

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

**Effektivität von *Key-Feature*-Prüfungen beim
Erwerb der Kompetenz *Clinical Reasoning* in
der medizinischen Ausbildung**

INAUGURAL - DISSERTATION
zur Erlangung des Doktorgrades
der Medizinischen Fakultät der
Georg-August-Universität zu Göttingen

vorgelegt von
Jil Cathérine Andresen

aus
Hannover

Göttingen 2019

Dekan: Prof. Dr. med. W. Brück

Referent/in: Prof. Dr. Tobias Raupach, MME

Ko-Referent/in: PD Dr. Sabine Sennhenn-Kirchner, MME

Drittreferent/in: Prof. Dr. Thomas Meyer

Datum der mündlichen Prüfung: 20.08.2020

Hiermit erkläre ich, die Dissertation mit dem Titel "Effektivität von *Key-Feature-Prüfungen* beim Erwerb der Kompetenz *Clinical Reasoning* in der medizinischen Ausbildung" eigenständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Göttingen, den

.....
(Unterschrift)

Die Ergebnisse dieser Arbeit wurden teilweise veröffentlicht:

Andresen J, Brown J, Meyer K, Anders S, Raupach T: Enhancing acquisition of complex cognitive skills using formative key feature examinations in undergraduate medical education. Vortrag im Rahmen der Jahrestagung der Gesellschaft für Medizinische Ausbildung, Universitätsklinikum Hamburg-Eppendorf, 25.09.2014-27.09.2014

Andresen J, Meyer K, Anders S, Raupach T: Unravelling the testing effect: a qualitative study on the effectiveness of formative key feature examinations as an instructional method. Vortrag im Rahmen der Jahrestagung der Gesellschaft für Medizinische Ausbildung, Universitätsklinikum Hamburg-Eppendorf, 25.09.2014-27.09.2014

Raupach T, Andresen J, Meyer K, Anders S: Improving diagnostic and therapeutic decisionmaking skills by test-enhanced learning in future physicians: A randomised controlled cross-over trial. Flipped Classroom Session im Rahmen der Association for Medical Education in Europe-Konferenz, Glasgow, 05.09.2015-09.09.2015

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

In dieser Arbeit wurde aus Gründen der besseren Lesbarkeit ausschließlich die männliche Form für Personenbezeichnungen gewählt. Sämtliche Personenbezeichnungen gelten gleichwohl für beide Geschlechter.

Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis	V
Abkürzungsverzeichnis	VI
1 Einleitung	1
1.1 Der Nationale Kompetenzbasierte Lernzielkatalog Medizin	2
1.2 <i>Clinical Reasoning</i> (CR)	2
1.3 Der Zyklus nach Kern	4
1.4 Entwicklung einer klinischen Kompetenz anhand der Miller-Pyramide	5
1.5 Lehrformen	7
1.5.1 Vorlesungen	7
1.5.2 Seminare	7
1.5.3 Praktische Lehrformen	8
1.5.4 Problembasiertes Lernen	9
1.6 Fragetypen	10
1.7 Prüfungskonsequenz	13
1.8 Lernen durch Testen (<i>Test-enhanced Learning</i>)	14
1.8.1 Mechanismen des <i>Testing Effects</i>	17
1.8.2 Lernen durch Testen im Medizinstudium	18
1.8.3 <i>Recognition Tests</i> und <i>Production Tests</i>	20
1.9 Lehrmethoden zur Vermittlung von CR	21
1.10 <i>Key Feature</i>	22
1.11 Studienfragen	24
2 Methoden	26
2.1 Studiendesign	26
2.2 Humanmedizinstudium an der Georg-August-Universität Göttingen	30
2.2.1 Probandenrekrutierung	30
2.3 Präsentationsformate	31

2.3.1	Konstruktion der Fälle.....	31
2.3.2	Fragenfall.....	31
2.3.3	Lesefall	32
2.4	Prüfungsrahmen	33
2.5	Praktische Durchführung.....	34
2.6	Fokusgruppengespräche	36
2.7	Auswertung.....	38
2.7.1	Quantitative Datenanalyse.....	38
2.7.2	Qualitative Datenanalyse.....	40
3	Ergebnisse.....	41
3.1	Stichprobe.....	41
3.2	Deskriptive Ergebnisse.....	42
3.2.1	Beschreibung der Stichprobe.....	42
3.2.2	Anwesenheit	42
3.2.3	Gruppenweise Betrachtung der Ergebnisse.....	42
3.2.4	Itemkennwerte und Cronbachs α	46
3.2.5	Fokusgruppen	48
3.3	Beantwortung der Studienfragen	48
3.3.1	Studentischer Lernerfolg in Abhängigkeit vom Präsentationsformat.....	48
3.3.2	Einfluss der Lernzeit auf den primären Endpunkt.....	49
3.3.3	Zeitliche Entwicklung der Testgütekriterien.....	50
3.3.4	Studentisches Antwortverhalten in Abhängigkeit von der Leistung in vorangegangenen E-Fallseminaren.....	51
3.3.5	Ergebnisse der qualitativen Begleitstudie	53
3.3.6	Theorien zu den Mechanismen des <i>Testing Effects</i>	60
4	Diskussion.....	62
4.1	Die wesentlichen Ergebnisse	62
4.2	Studiendesign.....	63
4.3	Gemeinsamkeiten zu vorherigen Studien.....	65

4.4	Computergestütztes Lernen im Medizinstudium	66
4.5	Stärken der Studie	67
4.6	Limitationen der Studie	68
4.7	Fokusgruppengespräche	70
4.7.1	Struktur der Fokusgruppen	71
4.8	Theorien zum <i>Testing Effect</i>	72
4.9	Ausblick und Schlussfolgerung	73
5	Zusammenfassung	75
6	Literaturverzeichnis	77
7	Anhang	85

Abbildungsverzeichnis

Abbildung 1: Zyklus nach Kern	4
Abbildung 2: Miller-Pyramide	6
Abbildung 3: Studiendesign	27
Abbildung 4: Entwicklung der Stichprobe.....	41
Abbildung 5: Durchschnittlich erreichte Punktzahlen im Eingangs-, Ausgangs- und Retentionstest	49
Abbildung 6: Zeitlicher Verlauf der gemittelten Itemschwierigkeiten der Interventions- items.....	51
Abbildung 7: Baumdiagramm der Häufigkeiten richtig bzw. falsch beantworteter Inter- ventionsitems.....	52

Tabellenverzeichnis

Tabelle 1: Theorien zu den zugrundeliegenden Mechanismen des <i>Testing Effects</i>	18
Tabelle 2: Curriculare Lehre mit Inhalten nach Lernzielkatalog, Inhalte der E-Fallseminare mit jeweiligem Präsentationsformat für die Gruppe A bzw. Gruppe B.....	29
Tabelle 3: Interventions- und Kontroll-Items für die Gruppe A.....	35
Tabelle 4: Beschreibung der Stichprobe insgesamt und randomisiert auf die Gruppen A und B	42
Tabelle 5: Prozentsatz der Studierenden mit einer richtigen Antwort in den Interventions-items der Gruppe A im Eingangs-, Ausgangs- und Retentionstest.....	44
Tabelle 6: Prozentsatz der Studierenden mit einer richtigen Antwort in den Interventions-items der Gruppe B im Eingangs-, Ausgangs- und Retentionstest.....	45
Tabelle 7: Mittelwerte der Itemschwierigkeiten und Itemtrennschärfen für die E-Fallseminare und Darstellung des Cronbachs α	47
Tabelle 8: Beschreibung der Fokusgruppen	48
Tabelle 9: Aussagen der Studierenden entsprechend der <i>Elaborate Retrieval Theory</i>	60
Tabelle 10: Aussagen der Studierenden entsprechend der <i>Transfer Appropriate Processing Theory</i>	61
Tabelle 11: Aussagen der Studierenden entsprechend der <i>Unspecific Goal Perspective</i> ..	61

Abkürzungsverzeichnis

ACE	<i>Angiotensin Converting Enzyme</i>
BGA	Blutgasanalyse
DIPS	Digitales Prüfungs- und Schulungszentrum
CR	<i>Clinical Reasoning</i>
E-Fallseminare	elektronische Fallseminare
EKG	Elektrokardiographie
HNO	Hals-Nasen-Ohren-Heilkunde
IMS	<i>Item Management System</i>
L	Lesen
LDAP-Server	<i>Lightweight-Directory-Access-Protocol-Server</i>
M 3.1	Modul „Erkrankungen des Herz-Kreislauf-Systems und der Lunge“
M 3.2	Modul „Erkrankungen der Niere und des Urogenitaltrakts“
M 3.3	Modul „Erkrankungen des Blutes, des Knochenmarks und Grundlagen der Tumorerkrankungen“
M 4.4	Modul „Erkrankungen der Augen, des Hals-Nasen-Ohrenbereichs, des Mundes und der Zähne“
NKLM	Nationaler Kompetenzbasierter Lernzielkatalog Medizin
PIN	Persönliche Identifikationsnummer
QTI-Datei	<i>Question-and-Test-Interoperability-Datei</i>
SSSS	viermal Studieren
SSST	dreimal Studieren, einmal Testen
STÄPS	Studentisches Trainingszentrum Ärztlicher Praxis und Simulation
STTT	einmal Studieren, dreimal Testen
T	Testen
TAN	Transaktionsnummer
UAK	Unterricht am Krankenbett
V. a.	Verdacht auf
W	Woche

1 Einleitung

Im Folgenden wird vorerst auf die Ziele der ärztlichen Ausbildung anhand der Approbationsordnung eingegangen und auf Kompetenzen, die ein Absolvent des Medizinstudiums erworben haben sollte, um mit der ärztlichen Weiterbildung fortzufahren. Insbesondere wird die Kompetenz *Clinical Reasoning* (CR) erläutert. Danach wird auf die systematische Vorgehensweise der Planung eines Curriculums eingegangen. Unterschiedliche Wissensniveaus und Lehrformen werden vorgestellt, bevor sich dem Themenkomplex „Prüfen“ gewidmet wird. Zunächst werden unterschiedliche Fragetypen erörtert und über Prüfungskonsequenzen berichtet. Anschließend werden der *Testing Effect* und dessen zugrundeliegende Mechanismen geschildert. Danach geht es um die Themenbereiche Lernen durch Testen im Medizinstudium, *Production Tests* und *Recognition Tests* und um Lehrmethoden zur Vermittlung von *Clinical Reasoning*. Schließlich wird der Begriff *Key Feature* vorgestellt und anschließend werden die Studienfragen dieser Arbeit präsentiert.

Die Ziele der ärztlichen Ausbildung sind in der Approbationsordnung für Ärzte niedergeschrieben. Ein Ziel, auf welches sich der Fokus dieser Arbeit besonders richtet, ist es „die für das ärztliche Handeln erforderlichen allgemeinen Kenntnisse, Fähigkeiten und Fertigkeiten in Diagnostik, Therapie, Gesundheitsförderung, Prävention und Rehabilitation [...] auf der Basis des aktuellen Forschungsstandes zu vermitteln“ (Bundesministerium für Gesundheit 2002: Approbationsordnung für Ärzte, S.1). Fachkompetenz und Handlungskompetenz sind hierbei neben Empathie und Patientenorientierung entscheidend, um eine optimale Patientenversorgung zu gewährleisten (Herzig et al. 2006). Zielsetzung dieser Arbeit ist es, auf der Grundlage aktueller Erkenntnisse aus der lernpsychologischen Forschung das Lernen von Kompetenzen für bestmögliche ärztliche Entscheidungsfindungen im Patienteninteresse durch ein innovatives Lehrformat für zukünftige Ärztinnen und Ärzte zu unterstützen.

In diesem Rahmen sollten zunächst einige Hintergrundinformationen zu Kompetenzen im Medizinstudium anhand des Nationalen Kompetenzbasierten Lernzielkatalogs Medizin gegeben werden.

1.1 Der Nationale Kompetenzbasierte Lernzielkatalog Medizin

Der Nationale Kompetenzbasierte Lernzielkatalog Medizin (NKLM) orientiert sich an der Approbationsordnung für Ärzte und beschreibt ein Absolventenprofil von Ärzten im Sinne eines Kerncurriculums für Medizin. Der NKLM legt besonderen Wert auf Kompetenzen, die ein Absolvent des Medizinstudiums haben sollte, um mit der ärztlichen Weiterbildung fortzufahren (MFT 2015). Kompetenzen verstehen sich im NKLM als „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 Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert 2001, S. 27-28). Als Beispiele für Kompetenzen werden im NKLM „eine situationsgerechte Krankheitsanamnese durchführen“ oder „die Prinzipien einer pharmakologischen Behandlung der Herzinsuffizienz erklären und eine geeignete Arzneitherapie vornehmen“ genannt (MFT 2015, S. 164, S. 214). Die Kompetenzen des NKLM wurden an das kanadische CanMEDS-Rahmenkonzept (Frank 2005) angelehnt, welches ein Grundgerüst von ärztlichen Kompetenzen beinhaltet und international eine große Anwendung in der medizinischen Ausbildung findet. Einerseits sollen die Kompetenzen dem ärztlichen Berufsalltag und andererseits dem gesellschaftlichen Anspruch gerecht werden. Insgesamt soll der NKLM den medizinischen Fakultäten als Orientierung bei der Planung ihrer Curricula dienen (MFT 2015). Die vorliegende Arbeit behandelt insbesondere die Kompetenz *Clinical Reasoning*.

1.2 *Clinical Reasoning* (CR)

CR ist eine essentielle Kompetenz in der medizinischen Ausbildung (Kunina-Habenicht et al. 2015). CR beschreibt das Treffen von fundierten Entscheidungen eines Arztes anhand von Wissen, Intuition, Erfahrung und Leitlinien über das diagnostische und therapeutische Verfahren auf Basis der Ausgangssituation und der Symptome eines Patienten unter Berücksichtigung von möglichen Differentialdiagnosen. In der Literatur werden aktuell zwei kognitive Ansätze beschrieben, die dem CR zugrunde liegen, namentlich der intuitive und der analytische Ansatz. In der Realität können beide Ansätze ineinander übergehen und voneinander abhängig sein (Croskerry 2009; Dawson 1993; Kassirer 2010; Stanovich 2004). Der intuitive Ansatz beschreibt den ersten Eindruck, den ein Arzt instinktiv ohne analytische Denkmuster erhält. Dieser kann durch das Erkennen sich wiederholender Situationen und durch Erfahrung verstärkt werden und scheint ohne große kognitive Anstrengung zu entstehen (Croskerry 2009; Dawson 1993; Kassirer 2010; Stanovich 2004).

Faktoren wie das äußere Erscheinungsbild des Patienten (Kleidung, Hygiene, Verhalten), die momentane Atmosphäre auf der Station (Anzahl der Patienten, Ressourcenknappheit, Dringlichkeit der Behandlung anderer Patienten) und die Ernsthaftigkeit der vermuteten Erkrankung haben Einfluss auf intuitive Entscheidungen (Croskerry 2009).

