeliegenden Anforderungsniveaus der Interventions- und Kontrollitems betrachtet werden (siehe Kapitel 2.4.1). Unter Einbeziehung aller Schwierigkeitsgrade ließ sich in der primären Analyse ein signifikanter Unterschied zwischen den *Overall*-Summenscores der Interventionsitems (T1) und der Kontrollitems (T3) feststellen. Dieser konnte in der Sensitivitätsanalyse allerdings nicht bestätigt werden. Eine mögliche Erklärung hierfür findet sich in Kapitel 4.6.2. Die Sensitivitätsanalyse deutete darauf hin, dass zum Semesterende den Teilnehmern die Beantwortung von neuen Fragen genauso leicht bzw. schwer fiel, wie zum Semesteranfang. Es ist möglich, dass die Kontrollitems von Beginn an schwerer waren als die Interventionsitems. Zur Einsortierung der Items in die verschiedenen Schwierigkeitsgrade wurde der Hammerindex von AMBOSS[®] als Orientierung verwendet. Die Lernplattform AMBOSS[®] stellt mit dem Hammerindex zu jedem Zeitpunkt lediglich eine Momentaufnahme des aktuellen Antwortverhaltens aller Nutzer dar. Allerdings können die Nutzer auch Fragen mehrfach beantworten, weshalb es zu einer Verzerrung der eigentlichen Anforderungsniveaus kommen kann. Da jedoch keine weiteren Details über die Nutzer von AMBOSS[®] bekannt sind, z. B. wann genau und wie oft diese die IMPP-Fragen beantworten, kann hier keine allgemeine Aussage darüber getroffen werden, inwiefern das einen Einfluss auf die Anforderungsniveaus der Items hatte. Dadurch, dass durch alle Nutzer der Hammerindex zustande kommt, ergibt sich eine sehr hohe Grundgesamtheit „n“ für jedes einzelne beantwortete Item. Aus diesem Grund wurde die ursprüngliche Einteilung der Anforderungsniveaus für die statistische Auswertung verwendet und es wird im Folgenden davon ausgegangen, dass die Anforderungsniveaus zwischen Interventions- und Kontrollitems gemäß der ursprünglichen Einteilung gleich waren.

Die *Overall*-Kontrollitems wurden nicht signifikant schlechter beantwortet als die *Overall*-Interventionsitems zum Zeitpunkt T3. So lässt sich vermuten, dass über die komplette Dauer des Semesters ein allgemeiner Lernfortschritt stattgefunden hat, der sich auch auf die Bearbeitung der Kontrollitems und damit auf neue Inhalte übertragen hat. Nun stellt sich die Frage, wodurch dieser allgemeine Lernfortschritt entstanden ist. Dieser konnte durch den Besuch von Vorlesungen, durch die Teilnahme an Klausuren oder den indirekten *testing effect* entstanden sein. Darüberhinausgehend konnte sich im *Exit Exam* bei einer regelmäßigen

Teilnahme an der Beantwortung der Fragenpakete oder durch die Beantwortung von MC-Fragen zur Vorbereitung von Klausuren bereits ein gewisser Trainingseffekt (*transfer-appropriate hypothesis*) bemerkbar machen (Morris et al. 1977). Interessant ist an dieser Stelle, dass der Lernerfolg bei Betrachtung der verschiedenen Schwierigkeitsgrade unterschiedlich stark gewesen zu sein schien.

Im Folgenden wird diskutiert, inwiefern den Teilnehmern die Beantwortung von leichten, mittelschweren und schweren Kontrollitems bei erstmaliger Bearbeitung zum Zeitpunkt T3 leichter bzw. schwerer fiel, im Vergleich zu den Interventionsitems zum Zeitpunkt T1 und inwiefern sich eventuell unter den drei Schwierigkeitsgraden Unterschiede im Lernerfolg feststellen ließen.

Smith et al. (2009) beobachteten in einem anderen Studienaufbau, dass der *testing effect* nicht nur auf die behandelten Fragen, sondern auch auf isomorphe Fragen übertragen werden kann und dadurch ein Effekt entsteht, der als indirekter *testing effect* verstanden werden kann. In dieser Studie wurden Studierenden während einer Vorlesung MC-Fragen präsentiert, die sie ohne die Präsentation der Ergebnisse mit ihren Mitstudierenden ausgiebig diskutieren durften. Wenig überraschend ist, dass beim wiederholten Abfragen von den anfangs 48% der Fragen, die falsch beantwortet worden sind, nun 42% richtig beantwortet wurden. Im Gegensatz dazu ist die Beobachtung, dass isomorphe Fragen, die die Studierenden im Anschluss ohne vorherige Diskussion beantworten sollten, mit einer Quote von 77% richtig beantwortet wurden, sehr wohl überraschend. Diese isomorphen und neuen Fragen wurden somit signifikant besser beantwortet als die vorherig diskutierten Fragen. Durch die Diskussion mit Kommilitonen wurde also ein Lernfortschritt bei gleichen Themen hervorgerufen (Smith et al. 2009). In der hier vorliegenden Studie wurden zum Zeitpunkt T3 die leichten Kontrollitems signifikant besser beantwortet als die leichten Interventionsitems zum Zeitpunkt T1. Die Bearbeitung neuer leichter Fragen schien den Studierenden somit leichter gefallen zu sein als zu Beginn des Semesters (T1). Bei den mittelschweren und schweren Kontrollitems fiel den Teilnehmern die Bearbeitung zum Zeitpunkt T3 nicht leichter und auch nicht schwerer als die Bearbeitung der Interventionsitems zum Zeitpunkt T1. Der beschriebene indirekte *testing effect* könnte in dieser Studie auf die leichten Kontrollitems gewirkt haben und einen Lernerfolg bei leichten Inhalten hervorgerufen haben. So hatten die Studienteilnehmer die Möglichkeit nach der Bearbeitung der Fragenpakete, die Richtigkeit ihrer eigenen Antworten zu überprüfen und dadurch ggf. Wissenslücken zu erkennen und aufzuarbeiten (Roediger und Karpicke 2006b). In der Fragebogen-Erhebung und den Interviews gaben die Teilnehmer an, sich durch die Bearbeitung der wöchentlichen Fragenpakete noch einmal intensiver mit den abgefragten Themen auseinander gesetzt zu haben. In den interdisziplinären Lehrveranstaltungen wurden darüber hinaus ebenfalls Examensfragen präsentiert und mit den Studierenden gemeinsam diskutiert. Auch dies kann Auswirkungen auf Abschlussprüfungen haben (Nelson et al. 2012). Im *Exit Exam* wurden ebenfalls Themeninhalte abgefragt, die bereits im Verlauf des Semesters in den Fragenpaketen, interdisziplinären Lehrveranstaltungen bzw. Vorlesungen behandelt worden sind.

Da sich die Studienteilnehmer schon mit diesen Themeninhalten beschäftigt hatten, kann es sein, dass es ihnen dadurch leichter fiel die Kontrollitems zum Zeitpunkt T3 zu beantworten. Zum Zeitpunkt T3 beantworteten die Teilnehmer die leichten bzw. mittelschweren Kontrollitems und die leichten bzw. mittelschweren Interventionsitems vergleichbar gut. Interventionsitems mit dem höchsten Schwierigkeitsgrad konnten im *Exit Exam* zum Zeitpunkt T3 signifikant besser beantwortet werden als die schweren Kontrollitems. An dieser Stelle kann die Überlegung aufgestellt werden, dass sich die oben beschriebenen Erklärungen für einen Lernerfolg bei neuen Inhalten nicht auf alle Schwierigkeitsgrade gleichermaßen übertragen lassen. So ließ sich ein indirekter *testing effect* bei leichten Themeninhalten vermuten. Dies traf allerdings nicht auf mittelschwere und schwere Inhalte zu.

Die hier aufgeführten Gründe machen deutlich, dass anhand der vorliegenden Daten keine eindeutige Aussage über einen allgemein größeren Lernerfolg bei wiederholten Inhalten im Vergleich zu neuen Inhalten geäußert werden kann. Erklärbar hierfür können ein allgemeiner Trainingseffekt der kognitiven Prozesse, nicht vergleichbare Anforderungsniveaus zwischen Kontroll- und Interventionsitems, der indirekte *testing effect* oder ein allgemeiner Lernzuwachs durch das Absolvieren des Semesters sein. Diese Effekte scheinen allerdings auf Items aufgeteilt nach Schwierigkeitsgrad einen unterschiedlich starken Einfluss zu haben. So werden schwierige Inhalte bei mehrfacher Wiederholung (Interventionsitems T3) signifikant besser beantwortet als schwierige neue Inhalte (Kontrollitems T3).

4.3 Entwicklung des Leistungsstandes bei mehrfacher Wiederholung (Studienfrage zwei)

Die Entwicklung des allgemeinen Leistungsstandes während Phase zwei des Projektes wurde in Kapitel 3.4.1 graphisch in Form von Wochenscores dargestellt. Hier war im Verlauf des Projektes keine klare Tendenz erkennbar, ob Studierende besser oder schlechter wurden. Mit Ausnahme der fünf Interventionsitems waren die vorgestellten Fragen allerdings auch immer neu für die Studierenden und repräsentierten den aktuellen Vorlesungsinhalt der jeweiligen Woche. Das *Exit Exam* behandelte lediglich Inhalte, die während des Semesters bereits gelehrt worden sind. Nun bestand dieses Fragenpaket zur Hälfte aus neuen Fragen (Kontrollitems). Wie bereits in Kapitel 4.2 erläutert, übertrug sich der Lernfortschritt nicht auf alle Fragen. Somit deutet sich an dieser Stelle der Arbeit kein allgemeiner Trainingseffekt bei der Beantwortung von MC-Fragen an.

Um den eventuell vorhandenen direkten *testing effects* nachweisen zu können, wurden die Leistungsunterschiede bei der Beantwortung der Interventionsitems zu den Zeitpunkten T1, T2 und T3 genauer betrachtet. Hier ließ sich ein signifikanter Leistungsunterschied bei den Interventionsitems unter Einbeziehung aller Schwierigkeitsgrade nachweisen. Dieser Unterschied fand sich zwischen den *Overall*-Summenscores zu T2 und T3 sowie zu T1 und T3. Dieser hielt auch der Sensitivitätsanalyse stand. Lediglich der exakte Zeitpunkt des Unterschieds verschob sich auf den Zeitraum zwischen T1 und T2. Erklären lässt sich dies

dadurch, dass in der primären Analyse nicht beantwortete Items mit null Punkten bewertet wurden und mehr Items zum Zeitpunkt T2 als zum Zeitpunkt T1 betroffen waren. In der Sensitivitätsanalyse wurden genau diese Items aus der Analyse genommen. Dadurch erhöhte sich der *Overall*-Summenscore der Interventionsitems zum Zeitpunkt T2 in der Sensitivitätsanalyse, wodurch der Unterschied zum Zeitpunkt T3 nicht mehr signifikant war.

Bei der Aufschlüsselung der Items nach Schwierigkeitsgraden waren statistisch signifikante Unterschiede im zeitlichen Verlauf in der primären Analyse sowie der Sensitivitätsanalyse nachweisbar. Der Zuwachs war bei schweren Fragen am größten. Auch in den Fokusgruppengesprächen konnten die Studienteilnehmer bestätigen, dass sie vor allem durch die Wiederholung von zuvor falsch beantworteten Fragen einen Lerneffekt feststellen konnten.

Die beobachteten signifikanten Unterschiede zwischen den Summescores zu den verschiedenen Zeitpunkten lassen sich ggf. mit dem direkten *testing effect* begründen. In der Theorie ist eine mögliche Erklärung für den direkten *testing effect* der sog. *retrieval effort*, der umso besser wirken kann, je aufwendiger das Wiederaufrufen der Gedächtnisinhalte gestaltet ist. Die hier vorliegenden Ergebnisse lassen vermuten, dass v. a. schwere Items besonders aufwendig abzurufen sind und so mit einem besonders großen *retrieval effort* in Zusammenhang stehen können.

4.4 Entwicklung der Teilnehmerzahl (Studienfrage drei)

Die Rücklaufquote hinsichtlich der Bearbeitung der Fragenpakete bis zum Ende des Projekts betrug 13,1%. In der EvaSys[®]-Fragebogen-Erhebung gaben 64,4% der Studienteilnehmer an, mindestens ein Fragenpaket beantwortet zu haben. Weniger als die Hälfte aller Teilnehmer besuchte im Gegensatz dazu eine interdisziplinäre Lehrveranstaltung (41,4%). Um die Entwicklung der Teilnehmerzahl nachvollziehen zu können, ließen sich objektive und subjektive Gründe feststellen. Letztere bestätigen teilweise die objektiven Gründe und wurden durch die qualitative Analyse herausgearbeitet.

4.4.1 Objektive Gründe für die Abnahme der Teilnehmerzahl

Während des Projektdurchlaufs ergaben sich mehrere Probleme, auf die die abnehmende Teilnehmerzahl zurückgeführt werden kann. So zeichnete sich zu Beginn des Projekts in allen Semestern eine Teilnahmebereitschaft von über 60% aller Studienteilnehmer ab. In den ersten Wochen ließ sich ein kontinuierlicher Verlust an Teilnehmern beobachten. Dieser Beobachtung könnte der sog. „Neuigkeitseffekt“ zugrunde liegen. So kann bei der Vorstellung von neuen Angeboten für eine kurze Dauer eine erhöhte Motivation hervorgerufen werden (Kerres 2013). Auch in anderen Studien wurden ebenfalls höhere Abbruchquoten bei *E-Learning*-Angeboten beobachtet (Lee und Choi 2011).

Im Verlauf des Projektes trat die DSGVO in Kraft, weshalb die verschiedenen Möglichkeiten, die Studierenden an die Teilnahme zu erinnern, ab der neunten Woche stark reduziert werden mussten.

Darüberhinausgehend gab es in der zehnten Woche ein technisches Darstellungsproblem. Normalerweise wurden neue Fragenpakete in der Auflistung aller Fragenpakete bei Stud.IP nach Namen sortiert angezeigt. Das zehnte Fragenpaket (Modul x.y_Woche 10 (online bis 20. Juni)) wurde somit an erster Stelle vor dem Fragenpaket der ersten Woche angezeigt und nicht (wie von den Studierenden erwartet) am Listenende, was dazu geführt haben könnte, dass unaufmerksame Teilnehmer dieses Fragenpaket übersahen. Dadurch lässt sich die Tatsache erklären, dass in der zehnten Woche ein Minimum an Teilnehmern das Fragenpaket bearbeitet hatte.

Der Zeitpunkt des *Exit Exams*, also die 14. Woche für das dritte, vierte und fünfte klinische Semester, fiel auf die letzte Vorlesungswoche vor der vorlesungsfreien Zeit. Somit spiegeln die 7% (viertes klinisches Semester) bis 18% (drittes klinisches Semester) der Teilnehmer am *Exit Exam* nicht die durchschnittliche Teilnehmerate wider. Wie in Kapitel 3.2 erläutert lag die durchschnittliche Teilnehmerate abhängig vom Semester zwischen 28% (fünftes klinisches Semester) und 53% (drittes klinisches Semester).

4.4.2 Subjektive Gründe für die Abnahme der Teilnehmerzahl

In den Gruppeninterviews und der EvaSys[®]-Fragebogen-Erhebung wurden die vermuteten objektiven Gründe bestätigt. Im Zusammenhang mit den niedrigen Besucherzahlen der interdisziplinären Lehrveranstaltungen wurde vor allem deutlich, dass die Veranstaltung nicht genug publik gemacht worden ist. In allen vier Interviewgruppen wurde betont, dass die Erinnerungen über *Facebook* oder per E-Mail einen wichtigen Aspekt eingenommen hatten. In einer Interviewgruppe ist zwei Teilnehmern auch die falsche Anordnung der Fragenpakete bei Stud.IP aufgefallen.

Abgesehen davon ergab sich dennoch die Frage, ob es noch weitere Gründe für die nachlassende Motivation gab. Wie in Kapitel 1.3 beschrieben, gibt es verschiedene Motivationsqualitäten, wie die intrinsische und extrinsische Motivation. Letztere kann durch eine externe Belohnung hervorgerufen werden (ten Cate et al. 2011). Diese bestand bei dem hier vorgestellten Projekt aus der Möglichkeit, leistungsunabhängig Büchergutscheine zu gewinnen und ein Zertifikat bei regelmäßiger Teilnahme zu erhalten. So wurde in allen Gruppeninterviews erwähnt, dass die Verlosung einen positiven Einfluss darauf hatte, die Fragenpakete zu bearbeiten.

Um allerdings ein tiefgreifendes Lernen von Studierenden zu erreichen (Seale et al. 2000), sollte das Ziel laut Fransson (1977) sein, bei Studierenden eine intrinsische Motivation aufzubauen. Grundlegend hierfür ist, dass Studierende der Lerninnovation einen subjektiven Wert und Nutzen zuordnen können (Pelaccia und Viau 2017). In den Gruppeninterviews waren nicht allen Teilnehmern die Ziele des Projektes bewusst. Zwar konnten alle vier

Gruppen erkennen, das *#clue* eine Vorbereitung für das Staatsexamen ermöglichen sollte, aber nur in einer Interviewgruppe wurde erkannt, dass vermehrt Querschnittsfragen präsentiert worden sind, die auch Lernstoff aus vergangenen Semestern beinhalteten. So sah der Großteil der Interviewteilnehmer in dem Projekt *#clue* keine Relevanz für aktuelle Klausuren. Auch im Zusammenhang mit den interdisziplinären Lehrveranstaltungen wurde noch einmal deutlich, dass auch hier der Fokus auf eine anstehende Klausur gelegt war. Dass Prüfungen das Gelernte besser reproduzieren lassen, als das erneute Lesen des Materials wurde schon in mehreren Studien nachgewiesen (Roediger und Karpicke 2006b; Karpicke und Roediger 2008). Studierende sind sich dem positiven Effekten von regelmäßigem Testen oft nicht bewusst (Karpicke und Roediger 2008; Schmidmaier et al. 2011), weshalb die Gefahr besteht, dass sie falsche Lernstrategien wählen (Karpicke et al. 2009). Diese Annahmen können wir in unserer Studie bestätigen. Ein Teil der Teilnehmer an den Gruppeninterviews konnte keinen wesentlichen Lernerfolg durch *#clue* feststellen. Einige Teilnehmer machten den Lernerfolg davon abhängig, wie intensiv sie sich selbst schon zuvor mit dem Lernstoff auseinandergesetzt hatten. So kam es zum Semesterende zu einer enormen Abnahme der Teilnehmerquote, obwohl *#clue* durch die Fragenpakete und die interdisziplinären Lehrveranstaltungen eine sinnvolle Unterstützung zur Klausurvorbereitung bot, indem examensrelevante Themen semesterbegleitend vertieft wurden. Aus diesem Grund ist es wichtig, den Studierenden die Vorteile von formativen Prüfungen und dem *testing effect* zu erläutern. Wenn die Studierenden die Wirkungsweise von regelmäßigen Tests auf die eigene Lernleistung verstanden haben, passen sie auch ihre Lernstrategien an (Dobson und Linderholm 2015).

Einen weiteren Einfluss auf die intrinsische Motivation haben motivierte Dozierende (Seale et al. 2000), die selbst von der Lerninnovation überzeugt sind und das Projekt bewerben. Dies war in den Gruppeninterviews der größte Einfluss der Teilnehmer, am Projekt teilzunehmen und könnte erklären, warum sich in den verschiedenen Semestern teilweise Unterschiede in der Teilnahmebereitschaft zeigten. Bei der organisatorischen Verbesserung des Projektes wurden regelmäßige und persönliche Hinweise von Lehrenden gefordert, um damit das Projekt in Erinnerung zu rufen. Vor allem die Einstellung der Erinnerungs-E-Mails führte bei den Teilnehmern zu einem Vergessen des Projekts. Bei der Zusammenstellung von Lerninnovationen, die die intrinsische Motivation fördern sollen, ist es zudem wichtig den Studierenden ein selbstständiges Planen zu ermöglichen (Kusurkar et al. 2011a; Kusurkar et al. 2011b). So besteht allgemein die Möglichkeit bei *E-Learning*-Angeboten selbst über den Ort und die Zeit zu bestimmen (Kerres 2013). Die zeitliche und räumliche Flexibilität für die Bearbeitung der Fragenpakete ermöglichte den Studierenden ein selbstständiges Planen und wurde positiv in den Gruppeninterviews bewertet. Allerdings forderten sie eine längere Verfügbarkeit der Fragenpakete über eine Woche hinaus, woraus sich schließen ließ, dass die Studierenden noch selbstbestimmter darüber entscheiden wollten, wann sie die Fragenpakete bearbeiten. Da die Studierenden, wie bereits erläutert, allerdings dazu neigen falsche Studienstrategien zu wählen, wurde die Verfügbarkeit seit dem

nachfolgenden Wintersemestern nur um drei Tage verlängert. Dadurch wird weiterhin eine semesterbegleitende Prüfungsvorbereitung gewährleistet und die Studierenden können dies nicht bis ans Semesterende aufschieben. Nur 41,4% der Studierenden, die an der Fragenbogen-Erhebung teilnahmen, gaben an, eine interdisziplinäre Lehrveranstaltung besucht zu haben. Als Grund dafür wurde u. a. in den Gruppeninterviews angegeben, dass die bei den Fragenpaketen bestehende Flexibilität hier nicht mehr gegeben war, wodurch die intrinsische Motivation, eine solche Veranstaltung zu besuchen, gesunken sein könnte.

Zudem sollten die Prüfungen ein gutes Anforderungsniveau mit einer anschließenden ausführlichen Bewertung aufweisen (Kusurkar et al. 2011b; Pelaccia und Viau 2017). Das Anforderungsniveau der Fragenpakete wurde sowohl in der Fragebogen-Erhebung als auch in den Gruppeninterviews als gut gewählt bewertet. Allgemein erkannten die Teilnehmer durch das Feedback ihre Wissenslücken und konnten dadurch ihren Wissens- und Leistungsstand überprüfen. Eine ausführliche Rückmeldung über die abgelegte Leistung fehlte allerdings den Interviewteilnehmern. So wurde in allen vier Interviewgruppen eine ausführliche Hintergrundinformation zur richtigen Antwort für eine weitere Steigerung des Lernerfolgs gefordert.

Somit sollte ggf. bei der Vorstellung von *#clue* darauf geachtet werden, verstärkt die persönlichen Vorteile (wie z. B. den *testing effect*) für die Studierenden, die sich aus der Bearbeitung der Fragenpakete ergeben, zu betonen. Dadurch ist denkbar, dass sie ihre Lernstrategien anpassen (Dobson und Linderholm 2015). Die Teilnahme selbst sollte nach wie vor freiwillig bleiben. Dennoch sollte eine regelmäßige Erinnerung gewährleistet werden. Gegebenenfalls kann durch die Vergabe von Leistungspunkten (siehe Kapitel 1.1) auch ein Anreiz geschaffen werden und somit eine direktere Relevanz erzielt werden. Denkbar wäre, dass hierdurch die intrinsische und extrinsische Motivation weiter gesteigert werden kann und die Teilnahmebereitschaft zukünftig steigt.

4.5 Beurteilung des Projektes *#clue* (Studienfrage vier)

Bei der Einführung neuer Bildungsmethoden sollten immer parallel Evaluationsinstrumente zum Einsatz kommen, da die Mitgestaltung durch Studierende wichtig ist (Thomas et al. 2016). Dies war bei *#clue* zum einen durch die studentische Gruppe möglich, die das Projekt parallel zum Ablauf evaluieren und Vorschläge einbringen konnte. Darüberhinausgehend erhielten die Studierenden am Ende des Semesters die Möglichkeit, ihre Meinung in EvaSys[®]-Fragebögen oder in Fokusgruppengesprächen zu äußern. Um den Lernerfolg oder allgemein die Bewertung von neuen *E-Learning*-Angeboten nachvollziehen zu können, sind laut Kerres (2013) u. a. folgende Schwerpunkte von Interesse: Qualität des *E-Learning*-Angebots, Emotionen und Lernmotivation, Lernverhalten, subjektiver Lernerfolg sowie Zufriedenheit (Kerres 2013).

4.5.1 Qualität des *E-Learning*-Angebots

Bezüglich der erlebten Qualität sind der Inhalt, das Format und die Umsetzung sowie die Kommunikation von Interesse (Kerres 2013). Im Vergleich zu anderen Möglichkeiten des Selbststudiums, wie z. B. AMBOSS[®], wurde von den Studierenden in allen Interviewgruppen *#clue* als eine gute Ergänzung angesehen. Einige beurteilten die vorhandenen Medien auch als ausreichend. An dieser Stelle ist allerdings zu betonen, dass AMBOSS[®], als weltweit zur Verfügung stehende Lernsoftware, seine Inhalte nicht auf die Vorlesungswoche der UMG abstimmt (Kuhn et al. 2018). Somit müssten Studierende ihre Fragenpakete selbst zusammenstellen. Gerade diese individuelle Anpassung an den Göttinger Modulplan wurde als sehr positiv empfunden. Darüberhinausgehend ist es wichtig, dass die Zusammenstellung der Fragen einen ausgewogenen Schwierigkeitsgrad aufweist (Kusurkar et al. 2011b; Pelaccia und Viau 2017). Die Auswahl der Fragen wurde unter diesem Aspekt als gut gewählt betrachtet. Auch der Umfang wurde als angemessen bewertet. Die inhaltliche Anpassung zu den Vorlesungsthemen hätte allerdings laut den Interviewteilnehmern teilweise noch besser sein können. Darüberhinausgehend wünschten sich die Interviewteilnehmer inhaltlich eine größere Variation der Themengebiete und eine noch stärkere Wiederholung von älteren Modulen. Zu selten werden Studierende in Vorlesungen aktiv mit einbezogen (Persike und Friedrich 2016; The Digital Turn 2016). Durch die Gestaltung der interdisziplinären Lehrveranstaltungen fand eine Diskussion von IMPP-Fragen zwischen verschiedenen Lehrenden und den Studierenden statt. *#clue* kann als *Blended-Learning*-Angebot eingestuft werden, da die IMPP-Fragen zuvor selbstständig zu Hause von den Studierenden bearbeitet wurden und die schwierigsten Fragen in Präsenzveranstaltungen besprochen wurden. V. a. hinsichtlich der technischen Umsetzung sahen die Studienteilnehmer allerdings noch Verbesserungsbedarf. Die Kommunikation über *social media* bewerteten die Studienteilnehmer als sehr wichtig, aber auch Werbung durch die Lehrenden und persönliche Hinweise von diesen sollten vermehrt werden.

4.5.2 Emotionen und Lernmotivation

Positive Emotionen können die Lernmotivation und tieferegreifende Lernstrategien von Studierenden fördern (Pekrun 2014). Vor allem positive Emotionen wie Ansporn kamen bei den Interviewteilnehmern während der Beantwortung der Fragenpakete auf. So hatte die Beantwortung der Fragenpakete überwiegend einen motivierenden Effekt auf die Interviewteilnehmer. Sie wurden durch die Fragenpakete dazu motiviert, vereinzelte Inhalte noch einmal nachzulesen.

Aber auch negative Emotionen, wie Frustration wurden in den Interviews geschildert. Zwar können diese auch einen negativen Effekt auf das Lernverhalten ausüben, lassen sich allerdings nicht immer vermeiden und können sogar bei selbstbewussten Studierenden die Motivation fördern (Pekrun 2014).

4.5.3 Lernverhalten

Formative Prüfungen sollen v. a. den Studierenden ein ausführliches Feedback geben (Black und Wiliam 1998; Krasne et al. 2006; Roediger und Karpicke 2006b) und ihnen eventuell bestehende Wissenslücken aufzeigen (Roediger und Karpicke 2006b; Cliff et al. 2008). In dem Projekt *#clue* sahen die Interviewteilnehmer eine Lernunterstützung. Zum einen hatten sie eine Rückmeldung zum aktuellen Wissens-/Leistungsstand und zum anderen wurden ihnen bestehende Wissenslücken aufgezeigt. Wie bereits in Kapitel 4.4.2 erläutert, empfanden die Studierenden das Feedback als nicht ausführlich genug. Darüberhinausgehend ist bei der Bewertung eines neuen Projekts v. a. der Lernaufwand für die Studierenden von Interesse (Kerres 2013). Diesen beurteilten die Interviewteilnehmer mit zehn Fragen pro Woche als angemessen.

4.5.4 Subjektiver Lernerfolg

Boeker und Klar (2006) beschreiben, dass der größte Lernerfolg bei Studierenden erreicht werden kann, wenn *E-Learning*-Angebote mit der regulären Lehre verknüpft werden. Bestehende Studien bestätigen, dass eine erste formative Prüfung möglichst zeitnah an die Behandlung der Vorlesungsinhalte anknüpfen sollte, um einen Lerneffekt in späteren Zeitpunkten erkennen zu können (Roediger und Karpicke 2006b). In allen vier Interviewgruppen konnten die Studierenden einen individuellen Lernerfolg für sich feststellen. Manche Studierende stellten auch fest, dass sie durch die regelmäßige Bearbeitung der Fragenpakete lernten, sich selbst besser einzuschätzen. Dies wird als *cognitive monitoring* beschrieben und ist im Zusammenhang mit formativen Prüfungen bereits bekannt (McDaniel und Little 2019).

4.5.5 Zufriedenheit

E-Learning-Angebote nur als freiwillige Zusatzangebote in die Lehre einzuführen sehen Boeker und Klar (2006) kritisch. Doch gerade die Möglichkeit einer freiwilligen Teilnahme wurde von den Studierenden in drei von vier Fokusgruppeninterviews geschätzt. Besonders die zeitliche und räumliche Flexibilität, die *E-Learning* Angebote gewährleisteten (Kerres 2013), wurde als positiv betrachtet. Als nachteiliger Aspekt wurde von einigen Interviewteilnehmern empfunden, dass die Fragenpakete noch keinen ausreichenden Mehrwert gegenüber anderen digitalen Lernformaten aufwiesen.

Zusammenfassend wurde das Projekt gut bewertet und die Studierenden würden auch im Folgesemester das Projektangebot nutzen.

4.6 Limitationen der Studie

4.6.1 Studiendesign

Bei der hier vorliegenden Studie handelt es sich um eine prospektive, nicht-randomisierte monozentrische Interventionsstudie, an der im Sommersemester 2018 alle Studierenden des dritten, vierten, fünften und sechsten klinischen Semesters der UMG teilnehmen konnten. Es handelt sich um eine monozentrische Einzelstudie, welche lediglich in einem Semester durchgeführt worden ist. In Folgestudien sollte untersucht werden, ob sich die Ergebnisse der Studienfragen eins und zwei bestätigen lassen und ob ggf. der direkte *testing effect* bei schweren Fragen auch longitudinal über mehrere Semester bestehen bleibt. Darüberhinausgehend war diese Studie nicht-randomisiert, da es keinen Vergleich zu den Studierenden gab, die nicht am Projekt teilgenommen haben. Um den Lernerfolg durch *E-Learning*-Angebote bestimmen zu können, beschreibt Kerres (2013), dass es sinnvoll ist, die Studienkohorte in zwei Gruppen aufzuteilen. Die eine Gruppe nimmt am konventionellen Unterricht teil, während die andere Gruppe das *E-Learning*-Angebot testet. Beide Gruppen müssen an einem Prä- und Posttest teilnehmen (Kerres 2013). Es war allerdings ethisch nicht vertretbar, manchen Studierenden gegen ihren Willen die Lehrinnovation vorzuenthalten. Zusätzlich wurden potenziellen Störfaktoren kaum dokumentiert (Nelson et al. 2012), Es ist beispielsweise denkbar, dass die Studierenden, die regelmäßig teilnahmen, allgemein bessere Lerner waren und dadurch einen Lernerfolg vorweisen konnten, der sich nicht auf die gesamte Studienpopulation übertragen lässt. Unterstützt wird diese Vermutung durch die Beobachtung der Wochenscores über alle Semester hinweg (siehe Abbildung 13). Hier waren die regelmäßigen Teilnehmer von Beginn an leistungsstärker und erreichten mit einer Ausnahme während des kompletten Semesters höhere Wochenscores. Die Ausnahme in Woche zehn fiel zeitlich auch mit der beobachteten minimalen Teilnahmebereitschaft zusammen.

Um Interventionsitems im *Exit Exam* präsentieren zu können, mussten diese im Verlauf der zehn bis 14 Wochen zweimal in den Fragenpaketen präsentiert werden und inhaltlich zu den Vorlesungswochen passen. Dennoch sollten sie einen gewissen Abstand zum *Exit Exam* aufweisen, um den Erinnerungseffekt an die Fragen zu minimieren. Dadurch war die thematische Auswahl an passenden Interventionsitems eingeschränkt und es fanden sich nur Interventionsitems, die die Vorlesungsinhalte der ersten Semesterwochen widerspiegelten. Der gewisse Abstand zum *Exit Exam* sollte mindestens zwei Wochen betragen. Im vierten und fünften klinischen Semester wurden die letzten Interventionsitems in Woche zwölf das zweite Mal präsentiert. Die Fragenpakete konnten immer eine ganze Woche lang bearbeitet werden. Falls ein Teilnehmer am letzten möglichen Tag das zwölfte Fragenpaket bearbeitet hat und das *Exit Exam* am ersten möglichen Tag der Freischaltung, ist es hierdurch möglich, dass nur eine Woche anstelle der zwei Wochen zwischen T2 und T3 lag. Diese Ungenauigkeit dürfte allerdings zu vernachlässigen sein, da insgesamt nur drei Interventionsitems betroffen

waren und diese nur einen kleinen Teil der Summenscores ausmachen, die in die statistische Analyse des *Exit Exams* eingegangen sind. Das *Exit Exam* selbst setzte sich aus fünf Kontroll- und fünf Interventionsitems zusammen. Bei der Zusammenstellung war es nicht möglich, dass die verschiedenen Semester die gleiche Anzahl an leichten, mittelschweren und schweren Items präsentiert bekommen haben. Somit setzten sich die Summenscores aus einer unterschiedlichen Anzahl an Items zusammen. So setzte sich beispielsweise der mittelschwere Summenscore des dritten klinischen Semesters aus einem einzigen Item zusammen, während das vierte klinische Semester zwei mittelschwere Items im *Exit Exam* präsentiert bekam. Darüberhinausgehend war im fünften und sechsten klinischen Semester jeweils ein Schwierigkeitsgrad im *Exit Exam* nicht vertreten. Das fünfte klinische Semester bekam keine schweren Interventionsitems präsentiert, während das sechste klinische Semester keine mittelschweren Interventionsitems präsentiert bekam. Um eine definitive Antwort zur ersten Studienfrage geben zu können, müssen die Anforderungsniveaus zwischen den Interventionsitems und den Kontrollitems übereinstimmen. Die Anforderungsniveaus zum Zeitpunkt T3 der Kontrollitems sind durch einen allgemeinen Lernzuwachs, den indirekten *testing effect* oder auch den allgemeinen Lernzuwachs während des Semesters beeinflusst. Die Erhebung der Anforderungsniveaus der Interventionsitems zum Zeitpunkt T1 fand darüberhinausgehend zu unterschiedlichen Zeitpunkten im Semester statt (siehe Abbildung 4: Verteilung der Interventionsitems.).