Der analytische Ansatz beschreibt Gedankenprozesse, die auf fundiertem Wissen, Studien, Logik und Kausalität basieren (Stanovich 2004). Differentialdiagnosen und Optionen werden berücksichtigt und mit dem Ziel verfolgt, die klinische Geschichte des Patienten logisch zu begründen (Kassirer 2010). Der analytische Ansatz findet insbesondere Verwendung, wenn ungewöhnliche Befunde erhoben werden oder Befunde auftreten, die nicht zu einer einfach erklärbaren Ursache passen (Croskerry 2009; Kassirer 2010; Mamede et al. 2007). In der Praxis kann sich ein analytischer Ansatz durch Wiederholung automatisieren und sich mit der Zeit zu einem intuitiven Ansatz verändern (Hogarth 2001). Das kann heißen, dass mit der Berufserfahrung Leitlinien verinnerlicht werden und instinktiv befolgt werden. Es kann daraus geschlossen werden, dass Berufseinsteiger durchschnittlich häufiger mit Hilfe des analytischen Ansatzes Entscheidungen treffen und Experten häufiger auf den intuitiven Ansatz zurückgreifen (Croskerry 2009). Weiterhin wird vermutet, dass „leichte Patientenfälle“ mit dem intuitiven Ansatz gelöst werden, während bei schwierigen Patientenfällen der analytische Ansatz Anwendung findet (Elstein und Schwarz 2002). Nachteile des intuitiven Ansatzes können sein, dass der erste Eindruck rasch entsteht, von Emotionen und Affekt geprägt sein kann (Vohs et al. 2007) und im Vergleich zum analytischen Ansatz mehr Fehler auftreten können (Kassirer 2010; Stanovich 2004). Fehler können beispielsweise entstehen, wenn der Patient atypische Symptome aufweist (Croskerry 2009) und daraus intuitiv falsche Schlussfolgerungen gezogen werden. Während der intuitive Ansatz für mehrere Merkmale einer Situation parallel empfänglich ist, verfolgt der analytische Ansatz einen Schritt nach dem anderen (Stanovich 2004). Ein Nachteil des analytischen Ansatzes ist, dass mehr Zeit (Dawson 1993) und Ressourcen beansprucht werden (Croskerry 2009). Insgesamt ist eine Kombination beider Ansätze realitätsgetreu und wichtig für CR. Eindeutige erste Eindrücke sind häufig richtig, aber das analytische Überdenken dieser ist sinnvoll, da es das Entstehen von Fehlern verhindern kann (Dawson 1993; Kassirer 2010).

Um bei einem Studienabschluss zu gewährleisten, dass die Studierenden über gewünschte Kompetenzen wie CR verfügen, ist eine zielgerichtete Curriculum-Planung erforderlich. Hierbei kann gemäß dem Zyklus nach Kern vorgegangen werden, der im Folgenden näher dargelegt wird.

1.3 Der Zyklus nach Kern

Kern et al. entwickelten 1998 eine systematische Vorgehensweise mit sechs Schritten für die Entwicklung eines Curriculums. Diese wird anhand der folgenden Abbildung 1 erklärt.

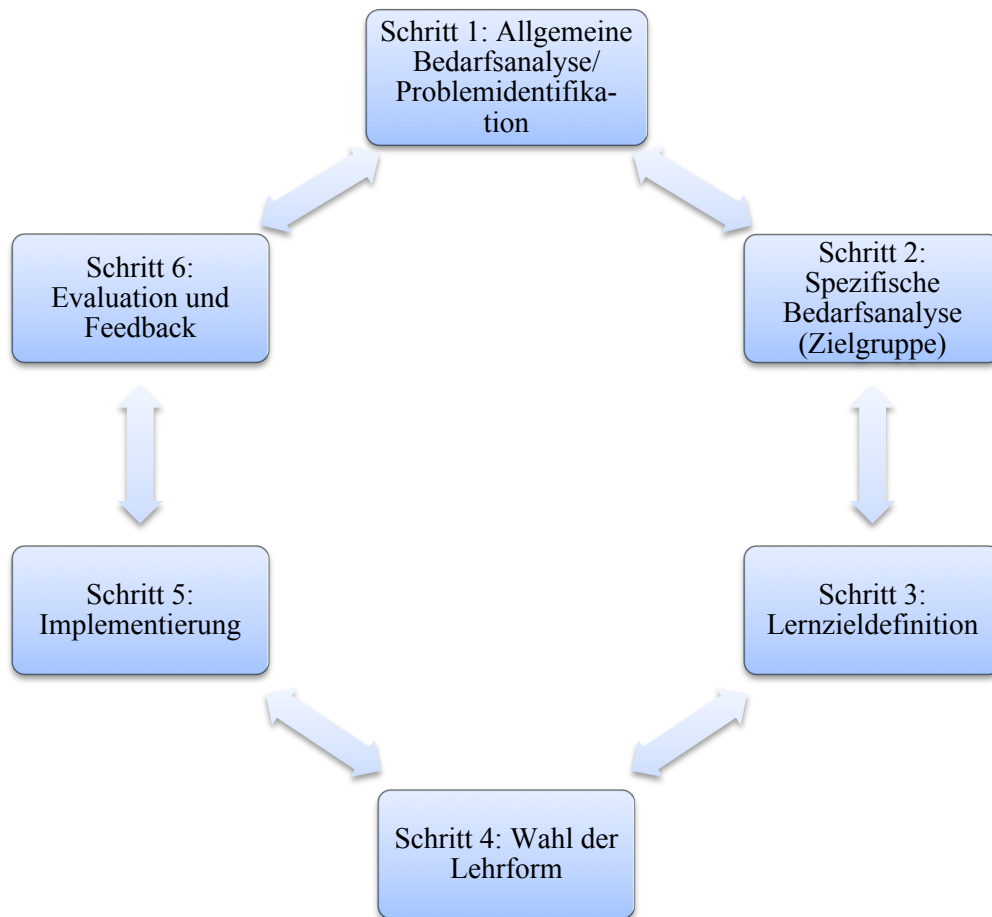


Abbildung 1: Zyklus nach Kern, modifiziert nach Kern et al. 1998

Zuerst müssen eine allgemeine Bedarfsanalyse und eine Problemidentifikation erfolgen (Schritt 1). Danach sollte eine spezifische Bedarfsanalyse für die Zielgruppe (die Studierenden) durchgeführt werden (Schritt 2). Dann werden Lernziele definiert (Schritt 3), die sich an dem zuvor identifizierten Bedarf orientieren und in einem nächsten Schritt entsprechende Lehrformen bzw. Unterrichtsmethoden gewählt (Schritt 4). Mit „Implementierung“ (Schritt 5) ist die praktische Umsetzung der Lehrformen wie beispielsweise die Stundenplanung, die Raumplanung, die Dozenteneinteilung oder die Vorbereitung der Ressourcen gemeint. Zuletzt (Schritt 6) sollten sowohl die Studierenden als auch das Curriculum geprüft werden. Durch Evaluation und Feedback kann ein bereits bestehendes Curriculum optimiert und es können Vergleiche zu anderen Curricula gezogen werden. Die Studierenden erhalten durch Prüfungen eine Rückmeldung. Aus der Abbildung 1 ist

ersichtlich, dass der Zyklus nicht endet, sondern sich weiterentwickelt. Das heißt, dass nach der Evaluation des Curriculums der Zyklus wieder mit einer neuen Problemidentifikation beginnen kann. Es ist außerdem erwähnenswert, dass die unterschiedlichen Schritte des Zyklus in der Praxis nicht unbedingt nacheinander folgen, sondern durchaus parallel stattfinden können und auch aufeinander Einfluss ausüben.

Lernziele (Schritt 3) werden nach Kern et al. (1998) in kognitive, affektive und psychomotorische Lernziele unterteilt. Kognitive Lernziele betreffen das Wissen, affektive Lernziele illustrieren u.a. professionelle Einstellungen, und psychomotorische Lernziele beziehen sich auf praktische Fertigkeiten. Lernziele beschreiben Kompetenzen, die im Verlauf eines Lernprozesses erworben werden sollen. In diesem Zusammenhang ist entscheidend, dass Lernziele und Lehrformen kongruent sein sollten. Kongruenz bedeutet hier, dass die Lehrform zum Erreichen des Lernziels geeignet ist. (Kern et al. 1998) Um eine passende Lehrform für das Erreichen eines Lernziels zu wählen, sollte bedacht werden, welches Wissensniveau (Wissen, Verstehen, Zeigen, Tun) einer Kompetenz erreicht werden soll. Bevor also in dieser Arbeit auf die Thematik „Lehrformen“ eingegangen wird, werden im nächsten Kapitel vorerst unterschiedliche Wissensniveaus anhand der Miller-Pyramide beleuchtet.

1.4 Entwicklung einer klinischen Kompetenz anhand der Miller-Pyramide

Lernziele können unterschiedliche Wissensniveaus abdecken. Nach Blooms Taxonomie kann das Wissen in vier Wissensniveaus (Faktenwissen, konzeptionelles Wissen, prozedurales Wissen und metakognitives Wissen) unterteilt werden (Krathwohl 2002). Faktenwissen bezeichnet den Bestand an theoretischen Fakten, Elementen und Details eines bestimmten Themengebietes. Das Wissen über die Zusammenhänge der Fakten, Klassifikationen oder Theorien wird unter dem konzeptionellen Wissen zusammengefasst. Das metakognitive Wissen bezieht sich unter anderem auf die eigenen Erkenntnisse (Krathwohl 2002). In dieser Arbeit soll sich insbesondere auf das prozedurale Wissen im Medizinstudium fokussiert werden. Prozedurales Wissen heißt, zu wissen, „wie etwas gemacht wird“ (Krathwohl 2002). Das heißt, über das Wissen von praktischen Abläufen, Untersuchungsmethoden und Kriterien zu verfügen, um zu bestimmen, wann welche Prozeduren durchgeführt werden (Krathwohl 2002). Konkret am Patientenfall kann das bedeuten, dass ein Studierender weiß, wie über angemessene Diagnostik entschieden, Befunde interpretiert und eine entsprechende Therapie durchgeführt werden sollten. Das prozedurale

Wissen lässt sich der Miller-Pyramide zuordnen (Kopp et al. 2006), auf welche im Folgenden eingegangen wird.

Anhand der Miller-Pyramide (Abbildung 2) (Miller 1990) lässt sich veranschaulichen, wie sich eine klinische Kompetenz von Medizinstudierenden und Ärzten entwickeln kann. Dies soll mit einem einfachen Beispiel – dem Legen einer Venenverweilkanüle – geschehen.

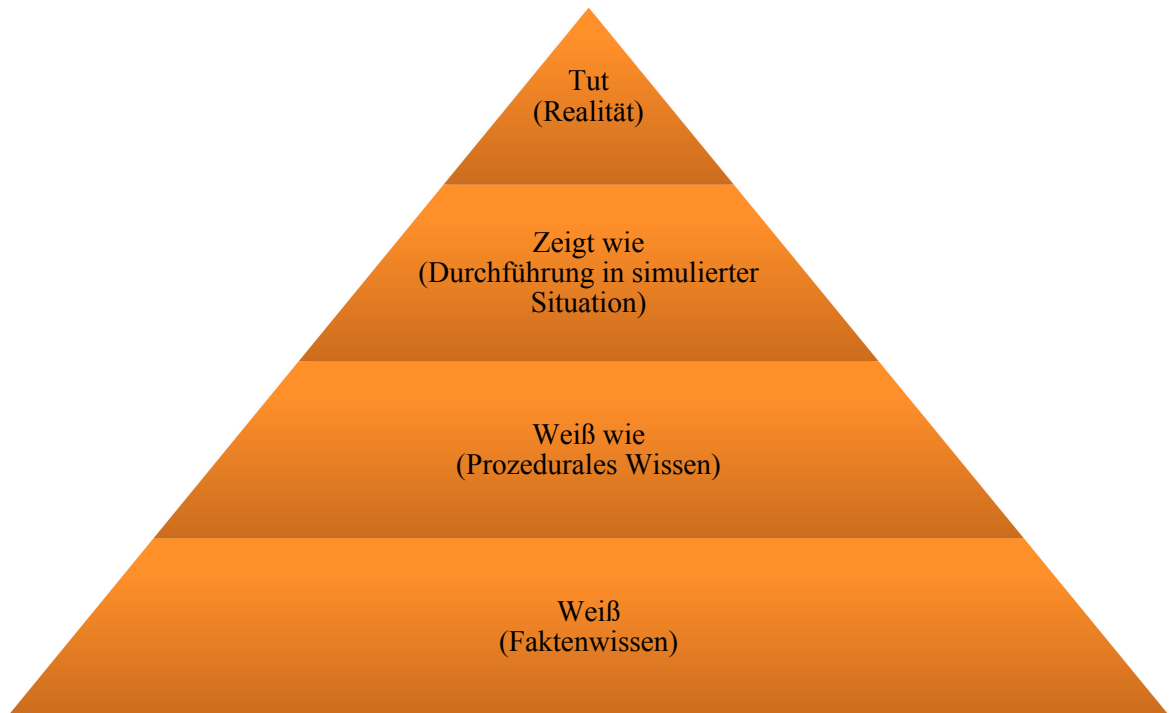


Abbildung 2: Miller-Pyramide, modifiziert nach Miller 1990

Die Basis der Pyramide bildet das Faktenwissen, welches ein Studierender haben muss, um eine bestimmte klinische Kompetenz entwickeln zu können (Beispiel: Indikation, Risiken und benötigte Utensilien für eine Venenverweilkanüle). Sobald der Studierende das Faktenwissen verinnerlicht hat, sollte er wissen, wie er es richtig anwenden kann. Der Studierende sollte also in der Lage sein, über das prozedurale Wissen zu verfügen (Beispiel: genauer Handlungsablauf des Legens einer Venenverweilkanüle). Das nächsthöhere Wissensniveau in der Miller-Pyramide beinhaltet, dass der Studierende zeigt, wie er die Tätigkeit in einer simulierten Situation durchführt (Beispiel: Legen einer Venenverweilkanüle an einer Puppe). Die Spitze der Miller-Pyramide illustriert, was der Studierende in einer realen Situation tut (Beispiel: Legen einer Venenverweilkanüle am Patienten). Es stellt sich nun die Frage, in welchem Rahmen Kompetenzen im Medizinstudium erworben werden können. Es unterscheiden sich diesbezüglich Lehrformen in ihrer Eignung.

1.5 Lehrformen

Es gibt verschiedene Lehrformen, von denen einige im Folgenden erörtert werden. Zu den im Medizinstudium häufig eingesetzten Lehrformaten gehören Vorlesungen, Seminare und der Unterricht am Krankenbett (UAK). Es kommen spezielle Lehrformen wie Famulaturen, Blockpraktika und das Praktische Jahr zum Einsatz. An einigen Fakultäten wird das problemorientierte Lernen als spezifische Lehrform eingesetzt; zudem werden zunehmend elektronische Lehrformate genutzt, die unter dem Oberbegriff „E-Learning“ firmieren. Eine der ältesten Lehrformen ist die Vorlesung.

1.5.1 Vorlesungen

Vorlesungen sind eine effektive Lehrform, um insbesondere großen Gruppen Lehrinhalte zu vermitteln. Gut geeignet sind Vorlesungen, um Kernkenntnisse zu vermitteln, Prinzipien zu erläutern, Interessen zu stimulieren und das Lernen der Studierenden auf Themen zu konzentrieren. Ein Nachteil ist, dass Studierende die erhaltenen Informationen nicht aktiv verarbeiten, sondern nur als passives Wissen wahrnehmen (Cantillon 2003). Außerdem ist der Bezug zur klinischen Praxis gering und Studierende erhalten wenig Feedback (Gordon 2003). Vorlesungen sind also nicht besonders gut dazu geeignet, um Fertigkeiten zu lehren, Einstellungen von Studierenden zu verändern oder das Nachdenken im klinischen Zusammenhang zu fördern (Cantillon 2003).

1.5.2 Seminare

Seminare umfassen nur eine kleine Gruppe von Studierenden und nicht das ganze Semester. Bei einer guten Umsetzung des Seminars können Studierende in einem engeren Kontakt mit dem Lehrenden stehen als bei Vorlesungen und können die Inhalte in Fachsprache diskutieren. So können Wissenslücken aufgedeckt werden und Studierende ihr Lernverhalten anpassen. Ausschlaggebend hierfür ist eine aktive Teilnahme, welche nicht immer gegeben ist. Das kann mehrere Gründe haben. Es kann einerseits an der Raumaufteilung liegen. Häufig sitzen Studierende in hintereinander aufgestellten Stuhlreihen. In diesem Fall haben die Studierenden untereinander keinen Augenkontakt, was zu weniger Beteiligung in Diskussionen führt. Andererseits kann eine schlechte Teilnahme durch das Verhalten des Lehrenden begründet sein. Viele Lehrende nehmen zu schnell die Rolle einer Autoritätsperson oder eines Experten ein und übernehmen so den hauptsächlichsten Redeanteil (Jaques 2003). Anstatt eines Dialogs unter den Studierenden entwickelt sich eher ein Vorlesungscharakter. Der dritte Grund für eine ungenügende Teilnahme sind die Studierenden selbst.