In weiteren Studien ist aus diesen Gründen zu empfehlen, ein *Entry Exam* einzuführen. Dadurch kann der Lernerfolg besser nachvollzogen werden (Kerres 2013). Zusätzlich kann zu Studienbeginn die Vergleichbarkeit der Anforderungsniveaus der Items vereinfacht werden und eine passende Auswahl an Interventionsitems und Kontrollitems getroffen werden. Die Summenscores der Kontrollitems wurden im *Exit Exam* zum Zeitpunkt T3 nach der erstmaligen Beantwortung bestimmt. Da zu diesem Zeitpunkt ein allgemeiner Lernzuwachs schon stattgefunden haben könnte, werden diese Items höchstwahrscheinlich immer besser beantwortet als bei einer erstmaligen Beantwortung dieser zum Zeitpunkt T1. Rowland (2014) stellte fest, dass es für die Interpretation von Testergebnissen sinnvoll ist, das Ausgangswissen zu kennen. So könnten der direkte und indirekte *testing effect* besser voneinander unterschieden werden. Allerdings besteht hier das Risiko, dass sich die Studierenden im *Exit Exam* an die Antworten der Kontrollitems erinnern. Dies sollte beim Studienaufbau berücksichtigt werden, indem z. B. im Anschluss an das *Entry Exam* weder die Fragen noch die richtigen Antworten einsehbar sind. Wie bereits oben beschrieben, ist es aus ethischen Gründen nicht möglich die Studienkohorte in eine eigentlich notwendige Kontrollgruppe und Experimentalgruppe aufzuteilen (Kerres 2013). Möglich ist beispielsweise die Studie zwei Semester hintereinander durchzuführen und so verschiedene Ausgangssituationen zu schaffen, die dann verglichen werden können. Eine weitere Möglichkeit besteht auch darin, ein Semester in zwei Kohorten aufzuteilen. Das *Exit Exam* kann für beide Kohorten gleich aussehen, allerdings werden die Interventions- und Kontrollitems zwischen den Kohorten ausgetauscht. Dieser Studienaufbau lässt sich allerdings erst bei stabileren Teilnehmerzahlen

verwirklichen. Um besser zwischen dem allgemeinen Lernzuwachs der Studierenden oder dem indirekten oder direkten *testing effect* unterscheiden zu können, wäre es sinnvoll die Ergebnisse der Modulabschlussklausuren auszuwerten. Diese sind verpflichtend für alle Studierenden, wodurch die motivierten und nicht motivierten Studierenden und ihre Leistungen besser miteinander verglichen werden könnten.

4.6.2 Limitationen durch die Erhebungsinstrumente

Bei der quantitativen Analyse finden sich einige Limitationen hinsichtlich der Datenaufbereitung. Die Tests konnten die Studierenden flexibel zu Hause alleine oder in Gruppen beantworten. Hierbei hatten sie prinzipiell auch die Möglichkeit, Hilfsmedien wie das Internet oder Lehrbücher zur Beantwortung zu verwenden. Zumindest in einer Interviewgruppe wurde angegeben, dass die Fragenpakete mit Hilfestellung bearbeitet wurden.

Ein grundsätzliches Problem in der durchgeführten Studie war der starke Rückgang der Teilnehmerzahl. Dadurch konnten unvollständige Datensätze von Studienteilnehmern nicht aus der Analyse genommen werden. Fehlende Daten wurden wie in Kapitel 2.4.1 erläutert mit null Punkten bewertet, da in der zweiten Forschungsfrage die Wirkungen des direkten *testing effects* untersucht wurden. Wenn also ein Teilnehmer ein Item zu einem bestimmten Zeitpunkt nicht beantwortet hat, konnte er zu diesem Zeitpunkt auch nicht vom *testing effect* profitieren. Je nachdem zu welchem Zeitpunkt die Antwort gefehlt hat, hatte dies verschiedene Effekte auf die statistischen Tests. Items, die zum Zeitpunkt T1 betroffen waren, wurden dadurch in der Intervention etwas besser dargestellt. Darüberhinausgehend wurden teilweise auch nicht beantwortete Interventionsitems von Teilnehmern zum Zeitpunkt T2 mit null Punkten bewertet, obwohl sie zum Zeitpunkt T1 die Antwort richtig wussten. Von einigen Studienteilnehmern wurden manche Items zum Zeitpunkt des *Exit Exams* das erste Mal bearbeitet. Aus diesen Gründen wurde eine Sensitivitätsanalyse durchgeführt, bei der entsprechende Items herausgenommen wurden. Bei der Bildung der neuen Summenscores für die betroffenen Studierenden wurde dann der Mittelwert aus den verbliebenen Items berechnet. Darüberhinausgehend sind die kompletten Datensätze der Kontrollitems in die Summenscores der Sensitivitätsanalyse eingeflossen. Dieser entstandene Messfehler sollte sich allerdings über alle Studienteilnehmer hinweg ausgeglichen haben. Durch die Bewertung von nicht beantworteten Items mit null Punkten in der primären Analyse ist ebenfalls zu erklären, dass in der Sensitivitätsanalyse kein signifikanter Unterschied zwischen den Summenscores der leichten Interventionsitems mehr zu finden ist. Der signifikante Unterschied fand sich in der primären Analyse zwischen T2 und T3. Die meisten der nicht beantworteten Items fanden sich zum Zeitpunkt T2, wodurch die *Overall-Interventionsitems* zum Zeitpunkt T2 in der primären Analyse vermeintlich schlechter beantwortet wurden als in der Sensitivitätsanalyse.

Eine grundsätzliche Fragestellung, die ebenfalls aufgeworfen wird, ist, wie viele Items überhaupt für die Bildung von einem sinnvollen Summenscore benötigt werden, um reliable Ergebnisse zu erhalten. Dies war in der hier vorliegenden Studie allerdings keine Forschungsfrage.

Auf die niedrige Teilnehmerzahl ist auch zurückzuführen, weshalb die durchgeführten Analysen nicht semestergetrennt durchgeführt werden konnten. So könnte die Fragestellung, ob sich in den verschiedenen Semestern unterschiedliche Lerneffekte finden lassen, ebenfalls in einer weiterführenden Studie aufgenommen werden. Da für die unterschiedlichen Semester allerdings individuell an den Modulplan angepasste Fragenpakete erstellt worden sind, dürfte diese Tatsache zu keiner allgemeinen Verfälschung der Ergebnisse geführt haben.

Eine weitere Einschränkung, die in nachfolgenden Studien berücksichtigt werden sollte, ist die kleine Stichprobe freiwilliger Studierender bei der EvaSys®-Fragebogen-Erhebung (insgesamt 135 Studierende aus drei verschiedenen Semestern entsprechen einer Rücklaufquote von 43%) und der Durchführung der Fokusgruppengespräche (insgesamt 20 Studierende aus vier verschiedenen Semestern) mit einer potenziellen Selektionsverzerrung. Zumindest bei letzterem konnte bei der Auswahl der Gesprächsteilnehmer durch die persönliche Rekrutierung eine unterschiedlich regelmäßige Teilnahme berücksichtigt werden, um eine breit gefächerte Meinung zu bekommen. Auf der anderen Seite kann die persönliche Rekrutierung dazu geführt haben, dass sich manche Teilnehmer verpflichtet gefühlt haben an den Interviews.

Aus den oben genannten Gründen wird die Frage aufgeworfen, ob die qualitativen Ergebnisse auf die gesamte Studienpopulation verallgemeinert werden können. Es bedarf Folgestudien, um die hier aufgeführten Ergebnisse zu verifizieren.

4.7 Weiterentwicklung von #clue und Ausblick

Der Masterplan Beschlusstext 2020 (2017) fordert über die horizontale Integration hinausgehend noch die Umsetzung der vertikalen Integration. Diesbezüglich ist eine Weiterentwicklung von #clue bezüglich vorklinischer Lehrinhalte denkbar. Neue Fragen lassen sich problemlos in ILIAS einarbeiten. So wird mittlerweile #clue bereits im vierten Semester der Vorklinik an der UMG angeboten.

MC-Prüfungen testen v. a. das vorhandene Faktenwissen von Studierenden. Aus diesem Grund sollen in zukünftigen M1- und M2-Prüfungen andere Fragenformate zum Einsatz kommen, die vermehrt auf die Überprüfung von Fertigkeiten ausgerichtet sind. (Deutscher Wissenschaftsrat 2014; Bundesministerium für Gesundheit et al. 2017). Um nach dem Ansatz des *constructive alignments* (Biggs 1996) zu handeln, sollten somit bei #clue geeignetere Prüfungsmethoden zum Einsatz kommen, die vornehmlich den Kompetenzerwerb testen (Bundesministerium für Gesundheit et al. 2017). Hierfür eignen sich beispielsweise *Key-*

Feature-Fälle, die noch in das Projekt integriert werden sollten, um zum einen den Kompetenzerwerb der Studierenden testen zu können und zum anderen auch weiterhin die Studierenden optimal auf das Staatsexamen vorbereiten zu können. Zusätzlich kann dadurch der direkte *testing effect* noch verstärkt werden (Foos und Fisher 1988; Roediger und Karpicke 2006b; McDaniel et al. 2007). Darüberhinausgehend würde die Präsentation der *Key-Feature-Fälle* in Fragenpaketen für einen höheren Praxisbezug im klinischen Alltag sorgen und eine bessere Abgrenzung zu AMBOSS® schaffen. Da allerdings für Kurzantwort-Formate die Studierenden mehr Zeit für eine Antwortsuche aufbringen müssen und somit die Beantwortung der Frage länger dauert, sollten diese nicht den Hauptbestandteil der Fragenpakete ausmachen, sondern lediglich eine Ergänzung darstellen. Mittlerweile ist erwiesen, dass durch strategisch gut konzipierte MC-Fragen vergleichbare, wenn nicht sogar bessere *testing effects* durch die indirekte Wirkung erzielt werden können wie bei Kurzantwortfragenformaten, aber im gleichen Zeitraum mehr Fragen behandelt werden können (McDaniel und Little 2019).

Wie in Kapitel 1.2.4 ausgeführt, besteht zurzeit eine besonders hohe Nachfrage nach digitalen Lehr- und Lernangeboten. *#clue* wurde als innovatives *E-Learning*-Angebot 2018 in die klinische Lehre der UMG eingeführt und seitdem evaluiert und weiterentwickelt. Die gewonnenen Erkenntnisse aus der hier vorgestellten Studie wurden zur Verbesserung und Weiterentwicklung des *E-Learning*-Angebots genutzt. Die neu entwickelte webbasierte App mit dem Namen „*DailyClue*“ bietet den Studierenden eine ansprechendere Nutzungsoberfläche. Trotz der zurzeit fehlenden Präsenzveranstaltungen erhalten die Nutzer immer eine aktuelle Rückmeldung über den selbst angeeigneten Wissensstand. Nun werden den Studierenden von montags bis freitags je zwei IMPP-Fragen präsentiert, die für zehn Tage freigeschaltet sind. Die beantworteten Fragen werden in einem persönlichen Archiv gespeichert und sind von den Nutzern jederzeit erneut einsehbar. Zusätzlich wird ein *Highscore* angezeigt, der den persönlichen Punktestand im Vergleich zu den Kommilitonen widerspiegelt. Die interdisziplinären Lehrveranstaltungen sollen ggf. als Online-Konferenzen stattfinden. Nach wie vor kommen regelmäßig Evaluationsinstrumente wie EvaSys®-Fragebögen oder Interviewgruppen zum Einsatz, um Anregungen für eine kontinuierliche Verbesserung zu entwickeln. Wenn *#clue* sich an der UMG erfolgreich etabliert hat, ist eine Umsetzung auch an anderen Universitäten denkbar.

5 Zusammenfassung

Die anstehenden Änderungen im Medizinstudium werden sich sowohl in der Lehre als auch in den Prüfungen widerspiegeln. Der Schwerpunkt liegt hierbei auf einer fächerübergreifenden und longitudinalen Verknüpfung der klinischen Module und einer Digitalisierung der Lehre. Durch das innovative Lehrprojekt *#clue* (*cross-linked undergraduate education*) wird Studierenden semesterbegleitend seit dem Sommersemester 2018 eine longitudinale Vorbereitung auf das Staatsexamen (M2-Prüfung) durch formative digitale Prüfungen angeboten. Ermöglicht wird dies durch den direkten *testing effect*, der die Retention von Wissen fördern kann. Somit bildet dieses *Blended-Learning*-Projekt eine Möglichkeit, um die horizontale und vertikale Integration in der medizinischen Ausbildung zu erleichtern. Dies wird zum einen durch die ehemaligen Examensfragen auf der *E-Learning*-Plattform ILLIAS erreicht und zum anderen durch die interdisziplinären Lehrveranstaltungen, die von Lehrpersonen aus verschiedenen Modulen fächerübergreifend und interaktiv gestaltet werden.

Im Rahmen der Studie wurden folgende Forschungsfragen beantwortet: Führt die mehrfache Wiederholung von Inhalten während des Semesters bezüglich dieser Inhalte zu einem größeren Lernerfolg als bei Inhalten, die nicht mit Fragen wiederholt wurden? Wie entwickelt sich der studentische Leistungsstand im Laufe des Semesters, wenn Inhalte anhand von Fragen mehrfach wiederholt werden? Wie entwickelt sich die Teilnehmerzahl bei einem freiwilligen Projekt, das zum Verbessern des eigenen Lernerfolgs angeboten wird? Wie beurteilen die Studierenden dieses innovative Projekt?

Im Verlauf des Semesters wiederholt aufgetretene Fragen wurden am Ende des Semesters nicht signifikant besser beantwortet als am Ende des Semesters erstmalig aufgetretene Fragen. Es kann somit keine eindeutige Aussage über einen allgemein größeren Lernerfolg bei wiederholten Inhalten im Vergleich zu neuen Inhalten geäußert werden. Erklärbar hierfür können nicht vergleichbare Anforderungsniveaus zwischen Kontroll- und Interventionsitems, der indirekte *testing effect* oder ein allgemeiner Lernzuwachs durch das Absolvieren des Semesters sein. Allerdings schien die Bearbeitung von Prüfungsfragen auf Items abhängig vom Schwierigkeitsgrad einen unterschiedlich starken Einfluss zu haben. So wurden Fragen zu schwierigen Inhalten bei mehrfacher Wiederholung am Semesterende signifikant besser bearbeitet als neue Fragen.

Unter Einbeziehung aller Schwierigkeitsgrade ließ sich bei mehrfachem Abfragen bei bestimmten Fragestellungen ein signifikanter Lernzuwachs feststellen, der sich ggf. auf den direkten *testing effect* zurückführen lässt. Dieser war bei schweren Fragen am höchsten. In der Theorie ist eine mögliche Erklärung für den direkten *testing effect* der sog. *retrieval effort*, der umso besser wirken kann, je aufwendiger das Wiederaufrufen der Gedächtnisinhalte gestaltet ist. Die hier vorliegenden Ergebnisse lassen vermuten, dass v. a. schwere Items besonders aufwendig abzurufen sind und so mit einem besonders großen *retrieval effort* in Zusammenhang stehen.

Die Teilnehmerzahlen nahmen bei diesem freiwilligen Projekt im Verlauf ab, wodurch sich eine Limitation für die Ergebnisse ergab. Abhängig vom Semester bearbeiteten durchschnittlich zwischen 28% (fünftes klinisches Semester) und 53% (drittes klinisches Semester) aller Studienteilnehmer die zur Verfügung stehenden Fragenpakete. Bei der Vorstellung von *#clue* in den Einführungsveranstaltungen sollte darauf geachtet werden, verstärkt die persönlichen Vorteile für die Studierenden, die sich aus der Bearbeitung der Fragenpakete ergeben, zu betonen. Dadurch ist denkbar, dass sie ihre Lernstrategien anpassen. Während des Semesters sollte eine regelmäßige Erinnerung gewährleistet werden. Um die Besucherzahlen der interdisziplinären Lehrveranstaltungen zu erhöhen, sollte auf diese vermehrt hingewiesen werden.

Das Projekt wurde von den Studierenden insgesamt als gut bewertet und würde auch in weiteren Semestern von diesen genutzt werden. Vor allem die individuelle Anpassung an den Göttinger Modulplan wurde als sehr positiv empfunden, wobei die Anpassung noch weiter verbessert werden kann. Die Zusammenstellung der Fragenpakete mit zehn Fragen wurde als angemessen beurteilt. Der Lernzuwachs ließ sich zwar bei schwierigen Fragen am besten feststellen, allerdings befürworteten die Studierenden die Variation der Schwierigkeitsgrade innerhalb von einem Fragenpaket. So könnten zu viele schwierige Fragen, eine demotivierende Wirkung auf die Studierenden haben. Um einen höheren Mehrwert gegenüber anderen digitalen Lehrformaten generieren zu können, sollten Hintergrundinformationen zu den einzelnen Fragen gegeben werden und die Fragenpakete durch vereinzelte *Key-Feature*-Fälle ergänzt werden.

Das Projekt *#clue* untersuchte die Wirkungen von Prüfungen anhand von digitalen, freiwilligen, formativen Prüfungen. Die Bearbeitung von Prüfungsfragen schien auf Fragen verschiedener Schwierigkeitsgrade unterschiedliche Einflüsse zu haben. V. a. bei schwierigen Fragen konnte ein Leistungszuwachs beobachtet werden, weshalb mit *#clue* vorzugsweise komplexere Themeninhalte trainiert werden können. Zukünftige Studien sollten untersuchen, ob dieser Leistungszuwachs auch longitudinal über mehrere Semester bestehen bleibt. *#clue* kann Studierende bei ihrer Vorbereitung fürs Staatsexamen langfristig unterstützen und somit ggf. eine Erleichterung darstellen. Eine Erweiterung für vorklinische Semester und eine Etablierung an anderen medizinischen Fakultäten ist möglich.

6 Anhang

Anhang 1: Leitfaden für die Fokusgruppeninterviews

Allgemeiner Leitfaden sowie Diskussionsfragen für die Fokusgruppen-Gespräche für das Projekt #clue

Frage: Möchten Sie sich duzen oder siezen?

Aufklärung der Gesprächsteilnehmenden:

1. Freiwillige Teilnahme am Fokusgruppengespräch
2. Abschrift des Interviews ohne Rückschlüsse auf Personen (keine Namen)
3. Spätere Löschung der Aufnahme nach Transkription
4. Transkription erfolgt durch UMG-Beschäftigte
5. Keine Weitergabe von Informationen an Externe

➔ Ausgabe der Probanden/-inneninformation zur Unterschrift

Gruppenregeln

1. Wir lassen einander ausreden.
2. Wir antworten nacheinander, da die Aufnahme sonst schwer verständlich wird.
3. Wenn ihr der Meinung anderer Gesprächsteilnehmenden zustimmt oder diese ablehnt, würden wir uns freuen, wenn ihr das verbal unterstreicht. Z.B. mit „ja genau“ oder „nein“.
4. Wir äußern uns nur konstruktiv und nicht beleidigend gegenüber den anderen Diskussionsteilnehmenden.
5. Es gibt bei den Antworten kein Richtig oder Falsch. Wir möchten die studentische Zufriedenheit mit dem Projekt evaluieren und daraus u.a. konkrete Verbesserungsvorschläge ableiten. Daher bitten wir um ehrliche und konstruktive Antworten.

1. Allgemeines

1. Was glaubt ihr, war das Ziel des Projekts #clue?
 - Inwieweit hattet ihr das Gefühl durch das Projekt beim Lernen unterstützt zu werden? Welchen Beitrag hat das Projekt zu eurem Lernerfolg geleistet?
 - Wie schätzt ihr euren persönlichen Lernerfolg ein?
2. Wie hat euch das Projekt #clue als Lehrinnovation gefallen?
 - Was war eure Motivation daran teilzunehmen?
 - Was ist euch von #clue besonders in Erinnerung geblieben?
 - Einfluss: Social-Media-Auftritt, wöchentliche Erinnerungsmails?
3. Würdet ihr im nächsten Semester wieder teilnehmen?
 - Wenn ja, warum bzw. warum nicht? Würdet ihr das Projekt an KommilitonInnen weiterempfehlen?

2. Wöchentliche Fragenpakete

1. Wie fandet ihr die Fragen inhaltlich?
 - Wie ernsthaft habt ihr die Fragen beantwortet?
 - Hatten die Fragenpakete eine motivierende Wirkung auf euch?
2. Habt ihr bemerkt, dass Fragen wiederholt aufgetaucht sind? Wenn ja, welchen Effekt hatte das für euch?
3. Wir haben die Beobachtung gemacht, dass die Teilnehmerzahlen im Verlauf kontinuierlich zurückgingen. Was glaubt ihr, woran das liegt?

3. Interdisziplinäre Lehrveranstaltungen

1. Was waren Gründe für euch zu einer interdisziplinären Lehrveranstaltung zu gehen?
2. Wir haben die Beobachtung gemacht, dass mehr Studierende die Fragenpakete beantworten als Teilnehmende die interdisziplinären Lehrveranstaltungen besucht haben. Was glaubt ihr, woran das liegt?
 - Wie könnte man eine außercurriculare Lehrveranstaltung gestalten, dass die Studierenden diese besuchen?

4. Verbesserungsvorschläge

1. Habt ihr noch Verbesserungsvorschläge für das Projekt? Wenn ja, was würdet ihr euch noch wünschen?
 - Was würdet ihr an der organisatorischen oder technischen Umsetzung verändern?
 - Was würdet ihr inhaltlich verändern?
2. Wie könnte sich euer Lernerfolg durch das Projekt weiter steigern lassen?

Anhang 2: Kodierleitfaden

- Markierte Bereiche sind induktiv gebildete Kategorien und Unterkategorien. Unmarkierte Bereiche sind deduktiv gebildete Kategorien
- Bei den Beispielen sind nicht alle Aussagen genannt; nur die, die am besten repräsentativ sind

Allgemeines: Viele Studierende haben mit dem Projekt #clue nur das Kreuzen der Fragenpakete verknüpft, weil sie nichts von den interdisziplinären Lehrveranstaltungen mitbekommen haben. Aus diesem Grund beziehen sich viele Punkte (wie z.B. Vorteile/Nachteile) der Kategorie „Allgemeines“ hauptsächlich auf die wöchentlichen Fragenpaketen.

Kategorien			Regeln	Beispiele
1. Ebene	2. Ebene	3. Ebene: 4. Ebene (5. Ebene)		
Ziel des Projektes			Aussagen, die nach der Frage getroffen wurden bzw. wenn der Student später noch einmal explizit das Wort „Ziel“ oder Synonyme hiervon nennt	
	Feedback an Lehre	Anpassung der Lehre		Gruppe01_Absatz10_Person2: „[...] Lehre besser auf mich zugeschnitten werden kann.“
	Lehre mit IMPP verknüpfen	Vorbereitung auf das Staatsexamen		Gruppe04_Absatz3_Person3: „So wie ich das verstanden habe, ging es darum, die Studierenden des klinischen Studienabschnitts mit wöchentlichen Fragen, die die Themen der Vorlesungen widerspiegeln, abzufragen, um die Studierenden besser auf das Staatsexamen vorzubereiten“ Gruppe01_Absatz115_Person4: „Dass man einfach auch, weil das IMPP noch einmal andere Fragen, eine andere Art hat zu fragen als hier einfach unsere Klausuren konzipiert sind und dass man da einfach schon ein Gefühl dafür bekommt, dass man, wenn man sich dann auf das Staatsexamen vorbereitet, schon mal ein bisschen mehr sensibilisiert ist für die Art der Fragen, die da einem begegnen.“

	Wiederholung von Lernstoff	Aktueller Lernstoff		Gruppe02_Absatz6_Person6: „Es war zumindest so grob, (.) zumindest für einzelne Module so gegliedert, dass Inhalte nach den Vorlesungen wiederholt wurden.“ Gruppe04_Absatz6_Person5: „[...] und das gleichzeitig unterstützend zu dem aktuellen Lernprogramm.“
		Lernstoff aus vergangenen Semestern		Gruppe02_Absatz4_Person1: „[...] ob die Lernleistung besser wird, wenn man die Studierenden regelmäßig mit den Lerninhalten konfrontiert und nicht ein Modul absolviert wird und dann mehrere Semester oder Jahre erfolgen bis zum Staatsexamen, wo dann eventuell das Wissen wieder verloren gehen kann.“
		Selbststudium fördern		Gruppe01_Absatz3_Person2: „Die Studenten dazu zu bringen, sich zu Hause noch einmal (.) aktiv mit der Materie auseinander zu setzen (...) [...]“
	Klausurvorbereitung			Gruppe02_Absatz5_Person2: „Und ja, auch so ein bisschen Klausurvorbereitung durch Multiple Choice.“
	Persönlichen Wissensstand überprüfen			Gruppe02_Absatz5_Person2: „Man hat mit diesem Medium auch die Möglichkeit selbst als Student sich zu überprüfen: ‚Wie ist mein momentaner Wissensstand?‘“
Lernunterstützung			Der Studierende erläutert, was das Projekt direkt gebracht hat. Allgemein oder sich selbst.	
	Examensvorbereitung	Schwerpunktsetzung im Staatsexamen		Gruppe04_Absatz17_Person3: „[...] dass man ein bisschen mehr die Schwerpunkte vom IMPP kennen lernt.“

	Persönliches Feedback	Wissenslücken finden		Gruppe03_Absatz15_Person3: „Oh, das habe ich jetzt noch nicht gewusst, aber ich weiß zu mindestens, dass ich das Thema schon einmal hatte und es wissen hätte können.“ Gruppe02_Absatz132_Person2: „[...] sondern dass man merkt: ‚Was habe ich denn falsch gemacht?‘ Ich glaube, dass das ein größerer Lernerfolg ist, wenn man lernt herauszufiltern: ‚Was kann ich nicht?‘ Und dafür finde ich clue relativ gut.“
		Wissen aufarbeiten	Aktiv Sachen nachschlagen bzw. sich bewusst machen, dass noch Nachholbedarf besteht	Gruppe01_Absatz27_Person3: „Und man kann zum Krankheitsbild noch ein paar Sachen nachlesen.“ Gruppe01_Absatz64_Person2: „[...] und dann macht man kurz den nächsten Tab bei Google auf und guckt sich das dann schnell an.“
		Zu älteren Modulen		Gruppe02_Absatz16_Person5: „Ja, ich fand das auch eher hilfreich, wenn man überlegt wie man die anderen Module noch in Erinnerung hat.“
		Zum aktuellen Wissens-/Leistungsstand		Gruppe04_Absatz8_Person2: „Ich glaube, dass es immer generell gut ist, wenn man den Vorlesungsstoff noch einmal wiederholt und anhand von Fragen abarbeitet. Von daher fand ich das schon unterstützend.“ Gruppe02_Absatz14_Person2: „Wenn man aber auf dem Wissensstand ist wie die Fragen, dann hat man eine sehr gute Rückmeldung, finde ich, ob man wenigstens so ungefähr das Thema verstanden hat.“
		Vergleich zu Mitstudierenden	Es war schon ein Vergleich möglich	Gruppe02_Absatz32_Person2: „Für mich war es tatsächlich eher der Vergleich. Also (.) vor allem wenn man danach noch einmal Rücksprache mit jemand anderem gehalten hat: ‚Und wie war es bei dir?‘“

	Wissen verfestigen		Möglich dadurch, dass aktuelle Themen abgefragt wurden	Gruppe01_Absatz13_Person1: „Dadurch, dass viele Fragen eben auf Themen bezogen waren, hat man schon das Wissen auch noch einmal verfestigt.“
	Lernerfolg	Generell: abhängig von der Lehre	Student trifft eine Aussage, die für mehrere gilt	Gruppe02_Absatz14_Person2: „In Bezug darauf finde ich, dass vor allem wenn die Lehre gut ist, wenn die Lehre sehr gut besucht ist und alle quasi dem Thema auch folgen, dass dann auch clue mehr Sinn hat.“
		Individuell: ja/teilweise (abhängig vom Zeitpunkt des Kreuzens; abhängig vom persönlichen Lernstatus)/nein (nicht klausurrelevant)	Studierender verwendet das Pronomen „ich“ und evaluiert in der Aussage den Nutzen des Projektes und woran das gelegen hat. Eventuell genannte Konsequenzen werden bei Lernunterstützung (s.o.) aufgeführt	Gruppe01_Absatz18_Person2: „Also ich persönlich würde sagen, das hat mich schon unterstützt.“ Gruppe03_Absatz11_Person2: „Also wenn man es gegen Ende der Woche gemacht hat, war es schon deutlich besser[...].“ Gruppe02_Absatz8_Person4: „Das war meiner Meinung nach extrem davon abhängig, wie sehr man schon die Gelegenheit hatte, sich selbst mit dem Stoff auseinanderzusetzen.“ Gruppe03_Absatz11_Person2: „Ja, bei mir hat das auch für Klausuren gar nichts geholfen.“
Motivation teilzunehmen			Alle Umstände, die dazu geführt haben, dass der Student an #clue teilgenommen hat	
	Sich verbessern			Gruppe01_Absatz19_Person5: „[...] ,okay, jetzt gibst du dir besonders viel Mühe‘. Weil wenn da am Ende steht: ‚Dieses Mal hast du eine Frage mehr richtig als letzte Woche‘([...])“
	Lernen, sich selbst einschätzen zu können			Gruppe02_Absatz26_Person2: „...und am Ende fand ich, wenn man so langsam reingekommen ist, hat man auch irgendwie gemerkt, wie weit man sich selbst überprüfen kann.“

	Sinnvolle Zeitinvestition			Gruppe01_Absatz29_Person2: „Also (...) das ist einfach wie gesagt schon ein guter kleiner Ansporn noch einmal etwas zu tun. Und wenn es nur ein paar Minuten zu Hause sind. Diese paar Minuten hätte man wahrscheinlich sonst nicht investiert.“
	Lehrperson		Durch Werbung von Dozenten hat man teilgenommen	Gruppe04_Absatz24_Person3: „Ich hätte noch hinzuzufügen, dass der Enthusiasmus von Professor Raupach, der uns das Projekt vorgestellt hat, deutlich dazu beigetragen hat, dass die Stimmung sehr positiv gegenüber dem Projekt war.“
	Verlosung	Einfluss/kein Einfluss		Gruppe04_Absatz17_Person3: „Man hat allerdings durch dieses Preisausschreiben einen Anreiz, daran teilzunehmen.“
	Studie			Gruppe02_Absatz24_Person3: „Unterstützung der Doktorarbeit.“
	Wiederholung alter Module		Wird in Zusammenhang mit „Motivation“ oder ähnlichem Synonym genannt	Gruppe02_Absatz25_Person5: „Wie gesagt, ich fand eigentlich das Wertvollste daran war, dass man durch die alten Fragen an alte Module erinnert wurde.“
	Regelmäßigkeit			Gruppe02_Absatz27_Person6: „Mir hilft ja immer dieses Regelmäßige.“
	Weil es alle gemacht haben			Gruppe02_Absatz75_Person2: „Einfach weil es alle gemacht haben.“
	Vergleichbarkeit		Wird in Zusammenhang mit „Motivation“ oder ähnlichem Synonym genannt	Gruppe02_Absatz75_Person2: „[...] weil es Vergleichbarkeit hatte...“
	Einblick Staatsexamen		Wird in Zusammenhang mit „Motivation“ oder ähnlichem Synonym genannt	Gruppe04_Absatz22_Person5: „Einfach auch, dass man sich selbst neben dem Studium schon einmal ein bisschen darauf vorbereitet, dass man irgendwann das Staatsexamen schreibt.“

Social-Media			Über E-Mails und Facebook wurden Erinnerungs-E-mails für Fragenpakete und Informationen zu interdisziplinären Lehrveranstaltungen versendet	
	Erinnerungs-E-Mails	Wegen E-Mail gekreuzt	Durch die Erinnerungs-E-Mails wurde man daran erinnert, noch einmal zu kreuzen	Gruppe01_Absatz40_Person4: „Also ich habe oft dann noch einmal die Fragen Dienstagnachmittag gemacht, weil dann die Erinnerungs-E-Mail kam.“
	facebook	Genutzt/nicht genutzt		Gruppe04_Absatz38_Person2: „Ich hab den facebook-Auftritt bis heute nicht bekommen // // Aber ich fände das generell eine gute Sache. Dem werde ich auf jeden Fall beitreten, da werde ich bestimmt dann daran erinnert.“
	Wegen Datenschutzgrundverordnung weggefallen	Erinnerungen zu unregelmäßig	Es muss nicht unbedingt das Wort Datenschutzverordnung fallen. Es reicht, wenn dem Studierenden aufgefallen ist, dass es keine Mails/Gruppemehr gab	Gruppe04_Absatz31_Person3: „Was mir sehr negativ aufgefallen ist, dass es dann irgendwann - vermutlich außerhalb der Handhabe der Studienleitung - ein Update gab der Privatsphären-Einstellungen oder der (...), ich weiß nicht genau, wie es heißt, aber es ist glaube klar, was gemeint ist// (...) // Ganz genau, die datenschutzrechtlichen Bestimmungen hatten sich geändert und dass es dann ziemlich abrupt zum Versiegen der Informationen der wöchentlichen Erinnerungen kam. Und dass sich das eben - ich glaube, da kann man auch ziemlich genau den Punkt festmachen, das behaupte ich jetzt einfach - wo die Teilnahme der einzelnen Studierenden zurückgegangen ist. Das ist mir negativ aufgefallen. Um es positiv zu formulieren: Ich fand es gut, dass am Anfang auf verschiedenen Kanälen, sowohl über social Media als auch über die Uni-Mails informiert wurde.“
		Als Projektende gewertet		Gruppe04_Absatz74_Person5: „[...] als es auch anfing, dass wir keine Mails mehr bekommen haben, dass das Projekt schon vorbei wäre.“

	Ist ein wichtiger Aspekt		Studierende machen deutlich, dass das ein wichtiger Aspekt ist	Gruppe02_Absatz35_Person4: „War richtig wichtig.“
Bewertung von #clue			Aussagen, die das gesamte Projekt bewerten	
	Vorteile		Alle Vorteile zu dem Projekt allgemein und den Fragenpaketen	
		Freiwillig		Gruppe04_Absatz17_Person3: „Abgesehen davon, finde ich das Konzept ziemlich gut, weil es nicht vorgeschrieben ist. Man kann freiwillig teilnehmen, man wird also nicht bestraft, wenn man es nicht macht.“
		neuer Blickwinkel		Gruppe04_Absatz10_Person5: „Und dass man vielleicht noch einmal einen anderen Blickwinkel auf die Themen erhält und nicht nur den Blickwinkel der Dozenten.“
		Ähnlich der Prüfungssituation		Gruppe01_Absatz33_Person4: „Also, () im Gegenteil, es ist eher genau die Situation, die man auch aus anderen Prüfungssituationen kennt.“
		Digitale Umsetzung		Gruppe02_Absatz120_Person5: „Man konnte es ja auch aufs Handy downloaden.“
		Unverfälschtes Ergebnis		Gruppe01_Absatz23_Person4: „Ich finde diese Veranstaltung, wenn man im Hörsaal sitzt, kann man sich auch leichter selbst betrügen. Wenn jemand neben einem sagt: ‚Das ist das und das‘, dann denkt man sich: ‚Ja, das stimmt natürlich, ja genau‘. Und bei clue sitzt man wirklich alleine und hat für sich ein persönliches Feedback.“
		Räumliche Flexibilität		Gruppe01_Absatz3_Person3: „[...] ungebunden vom Ort [...]“
		Zeitliche Flexibilität		Gruppe01_Absatz24_Person1: „Es ist auch ein bisschen flexibler, wenn man sich selber die Zeit aussuchen kann, in der man es dann wirklich macht.“

		An Göttinger Modulplan angepasst		Gruppe02_Absatz19_Person6: „Was ich gut fand im Vergleich zu zum Beispiel Amboss, wo man natürlich auch kreuzen kann und Staatsexamensfragen üben kann, ist es ja wirklich maßgeschneidert an dem Stoff, was man theoretisch schon können müsste [...]“ Gruppe03_Absatz48_Person2: „Ich fand es irgendwie cool, dass man sich so die Mühe gemacht hat, das auf die Woche anzupassen.“
	Nachteile			
		Vorhandene Medien reichen aus	Student nutzt Amboss oder ähnl. und macht deutlich, dass ihm das reicht/mehr bringt. Dopplung zur Beurteilung des Projektes als Lerninnovation möglich. Hier stellt der Student allerdings einen Vergleich zwischen dem Beantworten von Fragen auf den unterschiedlichen Plattformen und bewertet #clue nicht als „Lerninnovation“	Gruppe02_Absatz77_Person6: „/Ich kenne mein Altbewährtes, ich nutze Amboss. Es funktioniert. Wieso soll ich noch was Neues probieren?“ Gruppe03_Absatz78_Person4: „Andererseits werden viele, die dann Amboss und IMPP-Fragen nutzen, gemerkt haben, dass das dann doppelt gemoppelt ist.“
		Fragenpakete zu kurz frei geschaltet		Gruppe04_Absatz33_Person2: „Es war also sozusagen schon zu spät und dann konnte ich die Fragen nicht mehr beantworten und das war schade.“
		Fragen können nicht wiederholt werden		Gruppe02_Absatz50_Person4: „Weil ich finde es unpraktisch, dass man wirklich die Themen, dass man die Fragen nicht wiederholen kann.“
	Lerninnovation	Vergleich zu anderen Lehrformaten : gute Ergänzung		Gruppe03_Absatz21_Person1: „[...] dann ist es echt noch einmal ein anderes Lernformat, wo vielleicht auch andere Lerntypen einen Vorteil daraus ziehen können.“

		Keine Innovation	Der Studierende nutzt Amboss und erklärt, dass ihm das mehr bringt	Gruppe02_Absatz20_Person4: „Aber ja (...) die krasseste Innovation ever fand ich es jetzt nicht, muss ich sagen.“
	Wieder teilnehmen			Gruppe01_Absatz49_Person3: „Also ich auch. Ich muss auch dazu gestehen, das habe ich ja auch schon im Vorhinein gesagt, dass ich vielleicht eine Personbin, die nicht so wirklich regelmäßig und oft teilgenommen hat und das schon ein bisschen hat schleifen lassen. Aber ich glaube, dass ich im Nachhinein oder auch in die Zukunft geschaut, das mehr nutzen würde, als ich es dieses Semester getan habe.“
	Weiter empfehlen		Empfehlung des ganzen Projektes oder es nur auszuprobieren	Gruppe04_Absatz49_Person5: „Also ich würde es den Kommilitonen weiter empfehlen - natürlich nur unter der Voraussetzung, dass sie selbst sich darauf vorbereiten wollen. Aber ich würde definitiv eine Empfehlung aussprechen.“

Fragenpakete: alle Aspekte, die Bezug zu den Fragenpaketen nehmen, werden hier aufgenommen. Die Erinnerungs-E-Mails können hier erneut aufgeführt werden, falls hiermit der Rückgang der Teilnehmerzahlen begründet wird.