Es kann vorkommen, dass die Studierenden nicht miteinander diskutieren möchten, sondern nur die Fragen des Lehrenden beantworten und daraufhin ohne zu diskutieren die Lösung erhalten möchten. Außerdem kann ein Studierender das Gespräch extrem dominieren und dadurch eine Diskussion unterbinden (Jaques 2003). Im Gegensatz dazu kann es auch schüchterne Studierende geben, die sich nicht trauen an der Diskussion teilzunehmen.

1.5.3 Praktische Lehrformen

Im Medizinstudium wird versucht, möglichst viel klinische Anwendung der theoretischen Inhalte zu gewährleisten. Rahmenbedingungen dafür sind vor allen Dingen Untersuchungskurse, Blockpraktika, Famulaturen oder das Praktische Jahr. Studierende lernen hier den Stationsalltag, die Ambulanzen und die verschiedenen Krankheitsbilder anhand von Patienten kennen. Untersuchungskurse finden in Kleingruppen statt, während in Famulaturen und im Praktischen Jahr der Medizinstudierende meistens allein mit den Ärzten auf der Station oder in der Ambulanz arbeitet und folglich allein unterrichtet wird. Ein Vorteil besteht darin, dass die Lehre nicht auf eine Gruppe, sondern individuell auf das Wissen des Studierenden abgestimmt werden kann. Feedback kann durch den Lehrenden in regelmäßigen Abständen persönlich gegeben werden (Gordon 2003). Nicht nur während Famulaturen oder im Praktischen Jahr, sondern auch in Untersuchungskursen wird durch die aktive Teilnahme und die Relevanz für den eigenen Beruf die Motivation der Studierenden gesteigert. Die Lehrenden gelten in diesem Zusammenhang als Vorbilder hinsichtlich des professionellen Denkens und Handelns. Ein Vorteil dieser Lehrform besteht darin, dass Fertigkeiten wie das Erheben einer Anamnese, die körperliche Untersuchung, CR, Entscheidungsfindung, Empathie und Professionalität in der Praxis gelehrt und gelernt werden können. Es gibt einige Nachteile, worunter die Lehre unter diesen Bedingungen leiden kann. Beispielsweise können ein hoher zeitlicher Druck und oftmals eine Unterbesetzung des medizinischen Personals auf den Stationen bestehen. Außerdem sind trotz ausreichenden Fachwissens nicht alle Lehrenden in der medizinischen Didaktik gut genug geschult, um Studierende effektiv zu lehren (Spencer 2003). Das hat zur Folge, dass Studierende keine ausreichende Supervision erhalten und der positive und motivierende Effekt einer Supervision ausbleibt (Wimmers et al. 2006). Weiterhin besteht das Problem, dass der Umfang des Lerninhalts bei unterschiedlichen Studierenden nicht vergleichbar ist, da das Spektrum der behandelten Inhalte in der Umsetzung je nach Krankenhaus und Station variiert und von vielen Faktoren abhängt. Gründe dafür sind die unterschiedliche Aufenthaltsdauer von Patienten, die Anzahl der Neuaufnahmen (Wimmers et al. 2006) und teilweise

die Spezialisierung von Stationen, sodass häufig die Vielfalt an Patienten für Studierende begrenzt ist. Demzufolge ist diese Art der Lehrform nicht standardisiert, da jedem Studierenden ein anderer Umfang an Lerninhalten vermittelt wird.

1.5.4 Problembasiertes Lernen

Die nächste Lehrform, die beschrieben werden soll, ist das problembasierte Lernen. Problembasiertes Lernen wurde seit Mitte der 1960er Jahre zunehmend in das Curriculum medizinischer Universitäten integriert (Norman und Schmidt 1992). Die Terminologie des problembasierten Lernens wurde bisher mit gleichbedeutenden Bezeichnungen wie beispielsweise problemorientiertes Lernen oder fallbasiertes Lernen uneinheitlich verwendet (Reusser 2005). Die Idee hinter dieser Lehrform ist ein selbstbestimmtes, konstruktives und kontextbezogenes Lernen im Team zu generieren (Dolmans et al. 2005). In der medizinischen Ausbildung kann problembasiertes Lernen wie folgt aussehen. Einer kleinen Gruppe von fünf bis neun Studierenden (Barrows 1996) wird ein Patientenfall mit allgemeinen Informationen (Alter, Geschlecht etc.) und dem entsprechenden Problem (Symptome, Vorerkrankungen etc.) präsentiert. Es soll sich im Idealfall um Probleme handeln, die Studierende mit ihrem Wissen nicht tiefgehend erklären und lösen können. Während der Diskussion über die Problemlösung treten Fragen bzw. Hürden auf, die als eigene Lernziele für Studierende fungieren und anschließend erreicht werden können. Zudem können diagnostische und therapeutische Lösungswege vorgeschlagen und besprochen werden (Norman und Schmidt 1992). Die Gruppe lenkt sich dabei selbst. Idealerweise wird die Gruppe von Studierenden von einem Tutor moderiert, der sowohl Experte im diskutierten Fachgebiet ist, als auch weiß, wie studentisches Lernen vereinfacht werden kann (Dolmans et al. 2002). Vorrangige Ziele des problembasierten Lernens sind Fähigkeiten für Problemlösungen und Teamarbeit zu entwickeln (Barrows 1996). Zusätzlich soll der Erwerb, das Behalten und das Anwenden von Wissen gefördert werden (Norman und Schmidt 1992). Auch beim problembasierten Lernen besteht das Risiko, dass, wie weiter oben beschrieben, einige Studierende die Diskussion dominieren können und dadurch die Lerngeschwindigkeit einzelner Studierender unberücksichtigt bleibt. Somit profitieren nicht alle Studierende gleich von dieser Lehrmethode (Jaques 2003). Um problembasiertes Lernen zu standardisieren und den gleichen Inhalt auf eine einheitliche Art und Weise an alle Studierenden zu vermitteln, kann computergestütztes problembasiertes Lernen eine Alternative darstellen (Wahlgren et al. 2006). In Bezug auf den Erwerb von Faktenwissen, führte zusätzliches computergestütztes problembasiertes Lernen im Curriculum zu signifi-

kant besseren Prüfungsergebnissen. (Raupach et al. 2010) In Bezug auf den Erwerb von CR zeigten Raupach et al. (2009), dass computergestütztes problembasiertes Lernen gleich effektiv sein kann wie traditionelles problembasiertes Lernen. Jedoch war die Akzeptanz des traditionellen problembasierten Lernens unter den Studienteilnehmern höher. Es ist zu untersuchen, welche Lehrform sich besonders für computerbasiertes Lernen im Medizinstudium eignet. (Raupach et al. 2009)

Nachdem ein Überblick über einige Lehrformen gegeben wurde, wird zunächst beschrieben, welche Fragetypen für das Prüfen gelehrter Inhalte im Medizinstudium zum Einsatz kommen. Hierfür werden im Folgenden die Fragetypen erläutert.

1.6 Fragetypen

Insgesamt gibt es viele verschiedene Fragetypen, die in schriftlichen Prüfungen des Medizinstudiums verwendet werden. An dieser Stelle soll aber nur auf Einfachwahl, PickN, Kprim, Kurzantwortfragen, *Long-Essay*-Fragen und *Long-Menu* eingegangen werden. Alle der genannten Fragetypen finden im Medizinstudium an der Georg-August-Universität Verwendung.

Beim Fragetyp Einfachwahl, auch als Typ A bezeichnet, wird eine Frage gestellt oder eine unvollständige Aussage getätigt, auf die fünf auswählbare Antworten oder Ergänzungen folgen. Aus diesen fünf Antworten/Ergänzungen muss vom Prüfling entweder die richtige oder am besten zutreffende Antwort/Ergänzung (Typ A_{pos}) oder die falsche oder am wenigsten zutreffende Antwort/Ergänzung (Typ A_{neg}) ausgewählt werden (Krebs 2004). Die negative Einfachwahl (Typ A_{neg}) sollte in Prüfungen nur wenig Anwendung finden. Das zweite Staatsexamen des Medizinstudiums in Deutschland besteht ausschließlich aus dem Fragetyp Einfachwahl (Brüstle 2011). Häufig wird in diesem Rahmen die Bezeichnung „Einfachwahl“ missverständlich als Multiple-Choice verwendet. Unter dem Begriff „Multiple-Choice“ werden nämlich unterschiedliche standardisierte schriftliche Fragetypen, wie beispielsweise die in dieser Arbeit erwähnten Fragetypen Typ A_{pos} , Typ A_{neg} , PickN und Kprim, zusammengefasst. Das Prinzip von Multiple-Choice ist, dass eine oder mehrere Antworten auf eine Frage auszuwählen sind oder als falsch bzw. richtig bewertet werden müssen. (Krebs 2004) Ähnlich wie der Fragetyp Einfachwahl ist der Fragetyp PickN konstruiert. PickN besteht aus einer Frage und aus mehr als fünf und bis zu 26 Antwortmöglichkeiten. Aus den Antwortmöglichkeiten sollen von den Prüflingen mehrere (zwei bis fünf) ausgewählt werden (Brüstle 2011; Krebs 2004). Die Anzahl an auszuwählenden Antworten wird bei diesem Fragetyp angegeben (Brüstle 2011). Beim Fragetyp Kprim

folgen einer Frage oder einem unvollständigen Satz vier Antwortmöglichkeiten bzw. Ergänzungen. Jede der vier Antwortmöglichkeiten/Ergänzungen muss vom Prüfling als richtig oder falsch bewertet werden. In der Fragestellung wird nicht erwähnt, wie viele Richtig- oder Falschantworten zu erwarten sind (Krebs 2004). Eine Limitation dieses Fragetyps ist, dass, wenn ein Prüfling beispielsweise eine falsche Antwort als falsch bewertet, zwar ersichtlich ist, dass der Studierende weiß, dass die Antwort falsch ist, jedoch nicht erkennbar ist, ob der Prüfling auch die richtige Antwort kennt (Schuwirth und van der Vleuten 2003). Vorteile der Multiple-Choice-Fragetypen wie Einfachwahl, PickN und Kprim sind, dass die Korrektur von großen Prüfungsgruppen durch Maschinen ökonomisch und ressourcenarm und ohne subjektiven Einfluss erfolgen kann. Außerdem besteht durch die geringe Bearbeitungszeit die Möglichkeit eine hohe Anzahl von Fragen pro Prüfungsstunde zu stellen (Bloch et al. 1999). Dadurch kann eine inhaltlich breite Abdeckung eines Fachbereiches erfolgen. Ein Kritikpunkt beim Stellen von Multiple-Choice-Fragetypen wie Einfachwahl, PickN und Kprim ist das ungewollte Hinweisen auf richtige bzw. falsche Antworten durch die vorgegebenen Antwortmöglichkeiten (Auftreten des sog. *Cueing Effects*) und die daraus resultierende Verzerrung von Leistungen (Schuwirth et al. 1996a). Dementsprechend ist die exakte Formulierung von Antwortmöglichkeiten ohne Lösungshinweise aufwendig, aber dennoch notwendig (Bloch et al. 1999). Ein gemeinsamer Nachteil der Fragetypen Einfachwahl, PickN und Kprim ist, dass alle drei Fragetypen zum schlichten Abfragen von Faktenwissen verleiten (Krebs 2004). Nur begrenzt kann geprüft werden, ob der Prüfling über die Kompetenz verfügt, klinisch zu denken und medizinische Probleme zu lösen (Bloch et al. 1999). Bloom's überarbeitete Taxonomie klassifiziert kognitive Fähigkeiten in sechs hierarchisch angeordnete Stufen: Erinnern, Verstehen, Anwenden, Analysieren, Evaluieren und Erschaffen (Krathwohl 2002). Mit Fragetypen kann auf den unterschiedlichen Taxonomiestufen geprüft werden (Schurter 2009a). Da mit den Fragetypen Einfachwahl, PickN und Kprim häufig lediglich Faktenwissen abgefragt wird, prüfen diese drei Fragetypen zumeist auf einer niedrigen taxonomischen Stufe. Der nächste Fragetyp, welcher in diesem Rahmen vorgestellt werden soll, ist die Kurzantwortfrage. Die Kurzantwortfrage ist ein Fragetyp, der mit einzelnen kurzen Begriffen oder wenigen ganzen Sätzen beantwortet wird, die vom Prüfling aktiv formuliert werden müssen (Schurter 2009a). Mit diesem Fragetyp kann geprüft werden, wie Prüflinge Inhalte verstehen, interpretieren und Faktenwissen anwenden können. Problemlösungsfähigkeiten können mit Kurzantwortfragen besser geprüft werden als mit Einfachwahl oder PickN-Fragen (Bloch et al. 1999). Damit können Kurzantwortfragen auf einer höheren taxonomi-

schen Stufe als die Fragetypen Einfachwahl, PickN und Kprim prüfen. Ein nicht zu unterschätzender Nachteil ist die aufwendige Korrektur von Kurzantwortfragen, da meist händisch korrigiert wird und durch die selbst formulierten Antworten nicht immer eindeutig ist, wie viele Punkte zu vergeben sind. Folglich besteht das Risiko, dass die Objektivität, die Validität und die Reliabilität der Prüfung darunter leiden. Zur Reduktion dieses Risikos soll ein eindeutiger, auf Lernziele ausgerichteter Lösungsschlüssel verwendet werden und die Korrektur durch mehrere Personen erfolgen. Außerdem soll „quer korrigiert“ werden und nach Korrektur der ersten Frage aller Prüflinge sollen die Klausurbögen durchmischt werden, damit für die nächste Frage in anderer Reihenfolge korrigiert wird und zu strenge bzw. zu milde Bewertungen vermindert werden oder sich auf die Prüflinge verteilen (Schurter 2009b). Diese aufwendige Art zu korrigieren kann sehr zeit- und ressourcenintensiv sein.

Eine *Long-Essay*-Frage wird als offene Fragen gestellt und bedingt eine Antwort, die als längerer Text formuliert wird. In Textform werden Zusammenhänge erläutert oder Begründungen von Entscheidungen dargestellt und das dargelegte Wissen in diesem Rahmen bewertet. Es wird also kein Faktenwissen, sondern eher ein Zusammenhangswissen auf einer höheren taxonomischen Stufe geprüft. Durch die im Verhältnis zu Fragetypen wie Einfachwahl oder PickN-Fragen lange Bearbeitungszeit einer *Long-Essay*-Frage können nur wenige Fragen pro Zeiteinheit gestellt werden. Somit ist aufgrund der geringen Anzahl an Fragen pro Zeiteinheit der Arbeitsaufwand für die Fragenerstellung im Vergleich zu anderen Fragetypen geringer, die benötigte Zeit für Korrekturen dafür umso höher. Aus denselben Gründen wie bei den Kurzantwort-Fragen ist die Korrektur der Antworten von *Long-Essay*-Fragen kritisch. Die „quere“ Bewertung von unterschiedlichen Personen anhand von Bewertungslisten wird empfohlen (Schulze und Drolshagen 2006). Da die Korrektur von *Long-Essay*-Fragen einen zeitlich hohen und personell großen Aufwand bedeutet, sollte dieser Fragetyp nur genutzt werden, wenn Multiple-Choice-Fragetypen oder Kurzantworten nicht angemessen sind (Schuwirth und van der Vleuten 2003).