Kategorien			Regeln	Beispiele
1. Ebene	2. Ebene	3. Ebene: 4. Ebene (5. Ebene)		
Bewertung der Zusammenstellung der Fragenpakete			Alle Anmerkungen zu den Fragenpaketen selbst können hier untergebracht werden. Hier wird nur der IST-Zustand betrachtet. Was gefehlt hat/sich gewünscht wird → siehe Verbesserungsvorschläge	
	Umfang angemessen		Bezug zu Anzahl der Fragen	Gruppe04_Absatz139_Person5: „Ich fand den Umfang von zehn Fragen gut.“
	Inhaltliche Passung	gut		Gruppe02_Absatz57_Person2: „Ja, ich fand sie (...) etwas anders als aus anderen Quellen, aber das hat es auch irgendwie interessant gemacht.“ Gruppe04_Absatz60_Person3: „Aber wie vorhin schon gesagt wurde, fand ich es ziemlich gut, dass man so auch Aspekte, die so in der universitären Lehre hier nicht schwerpunktmäßig behandelt wurden, sozusagen wieder aufgegriffen wurden [...]“
	Zeitliche Passung	könnte teilweise besser sein		Gruppe03_Absatz64_Person4: „Also gerade zu Beginn eines Moduls ist mir aufgefallen, dass sich die Fragen nicht komplett mit den Lehrinhalten einer Woche gedeckt haben.“
	Schwierigkeitsgrad	Unterschiedlichen Schwierigkeitsgrad bemerkt; Gut gewählt		Gruppe01_Absatz55_Person2: „Sie waren mal sehr einfach und mal sehr schwierig zu beantworten. Aber insgesamt glaube ich, waren sie auf einem relativ gutem Niveau.“

	Wiederholung von einzelnen Fragen	Bemerk: ja/nein		Gruppe04_Absatz70_Person5: „Teilweise ja. Also bei ein paar Fragen ist es mir aufgefallen - zumindest in dem chirurgischen, also unfallchirurgischen Teil. Ansonsten (...) es waren aber, glaube ich, immer nur einzelne Fragen, die sich immer mal wieder wiederholt haben.“
		Effekt: Frage gefestigt(Lernzuwachs bei schwierigen Fragen)/bei leichten Fragen keinen Effekt/negative Emotionen	Emotionen und Folgen, die im Zusammenhang mit wiederholt auftretenden Fragen genannt werden	Gruppe01_Absatz58_Person1: „Wenn man die Fragen schon mal gekreuzt hat oder schon mal konnte, hatte das keinen so großen Effekt. Wenn man sie schon einmal nicht konnte, hatte das schon einen bisschen größeren Lerneffekt.“ Gruppe04_Absatz72_Person4: „Da habe ich jedenfalls sehr viel wiedererkannt und wusste teilweise noch, was damals die richtige Lösung war.“ Gruppe02_Absatz71_Person2: „Entweder sehr starkes Argernis, weil man sie wieder nicht geschafft hat oder auch tatsächlich einen Motivationseffekt, weil man dann gemerkt hat: ‚Ah, ich habe was wiedererkannt.‘ Das ging, glaube ich, in beide Extreme.“
Rückgang der Teilnehmerzahl			Alle Punkte, die sich die Studierenden vorstellen können, die Einfluss auf die Teilnehmerzahl gehabt haben können	
	Allgemeine Gründe		Semesterunabhängige Gründe	
		Mangelndes Durchhaltevermögen	Nachlassende Motivation fällt hier auch darunter	Gruppe01_Absatz66_Person2: „Je länger sie dauern und wenn sie nicht primär etwas Angenehmes sind, dann macht man sie nur mit eisernem Willen und Disziplin bis zum Ende durch“

		Nicht akute Relevanz		Gruppe02_Absatz75_Person2: „[...] wenn man gedacht hat: ‚Ich bin selbst sowieso noch nicht bei dem Thema, es hat noch keinen Lerneffekt für mich jetzt, diese Fragen zu beantworten.‘ Dann hat man es nicht mehr gemacht.“ Gruppe04_Absatz74_Person5: „[...] durch die Klausuren, die zwischendurch noch kamen, wo sich viele gesagt haben: ‚Okay, ich konzentriere mich lieber auf das, weil das wird eben auch benotet!‘“
		Freiwillig		Gruppe04_Absatz74_Person5: „Und das andere ist eine freiwillige Sache.‘ Und ich glaube, daran könnte es gelegen haben.“
		Datenschutzgrundverordnung: vergessen		Gruppe02_Absatz75_Person2: „[...] man wird auch nicht mehr dran erinnert. Es interessiert im Endeffekt/, jedenfalls hat man auch keine Rückmeldung darüber bekommen, dass es jemanden interessieren würde.“ Gruppe04_Absatz74_Person5: „Definitiv an der Datenschutzverordnung. Das war zumindest bei mir der Grund, also ich dachte eigentlich während des Semesters, als es auch anfing, dass wir keine Mails mehr bekommen haben, dass das Projekt schon vorbei wäre. Also ich habe das gar nicht so ganz mitbekommen, dass das daran lag. Und dadurch, denke ich, dass das daran lag, dass der Hauptteil, der die Fragen nicht mehr beantwortet hat, [...]“
		Einmal vergessen -> Motivation verloren		Gruppe03_Absatz79_Person2: „Ich glaube, vielleicht war auch der erste Knick, wenn man gemerkt hat, dass man es das erste Mal verpasst hat //“

		Nicht genügend Mehrwert		Gruppe03_Absatz78_Person4: „[...] warum bei einer Falschantwort etwas auch falsch ist. Das sind so die Gründe, warum ich bei mir selbst dann auch gemerkt habe, warum ich nicht mehr jede Woche gekreuzt habe.“
		Zeitmangel		Gruppe03_Absatz79_Person2: „// und ich finde auch, dass es zum Ende des Semesters dann auch stressiger wird.“
		andere Prioritäten	Freizeit	Gruppe02_Absatz82_Person1: „[...] aber auch jahreszeitenbezogen die Teilnehmrare unterschiedlich sein kann. Man fängt im April an, da war es draußen noch nicht so wirklich klasse und haben im Juli mit dem Semester, mit den Klausuren aufgehört und dann verbringen die meisten Leute ihre Freizeit draußen und werden nicht daran teilnehmen.“ Gruppe01_Absatz67_Person5: „Das ist da nicht so, man muss das aktiv wollen und sich aktiv hinsetzen und das machen und ich glaube, je näher es dann zu den Klausuren rückt, desto mehr verbringt man seine Zeit irgendwie mit den Dingen, wo man denkt, die sind jetzt essenzieller oder wichtiger in dem Moment.“

	semesterbezogen	Abschlussklausur		Gruppe01_Absatz69_Person4: „Ich glaube schon, dass gegen Ende des Semesters immer eine etwas umfangreichere Klausur kommt, wo noch Querschnittsfächer und so mit drin sind und da muss man sich schon noch einmal ein bisschen mehr angucken und teilweise auch noch einmal Sachen aus Modulen, die am Anfang des Semesters waren, teilweise noch in einem anderen Semester stattgefunden haben. Und ich glaube da ist es dann schon so, dass man sich sehr viel mit dieser Klausur beschäftigt.“
		Lehrende haben zu wenig Werbung gemacht		Gruppe03_Absatz83_Person1: „In 3.1 wird es auch noch viel häufiger erwähnt // // Da wird immer in der Vorlesung gesagt: ‚Und denkt an die clue-Fragen!‘ Und dann in 4.3 wurde es kein einziges Mal erwähnt.“
		Qualität der Lehre	Wenn die Vorlesungen alleine nicht ausgereicht haben, um die Fragen zu beantworten	Gruppe02_Absatz80_Person4: „Also bei uns, wir waren jetzt im dritten klinischen Semester, also im Siebten, war es auch einfach meiner Meinung nach ganz klar so, dass die Lehre im ersten Modul, also im Kardio-Modul viel besser und viel mehr darauf abgezielt war, dass man die Themen ganzheitlich verstanden hat. Also im Kardio-Modul - wie auch schon vorher gesagt - war es ganz normal, dass man ganz viele verschiedene Herangehensweisen an Themen hatte und darauf wurde auch aufgebaut und immer wieder hingewiesen. Später in Nephro und Uro, und vor allem auch in Onko war es so, dass die Vorlesungen bei weitem nicht so zahlreich besucht waren und dass die Lehre allgemein auch – mit Verlaub – auch einfach SCHLECHTER war und man dann viel weniger Motivation hatte, sich noch zusätzlich damit zu beschäftigen. Am Ende irgendwie hat nur noch gezählt, dass man sich den Stoff selbst beibringt für die Klausur (...) aber da war man gar nicht auf dem Level während der Module noch Fragen dazu zu beantworten. Also ich zumindest nicht.“

Ernsthaft beantwortet	Ja		Aussagen, die nach der Frage getroffen wurden	Gruppe04_Absatz62_Person2: „Nach meinem besten Wissen und Gewissen.“ Gruppe04_Absatz63_Person5: „Aber ansonsten habe ich mich schon hingesezt und versucht, mich wirklich auf die Fragen zu konzentrieren und die auch gewissenhaft zu beantworten.“
	teilweise	Mit Hilfestellung		Gruppe01_Absatz60_Person4: „Und es gab aber auch so Sachen, wo ich wirklich gedacht habe: ‚Okay, da habe ich überhaupt keine Ahnung mehr davon.‘ Vor allem im Modul 6.x, wo noch einmal Sachen dran waren, wo es, wenn man nicht unbedingt in der Vorlesung dazu war, wirklich schon sehr lange her war. Da gebe ich auch zu, dass ich mir einen kurzen Abriss durchgelesen habe, bevor ich mir dann die Frage noch einmal durchgelesen habe.“
		Ohne Hilfestellung		Gruppe03_Absatz70_Person3: „Nachgelesen habe ich jetzt nichts, weil man das sonst ja auch nicht kann.“
		Geraten		Gruppe03_Absatz70_Person3: „Wenn ich jetzt das Gefühl hatte, dass ich überhaupt nicht weiß, wovon geredet wird, habe ich einfach irgend etwas gekreuzt.“ Gruppe03_Absatz71_Person1: „Manchmal waren auch Fragen dabei, wo ich dann geraten habe.“
	nein	Keine Werbung		Gruppe02_Absatz60_Person3: „Zu Anfang noch sehr ernsthaft, als da die Werbetrömmel noch gut gerührt wurde während dem Kardio-Modul. Danach nicht mehr wirklich.“

Wirkung auf einen selbst			Wenn das Beantworten der Fragenpakete eine Auswirkung auf den Studenten hatte, unabhängig ob positiv oder negativ	
	Emotionen	Positive Emotionen: Freude, Ansporn, beruhigend, Spaß		Gruppe01_Absatz19_Person5: „Also wird der eigene Ansporn dadurch gefördert.“ Gruppe04_Absatz67_Person2: „Also bei mir war das glaube ich eher nicht so der Ehrgeiz, sondern mehr, dass ich das generell, wenn es nicht auf eine Note hinaus läuft, eigentlich ganz spaßig finde, solche Fragen zu beantworten. Das ist immer wie ein Quiz oder so.“ Gruppe02_Absatz31_Person4: „Manchmal (.) hat es echt total Spaß gemacht.“ Gruppe04_Absatz65_Person3: „Ich fand es befreiend.“
		Negative Emotionen: Frustration, Ärger		Gruppe02_Absatz31_Person4: ZWIESPÄLTIGE EMOTIONEN. Also wirklich tatsächlich. Manchmal (.) hat es echt total Spaß gemacht, wenn man gut war und der Medizinstudent ist ja auch immer ehrgeizig, aber gleichzeitig natürlich auch Frustration, wenn man irgendwie gemerkt hat, dass man einfach nicht so erfolgreich ist, wie man vielleicht gehofft hatte in der Woche.“

	Motivierend	Ja: nacharbeiten	Studierender erwähnt, dass ihn das Bearbeiten der Fragen dazu gebracht hat, noch einmal etwas nachzulesen	Gruppe03_Absatz70_Person3: „Bei manchen Sachen habe ich hinterher schon noch einmal nachgelesen.“ Gruppe01_Absatz63_Person2: „Also ich muss sagen, ich finde, immer den größten motivierenden Effekt hat eine Frage, wenn sie wirklich etwas anschneidet, wo man denkt: ‚Das hast du ja noch nie gehört... Was ist das?‘. Und dann: ‚Wieso fragen die das jetzt hier? Scheint ja anscheinend doch wichtig zu sein‘, und dann macht man kurz den nächsten Tab bei Google auf und guckt sich das dann schnell an. Also das hat schon einen Effekt.“
		Nein: erst kurz vor der Klausur die Fragenpakete bearbeiten, wenn man zu wenig weiß		Gruppe01_Absatz64_Person3: „Ich muss dazu jetzt sagen, das ist jetzt wirklich nur auf mich persönlich bezogen, als ich die ersten zwei Pakete gemacht hatte, dass es auf mich so einen etwas demotivierenden Aspekt hatte in dem Sinne, dass wenn ich dann das Gefühl hatte: ‚Okay, ich weiß da sehr viele Fragen nicht‘, dass es mich jetzt ein bisschen demotiviert hat, das weiter zu machen. Was vielleicht einfach daran liegt, dass ich generell eine Personbin, die sich mit so etwas wie (...) Klausurfragen ziemlich am Ende von einem Lernblock beschäftigt. Dass ich am Anfang wirklich erst lerne und am Ende dann mich mit den Fragen beschäftige. Genau. (...) Also das (.) vielleicht ging es anderen ja auch so. Also das weiß ich nicht.“ Gruppe02_Absatz65_Person6: „Aber da waren besonders für mich persönlich im Orthopädie Modul zum Beispiel, da konnte ich insgesamt sehr, sehr wenig beantworten und das fand ich dann irgendwann (.) genau klickt man durch: ‚Weiß ich nicht, weiß ich nicht, weiß ich nicht.‘. Aber da hatte man dann auch wenig Motivation.“

Interdisziplinäre Lehrveranstaltungen: Alle Punkte zu den interdisziplinären Lehrveranstaltungen und ob/wie sie wahrgenommen worden sind. Anregungen/Wünsche/Vorstellungen zur Gestaltung finden sich bei Verbesserungsvorschlägen.

Kategorien	Regeln	Beispiele
1. Ebene	2. Ebene	3. Ebene: 4. Ebene (5. Ebene)
Gründe zu besuchen	Alles was dazu geführt hat, dass der jeweilige Student an einer interdisziplinären Lehrveranstaltung teilgenommen hat; falls er diese nicht wahr genommen hat, können auch Gründe aufgeführt sein, warum er gerne eine besuchen würde	
	Einblick in ärztlichen Alltag	Gruppe01_Absatz85_Person5: „Das konfrontiert einen ja auch so ein bisschen mit den Problemen, die man später im ärztlichen Alltag auch hat. Dass es für ein Problem vielleicht nicht nur eine Lösung gibt, sondern die eine Fachdisziplin hält die Lösung vielleicht für den richtigen Weg, die andere eine andere Lösung und dann gibt es vielleicht auch mal in der interdisziplinären Lehrveranstaltung eine Diskussion, was jetzt richtiger (..) oder richtig/falsch ist und ich denke, das wird einem später dann auch überall begegnen – egal welche Richtung man dann einschlägt. Da kriegt man das vielleicht auch mal ein bisschen mit, dass es nicht nur eine Lösung gibt.“
	Realitätsnah	Gruppe01_Absatz84_Person2: „Alles wird so ein bisschen (.) relativiert. Und das ist ja viel realistischer als die Sicht, die man sonst nur aus einer einzigen Position heraus hätte.“
	Vermittlung neuer Sichtweisen	Gruppe01_Absatz84_Person2: „Aber was einem als Student zum Zeitpunkt des Studierens glaube ich, immer

				nicht so richtig klar ist, (...) ist, dass man immer das Fach beigebracht bekommt, das immer komplett aus der Sicht des Fachs, (...) also komplett aus der Sicht des – Entschuldigung für den Ausdruck, aber – „FACHIDIOTEN“ oder komplett jetzt aus der Sicht des Kardiologen, (...) des Viszeralchirurgen oder sonst irgendwem. Und man (...) kriegt dadurch einfach noch einmal eine komplett andere Sicht.“
	Vermittlung eines allgemeinen Blickes		Dozenten führen einen Diskurs und zoomen etwas raus	Gruppe01_Absatz84_Person2: „Und deshalb mag das (...) ist das gut, dass es gerade gegen Ende des Studiums hin so ein bisschen (...) mehr rauszoomt aus dem Ganzen und einen allgemeineren Blick vermittelt. So habe ich das aufgefasst.“ Gruppe01_Absatz87_Person3: „[...] die so den Blick auf das komplette Ganze einem öffnen.“
	Anwendung des eigenen Wissens			Gruppe01_Absatz86_Person4: „Und je nachdem, wie es konzipiert war, wenn es dann auch noch interaktiv war, dass man wirklich auch, das was man selbst schon wusste, auch einbringen konnte, dann ist zum einen, wie schon gesagt wurde (...). ist es einfach ein Abbild der Realität auf der einen Seite, aber man kann halt auch noch einmal das, was man gelernt hat, in einem ganz anderen Kontext anwenden als wenn man nur am Ende Fragen dazu beantwortet.“
	Interesse			Gruppe03_Absatz20_Person3: „Die fand ich sehr gut. Die waren ja, wie gesagt, interdisziplinär mit mehreren Fachrichtungen. Das fand ich erstens sehr interessant und zweitens oft auch sehr aufschlussreich. Da waren auch oft mit Sachen dabei, die ich so jetzt nicht verstanden hatte und das fand ich auf jeden Fall sehr gut.“

	Klausurtipps		Student macht deutlich, dass er sich Tipps erhofft oder dass Veranstaltung relevant für Klausur ist	Gruppe02_Absatz86_Person4: „Also ehrlicherweise hat man sich schon auch Klausurtipps erhofft.“
	Wiederholung vergangener Module			Gruppe02_Absatz90_Person2: „[...] Das fand ich schön, dass man noch einmal eine Wiederholung aus dem Fach hatte, was man jetzt schon ein Semester nicht mehr hatte.“
	Besprechung von schwierigen Fragen			Gruppe03_Absatz22_Person3: „Was mich auch sehr angesprochen hat, ist, dass die Fragen, die viele falsch beantwortet hatten, dann teilweise - ich sag mal - diskutiert wurden.“
Gründe nicht zu besuchen			Hier werden alle Gründe aufgeführt, die dazu geführt haben, dass die Studierenden nicht teilgenommen haben	
	Nicht genug publik gemacht	Steht nicht im Stundenplan	Wenn der Student nicht davon wusste, dass eine Veranstaltung stattfindet, wird dies unter „nicht genug publik gemacht“ aufgeführt	Gruppe01_Absatz87_Person3: „Aber ich muss ehrlich gesagt auch dazu sagen, also ich habe jetzt das vierte Klinische hinter mir, dass mir auch nicht bewusst gewesen ist, dass es so interdisziplinäre Veranstaltungen gegeben hätte, wenn es die gab. Also (...) da hätte ich gerne teilgenommen, aber ich wusste auch nicht, dass ich die verpasst habe.“
	unflexibel	Ungünstiger Termin, vorgegebener Ort		Gruppe01_Absatz89_Person5: „Es ist halt wieder ein Präsenztermin. Der findet zu der und der Uhrzeit an dem und dem Ort statt und man ist nicht so flexibel wie mit den Fragepaketen.“

	Kein ausreichender Mehrwert		Nicht klausurrelevant, kein direkter und unmittelbarer Lerneffekt	Gruppe01_Absatz91_Person2: „Eine gute Veranstaltung ist dann gut besucht, wenn sie auch gut ist und wenn die Veranstaltung einfach (...) aus studentischer Sicht oder aus der Mehrheit der studentischen Sicht oder Mehrheit der Studenten nicht als würdig empfunden wird, dahinzugehen, weil die einfach nicht so den Mehrwert bietet. Dann kann das vielleicht auch ein Grund sein, warum die nicht so gut besucht war.“
	Mangelndes Interesse			Gruppe01_Absatz91_Person2: „Oder auch vielleicht wirklich mangelndes Interesse von studentischer Seite aus. Das (...) muss man, kann man ruhig auch mal versuchen, oder kann man den Leuten eventuell auch ankreiden.“
	Prioritäten anders gesetzt		Mögliche Punkte können hier sein: Freizeit, für Klausur lernen etc.	Gruppe01_Absatz97_Person2: „[...] aber ganz rational keine Zeit dafür hat, weil man will vielleicht auch noch einmal eine Runde joggen gehen oder Fußballspiel oder so was machen, dann muss man irgendwo Abstriche machen. Und ja (...) so wichtig ist es dann am Ende dann doch nicht. Deshalb, aber das ist am Ende ein grundsätzliches Problem. Das wird man schwierig lösen können.“

Verbesserungsvorschläge: Hier werden Wünsche/Vorstellungen jeglicher Art aufgeführt. Dem Studenten sind keine Grenzen gesetzt. Wenn ein Student seine Wünsche zur Gestaltung einer interdisziplinären Lehrveranstaltung/der Fragenpakete aufführt, selbst aber nicht teilgenommen hat, sind diese dennoch hier aufgeführt.

Kategorien	Regeln	Beispiele
1. Ebene	2. Ebene	3. Ebene: 4. Ebene (5. Ebene)
Fragenpakete	Alle Verbesserungsvorschläge, die im Verlauf zu den Fragenpaketen genannt worden sind	
	organisatorisch	Alles, was sich durch die Änderungen der Grundeinstellungen verwirklichen lässt
		Verfügbarkeit der Fragenpakete: mehr als eine Woche freigeschaltet lassen
		Begrenzte Bearbeitungsdauer von 50 Minuten: abschaffen/kürzer
		Gruppe02_Absatz747_Person3: „Aber wie schon B4 gesagt hat, wenn man doch noch länger braucht um den Stoff zu wiederholen, dass man dann einfach sagt: ‚Drei Wochen später, jetzt habe ich das Thema bearbeitet. Jetzt kreuze ich die entsprechenden Fragen.‘. Das wäre einfach auch schön, oder wäre eine Motivation für mich, das nächstes Semester zu nutzen.“
		Gruppe01_Absatz103_Person5: „[...] und dass dann da so eine Zeitsperre war, das fand ich ziemlich (...) das hat mich in dem Moment so ein bisschen genervt, weil ich dachte: ‚Ich wollte es ja jetzt wirklich machen und dann ging es nicht.‘ [...]“

		Erinnerungen (früher als einen Tag vor Ablauf, gleichmäßig, persönliche Hinweise von aktuellen Lehrenden)	Gruppe04_Absatz34_Person4: „Es wäre vielleicht gut, wenn das zwei bis drei Tage vorher ist. Dann hat man noch etwas mehr Zeit.“ Gruppe02_Absatz112_Person2: „Eine mögliche gleichmäßige Erinnerung in allen Modulen.“ Gruppe01_Absatz107_Person3: „Und genau noch eine andere Sache wäre, die ich mir vorstellen kann, wo man vielleicht noch mehr Leute motivieren kann, das zu machen, ist, wenn man so die Dozenten in einem Modul so ein bisschen dazu anregen würde, dass sie, wenn sie eine Vorlesung zu einem Thema haben, am Ende auch noch einmal auf clue oder so hinweisen.“
	technisch		Alles was programmiert werden müsste oder das Programm ILIAS betrifft
		Eine App entwickeln	Gruppe01_Absatz104_Person2: „[...] wenn man einfach eine kleine App hätte [...]“
		Quick-Evaluation pro Fragenpaket	Gruppe04_Absatz124_Person3: „Ansonsten könnte man sich überlegen, pro Frage oder pro Fragenpaket noch eine Art Quick-Evaluation hinten dran zu hängen, wie sinnvoll das Fragenpaket war, einfach mit drei Kästchen. Wie sinnvoll war das Fragenpaket? Hat es den Vorlesungsinhalt wiedergespiegelt? Aber auch da weiß ich nicht, ob das ankommen würde.“
		Forum zum gemeinsamen Diskutieren	Gruppe04_Absatz110_Person3: „[...] eine Art Forum, so dass man nach Beantwortung der Fragen anonym noch Kommentare dazu posten kann, die auch von den anderen eingesehen werden können.“

		ILIAS: Layout (Antwortpunkte ordnen; „bestätigen&weiter-Button“ abschaffen; runterscrollen zur richtigen Antwort), aufrufen zu umständlich;	Hinweise zur Benutzeroberfläche/Programm ILIAS	Gruppe04_Absatz113_Person3: „Was mir noch extrem am Herzen liegt und ich glaube, dass das eine der maßgeblichsten Hürden ist, man muss sich, um die Fragen zu beantworten - das geht zwar auch am Handy - man muss sich allerdings zuerst beim Stud.IP anmelden, und dann durchklicken und das ist noch relativ kompliziert. Falls es in irgendeiner Form möglich wäre, dass man per Mail oder in Stud.IP direkt einen Link bekäme, dass ich praktisch mit einem Klick zu den Fragen komme und nicht noch auf (...) die Untergruppen und ‚Test starten‘ klicken muss. Ich glaube, das würde es einfacher machen.“ Gruppe01_Absatz103_Person5: „...und dann immer auf ‚bestätigen und weiter drücken‘, dass dann Leute, die das wirklich nervt, sagen: ‚Ich habe da keine Lust mehr drauf, ich lasse es sein.‘. Das ist ein bisschen schade, (...) wenn das der Grund ist.“
		Vergleich zu Mitstudierenden		Gruppe03_Absatz124_Person2: „Da weiß man dann hinterher: ‚Bin ich die einzige, die so denkt oder machen das andere auch?‘ Gruppe03_Absatz125_Person3: „Da konnte man auch immer sehen, wer wie viel angekreuzt hat. Das ist auch eine gute Idee.“
		Sitzung bei AMBOSS erstellen		Gruppe01_Absatz102_Person4: „[...] es gibt ja angeblich bei ‚Amboss‘ bei diesem online-Lehrprogramm auch fürs Staatsexamen, Möglichkeiten, wo Dozierende Sitzungen erstellen für alle, die einen Zugang haben und wir haben ja alle einen Zugang.“

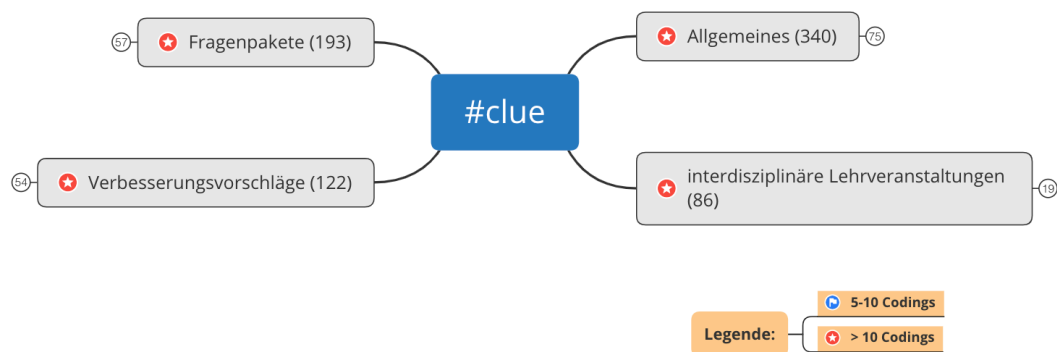
		Module und Fragen aktiv selbst auswählen		Gruppe02_Absatz110_Person2: „[...] dass man sagt, dass man die Fragen so ein bisschen suchen kann. Dass man Module auswählen kann und sagen kann; Hier zu dem Paket möchte ich noch einmal Fragen machen oder zu dem noch einmal oder auch vom Jahr davor, wenn man zum Beispiel wiederholende Fragen hat, die dann vom Vorsemester sind“, dass man dann auf das Vorsemester nochmal zurückgreifen kann [...].“
		Anzeige der Fragenpakete		Gruppe03_Absatz111_Person4: „Mir ist auch noch einmal aufgefallen, ziemlich zum Schluss des Semesters, diese Wochen werden ja dann immer noch einmal aufgeführt und irgendwann war da Woche zehn, stand dann nicht unter Woche neun, sondern zwischen vier und fünf. Und das war das erste Mal, dass ich das Fragenpaket verpasst habe. Aber ich habe nicht genau geguckt. Dass man vielleicht darauf achtet, dass das in der richtigen Reihenfolge ist.“
	Inhaltlich		Alles was die Auswahl der Fragen selbst betrifft	
		Mehr Themengebiete pro Fragenpaket abdecken		Gruppe04_Absatz57_Person5: „Was mir noch aufgefallen ist, ist, dass bestimmte Fragenpakete - bestimmt ein Drittel oder fast die Hälfte der Fragen - teilweise den gleichen Komplex umfassen haben oder das gleiche Themengebiet. Und wenn man sich mit dem Themengebiet noch nicht beschäftigt hatte, konnte man drei, vier Fragen gar nicht beantworten. Wenn man ein, zwei Begriffe aus der Frage noch gar nicht kannte oder es noch gar nicht behandelt hat (...) und das fand ich dann ein bisschen schade, weil dadurch dann eben auch ein bisschen das Ergebnis etwas verfälscht wurde. Und da könnte man vielleicht noch einmal ein bisschen auf Variation achten.“

		Vereinzelte Fragen von Lehrenden		Gruppe03_Absatz118_Person4: „Das wäre jetzt so eine Idee von mir gewesen. Also jetzt nicht alle zehn Fragen durch Fragen von Dozierenden ersetzen, aber vielleicht so zwei.“
		Wiederholung von älteren Modulen verstärken		Gruppe01_Absatz122_Person5: „Dass sich das einfach mehr wiederholt.“
	Schwierigkeitsgrad	angeben		Gruppe02_Absatz122_Person2: „Vielleicht – das ist jetzt auch wieder Design und Inhalt – aber ob man da vielleicht auch so ein bisschen (...), ob man da den Schweregrad angibt.“
		Projekt auch in anderen Fächern anbieten		Gruppe01_Absatz50_Person2: „[...] und auch gerne für andere Fächer übernehmen (...) habe ich auch nichts dagegen oder hätte ich nichts dagegen gehabt.“
Interdisziplinäre Lehrveranstaltung			Alle Verbesserungsvorschläge, die zu den interdisziplinären Lehrveranstaltungen genannt worden sind	
	Vorstellungen zur Gestaltung	Abwechslungsreich: Bier und Brezel, motivierte Lehrende	Format: Bier und Brezel, Vorlesung des Augenmoduls werden genannt	Gruppe02_Absatz100_Person5: „Beim Augenmodul sind die ja sehr erfolgreich damit gewesen mit dem Modell. Dass sie Bier und Brezel anbieten, das abends machen und dann eine tolle Vorlesung halten. Das Konzept ist da wirklich vergleichbar.“
		Informativ		Gruppe01_Absatz93_Person1: „Möglichst informativ.“
		Relevanz: für aktuelle Klausur	Für aktuelle Klausur	Gruppe02_Absatz101_Person3: „Durchaus der Anreiz, dass doch Themen besprochen werden, die einfach klausurrelevant sind.“
		Längerfristige Planung (im Stundenplan aufnehmen)	Nicht spontan an die Tafel schreiben/per E-Mail/facebook ankündigen; auch nicht spontan beorderte Dozierende gehören dazu	Gruppe03_Absatz27_Person1: „Es wäre cool, wenn die im Stundenplan mit stehen würden.“

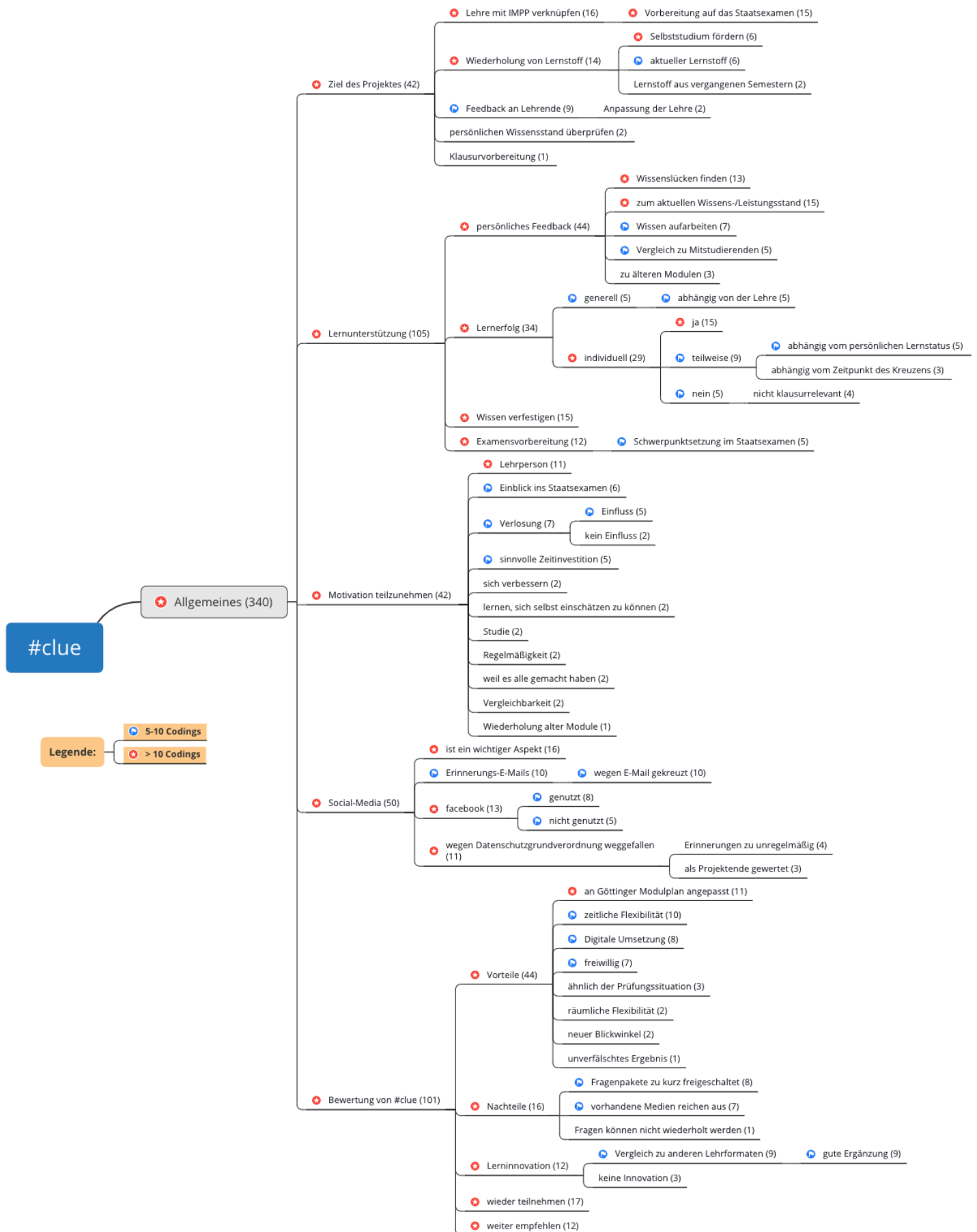
		Live übers Internet übertragen		Gruppe04_Absatz100_Person3: „Eventuell noch übers Internet zu übertragen [...]“
		Für mehrere Semester anbieten		Gruppe01_Absatz98_Person5: „Vielleicht könnte man auch solche interdisziplinären Veranstaltungen nicht nur für das betreffende Semester machen, sondern wie es die Augenheilkunde mit der HNO und MKG zusammen macht. Die machen ja so eine Lehrveranstaltung, wo sie sagen, alle ab dem Semester, wo es gelehrt wird, können kommen und da kommen dann auch so viele freiwillig, die sich dann in dem Semester befinden, wo sie Examen machen. Ich war jetzt zweimal da und es war beide Male gut besucht. Selbst wenn die, die in dem Semester gerade sind, nicht kommen, hat es trotzdem einen Lerneffekt für viele andere gehabt. Ich schätze mal, wenn wir jetzt von unseren eigenen Semestern nicht wussten, dass es stattgefunden hat, dann war es auch nicht semesterübergreifend. Vielleicht wäre das einfach eine Möglichkeit, um Leute zusammen zu kriegen.“
Motivation steigern			Alle Einflüsse für die Studenten auf ihre Motivation.	
		Unmittelbare Relevanz		Gruppe01_Absatz107_Person3: „Das, was wir eben hatten, dass das dieses Gefühl nochmal vermittelt: ‚Okay, das hat jetzt auch unmittelbare Relevanz für dieses Modul und für diese Klausur.‘ Und dann glaube ich, dass man damit auch noch einmal mehr Leute motivieren könnte.“