Der computergestützte Fragetyp *Long-Menu* wurde entwickelt, um den zuvor beschriebenen *Cueing Effect* zu verhindern und Bewertungen computergestützt zu vereinfachen. Auf eine Frage folgt ein leeres Texteingabefeld, in das der Prüfling die ersten Buchstaben seiner Lösung eintippen muss. Der Computer vergleicht die eingetippten Buchstaben mit einer alphabetischen Liste von mehr als 2500 Wörtern (Schuwirth et al. 1996b). Alle Alternativen, die diese Buchstabenfolge enthalten, werden dem Prüfling angezeigt. Aus diesen Alternativen kann der Prüfling seine Antwort auswählen und bestätigen. Die *Long-*

Menu-Liste enthält sowohl richtige Antworten inklusive möglicher Synonyme als auch Distraktoren (Rotthoff et al. 2006). Das Erstellen einer solchen vollständigen *Long-Menu*-Liste mit Antworten, Synonymen und Distraktoren stellt für den Autor eine Herausforderung dar (Fischer et al. 2005; Rotthoff et al. 2006). Bei einer vollständigen Antwortliste stellt die *Long-Menu*-Frage einen äquivalenten Fragetyp zum offenen Fragetyp dar (Schuwirth et al. 1996b).

Nachdem nun ein Überblick über einige Lehrformen und Fragetypen und deren Nutzen gegeben wurde, ist in diesem Zusammenhang entscheidend, dass Prüfungen und deren Konsequenz einen großen Einfluss auf das Lernen ausüben können.

1.7 Prüfungskonsequenz

Vorerst soll das häufig zitierte Axiom ‚*assessment drives learning*‘ vorgestellt werden, welches beschreibt, dass Prüfungen einen Einfluss auf das Lernen haben können. Newble und Jaeger beobachteten, dass die alleinige Umstellung eines Prüfungsformats die Lernaktivitäten von Medizinstudierenden veränderte (Newble und Jaeger 1983). Mit der Intention, Medizinstudierende besser auf ihren Berufsalltag vorzubereiten, wurden damals curriculare Änderungen (mehr praxisnahe Lehre auf Station anstatt der Lehre theoretischer Inhalte) im Medizinstudium an der Universität in Adelaide vorgenommen bei zunächst gleichbleibender Abschlussprüfung mit Multiple-Choice-Fragetypen. Die Studierenden verbrachten trotzdem weiterhin mehr Zeit mit dem Lernen von theoretischen Inhalten als mit dem Lernen klinischer Fertigkeiten, da für sie das Bestehen der Abschlussprüfung mit Multiple-Choice-Fragetypen von zentraler Bedeutung war. Nach Einführung eines klinischen Prüfungsformats für die Abschlussprüfung, gewann das Lernen von praktischen Fertigkeiten für die Medizinstudierenden stark an Bedeutung und so änderte das neue Prüfungsformat das studentische Lernverhalten. Es gab Hinweise darauf, dass insbesondere die Prüfungskonsequenz das Lernen beeinflussen kann. Im Hinblick auf die Prüfungskonsequenz spielen die Begriffe „summativ“ und „formativ“ eine wichtige Rolle, auf die im Folgenden eingegangen werden soll. Summative Prüfungen sind Prüfungen, aus denen Konsequenzen resultieren. Mögliche Prüfungskonsequenzen sind das Bestehen der Prüfung gegebenenfalls mit einer Notenvergabe oder das Nicht-Bestehen mit der Folge, die Prüfung entweder zu wiederholen oder in extremen Fällen aus dem jeweiligen Kurs auszuschneiden. Summative Prüfungen werden eingesetzt, um vorheriges Lernen zu prüfen und daher wird ihre Funktion auch als „Testen des Lernens (bzw. des Gelernten)“ (engl. *testing of learning*), bezeichnet (Roediger und Karpicke 2006a). Raupach et al. beschrieben 2013, dass summa-

tive Prüfungen einen starken Lernanreiz generieren und sich auf das Lernverhalten auswirken. In der Studie von Raupach et al. (2013) ging es um die Auswertung der Elektrokardiographie. Sowohl verschiedene Lehrintensitäten als auch unterschiedliche Prüfungskonsequenzen wurden miteinander verglichen. Gemessen wurde das Ergebnis anhand schriftlicher Freitextprüfungen. Die Ergebnisse zeigten, dass summative Prüfungen einen größeren Einfluss auf das studentische Lernverhalten haben als innovative Lehrformate. Es wurde eine starke Assoziation zwischen summativen Prüfungen und guten Prüfungsergebnissen gezeigt, während die Lehrintensität unabhängig von der Prüfungsleistung war. (Raupach et al. 2013). Die Prüfung mit ihren potenziellen Konsequenzen kann – unabhängig von der Quantität und Qualität der Lehre – ausreichen, dass sich der Prüfling vor der Prüfung intensiv mit dem Lerninhalt beschäftigt. Allgemein werden zur Zeit hauptsächlich summative Prüfungen in Ausbildungen eingesetzt (Roediger und Karpicke 2006a).

Formative Prüfungen hingegen ziehen keine negativen Konsequenzen nach sich. Deshalb erzeugen sie einen geringeren Lernanreiz als summative Prüfungen (Raupach et al. 2013). Formative Prüfungen generieren Feedback und zeigen damit auf, welche Inhalte der Prüfling schon beherrscht und welche Inhalte noch gelernt werden müssen, um ein Thema zu verstehen (Black und Wiliam 1998). Sie generieren Rückmeldungen, auf deren Grundlage Studierende ihren Schwerpunkt beim Lernen neu definieren können (Hudson und Bristow 2006). Deswegen wird in Bezug auf formative Prüfungen häufig von einem Testen, welches das Lernen fördert (engl. *testing for learning*), gesprochen (Roediger und Karpicke 2006a). Nachdem über die Prüfungskonsequenz geschrieben wurde, werden nun die Prozesse während einer Prüfung betrachtet.

1.8 Lernen durch Testen (*Test-enhanced Learning*)

Das Testen von Lerninhalten kann einen größeren positiven Effekt auf die Langzeiterinnerung dieser Lerninhalte haben als das erneute Studieren von Lerninhalten, auch wenn das Ergebnis des Tests unbefriedigend ausfällt und kein Feedback gegeben wird. Dieses Phänomen wird auch als *Testing Effect* bezeichnet (Roediger und Karpicke 2006a) und wurde bereits in zahlreichen Studien bestätigt (Hogan und Kintsch 1971; Karpicke und Roediger 2008; Thompson et al. 1978). Es wird zwischen dem direkten und dem indirekten *Testing Effect* unterschieden. Der indirekte *Testing Effect* beschreibt den Effekt, der das Lernverhalten beeinflusst. Dies bedeutet beispielsweise, dass bewertete Prüfungen den Studierenden dazu veranlassen, vor der Klausur möglichst viele Inhalte zu lernen oder Feedback dazu führen kann, dass Studierende Lerndefizite erkennen und ihr zukünftiges Lernen

daraufhin verändern und anpassen. Der direkte *Testing Effect* beschreibt, dass während der Teilnahme an einer Prüfung, die langfristige Speicherung von Inhalten gefördert wird (Roediger und Karpicke 2006a). Ursprünglich wurde sich auf zwei mögliche Hypothesen für die Erklärung des direkten *Testing Effects* fokussiert – die *Amount-of-Processing*-Hypothese und die *Retrieval*-Hypothese (Dempster 1996). Nach der *Amount-of-Processing*-Hypothese führt eine vermehrte Exposition gegenüber dem Lerninhalt durch das Testen zu einem größeren Lernerfolg – unabhängig davon, in welchem Rahmen diese Exposition stattfindet. Während hier also die Lernzeit der wesentliche Faktor für das Lernergebnis ist, vermuten Verfechter der *Retrieval*-Hypothese, dass durch das aktive Reproduzieren von Inhalten aus dem Gedächtnis die Retention dieser Lerninhalte im Gedächtnis gefördert wird (Dempster 1996; Larsen et al. 2008; Roediger und Karpicke 2006b). Es wurden einige Studien durchgeführt, um die beiden Hypothesen zu testen. Hierbei wurde mehrfach gezeigt (Carrier und Pashler 1992; Glover 1989), dass im gleichen Zeitumfang wiederholtes Testen zu einem größeren Lernerfolg führte als wiederholtes Lesen. Die Erkenntnisse sollen am Beispiel der Studie von Roediger und Karpicke (2006b) erklärt werden, in der zwei Experimente durchgeführt wurden (Roediger und Karpicke 2006b). Das erste Experiment umfasste eine Stichprobe von 120 Studierenden der Washington University und fand in zwei Phasen statt. Das Arbeitsmaterial bestand aus zwei unterschiedlichen Texten mit ähnlicher Wortanzahl. In der ersten Phase wurden den Studierenden nacheinander die zwei Texte ausgehändigt, die die Studierenden jeweils sieben Minuten lang studierten. Danach sollte ein Text sieben Minuten lang erneut studiert und der andere Text in einem Test von sieben Minuten schriftlich frei wiedergegeben werden. In der zweiten Phase wurden drei Retentionstests mit derselben Anforderung wie in Phase 1 (schriftliche freie Wiedergabe) nach fünf Minuten, zwei Tagen und nach einer Woche durchgeführt. Die Studierenden erhielten kein Feedback zu den Tests. Für die Bewertung der Tests wurden in jedem der zwei Texte 30 Merkmale definiert, die 30 Punkten entsprachen. Für jedes richtig wiedergegebene Textmerkmal gab es einen Punkt. Unterschiedliche Effekte wurden beobachtet. Der kurzfristige Retentionstest nach fünf Minuten zeigte, dass nach zweimaligem Studieren des Textes kurzfristig bessere Ergebnisse erzielt wurden als nach einmaligem Studieren und anschließendem Test mit schriftlicher freier Wiedergabe. Die Retentionstests nach zwei Tagen und nach einer Woche ergaben, dass langfristig gesehen Testen zu einem besseren Behalten von Lerninhalten führte als erneutes Studieren. Durch die einheitliche Expositionszeit von sieben Minuten konnte gezeigt werden, dass der direkte *Testing Effect* nicht durch eine zusätzliche Exposition mit dem Lerninhalt zustande kommt, son-

dern andere Prozesse dem direkten *Testing Effect* zugrunde liegen (siehe auch (Glover 1989; Larsen et al. 2008)). Somit erschien die *Amount-of-Processing*-Hypothese eher unwahrscheinlich und die *Retrieval*-Hypothese als Erklärung für den direkten *Testing Effect* rückte in den Fokus der Forschung.

Im zweiten Experiment wurden insbesondere die Effekte des wiederholten Testens in der Interventionsphase im Vergleich zum wiederholten Studieren in der Interventionsphase (also vor den finalen Retentionstests) untersucht. Das Arbeitsmaterial bestand aus denselben Texten und 30 Merkmalen wie im ersten Experiment. Die Stichprobe von 180 Studierenden der Washington University wurde in drei Gruppen unterteilt. Entweder studierten die Studierenden einen Text viermal (SSSS) oder studierten den Text dreimal und wurden einmal darüber in Form von schriftlicher freier Wiedergabe getestet (SSST) oder studierten einmal den Text und wurden dreimal darüber getestet (STTT). Danach nahm jeweils die Hälfte der drei Gruppen an einem finalen Retentionstest nach fünf Minuten teil. Die andere Hälfte wurde nach einer Woche getestet. Die Ergebnisse des ersten Experiments wurden bestätigt. Im finalen Test nach fünf Minuten erzielte die Gruppe SSSS das beste Ergebnis. Die Gruppe SSST schnitt schlechter ab als die Gruppe SSSS und das schlechteste Ergebnis erzielte die Gruppe STTT. Im finalen Test nach einer Woche erreichte hingegen die Gruppe STTT die besten Ergebnisse. Die Gruppe SSST erzielte etwas weniger Punkte und die Gruppe SSSS schnitt am Schlechtesten ab (Roediger und Karpicke 2006b). Zusammenfassend konnte gezeigt werden, dass wiederholtes Testen einen starken Einfluss auf den langfristigen Lernerfolg hat (siehe auch (Thompson et al. 1978; Wheeler et al. 2003)) und wiederholtes Studieren eher die kurzfristige Retention fördert. In der Übersichtsarbeit von Roediger und Karpicke (2006a) wurde unter anderem zusammengefasst, dass häufiges und regelmäßiges Testen einen größeren positiven Effekt auf den langfristigen Lernerfolg hat als geballtes Testen (Roediger und Karpicke 2006a). Ferner stellten Karpicke und Roediger im Jahr 2008 fest, dass wiederholtes Testen von Inhalten das Vergessen dieser Inhalte mehr als das einmalige Testen der Inhalte verlangsamte. Nachdem ein Lerninhalt richtig wiedergegeben wurde, hatte wiederholtes Studieren des Lerninhaltes im Gegensatz zu wiederholtem Testen keinen positiven Effekt auf die langfristige Retention (Karpicke und Roediger 2008). In der vorgestellten Studie von Karpicke und Roediger (2006b) wurde das Testen ohne Feedback durchgeführt. In einer anderen Studie ergab sich, dass mit der Gabe von Feedback ein sogar noch stärkerer Effekt des Testens auf das Lernen ausgeübt werden kann (Kang et al. 2007). Einen besonders positiven Einfluss kann Feedback haben, welches direkt nach dem Test und nicht verspätet folgt (Roediger und Karpicke 2006a).

In den letzten Jahren wurde zu drei Theorien geforscht, von denen angenommen wird, dass sie die zugrundeliegenden Mechanismen des *Testing Effects* beschreiben. Das nächste Kapitel soll einen Überblick über diese Theorien geben.

1.8.1 Mechanismen des *Testing Effects*

Im Folgenden werden drei Theorien zu den zugrundeliegenden Mechanismen des *Testing Effects* erläutert. Sie werden *Elaborate Retrieval Theory* (Carpenter 2009), *Transfer Appropriate Processing Theory* (Morris et al. 1977) und *Unspecific Goal Perspective* (Paas und Kischner 2012) genannt und sind in Tabelle 1 dargestellt. In der vorliegenden Arbeit wurden diese drei Theorien anhand von Aussagen von Studierenden untersucht.

Die *Elaborate Retrieval Theory* umfasst genau genommen zwei unterschiedliche Prozesse. Der eine (*Associative Memory Theory*) besteht darin, dass bei einem Test durch das Abrufen einer Information aus dem Gedächtnis nicht nur diese Information, sondern auch zusammenhängende Informationen im Gedächtnis aktiviert werden. Auf den zweiten Prozess (*Degree of Semantic Elaboration*) wird im nachfolgenden Kapitel 1.8.3 genauer eingegangen. Es wird postuliert, dass durch den mentalen Aufwand für den Abruf von Informationen, die Lerninhalte besser gespeichert werden (Endres und Renkl 2015). Nach der *Transfer Appropriate Processing Theory* werden Lerninhalte durch eine ähnliche Aufgabenstellung in der Übungsphase wie im finalen Test durch Transfer gefestigt (Morris et al. 1977). Bei Roediger (2006) konnte beispielsweise gezeigt werden, dass bei identischer Aufgabenstellung der vorherigen Tests und des finalen Tests die Fähigkeiten zur Bewältigung der Aufgabe trainiert wurden und das Behalten von Inhalten gestärkt wurde (Roediger und Karpicke 2006b). Bestätigen konnte dies eine Studie von McConnell et al. (2015), in der die Fragetypen, die in der Übungsphase und im finalen Test einheitlich eingesetzt wurden, eine größere Retention hervorriefen, als wenn sich bei gleichen Lernzielen die Aufgabenstellung in der Übungsphase und im finalen Test unterschieden (McConnell et al. 2015). Nach der dritten in der Literatur diskutierten Theorie (*Unspecific Goal Perspective*) fördern Aufgaben mit einer größeren thematischen Bandbreite die Retention stärker als auf Details fokussierte Fragen (Paas und Kischner 2012). So konnten zum Beispiel Wirth et al 2009 feststellen, dass Studierende mit weniger spezifischen Aufgaben eine bessere Leistung im Lösen von Problemen erzielten als Studierende mit spezifischen Aufgaben (Wirth et al. 2009).