	Noch detailliertes Feedback der eigenen Leistung			Gruppe02_Absatz123_Person1: „Ich glaube, ein ganz großer Vorteil von Amboss und ähnlichen Programmen ist, dass sie sehr gutes und sehr detailliertes Feedback über die eigene Leistung geben. Über alle möglichen Statistiken, wie man was wann wo gekreuzt hat und welche Leistung man gebracht hat im zeitlichen Verlauf und ich glaube so ein Mechanismus mit diesem Feedback, wenn da so ein Pop-Up kommt: ‚Ja, das war richtig.‘ oder ‚Das hast du vor paar Wochen oder Monaten schon mal richtig gekreuzt und jetzt wieder. Sehr schön!‘, das würde die Leute sehr motivieren. Das kennt man ja auch als verhaltenspsychologischen Aspekt.“
Lernerfolg steigern			Alle Faktoren, zu einer weiteren Steigerung des Lernerfolgs führen können	
	Noch besser auf den Lehrplan abstimmen			Gruppe03_Absatz114_Person3: „[...] und wenn das möglich ist, dass man das noch etwas besser mit dem Lehrplan abstimmen kann.“
	Detaillierteres Feedback	Hintergrundinformationen zur Antwort		Gruppe01_Absatz102_Person4: „[...] paar Hintergrundinformationen zu den Fragen direkt erhält [...]“
	Von Beginn an fachübergreifend anbieten			Gruppe01_Absatz114_Person2: „Das ganze einfach fachübergreifend machen, von Beginn an. (...) Das wäre auf jeden Fall schon einmal ein Punkt.“
	Fächerwiederholungen verstärken			Gruppe03_Absatz128_Person1: „[...] das vielleicht immer mal wieder mehr als zehn Fragen kommen, vielleicht auch zwölf oder so und dann noch welche aus den Modulen von davor. So wie das gemacht wird bei E-Fall Seminaren.“
	Regelmäßigere Teilnahme			Gruppe04_Absatz136_Person3: „Regelmäßigere Teilnahme, auch hier Informationspolitik.“

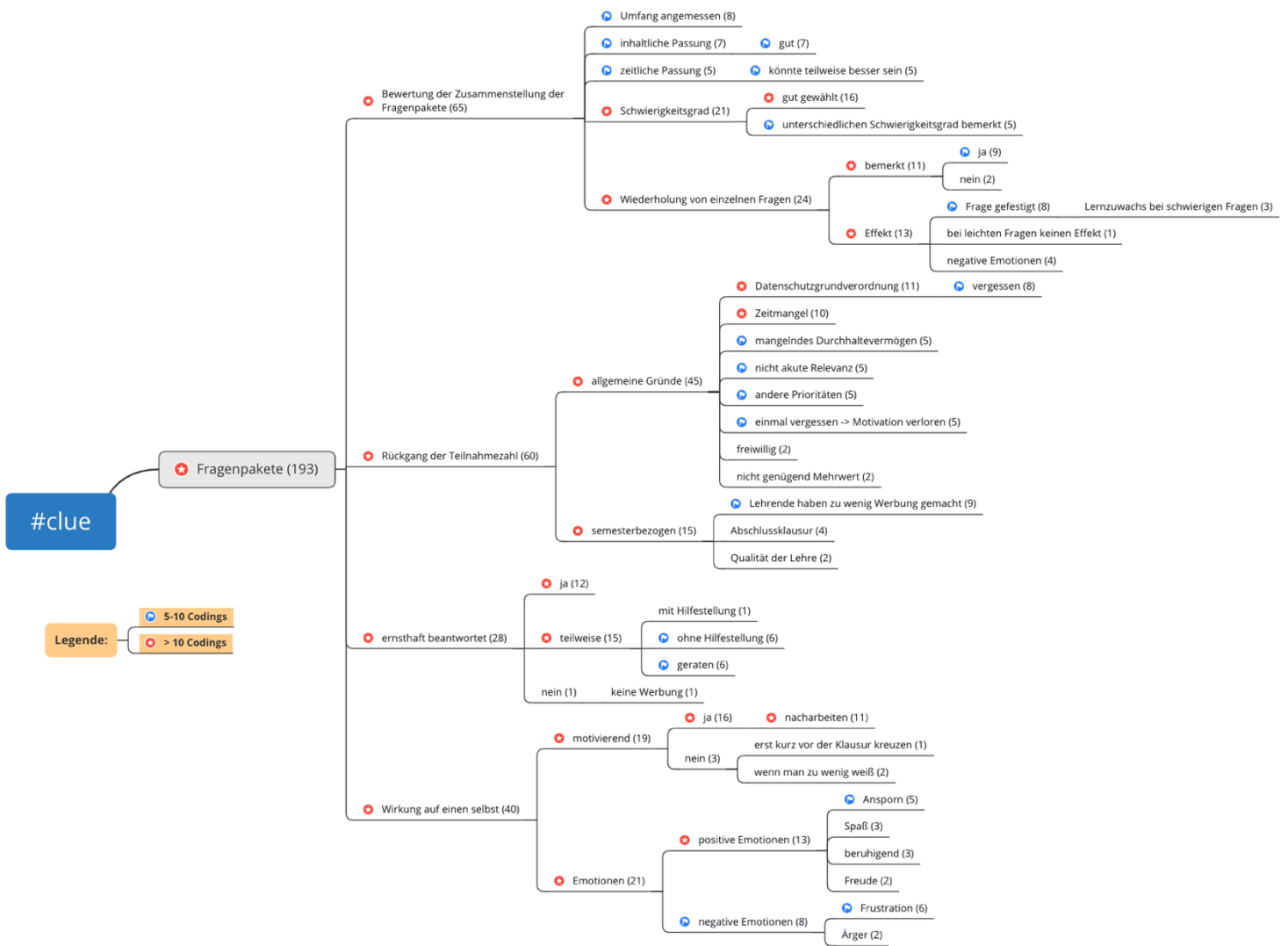
Anhang 3: MindMap „Übersicht“



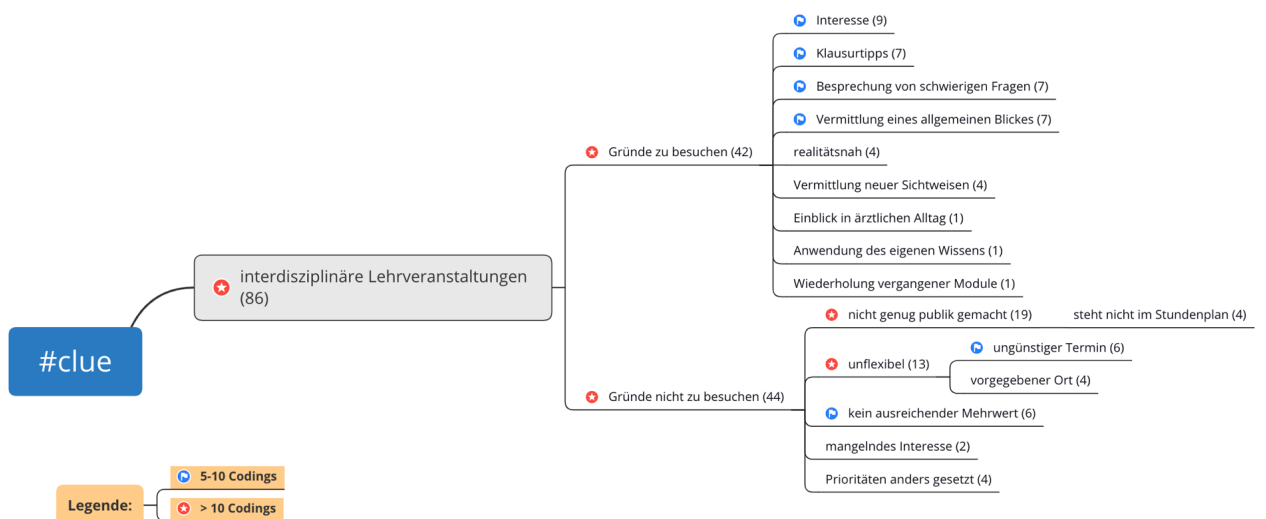
Anhang 4: MindMap „Allgemeines“



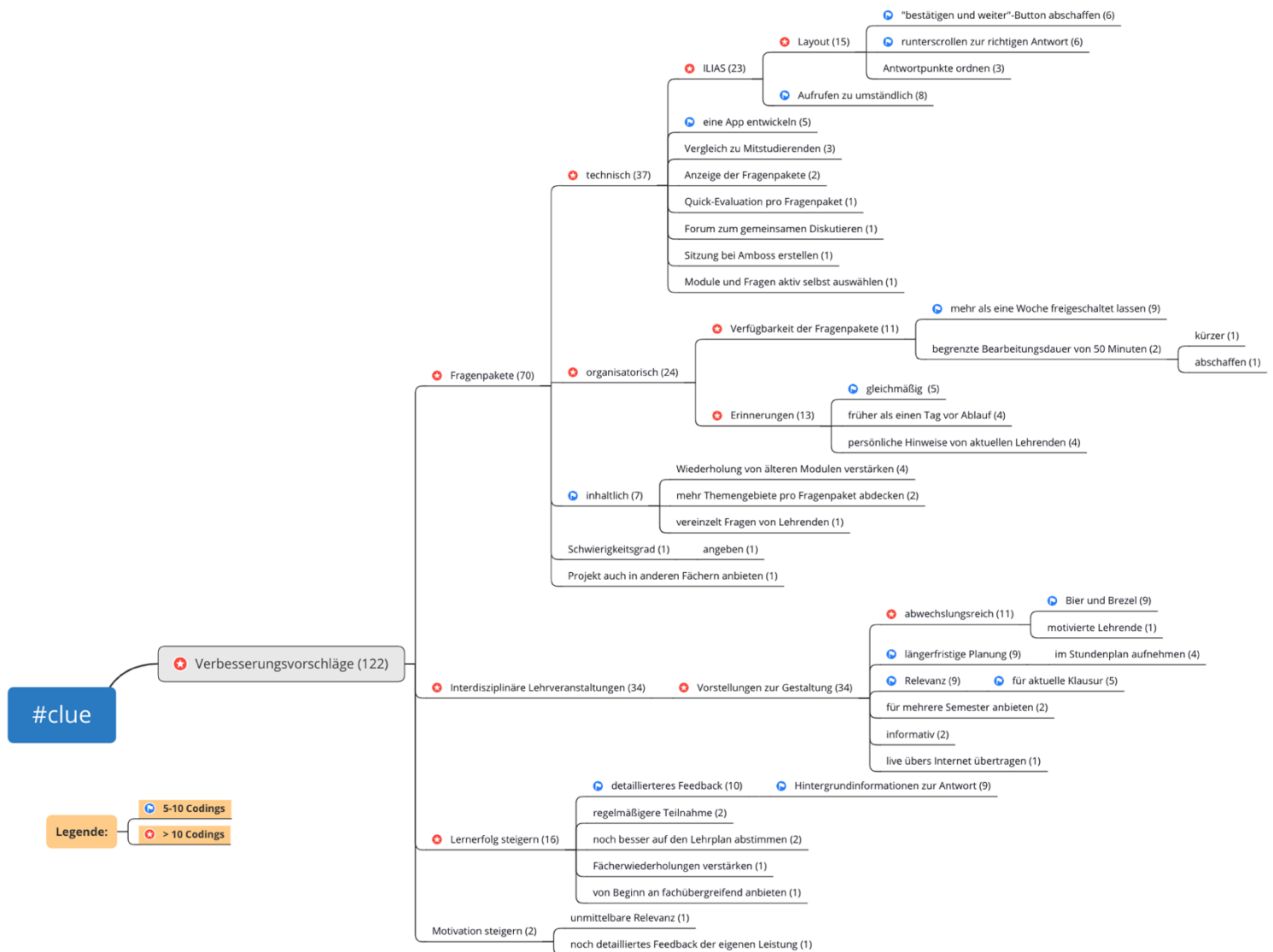
Anhang 5: MindMap „Fragenpakete“



Anhang 6: MindMap „interdisziplinäre Lehrveranstaltungen“



Anhang 7: MindMap „Verbesserungsvorschläge“



Anhang 8: Verbesserungsvorschläge #clue

Verbesserungsvorschläge #clue

- **Fragenpakete**
 - **Technisch**
 - ILIAS:
 - **Erinnerungs-E-Mails früher vor Ablauf und eventuell mit Link**, der in das Fragenpaket auf ILIAS direkt führt (über e-campus zu umständlich)
 - Layout verschönern: „bestätigen und weiter“-Button; umständliches scrollen, wenn die Lösungen eingesehen werden möchte
 - High-Score-System oder zumindest angeben, zu wie viel Prozent welche Antwort gekreuzt worden ist → Studenten wünschen sich, dass sie sich mit anderen vergleichen können. Hierbei geht es nicht um das „Ich will der Beste sein“-Denken, sondern viel mehr darum, zu sehen: „Ich war nicht der einzige, der diese Frage nicht beantworten konnte.“
 - **Quick-Evaluation pro Fragenpaket**: Am Ende von jedem Fragenpaket kann der Student es mit 2 kurzen Fragen bewerten (z.B.: Wie sinnvoll war das Fragenpaket? Wie gut hat es den Vorlesungsinhalt widerspiegelt?)
 - Forum, in dem anonym noch Fragen diskutiert werden können, auf die Kommilitonen/Dozenten antworten können
 - Ganz langfristig gesehen: eine App entwickeln
 - **Organisatorisch**
 - Fragenpakete für 2 Wochen oder für ein ganzes Modul **länger online stellen**, immer Mittwochs das neue Fragenpaket (zwei Fragenpakete parallel bearbeitbar)
 - **E-Mails kommen von einer eigenen E-Mailadresse von #clue**
 - Falls die Fragenpakete nach wie vor nur 1 Woche online bleiben sollten: **Erinnerungsmails schon Sonntag/Montagmorgen**
 - Mehr und regelmäßige **Hinweise auf #clue-Fragenpakete von Dozenten**, die auch aktuell Vorlesungen halten
 - **Kurze Kommentare** von Dozenten zu den schwierigsten Fragen
 - **Inhaltlich**
 - Mehr Fragen aus **alten/vergangenen Modulen**: als Idee → die 10. Frage könnte immer ein Rückblick zu der jeweiligen Woche des vorausgegangenen Semesters sein; hier könnte die Frage gewählt werden, die die Studenten am schlechtesten beantwortet haben
 - Mehr Themenvariationen in den einzelnen Wochen
 - 1-2 Fragen von den Dozenten aus einer Woche anfordern → Relevanz für aktuelle Klausur steigt und bietet gegenüber Amboss einen noch größeren Mehrwert
 - Themen teilweise besser auf Vorlesungen anpassen (explizit genannt: Modul 5.1, Modul 4.2)
 - **Schwierigkeitsgrad**
 - **Angeben**, damit sich der Student und sein Können besser einschätzen kann
- **Interdisziplinäre Lehrveranstaltung**
 - **Termin der Veranstaltungen für alle Semester bekannt geben**
 - **Interdisziplinäre Lehrveranstaltung am Ende eines Fragenpaketes ankündigen**
 - Relevanz für aktuelle Klausur schaffen
 - Bier und Brezel Konzept
 - Live über das Internet übertragen
- **Allgemeine Ideen**
 - Posterdruck für Werbung

7 Literaturverzeichnis

Anderson J (2004): Medical Teacher 25th anniversary series multiple-choice questions revisited. *Med Teach* 26, 110–113

Appel J, Friedman E, Fazio S, Kimmel J, Whelan A (2002): Educational assessment guidelines: A clerkship directors in internal medicine commentary. *Am J Med* 113, 172–179

Approbationsordnung für Ärzte vom 27. Juni 2002 (Bundesgesetzblatt I S. 2405), die zuletzt durch Artikel 3 des Gesetzes vom 16. März 2020 (Bundesgesetzblatt I S. 497) geändert worden ist

Arnold P, Kilian L, Thillosen A, Zimmer G: Handbuch E-Learning: Lehren und Lernen mit digitalen Medien. 3. Auflage; W. Bertelsmann Verlag GmbH & Co. KG, Bielefeld 2013

Ben-David MF (2000): The role of assessment in expanding professional horizons. *Med Teach* 22, 472–477

Biggs J (1996): Enhancing teaching through constructive alignment. *High Educ* 32, 347–364