Tabelle 1: Theorien zu den zugrundeliegenden Mechanismen des *Testing Effects*

Theorien		Definition
<i>Elaborate Retrieval Theory</i>	<i>Associative Memory Theory</i>	Der Abruf einer Information führt im Gedächtnis zur Aktivierung sowohl von dieser Information als auch von einer zusammenhängenden Information
	<i>Degree of Semantic Elaboration</i>	Mentale Anstrengung wird für den Abruf von Inhalten benötigt
<i>Transfer Appropriate Processing Theory</i>		Testen stärkt kognitive Prozesse durch Transfer, wenn die Aufgaben in der Übungsphase den Aufgaben im finalen Test ähneln.
<i>Unspecific Goal Perspective</i>		Das Lösen von thematisch breiteren Aufgaben führt zum besseren Behalten von Lerninhalten als das Lösen von sehr spezifischen Aufgaben.

1.8.2 Lernen durch Testen im Medizinstudium

Da viele Studien mittels Wortlisten (Hogan und Kintsch 1971, Karpicke und Roediger 2007, Thompson et al. 1978) oder in simulierten Klassenzimmern (Butler und Roediger 2007) durchgeführt wurden, empfahlen Larsen et al. (2008) Studien zu konzipieren, die in reale Lehrsituationen integriert werden, um die Bedeutung des *Testing Effects* im Medizinstudium zu untersuchen. Die folgende daraufhin konzipierte Studie von Schmidmaier et al. 2011 soll in diesem Rahmen als Beispiel dienen. Ziel der Studie war es, den Effekt des Lernens von Faktenwissen mit Karteikarten bei Medizinstudierenden im klinischen Studienabschnitt zu untersuchen. In erster Linie wurde analysiert, ob wiederholtes Testen einen positiveren Effekt auf den kurzfristigen (nach einer Woche) und langfristigen Lernerfolg (nach sechs Monaten) hat als wiederholtes Studieren der Karteikarten. Die Ergebnisse zeigten, dass nach einer Woche das wiederholte Testen zu einem besseren Lernerfolg führte als das wiederholte Studieren der Karteikarten. Nach sechs Monaten ohne zwischenzeitliche Interventionen nahm insgesamt die studentische Leistung im Behalten des Faktenwissens ab, wobei sich kein Leistungsunterschied zwischen wiederholtem Testen und Studieren detektieren ließ (Schmidmaier et al. 2011). Studien fokussierten sich besonders auf das Auswendiglernen von Faktenwissen (Dobson und Linderholm 2015; Schmidmaier et al. 2011). Die daraus gewonnenen Erkenntnisse können zwar im Medizinstudium ange-

wendet werden, jedoch sind, wie schon beschrieben, in der ärztlichen Ausbildung das Lehren und Lernen von Kompetenzen und die Anwendung von Wissen im Medizinstudium von besonderer Bedeutung. Hier stellte sich die Frage, welche Fragetypen dafür am besten geeignet sind.

Generell konnte in der Literatur bestätigt werden, dass sowohl Einfachwahl- als auch Kurzantwortfragen zu einem besseren Behalten von Inhalten führen als das Studieren von Inhalten, wobei Kurzantwortfragen zu einem noch besseren Lernerfolg führen als Einfachwahlfragen (Butler und Roediger 2007; McDaniel et al. 2007a; McDaniel et al. 2007b; Roediger und Karpicke 2006a). Fragetypen wie Einfachwahl, PickN oder Kprim können sich auch negativ auf den Lernerfolg auswirken, da neben Richtigantworten auch die Falschantworten zu einer Frage aufgeführt sind. Wenn Studierende beispielsweise eine Falschantwort als richtig bewerten, kann sich dadurch, dass das Abfragen der Inhalte das Langzeitwissen fördert, nun die Falschantwort fälschlicherweise als die korrekte Antwort gemerkt werden. Auch wenn Falschantworten richtigerweise vom Studierenden als falsch eingeschätzt werden, kann es sein, dass der Studierende nach einiger Zeit die Antwort als richtig einschätzt, da allein das Abfragen von Inhalten das Langzeitwissen stärkt. Hinweise darauf geben unter anderem Studien wie Hasher et al. (1977) oder Roediger und Marsh (2005). In diesem Zusammenhang zeigten Butler und Roediger (2008), dass ein Feedback nach dem Beantworten der Frage diesen negativen Effekt ausgleichen kann (Roediger und Karpicke 2006a). Demzufolge sollten Studierende nach Tests sofortiges Feedback erhalten (Roediger und Karpicke 2006a).

Larsen et al. (2008) regten an, dass das Testen klinischer Anwendung von Wissen anhand von Schauspielpatienten oder standardisierten Patienten geprüft werden sollte, da diese realen Patientenfällen ähneln. Im Jahre 2012 wurden in einer Studie von Larsen et al. (2012) drei verschiedene Präsentationsformate verglichen – das Lesen eines Arbeitsblatts, ein Test mit Kurzantwortfragen und ein Test mit Anamnesegesprächen von Schauspielpatienten. Die Ergebnisse der Studie zeigten, dass Studierende durch Tests mit Schauspielpatienten ein gleich gutes Ergebnis wie mit schriftlichen Tests erzielten und dass die Studierenden mit beiden Präsentationsformaten einen besseren Lernerfolg gegenüber dem Lesen des Arbeitsblatts aufwiesen. Insbesondere zeigten diese Ergebnisse, dass testbasiertes Lernen auch auf die klinische Praxis angewendet werden kann und vielversprechend für die medizinische Ausbildung ist (Larsen et al. 2012). Jedoch wirft das Prüfen anhand von Schauspielpatienten die Problematik auf, dass es ein ressourcenintensives Prüfungsformat in der medizinischen Ausbildung darstellt. Denn der Einsatz von Schauspielpatienten ist

zeitaufwendig und personalintensiv. In diesem Zusammenhang scheint es sinnvoll, die Funktionen von Fragetypen übergeordnet zu betrachten. Dabei kann eine Unterscheidung von *Recognition Tests* und *Production Tests* hilfreich sein, welche im folgenden Kapitel erklärt werden.

1.8.3 *Recognition Tests* und *Production Tests*

In der Literatur gab es viele Hinweise darauf, dass Tests, die die eigenständige Produktion einer Antwort fordern (sog. *Production Tests*) zu einem besseren langfristigen Behalten von Lerninhalten führen als Tests, die das Wiedererkennen von Antworten bedingen (sog. *Recognition Tests*) (Butler und Roediger 2007; Kang et al. 2007; McDaniel et al. 2007b; Roediger und Karpicke 2006a). Als Erklärung dafür wurde angenommen, dass der *Testing Effect* stärker ausgeprägt ist, wenn der Abrufprozess von Inhalten aus dem Gedächtnis aufwendig ist (engl. *effortful retrieval*) (Bjork 1975; Kang et al. 2007; McConnell et al. 2015). *Production Tests* beinhalten Fragetypen wie Kurzantwortfragen oder *Long-Essay*-Fragen, welche einen aufwändigen Abrufprozess verlangen, weil die Antworten von dem Studierenden produziert werden müssen. *Recognition Tests* umfassen Fragetypen wie Einfachwahl, PickN oder Kprim Fragen, bei denen ein verhältnismäßig einfacher Abrufprozess erfordert ist, da Antworten nicht produziert, sondern als richtig bzw. falsch wiedererkannt werden müssen (Larsen et al. 2008; Roediger und Karpicke 2006a). In Bezug auf den *Testing Effect* heißt das, dass beispielsweise die zu den *Production Tests* gehörende Kurzantwortfrage einen stärker ausgeprägten positiven Effekt auf das Behalten von Lerninhalten hat als der zu den *Recognition Tests* gehörende Fragetyp Einfachwahl, weil der Studierende sich stärker anstrengen muss die Antwort für eine Kurzantwortfrage zu verfassen als eine richtige Antwort aus den Antwortmöglichkeiten der Einfachwahl-Frage wiederzuerkennen (McConnell et al. 2015). Bezüglich Bloom's Taxonomie (siehe oben) prüfen *Recognition Tests* demnach auf einem niedrigen taxonomischen Niveau, während *Production Tests* auf einem höheren taxonomischen Niveau prüfen. Für Ausbildungszwecke sollten Fragetypen gewählt werden, die eine mentale Anstrengung erfordern (Endres und Renkl 2015) und dementsprechend auf einem höheren taxonomischen Niveau testen. Nachdem die Forschung zum Thema „Lernen durch Testen“ vorgestellt wurde, soll der Fokus erneut auf die Kompetenz CR gelegt werden. Im folgenden Kapitel ist zusammengefasst, wie die Lehre von CR bisher ausgerichtet wurde.

1.9 Lehrmethoden zur Vermittlung von CR

Das Erwerben der Kompetenz CR und deren Fähigkeiten, wie über die Diagnostik und Therapie von Patienten zu entscheiden, hat Auswirkungen auf das Wohlergehen von Patienten. Es ist also unumgänglich, die Lehre von CR in einem höchstmöglichen Standard in die Ausbildung von Medizinstudierenden zu integrieren. Wie schon angedeutet, ist es hierbei wesentlich, dass eine Kongruenz zwischen Lernziel und Lehrmethode herrscht (Kern et al. 1998). Das heißt, dass die Lehrmethode eingesetzt werden muss, die die Studierenden am besten dabei unterstützt, die Kompetenz zu erwerben. CR kann auf unterschiedliche Weise gelehrt werden. Eine Methode ist, dass Lernende in Kleingruppen mit realen Patientenfällen konfrontiert werden. Anstatt realer Patientenfälle können auch Fallbeispiele ermöglichen, Fähigkeiten, die für das CR benötigt werden, zu entwickeln. Die Fallbeispiele sollten in derselben chronologischen Reihenfolge ablaufen wie die realen Patientenfälle und an den Wissensstand der Lernenden angepasst sein. Die Kleingruppen können von einem Experten interaktiv gestaltet werden. Dies hat den Nachteil, dass vor allen Dingen bei vielen Teilnehmern und folglich vielen Kleingruppen diese Methode sehr ressourcenintensiv sein kann (Kassirer 2010). Um die Kompetenz CR zu erwerben, ist die fundierte Wissensbasis eine entscheidende Komponente (Norman 2005; Schuwirth 2009). Zusätzlich spielt das bewusste, intensive Üben (engl.: *deliberate practice*) eine große Rolle, um Expertise auf einem Gebiet zu erlangen (Ericsson 2004). Hierfür ist es sinnvoll, Lernende mit ähnlichen Fallbeispielen mit Feedback zu konfrontieren, um einen Transfer von Lösungskonzepten auf andere Patientenfälle herzustellen und zu vereinfachen (Norman 2005). Weitere Faktoren, die zum CR beitragen, sind metakognitive Prozesse wie beispielsweise das reflektierte Lösen von Problemen (engl.: *reflective coping* (Greenglass 2002)). Es beschreibt eine Herangehensweise für das Lösen von Problemen (z. B. eines Patientenfalls), die sich durch ausgiebiges Überlegen und das kognitive Simulieren von potenziellen Lösungsansätzen und deren Folgen auszeichnet (Dunphy et al. 2010). In der Literatur werden zwei Methoden diskutiert, wie klinische Informationen eines Patientenfalls dem Studierenden präsentiert werden sollten, um CR zu lehren. Die eine Methode nennt sich *Whole Case Approach* und die andere *Serial Cue Approach* (Schmidt und Mamede 2015). Bei der Methode *Whole Case Approach* werden den Studierenden bereits zu Beginn alle klinischen Informationen und Befunde eines Patienten mitgeteilt, woraufhin die Diagnose gestellt werden soll. Bei der weiter verbreiteten Methode *Serial Cue Approach* wird ein Patientenfall simuliert und folglich werden die klinischen Informationen eines Patientenfalls wie im Berufsalltag nach und nach mitgeteilt (Schmidt und Mamede 2015). Der Studierende wird

dabei in die Rolle eines Arztes versetzt. Zuerst werden die dominanten Symptome geschildert, woraufhin der Studierende weitere Informationen durch Anamnese und Diagnostik erheben muss. Die Befunde der angeforderten Diagnostik werden daraufhin mitgeteilt. Schmidt und Mamede (2015) betonten, dass für das Finden der passenden Lehrmethode darauf geachtet werden muss, in welchem Studienabschnitt sich die Studierenden befinden. Die Methode *Serial Cue Approach* sei besonders dann sinnvoll, wenn die Studierenden schon klinisches Vorwissen zu den Krankheitsbildern der Patientenfälle hätten (Schmidt und Mamede 2015).

Insgesamt ist die Forschung zum Lehren des CR bei weitem noch nicht abgeschlossen (Kassirer 2010; Norman 2005; Schmidt und Mamede 2015). Es bestand die Aufgabe, einen Fragetyp zu finden, der CR prüfen kann (van der Vleuten und Newble 1995). Im Allgemeinen wird in der medizinischen Ausbildung am häufigsten der Fragetyp Einfachwahl verwendet (Larsen et al. 2008). Jedoch stellte sich heraus, dass Einfachwahl-Fragen begrenzt dafür geeignet sind, die klinischen Gedankenprozesse eines Medizinstudierenden widerzuspiegeln (Bloch et al. 1999). Für klinische Gedankenprozesse und somit auch für die Kompetenz CR ist das prozedurale Wissen ein essentieller Bestandteil. Zur Prüfung prozeduralen Wissens bzw. komplexer kognitiver Prozesse sind insbesondere *Key-Feature-Fragen* geeignet (Kopp et al. 2006; Hrynchak et al. 2014), auf welche im nächsten Kapitel eingegangen werden soll.

1.10 Key Feature

Der Begriff *Key Feature* wurde von Bordage und Page 1987 erstmalig eingeführt und ist definiert als ein Knotenpunkt in der Lösung eines Problems und fokussiert sich auf den Punkt in der Problemlösung, an dem Prüflinge am ehesten einen Fehler begehen und der einen schwierigen Aspekt der Identifikation und der Handhabung eines Problems in der Praxis darstellt (Page et al. 1995). Die Definition von *Key Feature* impliziert, dass nicht alle Schritte einer Problemlösung gleichbedeutend sind und es sinnvoll erscheint, sich in einer Prüfung auf die entscheidenden Schritte zu konzentrieren (Page et al. 1995). Ursprünglich wurden *Key Features* entwickelt, um in Kanada ein valides schriftliches medizinisches Abschlussexamen zu schaffen, welches die Kompetenz CR prüft (Hrynchak et al. 2014; Page und Bordage 1995). Nach sechs Jahren Forschung (Page und Bordage 1995) fanden *Key Features* im Jahre 1992 das erste Mal Anwendung im kanadischen Abschlussexamen (Page et al. 1995). Seither wurden *Key Features* in weiteren medizinischen

Universitäten beispielsweise in Kanada oder Australien implementiert (Hrynchak et al. 2014).

Folgende Gründe sprechen dafür, CR mit Hilfe von *Key Features* zu prüfen: Durch die Konzentration auf Knotenpunkte in der Problemlösung sind die zu bearbeitenden Patientenfälle kurz, was wiederum eine höhere Anzahl an Patientenfällen pro Test ermöglicht. Des Weiteren können *Key Features* in Verbindung mit unterschiedlichen Fragetypen verwendet werden und sorgen so für eine flexible Anwendung (Page et al. 1995). Fragen mit *Key Features* sind üblicher Weise wie folgt aufgebaut: Es gibt einen Fragenstamm, der entscheidende und nebensächliche Informationen enthält. Diese Informationen sind meistens die Ausgangssituation, in die sich der Prüfling versetzen soll (Famulant, Arzt auf Station, Notaufnahme, Hausarztpraxis) und das Alter, das Geschlecht, der Allgemeinzustand und die Symptome des Patienten („Fallvignette“). Auf den Fragenstamm folgen einige Fragen, die auf die *Key Features*, also die Knotenpunkte des Falls, abzielen (Hrynchak et al. 2014; Kopp et al. 2006). Fotos von Untersuchungsergebnissen, Videos oder Tonaufnahmen können die Patientenfälle noch realitätsgetreuer erscheinen lassen (van Bruggen et al. 2012). Eine Literaturübersicht von Hrynchak et al. 2014 konnte bekräftigen, dass *Key Features* eine Möglichkeit darstellen, CR zu prüfen.