Bjork RA, Bjork EL: A new theory of disuse and an old theory of stimulus fluctuation. in: Healy A, Kosslyn S, Shiffrin R (Hrsg.): Essays in honor of William K. Estes: From learning processes to cognitive processes. 2. Auflage; Erlbaum, Hillsdale 1992, 35–62

Black P, Wiliam D (1998): Assessment and classroom learning. *Assess Educ* 5, 7–74

Boeker M, Klar R (2006): E-Learning in der ärztlichen Aus- und Weiterbildung: Methoden, Ergebnisse, Evaluation. *Bundesgesundheitsbl* 49, 405–411

Burgess A, Mellis C (2015): Feedback and assessment for clinical placements: achieving the right balance. *Adv Med Educ Pract* 6, 373–381

Butler AC, Roediger HL (2008): Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem Cognit* 36, 604–616

Cantillon P, Irish B, Sales D (2004): Using computers for assessment in medicine. *BMJ* 329, 606–609

Cerutti B, Blondon K, Galetto A (2016): Long-menu questions in computer-based assessments: a retrospective observational study. *BMC Med Educ* 16, 55

Cilliers FJ, Schuwirth LW, Adendorff HJ, Herman N, van der Vleuten CP (2010): The mechanism of impact of summative assessment on medical students' learning. *Adv Health Sci Educ Theory Pract* 15, 695–715

- Cliff W, Freeman S, Hansen PA, Kibble JD, Peat M, Wenderoth MP (2008): Is formative assessment an effective way to improve learning? A symposium at experimental biology 2008. *Adv Physiol Educ* 32, 337–338
- Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM (2008): Internet-based learning in the health professions: A meta-analysis. *JAMA* 300, 1181–1196
- Dahle LO, Brynhildsen J, Fallsberg MB, Rundquist I, Hammar M (2002): Pros and cons of vertical integration between clinical medicine and basic science within a problem-based undergraduate medical curriculum: examples and experiences from Linköping, Sweden. *Med Teach* 24, 280–285
- Deci EL, Ryan RM: *Intrinsic motivation and self-Determination in human behavior*. Springer Science+Business Media, New York 1985
- Dobson JL, Linderholm T (2015): Self-testing promotes superior retention of anatomy and physiology information. *Adv in Health Sci Educ* 20, 149–161
- Dresing T, Pehl T: *Praxisbuch Interview, Transkription & Analyse: Anleitungen und Regelsysteme für qualitativ Forschende*. 5. Auflage; Eigenverlag, Marburg 2013
- Entwistle NJ, Thompson J, Wilson JD (1974): Motivation and study habits. *High Educ* 3, 379–396
- Farmer EA, Page G (2005): A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ* 39, 1188–1194
- Foos PW, Fisher RP (1988): Using tests as learning opportunities. *J Educ Psychol* 80, 179–183
- Fransson A (1977): On qualitative differences in learning: IV – effects of intrinsic motivation and extrinsic test anxiety on process and outcome. *Br J Educ Psychol* 47, 244–257
- Graber ML, Franklin N, Gordon R (2005): Diagnostic error in internal medicine. *Arch Intern Med* 165, 1493–1499
- Grundman JA, Wigton RS, Nickol D (2000): A controlled trial of an interactive, web-based virtual reality program for teaching physical diagnosis skills to medical students: *Acad Med* 75, 47–49
- Güntert A, Wanner E, Brauer HP, Stobrawa FF: *Approbationsordnung für Ärzte (ÄAppO) Bundesärzterordnung (BÄO): Mit Erläuterungen und praktischen Hinweisen*. Deutscher Ärzte-Verlag, Köln 2003
- Hatala R, Norman GR (2002): Adapting the key features examination for a clinical clerkship. *Med Educ* 36, 160–165
- Hodson P, Connolly M, Saunders D (2001): Can computer-based learning support adult learners? *J Furth High Educ* 25, 325–335

- Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR, for the International CBME Collaborators (2010): The role of assessment in competency-based medical education. *Med Teach* 32, 676–682
- Hopstock LA (2008): Motivation and adult learning: A survey among hospital personnel attending a CPR course. *Resuscitation* 76, 425–430
- Hrynychak P, Glover Takahashi S, Nayer M (2014): Key-feature questions for assessment of clinical reasoning: a literature review. *Med Educ* 48, 870–883
- Hudson JN, Bristow DR (2006): Formative assessment can be fun as well as educational. *Adv Physiol Educ* 30, 33–37
- Joughin G (2010): The hidden curriculum revisited: a critical review of research into the influence of summative assessment on learning. *Assess Eval High Educ* 35, 335–345
- Karpicke JD (2009): Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *J Exp Psychol Gen* 138, 469–486
- 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
- Karpicke JD, Blunt JR (2011): Retrieval practice produces more learning than elaborative studying with concept mapping. *Science* 331, 772–775
- Karpicke JD, Butler AC, Roediger HL (2009): Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory* 17, 471–479
- Kerres M: *Mediendidaktik: Konzeption und Entwicklung mediengestützter Lernangebote*. 4. Auflage; Oldenbourg Verlag München 2013
- Kolhe SK, Kadam SS, Narkhede JP, Kulkarni V (2018): Integrated teaching in medical curriculum – undergraduate students perception. *J Res Method Educ* 8, 1–6
- Kopp V, Möltner A, Fischer MR (2006): Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. *GMS Z Med Ausbild* 23, 3, Doc. 50
- Krasne S, Wimmers PF, Relan A, Drake TA (2006): Differential effects of two types of formative assessment in predicting performance of first-year medical students. *Adv Health Sci Educ Theory Pract* 11, 155–171
- Kromann CB, Jensen ML, Ringsted C (2009): The effect of testing on skills learning. *Med Educ* 43, 21–27

- Kuhn S, Frankenhauser S, Tolks D (2018): Digitale Lehr- und Lernangebote in der medizinischen Ausbildung: Schon am Ziel oder noch am Anfang? *Bundesgesundheitsbl* 61, 201–209
- Kusurkar RA, Ten Cate ThJ, van Asperen M, Croiset G (2011a): Motivation as an independent and a dependent variable in medical education: A review of the literature. *Med Teach* 33, 242–262
- Kusurkar RA, Croiset G, Ten Cate OThJ (2011b): Twelve tips to stimulate intrinsic motivation in students through autonomy-supportive classroom teaching derived from Self-Determination Theory. *Med Teach* 33, 978–982
- Kusurkar RA, Ten Cate ThJ, Vos CMP, Westers P, Croiset G (2013): How motivation affects academic performance: a structural equation modelling analysis. *Adv Health Sci Educ Theory Pract* 18, 57–69
- Larsen DP, Butler AC, Roediger III HL (2008): Test-enhanced learning in medical education. *Med Educ* 42, 959–966
- Larsen DP, Butler AC, Roediger III HL (2009): Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Med Educ* 43, 1174–1181
- Lee Y, Choi J (2011): A review of online course dropout research: implications for practice and future research. *Educ Technol Res Dev* 59, 593–618
- Little JL, Bjork EL (2012): The persisting benefits of using multiple-choice tests as learning events. *Cogsci* 34, 683–688
- Little JL, Bjork EL, Bjork RA, Angello G (2012): Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychol Sci* 23, 1337–1344
- Little JL, Frickey EA, Fung AK (2019): The role of retrieval in answering multiple-choice questions. *J Exp Psychol Learn Mem Cogn* 45, 1473–1485
- Mayer RE, Stull A, DeLeeuw K, Almeroth K, Bimber B, Chun D, Bulger M, Campbell J, Knight A, Zhang H (2009): Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemp Educ Psychol* 34, 51–57
- Mayring P: *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. 11. Auflage; Beltz Verlag, Weinheim und Basel 2010
- McDaniel MA, Little JL: Multiple-choice and short-answer quizzing on equal footing in the classroom: potential indirect effects of testing. In: Dunlosky J, Rawson KA (Hrsg.): *The Cambridge Handbook of Cognition and Education*. 1. Auflage; Cambridge University Press 2019, 480–499

- McDaniel MA, Roediger HL, McDermott KB (2007): Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev* 14, 200–206
- McDaniel MA, Wildman KM, Anderson JL (2012): Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *J Appl Res Mem Cogn* 1, 18–26
- McLean M (2001): Introducing a reward system in assessment in histology: A comment on the learning strategies it might engender. *BMC Med Educ* 1, 7
- Morris CD, Bransford JD, Franks JJ (1977): Levels of processing versus transfer appropriate Processing. *J Verbal Learning Verbal Behav* 16, 519–533
- Nelson C, Hartling L, Campbell S, Oswald AE (2012): The effects of audience response systems on learning outcomes in health professions education. A BEME systematic review: BEME Guide No. 21. *Med Teach* 34, 386–405
- Newble D (2016): Revisiting ‘The effect of assessments and examinations on the learning of medical students’. *Med Educ* 50, 498–501
- Newble DI, Jaeger K (1983): The effect of assessments and examinations on the learning of medical students. *Med Educ* 17, 165–171
- Newble DI, Entwistle NJ (1986): Learning styles and approaches: implications for medical education. *Med Educ* 20, 162–175
- Niegemann HM, Hessel S, Hochscheid-Mauel D, Aslanski K, Deimann M, Kreuzberger G (Hrsg.): *Kompendium E-Learning (X.media.press)*. Springer, Berlin 2004
- Nikendei C, Jünger J (2006): OSCE – praktische Tipps zur Implementierung einer klinisch-praktischen Prüfung. *GMS Z Med Ausbild* 23, Doc. 47
- Page G, Bordage G (1995): The medical council of canada’s key features project: A more valid written examination of clinical decision-making skills. *Acad Med* 70, 104–110
- Pekrun R: *Emotions and learning*. 24. Auflage; UNESCO’s International Bureau of Education 2014, 2–31
- Pelaccia T, Viau R (2017): Motivation in medical education. *Med Teach* 39, 136–140
- 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
- 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
- Roediger HL, Marsh EJ (2005): The positive and negative consequences of multiple-choice testing. *J Exp Psychol Learn Mem Cogn* 31, 1155–1159

- Roediger HL, Karpicke JD (2006a): Test-enhanced learning: Taking memory tests improves long-term retention. *Psychol Sci* 17, 249–255
- Roediger HL, Karpicke JD (2006b): The power of testing memory: Basic research and implications for educational practice. *Perspect Psychol Sci* 1, 181–210
- Rowland CA (2014): The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin* 140, 1432–1463
- Ryan RM, Deci EL (2000): Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemp Educ Psychol* 25, 54–67
- Schmidmaier R, Ebersbach R, Schiller M, Hege I, Holzer M, Fischer MR (2011): Using electronic flashcards to promote learning in medical students: retesting versus restudying. *Med Educ* 45, 1101–1110
- Schuwirth LWT, van der Vleuten CPM (2004): Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ* 38, 974–979
- Schuwirth LWT, Vleuten CPM, Donkers HJLM (1996a): A closer look at cueing effects in multiple-choice questions. *Med Educ* 30, 44–49
- Schuwirth LWT, Vleuten CPM, Stoffers HEJH, Peperkamp AGW (1996b): Computerized long-menu questions as an alternative to open-ended questions in computerized assessment. *Med Educ* 30, 50–55
- Seale JK, Chapman J, Davey C (2000): The influence of assessments on students' motivation to learn in a therapy degree course. *Med Educ* 34, 614–621
- Smith MK, Wood WB, Adams WK, Wieman C, Knight JK, Guild N, Su TT (2009): Why Peer Discussion Improves Student Performance on In-Class Concept Questions. *Science* 323, 122–124
- Stipek DJ: *Motivation to learn: From theory to practice*. Prentice-Hall, New Jersey 1988
- ten Cate OThJ, Kusurkar RA, Williams GC (2011): How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE Guide No. 59. *Medical Teacher* 33, 961–973
- Thomas PA, Kern DE, Hughes MT, Chen BY (Hrsg.): *Curriculum development for medical education: a six-step approach*. 3. Auflage; Johns Hopkins University Press, Baltimore 2016
- Tolks D: eLearning in der medizinischen Aus-, Weiter- und Fortbildung. In: Fischer F, Krämer A (Hrsg.): *eHealth in Deutschland: Anforderungen und Potenziale innovativer Versorgungsstrukturen*. Springer-Verlag, Berlin und Heidelberg 2016, 223–239

van Bruggen L, Manrique-van Woudenberg M, Spierenburg E, Vos J (2012): Preferred question types for computer-based assessment of clinical reasoning: a literature study. *Perspect Med Educ* 1, 162–171

Weinert FE: Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. 2. Auflage; Beltz Verlag, Weinheim und Basel 2002

Wheeler M, Ewers M, Buonanno J (2003): Different rates of forgetting following study versus test trials. *Memory* 11, 571–580

Williams GC, Saizow RB, Ryan RM (1999): The importance of self-determination theory for medical education. *Acad Med* 74, 992–995

Woltering V, Herrler A, Spitzer K, Spreckelsen C (2009): Blended learning positively affects students' satisfaction and the role of the tutor in the problem-based learning process: results of a mixed-method evaluation. *Adv Health Sci Educ Theory Pract* 14, 725–738

Yang A, Goel H, Bryan M, Robertson R, Lim J, Islam S, Speicher MR (2014): The Picmonic® Learning System: enhancing memory retention of medical sciences, using an audiovisual mnemonic Web-based learning platform. *Adv Med Educ Pract* 5, 125–132

Internetquellen:

AMBOSS: Campuslizenzen. <https://www.amboss.com/de/campuslizenz/uni-uebersicht>; Abgerufen am 12.05.2020

AMBOSS: Handbuch – Kreuzmodus. https://www.amboss.com/de/wissen/Handbuch_-_Kreuzmodus#xid=cn0aHg&anker=Zf3e38058d1b0f43e0ed1238e1de09a9c; Abgerufen am 12.05.2020

AMBOSS: Über uns. <https://www.amboss.com/de/ueber-uns>; Abgerufen am 12.05.2020

Masterplan Beschlusstext 2020. Bundesministerium für Gesundheit, Bundesministerium für Wissenschaft und Forschung, Kulturministerkonferenz, Gesundheitsministerkonferenz, Vertretern aus den Koalitionsfraktionen des Deutschen Bundestages, Berlin 2017. https://www.bmbf.de/files/2017-03-31_Masterplan%20Beschlusstext.pdf; Abgerufen am 29.01.2020

Deutsche Hochschulmedizin e.V.: Empfehlungen III der Deutschen Hochschulmedizin zum Studium der Humanmedizin und Zahnmedizin während der aktuellen COVID-19-Pandemie. https://medizinische-fakultaeten.de/wp-content/uploads/2020/03/Covid-19-Empfehlungen-DHM-III-zum-Medizinstudium_2020-03-25.pdf; Abgerufen am 12.08.2020

Deutscher Wissenschaftsrat: Empfehlungen zur Weiterentwicklung des Medizinstudiums in Deutschland auf Grundlage einer Bestandsaufnahme der humanmedizinischen Modellstudiengänge. Dresden 2014. https://www.wissenschaftsrat.de/download/archiv/4017-14.pdf?__blob=publicationFile&v=3; Abgerufen am 30.01.2020

Deutscher Wissenschaftsrat: Neustrukturierung des Medizinstudiums und Änderung der Approbationsordnung für Ärzte – Empfehlungen der Expertenkommission zum Masterplan Medizinstudium 2020. Dresden 2018. https://www.wissenschaftsrat.de/download/archiv/7271-18.pdf?__blob=publicationFile&v=1; Abgerufen am 02.02.2020

Göttinger Lernzielkatalog. https://www.umg.eu/fileadmin/user_upload/Go_ttinger_Lernzielkatalog_Stand_19-07-2017.pdf; Abgerufen am 02.04.2020

Michel LP, Goertz L, Radomski S, Fritsch T, Baschour L (2015): Digitales Prüfen und Bewerten im Hochschulbereich: Im Auftrag der Themengruppe "Innovationen in Lern- und Prüfungsszenarien" koordiniert vom CHE im Hochschulforum Digitalisierung. https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD%20AP%20Nr%2015_Digitale%20Lernszenarien.pdf; Abgerufen am 15.04.2020

Persike M, Friedrich JD (2016): Lernen mit digitalen Medien aus Studienperspektive: Sonderauswertung aus dem CHE Hochschulranking für die deutschen Hochschulen. https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_Nr_17_Lernen_mit_digitalen_Medien_aus_Studierendenperspektive.pdf; Abgerufen am 15.04.2020

The Digital Turn – Hochschulbildung im digitalen Zeitalter Hochschulforum Digitalisierung beim Stifterverband für die Deutsche Wissenschaft e.V., Berlin: Hochschulforum Digitalisierung. 2016 <https://hochschulforumdigitalisierung.de/sites/default/files/dateien/Abschlussbericht.pdf>; Abgerufen am 14.04.2020

Danksagung

Ich möchte allen danken, die mich bei der Erstellung dieser Arbeit unterstützt haben.

Der größte Dank geht hierbei an meinen Doktorvater Prof. Dr. med. Tobias Raupach, MME für die Überlassung des Dissertationsthemas, die engagierte Betreuung und zuverlässige Unterstützung während der letzten Jahre. Vielen Dank für die zahlreichen und wertvollen Anregungen bei der Studiendurchführung sowie der schriftlichen Ausarbeitung und die geduldige Beantwortung meiner Fragen.

Allen Mitarbeitern, sowie Doktoranden des Bereichs Medizindidaktik und Ausbildungsforschung der UMG danke ich für ihre Hilfsbereitschaft. Ein besonderer Dank gilt hier Amelie Friederike Büttcher für die Unterstützung bei der Durchführung der Studie und die wertvollen Ratschläge sowie Susanne Borgmann und Dr. med. Angéline-Charline Middeke für die Durchführung der Fokusgruppengespräche. Darüberhinausgehend danke ich Simon Zegota, der mir bei allen Fragen zur statistischen Auswertung weiterhalf. Zusätzlich bedanke ich mich bei Binia-Laureen Grebener für den motivierenden Zuspruch und die Hilfe bei der Verfassung der Arbeit.

Darüber hinaus gilt mein Dank Herrn Prof. Dr. Andreas Schaper als Zweitbetreuer meiner Dissertation für die anregenden Diskussionen während der Fortschritttreffen.

Herzlich danken möchte ich auch den Studierenden der UMG für die Teilnahme an dieser Studie.