Welche Fragetypen für weitere Studien gewählt werden, um mit *Key Features* das CR zu prüfen, muss im Hinblick auf die vorherige Forschung zu den Fragetypen abgewogen werden (Hrynchak et al. 2014). In der Literatur wird vermutet, dass mit einer Kombination aus *Key Features* und dem computergestützten Fragetyp *Long-Menu* das Prüfen von CR nah an der klinischen Realität durchgeführt werden kann (Rotthoff et al. 2006). Eine Studie von Fischer et al. 2005 zeigte, dass elektronische *Key Features* im Fragetyp *Long-Menu* eine zuverlässige Prüfungsmethode darstellen können, um das CR zu prüfen. Limitationen waren, dass die Stichprobe mit 37 Medizinstudierenden eine relativ kleine war und zu wenige *Key Features* geprüft wurden. Deshalb wurde empfohlen, Studien mit größeren Stichproben und mit einer größeren Anzahl an besser überarbeiteten *Key Features* unter der Verwendung des *Long-Menu*-Fragetyps zu konzipieren (Fischer et al. 2005). Rotthoff et al. 2006 gaben außerdem die Empfehlung, nicht nur eine *Long-Menu*-Liste zu verfassen, sondern eine Liste für jeden Bereich (Diagnosen, die Diagnostik oder Therapie) zu erstellen, um einen besseren Überblick zu erhalten. Eine Pilotierung von Fragen mit *Key Features* wird empfohlen, da dies eine Möglichkeit zum Überarbeiten der Fragen und Antwortlisten generiert (Page et al. 1995). So können fehlende Synonyme und Distraktoren den Antwortlisten hinzugefügt werden.

Auf Grundlage der bisherigen Forschung wurde unter Berücksichtigung des *Testing Effects* eine Studie entwickelt, die CR durch *Key Features* im Fragetyp *Long-Menu* mittel- und langfristig untersuchte. Es stellte sich die Frage, ob wiederholte *Key-Feature*-Prüfungen (Fragenfälle) einen größeren Lernerfolg hinsichtlich komplexer kognitiver Fertigkeiten erzeugen als wiederholtes Studieren der Inhalte anhand von „Lesefällen“. Die Studie wurde im Hinblick auf den langfristigen Erwerb von prozeduralem Wissen entworfen und in den Ablauf eines Semesters des Medizinstudiums der Georg-August-Universität Göttingen eingegliedert. Durch eine vorher durchgeführte Pilotstudie konnten die *Key-Feature*-Fragen und die *Long-Menu*-Listen überarbeitet werden. Die Pilotstudie wurde im Sommersemester 2013 im 3. klinischen Semester der Medizinischen Fakultät der Georg-August-Universität Göttingen longitudinal vernetzt und im Rahmen von curricular integrierten Pflichtveranstaltungen durchgeführt. Von den Studierenden wurden Fragenfälle mit *Key Features* computergestützt bearbeitet. Thematisch beruhten die Fragenfälle auf den zuvor gelehrteten Inhalten der modularen Lehre. Bei der Datenanalyse wurden insbesondere die Gütekriterien (Itemschwierigkeiten, Itemtrennschärfen und die interne Konsistenz) berechnet. Die Ergebnisse wurden für die Verbesserung der *Key-Feature*-Fälle und zur Erweiterung der *Long-Menu*-Listen für die Studie der vorliegenden Arbeit genutzt. Im Rahmen von Fokusgruppengesprächen wurden studentische Beurteilungen eingeholt und qualitativ analysiert.

Die Verwendung eines sofortigen Feedbacks beruhte einerseits auf der bereits genannten Literatur und andererseits auf den Rückmeldungen der Studierenden der Pilotstudie. Es ergaben sich folgende Studienfragen, die im Rahmen der vorliegenden Arbeit beantwortet werden.

1.11 Studienfragen

- 1) Erzeugt die Bearbeitung von wiederholten *Key-Feature*-Fragen einen größeren Lernerfolg hinsichtlich differenzialdiagnostischer und therapeutischer Fähigkeiten als das Lesen von wiederholten inhaltlich gleichen Lesefällen? Hierbei wird unterschieden zwischen dem
 - a) mittelfristigen und dem
 - b) langfristigen Lernerfolg.
- 2) Welchen Einfluss hat die Lernzeit auf den Zusammenhang zwischen Präsentationsformat (Fragenfälle versus Lesefälle) und Lernerfolg?

- 3) Wie entwickeln sich die Itemschwierigkeiten und Itemtrennschärfen sowie die interne Konsistenz von Fragenfällen im Laufe des Semesters?
- 4) Lassen sich distinkte Muster des Leistungserhaltes bzw. -verlustes bei wiederholter Bearbeitung von *Key-Feature*-Fragen identifizieren?
- 5) Wie beurteilen Studierende die elektronischen Fallseminare?
- 6) Inwieweit wurden die Mechanismen des *Testing Effects* in Fokusgruppengesprächen bekräftigt?

2 Methoden

Im Folgenden wird zunächst das Design der Studie beschrieben. Außerdem wird in Bezug auf die Integration der Studie ein Überblick über das Curriculum des Medizinstudiums an der Georg-August-Universität Göttingen gegeben. Daraufhin werden der zeitliche Rahmen und die Probandenrekrutierung erläutert und die beiden Prüfungsformate Fragenfall und Lesefall vorgestellt. Danach wird auf den Prüfungsrahmen näher eingegangen. Nach der Erläuterung der praktischen Durchführung und des Zeitplans der Studie werden die quantitativen und qualitativen Analysemethoden dargestellt. Die Studie wurde von der Ethik-Kommission der Medizinischen Fakultät der Georg-August-Universität Göttingen bewilligt. (Antragsnummer: 2/10/13; Votum vom 7.11.2013)

2.1 Studiendesign

Die Studie wurde im Wintersemester 2013/2014 als randomisierte, kontrollierte Studie im *Crossover*-Design im Studiengang Humanmedizin an der Georg-August-Universität Göttingen durchgeführt. Abbildung 3 soll zur Veranschaulichung des Studiendesigns dienen. Die Studierendekohorte wurde nach Geschlecht und Klausurleistungen des Vorsemesters stratifiziert und nach dem Zufallsprinzip in zwei Studiengruppen (Gruppe A und Gruppe B) eingeteilt. Die Studie war so konzipiert, dass die Lehrinhalte zuerst in den modulbegleitenden Vorlesungen und Seminaren gelehrt wurden und danach im Rahmen der Studie in wöchentlichen elektronischen Fallseminaren (E-Fallseminaren) mit zwei unterschiedlichen Präsentationsformaten wiederholt wurden. Die Lehrinhalte beider Präsentationsformate waren identisch.

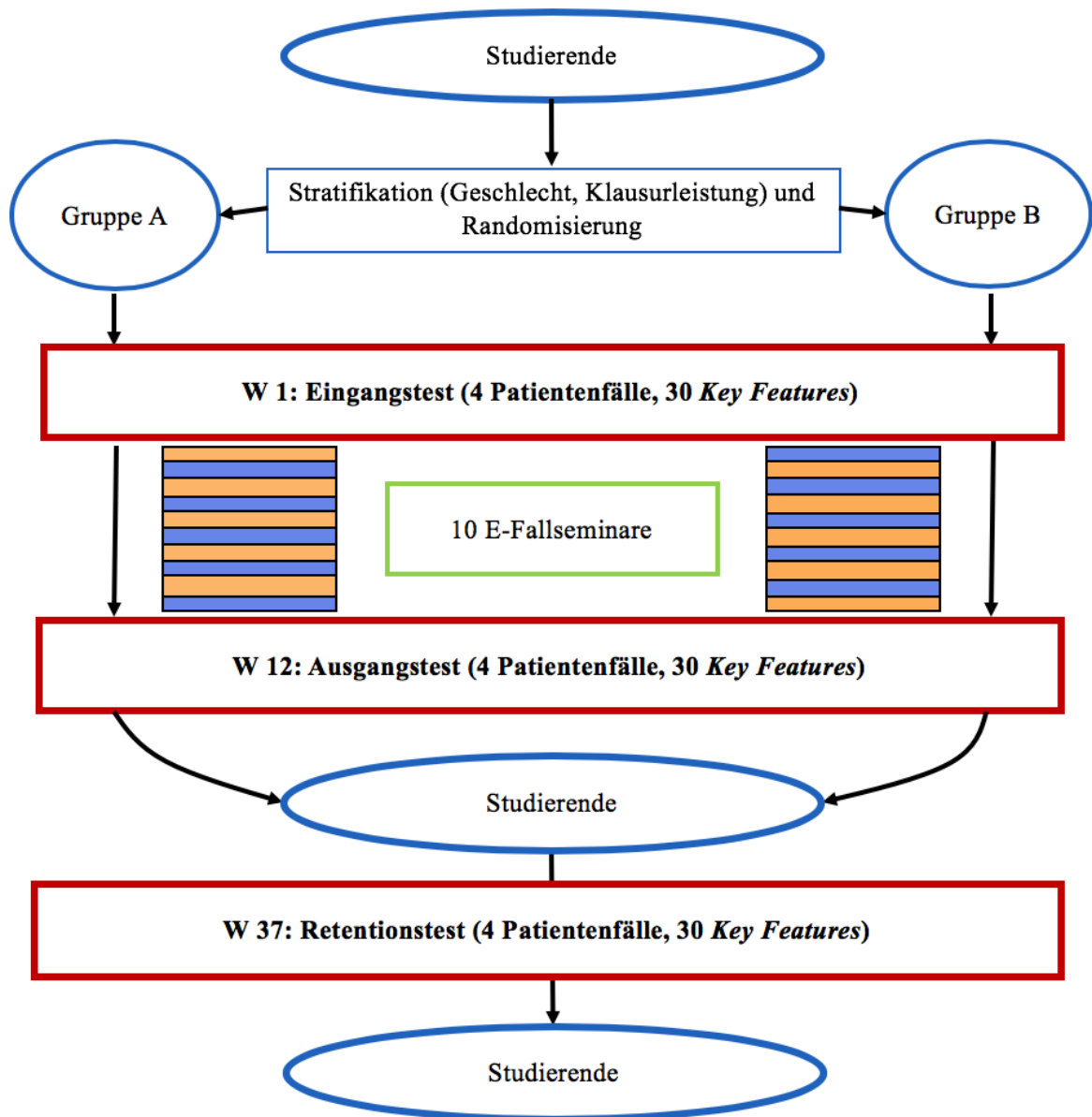


Abbildung 3: Studiendesign, W=Woche

Zur Objektivierung des mittelfristigen Lernerfolgs wurde in der ersten Semesterwoche ein Eingangstest und in der vorletzten Semesterwoche ein Ausgangstest durchgeführt. Zur Objektivierung des langfristigen Lernerfolgs fand ein halbes Jahr nach dem Ausgangstest ein Retentionstest statt. Der Eingangs-, der Ausgangs- und der Retentionstest enthielten für alle Studierenden die gleichen 30 *Key-Feature*-Fragen. *Crossover*-Design heißt, dass jede der beiden Studiengruppen immer abwechselnd Testgruppe bzw. Kontrollgruppe war. Dieses Prinzip beinhaltet folglich, dass jeder Studienteilnehmer durch die Durchführung beider Prüfungsformate auch seine eigene Kontrolle war. Das *Crossover*-Design dieser Studie wird im Folgenden beschrieben: Ein E-Fallseminar bestand entweder aus *Key-Feature*-Fragen (Fragenfälle) oder aus reinen Lesefällen. Im Anhang 2 und 3 befindet sich für beide

Präsentationsformate ein Beispiel. In der zweiten Semesterwoche beantwortete eine Studiengruppe Fragenfälle, während die andere Studiengruppe parallel im anderen Raum die inhaltlich gleichen Fälle las. Im nächsten E-Fallseminar wurden die Präsentationsformate getauscht, sodass die Studiengruppe, die beim letzten E-Fallseminar Fragenfälle beantwortet hatte, nun die reinen Lesefälle las. Folglich ergab sich ein Wechsel der Präsentationsformate zu jedem E-Fallseminar bis einschließlich Semesterwoche 11. Dieses Prozedere ist in Abbildung 3 durch sich abwechselnde orangene (Fragenfälle) und blaue Streifen (Lesefälle) gekennzeichnet.

In Tabelle 2 sind die jeweils geprüften Inhalte und Präsentationsformate dargestellt. Nachdem ein Thema im Rahmen des Moduls gelehrt worden war, wurde das Thema in den E-Fallseminaren zwischen dem Eingangs- und dem Ausgangstest durch Lesefälle oder Fragenfälle von den Studierenden wiederholt. Beispielsweise wurden die Inhalte des Oberthemas „Koronare Herzkrankheit“ in der ersten Semesterwoche innerhalb des Moduls 3.1 gelehrt. In der zweiten Semesterwoche bezog sich das E-Fallseminar auf das Oberthema „Koronare Herzkrankheit“. Das heißt, dass Gruppe A in Woche 2 die Fragenfälle zum Oberthema „Koronare Herzkrankheit“ bearbeitete und Gruppe B in Woche 2 die Lesefälle zum Oberthema „Koronare Herzkrankheit“ las. Zu jedem E-Fallseminar wechselte das Präsentationsformat für die jeweilige Gruppe. Folglich wurden die Studierenden der Gruppe A in Woche 3 mit Lesefällen zum Oberthema „Herzinsuffizienz“ konfrontiert, während Gruppe B in Woche 3 die *Key-Feature-Fragen* zum Oberthema „Herzinsuffizienz“ beantwortete. Wie an Tabelle 2 sichtbar ist, zog sich dieses Prozedere bis inklusive Woche 11 durch. Im Anhang 1 befindet sich der genaue Zeitplan mit Daten der stattgefundenen E-Fallseminare einschließlich Themen und Fällen.

Tabelle 2: Curriculare Lehre mit Inhalten nach Lernzielkatalog, Inhalte der E-Fallseminare mit jeweiligem Präsentationsformat für die Gruppe A bzw. Gruppe B, W=Woche, T=Testen (Fälle mit *Key-Feature-Fragen*), L=Lesen (reine Lesefälle)

W	Lehrinhalt		Präsentationsformat	
	Formale Lehre	Oberthema der E-Fallseminare	Gruppe A	Gruppe B
1	Koronare Herzkrankheit (KHK)	Eingangstest	T	
2	Herzinsuffizienz	Koronare Herzkrankheit	T	L
3	Vitien	Herzinsuffizienz	L	T
4	Arrhythmien	Vitien & Koronare Herzkrankheit	T	L
5	Respiratory disease	Arrhythmien & Herzinsuffizienz	L	T
6	Gefäßerkrankungen, Lungenembolie, entzündliche Herzerkrankungen	Koronare Herzkrankheit, Vitien, Lungenerkrankungen	T	L
7	Nephrotisches Syndrom	Herzinsuffizienz, Arrhythmien, Gefäßerkrankungen, entzündliche Herzerkrankungen	L	T
8	Homöostase	Nephrotisches Syndrom, Lungenerkrankungen, Vitien	T	L
9	Nierenversagen	Homöostase, Gefäßerkrankungen, Lungenembolie, entzündliche Herzerkrankungen, Arrhythmien	L	T
10	Anämie	Nierenversagen, Nephrotisches Syndrom, Lungenerkrankungen	T	L
11		Anämie, Homöostase, Gefäßerkrankungen, Lungenembolie, entzündliche Herzerkrankungen	L	T
12		Ausgangstest	T	
37		Retentionstest	T	

2.2 Humanmedizinstudium an der Georg-August-Universität Göttingen

Im Folgenden wird der Ablauf des Medizinstudiums an der Georg-August-Universität erläutert. An der Georg-August-Universität Göttingen ist der Studiengang Humanmedizin in zwei Studienabschnitte eingeteilt: die Vorklinik und die Klinik. Anschließend folgt das Praktische Jahr. Während der Vorklinik werden naturwissenschaftliche Grundlagen der Medizin vermittelt, wohingegen sich der klinische Abschnitt des Studiums mit der Lehre der unterschiedlichen klinischen Fachbereiche befasst (<https://www.umg.eu/studium-lehre/humanmedizin/>). Der klinische Abschnitt ist in Module unterteilt, wobei sich die einzelnen Fachbereiche auf verschiedene Module aufteilen. So ist der Fachbereich Innere Medizin auf Module aufgeteilt, die zum Teil in unterschiedlichen Semestern stattfinden (<https://www.umg.eu/studium-lehre/humanmedizin/klinik/module/>). Prüfungen finden meist am jeweiligen Ende eines Moduls statt. Die prüfungsrelevanten klinischen Lehr- und Lerninhalte sind dem Lernzielkatalog der Universitätsmedizin Göttingen zu entnehmen. (https://www.umg.eu/fileadmin/user_upload/Go__ttinger_Lernzielkatalog_Stand_19-07-2017.pdf) Die vorliegende Studie wurde in das dritte klinische Semester integriert und parallel zu den Modulen M3.1 („Erkrankungen des Herz-Kreislauf-Systems und der Lunge“), M3.2 („Erkrankungen der Niere und des Urogenitalsystems“) und M3.3 („Erkrankungen des Blutes, des Knochenmarks und Grundlagen der Tumorerkrankungen“) durchgeführt.

2.2.1 Probandenrekrutierung

Zur Teilnahme an der Studie wurden alle Studierenden des 3. klinischen Semesters eingeladen, die sich für die Module M3.1, M3.2 und M3.3 angemeldet hatten. Vier Wochen vor Semesterbeginn wurden die Studierenden per E-Mail über die Studie informiert. Die Teilnahme an den E-Fallseminaren war im Rahmen der modularen Lehre verpflichtend, die Teilnahme an der begleitenden Studie (d. h. die Bereitstellung der jeweils eingegebenen Daten) war freiwillig. An die E-Mail war eine Probandeninformation angehängt und für Studierende bestand die Möglichkeit, per Mail Rückfragen zu stellen. Am ersten Tag des Semesters wurden jedem Studierenden die vorab versandte Probandeninformation und eine Einverständniserklärung als Papierausdruck ausgehändigt. Bei Einverständnis für die Auswertung der persönlichen Daten unterschrieben die Studierenden die Einverständniserklärung.

2.3 Präsentationsformate

2.3.1 Konstruktion der Fälle

Die Thematik der E-Fallseminare richtete sich, wie zuvor beschrieben, nach den Lehrplänen der Module. Zu jedem E-Fallseminar wurden vier Fälle erstellt. Jeder Patientenfall wurde einmal als Fragenfall mit fünf *Key-Feature*-Fragen und einmal als Lesefall verfasst. Die Fragenfälle wurden im Rahmen der Pilotstudie realitätsgetreu von dem Projektleiter, den Modulkoordinatoren und den Assistenz- und Fachärzten konzipiert und überarbeitet. Zu jedem Fragenfall wurde entsprechend ein Lesefall verfasst. Nur das Präsentationsformat unterschied sich, der Inhalt blieb gleich. Für jede der fünf *Key-Feature*-Fragen wurde ein Feedback geschrieben, welches der Studierende nach Beantworten der Frage anklicken konnte, um es zu lesen. Für jeden Lesefall wurden Hintergrundinformationen verfasst, welche inhaltlich den Feedbacks des Fragenfalls entsprachen, um sicherzustellen, dass alle Studierenden die gleichen Informationen erhielten. Die Hintergrundinformationen konnten nach dem Lesen des Falls ebenfalls per Klick geöffnet werden.

Im Folgenden werden die beiden Präsentationsformate beschrieben (siehe auch Anhang 2 und 3).

2.3.2 Fragenfall

Ein Fragenfall bestand aus einer klinischen Fallvignette und fünf *Key-Feature*-Fragen mit einer Lösung und einem zugehörigen Feedback nach jeder Frage. Die klinische Fallvignette stellte die jeweils individuelle Ausgangssituation des Patienten dar. Hierzu gehörten unter anderem das Alter, das Geschlecht, die Vorgeschichte, die aktuellen Symptome oder der Medikamentenstatus. Außerdem wurde das Setting (Hausarztpraxis, Krankenhaus, Notfallwagen) und die dem Studierenden zugewiesene Funktion (Famulant, Hausarzt, Assistenzarzt) mitgeteilt. Danach folgte die erste *Key-Feature*-Frage im Fragetyp *Long-Menu*. Die Datenbank enthielt tausende Antwortmöglichkeiten, die bei jeder *Long-Menu*-Frage ausgewählt werden konnten. Zur Beantwortung der Frage musste der Studierende mindestens die ersten drei Buchstaben seiner Lösung in ein freies Feld eingeben, damit sich eine Dropdown-Liste von Antwortmöglichkeiten öffnete. Die Dropdown-Liste zeigte zur Übersichtlichkeit nur noch die Antwortmöglichkeiten an, in denen die drei Buchstaben vorkamen. Prinzipiell konnten die Studierenden somit zum Beispiel bei einem Fall mit naheliegender Verdacht auf ein Nierenversagen mit der Eingabe der ersten drei Buchstaben „Nie“ auch eine „Lobärpneumonie“ als wahrscheinlichste Diagnose auswählen. Je

nach Aufgabenstellung konnten mehrere Synonyme oder Bezeichnungen als korrekt gewertet werden. Beispielsweise war in der Prüfungssoftware hinterlegt, dass in einem bestimmten Fall mit Verdacht auf einen Herzinfarkt die Auswahl von „ACS“, „Akutes Koronarsyndrom“ ebenso korrekt war wie die hauptsächlich gesuchte Antwort. Wenn ein Studierender seine Antwort in der Liste nicht wiederfand, hatte er die Möglichkeit, in das Antwortfeld „weiß nicht“ einzugeben und dann in einem Freitextfeld selbst eine Antwort zu formulieren. Nachdem der Studienteilnehmer eine Antwort gewählt und diese bestätigt hatte, erschien ein neues Fenster mit der richtigen Lösung. Zusätzlich konnte das Feedback „Ausführliche Erläuterungen“ angeklickt werden. Hier wurde die richtige Antwort begründet und die richtigen Antwortmöglichkeiten aufgelistet. Wichtige theoretische und praktische Hinweise für die Lösung wurden erläutert und Hintergrundwissen erweitert. Außerdem wurde für häufige Falschantworten erläutert, warum diese nicht als richtig gewertet wurden. Die häufigsten Falschantworten konnten wir aus der Pilotstudie entnehmen. Danach folgte die nächste *Key-Feature-Frage*. Zum Teil wurden bei einigen Fällen auch Laborbefunde, Röntgenbilder oder EKGs eingebaut.

2.3.3 Lesefall

Jeder Lesefall entsprach inhaltlich exakt dem jeweiligen Fragenfall, der zum gleichen Zeitpunkt der anderen Gruppe präsentiert wurde. Ein Lesefall bestand aus einem Lesetext und aus einem aufrufbaren Fenster mit Hintergrundinformationen. Der Lesetext stellte den Ablauf des Falls dar. Genau wie in den Fragenfällen wurde je nach Fall mit den Symptomen und der Vorgeschichte des Patienten begonnen. Daraufhin wurden körperliche Untersuchungen und Diagnostik bis hin zur Therapie und möglichen Komplikationen beschrieben. Die anklickbaren Hintergrundinformationen enthielten die Informationen, die im zugehörigen Fragenfall in den einzelnen Feedbacks enthalten waren, um vergleichbare Bedingungen für die Studierenden und für die Datenanalyse zu schaffen.

Sowohl im Fragenfall als auch im Lesefall wurde jeweils nach einem Fall eine Einfachwahl-Frage gestellt, die sich auf den gleichen Inhalt bezog wie der präsentierte Fall. Hintergrund dieser Maßnahme war, dass die Studierenden in der Pilotstudie den Wunsch geäußert hatten, mit Einfachwahl-Fragen auf die Klausur vorbereitet zu werden. Es wurde angenommen, dass durch die Integration der Fragen die studentische Motivation gesteigert werden kann.

2.4 Prüfungsrahmen

Die computergestützten E-Fallseminare fanden im Digitalen Prüfungs- und Schulungszentrum (DIPS) der Universitätsmedizin Göttingen statt. Insgesamt befanden sich 150 Computer in den DIPS-Räumen 431 und 434. Um einen Eindruck von den Räumlichkeiten zu erhalten, sind im Anhang 4 zwei Fotos eines Raumes während eines E-Fallseminars aufgeführt.

Die Fälle wurden in das *Item Management System* (IMS) eingegeben, hochgeladen und als Klausur erstellt. Das IMS ist ein Prüfungssystem, mit dessen Hilfe Prüfungsfragen in verschiedenen Formaten konstruiert und zu Klausuren zusammengesetzt werden können. Des Weiteren können Items verwaltet und mit anderen Nutzern geteilt werden (<https://www.ucan-assess.org/cms/de/tools/item-and-exam-management/>). Um die Übersichtlichkeit zu verbessern, wurden unterschiedliche *Long-Menu*-Listen (z. B. Diagnosen, Diagnostik, Therapie) im IMS angelegt. Die Antwortmöglichkeiten wurden vor Beginn der Studie in die entsprechenden Listen eingetragen. Nach dem Eingeben der Fragen- bzw. der Lesefälle und dem Erstellen der Klausuren wurden diese als QTI-Datei (*Question-and-Test-Interoperability-Datei*) zum Geschäftsbereich G3-71 (Applikationen Forschung und Lehre) geschickt. Eine QTI-Datei ist ein Standardformat für den Datenaustausch von Items. Von den Mitarbeitern des Geschäftsbereichs G3-71 wurden die Klausuren in das Campus-Prüfungssystem importiert. Das Campus-Prüfungssystem ist ein Programm, welches für elektronische medizinische Prüfungen entwickelt wurde und die Anzeige verschiedener Fragetypen ermöglicht (<http://www.medizinische-fakultaet-hd.uni-heidelberg.de/CAMPUS-Lehr-und-Pruefungssystem.110009.0.html>). Es können wie in dieser Studie zum Beispiel *Key Features* im Fragetyp *Long-Menu* oder Einfachwahl-Fragen gestellt werden. Des Weiteren bietet das System einen sogenannten Exklusiv-Modus. Das heißt, dass die Studienteilnehmer während der Prüfung keine anderen Programme (wie z. B. das Internet) öffnen konnten. Was dieses System besonders auszeichnet ist, dass es bei Ausfall des Servers, der Datenbank oder des Netzwerkes zu keinem Prüfungsabbruch kommt und dass sämtliche Aktivitäten protokolliert werden, um eine rechtssichere Prüfung zu gewährleisten (<http://www.medizinische-fakultaet-hd.uni-heidelberg.de/CAMPUS-Lehr-und-Pruefungssystem.110009.0.html>). Danach wurde die in das Campus-Prüfungssystem importierte Prüfung auf die Zero-Clients (die Computer) des DIPSs der Universitätsmedizin Göttingen hochgeladen. Mit Hilfe der Management-Konsole wurden die persönlichen Identifikationsnummern (PINs) für den Zugang der Prüfung auf dem Server generiert. Zwar wurde nur eine PIN benötigt, um die Prüfung zu be-

ginnen, jedoch wurde eine Ersatz-PIN zur Sicherheit generiert, damit bei einem Ausfall der ersten PIN eine zweite PIN zur Verfügung stand.

Um technische Probleme frühzeitig zu erkennen, wurde jedes E-Fallseminar von dem Geschäftsbereich G3-71, Prof. Dr. Tobias Raupach, einer wissenschaftlichen Mitarbeiterin und der Autorin dieser Arbeit im Vorfeld des eigentlichen Termins an den Computern im DIPS-Raum als Testlauf geprüft. Durch direkte Rückkopplung an Mitarbeiter des Geschäftsbereichs G3-71 konnten identifizierte Fehler in der Technik, im Inhalt oder in der Eingabe der Fälle rechtzeitig behoben werden. Nach der Durchführung jedes E-Fallseminars wurden die Daten von einem Mitarbeiter des Geschäftsbereichs G3-71 aus dem Campus-Prüfungs-Programm als Excel-Tabelle exportiert und in verschlüsselter Form an den Studienleiter übermittelt.

2.5 Praktische Durchführung

Am Anfang jedes E-Fallseminars meldeten sich die Studienteilnehmer an den Computern über ihren eigenen Uni-Account an. Dies erfolgte über den LDAP-Server (*Lightweight-Directory-Access-Protocol-Server*) zum Abgleich der Identität. Danach wurde zeitgleich in beiden Räumen 431 und 434 die PIN freigegeben. Nach der Eingabe der PIN konnte jeder Studienteilnehmer zeitgleich mit den Lesefällen bzw. mit den *Key-Feature*-Fragen beginnen. Die Studienteilnehmer hatten maximal 45 Minuten Zeit die Prüfung durchzuführen. Das E-Fallseminar konnte individuell auch vor der maximalen Prüfungsdauer von 45 Minuten beendet werden.

Zur Objektivierung des Lernerfolgs wurde in der ersten Semesterwoche ein Eingangstest und in der vorletzten Semesterwoche ein Abschlusstest durchgeführt. Der Eingangstest beinhaltete für beide Studiengruppen 30 *Key-Feature*-Fragen. Diese 30 *Key-Feature*-Fragen wurden anhand ihrer in der Pilotstudie erreichten Item-Charakteristika (Itemschwierigkeiten, Itemtrennschärfen) ausgewählt. Wie in Tabelle 3 dargestellt ist, dienten von den 30 *Key-Feature*-Fragen in jeder Gruppe 15 Fragen als Interventions-Items (*Key-Feature*-Fragen) und 15 Fragen als Kontroll-Items (Lesefälle). Als Interventions-Items wurden die Items bezeichnet, die in den E-Fallseminaren als *Key-Feature*-Frage gestellt wurden und als Kontroll-Items wurden die Items bezeichnet, die in Form von Lesefällen präsentiert wurden. Tabelle 3 bezieht sich hierbei auf die Gruppe A.

Tabelle 3: Interventions- und Kontroll-Items für die Gruppe A

Gruppe A	
Interventions-Items	Kontroll-Items
Erkennen eines nephrotischen Syndroms anhand der Labor- und Urinbefunde	Ableiten der Verdachtsdiagnose „Akute Lungenembolie“ bei typischer Klinik und Risikoprofil
V. a. Systemischer Lupus erythematoses bei typischer Klinik und wegweisendem AK-Befund	Wells-Score zur Entscheidung für oder gegen eine D-Dimer-Bestimmung
Nierenbiopsie zur Diagnosesicherung	sofortiges Thorax-CT bei hoher Wahrscheinlichkeit und klinischer Stabilität
Behandlung der Lupus-Nephritis durch einen Steroidstoß	Echokardiographie oder Troponinbestimmung zur Risikostratifizierung
Verdachtsdiagnose COPD bei chronischem Husten und Auswurf bei einem starken Raucher mit expiratorischem Giemen	Indikation zur Fibrinolyse bei hämodynamischer Instabilität
Erkennen einer Obstruktion in der Lungenfunktion	Verdacht auf sekundäre Hypertonie bei Manifestation >60 J., diskrepanten Organschäden und Therapierefraktärität
Erkennen einer links-apikalen Lobärpneumonie im Röntgenbild	Verdacht auf diastolische Dysfunktion bei guter EF, NYHA II und erhöhtem BNP
CRB-65-Index zur Entscheidung über die Empfehlung zur stationären Aufnahme	ACE-Hemmer bei V. a. ACE-Hemmer-Husten absetzen
Blutgasanalyse (BGA) bei V. a. beginnende CO ₂ -Narkose	Krankenhaus-Einweisung bei Hyponatriämie <120 mM
Non-invasive Ventilation bei nachgewiesener Hyperkapnie	Thiazid-Diuretika als mögliche Auslöser einer Hyponatriämie
laborchemische Diagnose einer manifesten Hyperthyreose nach Kontrastmittelgabe	zentrale pontine Myelinolyse als Komplikation einer zu raschen Natrium-Substitution
Absetzen von Amiodaron bei manifester Hyperthyreose	Schellong-Test bei V. a. orthostatisch bedingte Synkope
klinischer Verdacht auf Lungenfibrose bei Dyspnoe und inspiratorischem Velcro-Knistern	Schädel-CT zur Diagnose eines Hirninfarktes
Amiodaron als Verursacher einer Lungenfibrose	Erkennen einer Tachyarrhythmia absoluta im EKG
Indikationsstellung zur Langzeit-Sauerstofftherapie aufgrund eines pO ₂ < 55 mmHg in der arteriellen BGA	CHA ₂ DS ₂ -VASc-Score zur Indikationsstellung zur dauerhaften Antikoagulation

Für Gruppe B galt, dass die Kontroll-Items der Gruppe A den Interventions-Items der Gruppe B entsprachen und die Interventionsitems der Gruppe A als Kontroll-Items der

Gruppe B fungierten. Die 30 Items wurden zwischen dem Eingangstest und dem Ausgangstest in jeweils zwei E-Fallseminaren wiederholt. Um eine Kontamination zwischen den beiden Gruppen zu vermeiden, war es wichtig, dass die Studierenden zu jedem E-Fallseminar im gleichen Raum saßen. Eine Studiengruppe wurde für die Bearbeitung aller E-Fallseminare dem Raum 431 zugeteilt, die andere Studiengruppe dem Raum 434. Das Präsentationsformat wechselte wöchentlich zwischen den beiden Räumen.

Am 20.01.2014 fand der Abschlusstest statt. Dass das E-Fallseminar aus einem Abschlusstest mit 30 *Key-Feature*-Fragen bestand, war den Studierenden unbekannt, um kein selbstständiges Wiederholen der Inhalte zu induzieren. Die 30 *Key-Feature*-Fragen entsprachen dem Eingangstest.

Im darauffolgendem Semester fand am 14.07.2014 eine computergestützte Übungsklausur für das Modul 4.4 („Erkrankungen der Augen, des Hals-Nasen-Ohrenbereichs, des Mundes und der Zähne“) in den DIPS-Räumen statt, in deren Rahmen ein unangekündigter Retentionstest durchgeführt wurde. Die Übungsklausur für das Modul 4.4 umfasste zwölf Einfachwahl-Fragen. Der Retentionstest bestand aus denselben 30 *Key-Feature*-Fragen des Eingangs- bzw. Abschlusstests. Der Eingangstest, der Abschlusstest und der Retentionstest waren formative Prüfungen.

2.6 Fokusgruppengespräche

Fokusgruppen sind spezielle Gruppen bezüglich ihrer Größe, Merkmale und Absicht. Sie bestehen gewöhnlich aus fünf bis zehn Teilnehmern, die ähnliche Merkmale in Bezug auf das Fokusgruppenthema aufweisen und von einem Moderator auf ein Thema fokussiert werden. Durch eine indirekte Fragetechnik entsteht eine Gruppendynamik, die Wahrnehmungen der Teilnehmer teilen lässt und neue Erkenntnisse zum Fokusgruppenthema hervorruft. Zuerst werden offene Leitfragen gestellt. Offene Leitfragen haben die Funktion, keine Antwortmöglichkeiten vorzugeben, sondern die eigene Meinung der Fokusgruppenteilnehmer zu erfragen. Im zeitlichen Verlauf können spezifischere Fragen folgen, um detaillierter auf Themengebiete einzugehen (Krueger und Casey 2015). Fokusgruppengespräche dienen als wichtige Methode in der qualitativen Forschung (Morgan 1996, Krueger und Casey 2015) und sind zudem eine nützliche Methode, um quantitative Methoden zu ergänzen (Clarke und Yaros 1988; Johnson und Onwuegbuzie 2004; Morgan 1996). Deshalb wurden zur Beantwortung der Studienfragen 5 und 6 nach dem Ausgangstest Fokusgruppengespräche geführt. Es wurde über den Semesterverteiler eine E-Mail geschickt, um zu erfragen, wer an diesen Gesprächen teilnehmen möchte. Über eine

Doodle-Umfrage konnten sich Interessierte für verschiedene Termine eintragen. Die Fokusgruppengespräche fanden im STÄPS (Studentisches Trainingszentrum Ärztlicher Praxis und Simulation) statt und wurden digital aufgenommen. Moderatorin der Gespräche war eine wissenschaftliche Mitarbeiterin, die nicht direkt in den Lehrbetrieb involviert war. Sie hatte die Aufgabe, Leitfragen zu stellen und je nach Gesprächsentwicklung und Diskussionsbereitschaft Zusatzfragen zu stellen oder wegzulassen. Dabei achtete die wissenschaftliche Mitarbeiterin darauf, dass nicht allzu sehr von den Leitfragen abgewichen wurde und trotzdem möglichst alle Aspekte angesprochen wurden. Es gab sieben offene Leitfragen mit jeweiligen Zusatzfragen. Auf die wichtigen Aspekte wurde mit Hilfe der sieben Leitfragen gelenkt. Als Gesprächseinstieg wurde die Frage „Was ist Ihnen von den E-Fallseminaren besonders in Erinnerung geblieben? Was war besonders hervorstechend?“ gestellt, um einerseits herauszufinden, auf was für Themen die Studierenden besonders großen Wert legten und um andererseits den Studierenden die Möglichkeit zu geben selbst zu entscheiden, in welchen Themengebieten am meisten Diskussionsbedarf bestand und in welche Richtung sie das Gespräch lenken wollten. Die Studierenden wurden außerdem zur praktischen Relevanz, zum Lernverhalten, zum Zweck, zur Zeiteinteilung der E-Fallseminare und zum persönlich empfundenen Unterschied zwischen den Präsentationsformaten Lesefall und Fragenfall befragt. Die Zusatzfragen sprachen einzelne Zusammenhänge noch detaillierter an und wurden bei Bedarf gestellt. Im Folgenden sind die Leitfragen aufgeführt. Der gesamte Fragebogen befindet sich im Anhang 5.

- 1) Was ist Ihnen von den E-Fallseminaren besonders in Erinnerung geblieben? Was war besonders hervorstechend?
- 2) Womit haben Sie während der E-Fallseminare die meiste Zeit verbracht?
- 3) Was war in Ihren Augen der Zweck der E-Fallseminare?
- 4) Wie relevant für die Praxis erschienen Ihnen die Inhalte der E-Fallseminare?
- 5) Inwiefern hat die Teilnahme an den E-Fallseminaren Ihr Lernverhalten beeinflusst?
- 6) (falls noch nicht ausreichend beantwortet) Im wöchentlichen Wechsel haben Sie entweder Fragenfälle oder Lesefälle bearbeitet. Was ist für Sie der wesentliche Unterschied zwischen den beiden Formaten?
- 7) Wie könnte das Lehrformat „E-Fallseminar“ in Ihren Augen weiter verbessert werden?

Als Aufwandsentschädigung erhielt jeder Teilnehmer bzw. jede Teilnehmerin einen Büchergutschein im Wert von 20 Euro.

2.7 Auswertung

Für den Einschluss in die Analyse war es unabdingbar, dass die Studierenden in jedem E-Fallseminar im gleichen Raum saßen, damit das Präsentationsformat für jeden Studierenden wöchentlich wechselte. Die einzelnen *Key-Feature*-Fragen entsprachen einzelnen Items. Somit wurden die Datenanalysen auf Fragen-Ebene und nicht auf Fall-Ebene durchgeführt.

2.7.1 Quantitative Datenanalyse

Alle drei studienbezogenen Prüfungen (Eingangstest, Ausgangstest und Retentionstest) bestanden aus jeweils 30 *Key-Feature*-Items, von denen für jeden Teilnehmer 15 Items als Interventionsitems und 15 Items als Kontrollitems fungierten. Nach dem schon beschriebenen *Crossover*-Design entsprachen also die 15 Interventionsitems der Gruppe A den 15 Kontrollitems der Gruppe B und andersherum.

Die deskriptive Analyse wurde im Hinblick auf die demographischen Daten, die E-Fallseminare und den Eingangs-, Ausgangs- und Retentionstest durchgeführt. Zur Auswertung nominalskalierten Daten wurden die Häufigkeiten berechnet, mit denen die unterschiedlichen Merkmalsausprägungen auftraten. Kreuztabellen wurden erstellt und die stochastische Unabhängigkeit zwei nominalskalierten Variablen (z. B. Geschlecht, Gruppenzugehörigkeit) wurde mit Hilfe des Chi-Quadrat-Tests geprüft. Metrische Daten wurden mit Mittelwert \pm Standardabweichung angegeben. Mittels des ungepaarten t-Tests wurden Unterschiede der Mittelwerte von verhältnisskalierten Merkmalen (z. B. Alter, Klausurleistung des Vorsemesters in Prozent) zwischen den Gruppen berechnet. Jedes E-Fallseminar wurde ausgewertet; dabei wurden für jede *Key-Feature*-Frage („Item“) die Itemschwierigkeit und die Itemtrennschärfe berechnet. Für die Berechnung der internen Konsistenz wurde der Cronbachs α -Koeffizient genutzt, wobei Werte von über 0,8 anzustreben sind (Bortz und Döring 2006).

Der primäre Endpunkt der Studie war der Unterschied der erreichten Prozentzahlen zwischen Interventionsitems und Kontrollitems im Retentionstest. Bei der gegebenen Stichprobe von 124 Studierenden war vor dem Hintergrund der in der Pilotstudie gemessenen Mittelwerte und Standardabweichungen ein Unterschied der Effektstärke 0,25 mit einer Power von 0,8 nachweisbar. Zur Berechnung der Effektstärke wurde Cohens d verwendet. Werte von 0,2 entsprechen einer kleinen, Werte von 0,5 einer mittleren und Werte von 0,8 einer großen Effektstärke (Cohen 1992).

Um Studienfrage 1 („Erzeugt die Bearbeitung von wiederholten *Key-Feature*-Fragen einen größeren Lernerfolg hinsichtlich differenzialdiagnostischer und therapeutischer Fähigkeiten als das Lesen von wiederholten inhaltlich gleichen Lesefällen?“) zu beantworten, wurde statistisch das Verfahren des sogenannten ‚dummy codings‘ durchgeführt. Für jeden Studierenden konnten – in Abhängigkeit von seiner Gruppenzugehörigkeit – jeweils 15 Interventions- und 15 Kontrollitems definiert werden. Für die aggregierte Analyse wurde für jeden Studierenden ein Summenscore für die Interventionsitems und ein separater Summenscore für die Kontrollitems gebildet. Für einen Studierenden der Gruppe A wurde der Summenscore aus den Leistungen in den Interventionsitems der Gruppe A gebildet, für einen Studierenden der Gruppe B aus den Interventionsitems seiner Gruppe. Das gleiche Prinzip galt für den Summenscore der Kontrollitems für jeden Studierenden. Entsprechend waren die Scores von der Gruppenzugehörigkeit individueller Studierender unabhängig. Somit konnte die Stichprobe insgesamt betrachtet werden. Mit einem gepaarten t-Test wurden Unterschiede der Summenscores für Interventions- und Kontrollitems untersucht. Um den mittelfristigen Lernerfolg in Studienfrage 1 beurteilen zu können, wurden Daten des Ausgangstests verwendet. Um den langfristigen Lernerfolg zu untersuchen, wurden die Daten des Retentionstests genutzt.

Um die Studienfrage 2 („Welchen Einfluss hat die Lernzeit auf den Zusammenhang zwischen Präsentationsformat (Fragenfälle versus Lesefälle) und Lernerfolg?“) zu beantworten, wurde eine lineare Regression durchgeführt. Zuerst wurde für jeden Studierenden die Differenz (A) zwischen der von den Studierenden verwendeten Durchschnittszeit für die Interventionsitems und der verwendeten Durchschnittszeit für die Kontrollitems berechnet. Zugleich wurde die Differenz (B) zwischen der im Mittel erreichten Punktzahl in den Interventionsitems und der durchschnittlich erreichten Punktzahl in den Kontrollitems bestimmt. Daraufhin wurde eine lineare Regression mit der unabhängigen Variable (Differenz A) und der abhängigen Variable (Differenz B) durchgeführt. Als weitere unabhängige Variablen gingen das Geschlecht und die Klausurleistung des Vorsemesters in die Analyse ein.

Um die Studienfrage 3 („Wie entwickeln sich die Itemschwierigkeiten und -trennschärfen sowie die interne Konsistenz von Fragenfällen im Laufe des Semesters?“) zu beantworten, wurden für alle E-Fallseminare die Mittelwerte der Itemschwierigkeiten und der Itemtrennschärfen sowie die interne Konsistenz (Cronbachs α) berechnet. In einem zweiten Schritt wurden die Veränderungen der Kennwerte Itemschwierigkeit und Trennschärfe bei wiederholtem Einsatz der Interventionsitems im Verlauf dargestellt.

Um die Studienfrage 4 („Lassen sich distinkte Muster des Leistungserhaltes bzw. -verlustes bei wiederholter Bearbeitung von *Key-Feature*-Fragen identifizieren?“) zu beantworten, wurden die Häufigkeiten berechnet, wie oft ein wiederholtes Interventionsitem richtig oder falsch beantwortet wurde, wenn zuvor das gleiche Item richtig oder falsch beantwortet wurde. Die Verläufe des Leistungserhaltes bzw. -verlustes wurden in einem Baumdiagramm dargestellt.

Die quantitative Datenanalyse wurde mit dem Statistikprogramm SPSS (Mac-Version 21.0.0.0, SPSS Inc., Chicago, Illinois, USA) durchgeführt. Das Signifikanzniveau betrug 5%.

2.7.2 Qualitative Datenanalyse

Um die Studienfrage 5 („Wie beurteilen Studierende die elektronischen Fallseminare?“) und die Studienfrage 6 („Inwieweit wurden die Mechanismen des *Testing Effects* in Fokusgruppengesprächen bekräftigt?“) zu beantworten, wurde eine qualitative Inhaltsanalyse nach Mayring durchgeführt (Mayring 2010). Die akustischen Aufzeichnungen der Fokusgruppengespräche wurden transkribiert. Mit Hilfe des Programms MAXQDA (VERBI GmbH, Charlottenburg Innovation Centre, Berlin) wurde der Text in Bezug auf häufig vorkommende Zusammenhänge und Schlagwörter von zwei unterschiedlichen Personen in Hauptkategorien (Codes) und Unterkategorien (Subcodes) kodiert. Anhand der Codes und Subcodes wurde zur Strukturierung und Visualisierung der qualitativen Daten eine Mind Map mit der Mind-Mapping-Software XMind (Version XMind 7.5, XMind Ltd, Hongkong) konstruiert. Zwei Personen entwarfen jeweils eine eigene Mind Map, exportierten diese als pdf-Datei und verglichen die beiden Mind Maps anschließend. Durch Diskutieren, gemeinsames Umstrukturieren, Hinzufügen und Zusammenfassen entstand die finale Mind Map im XMind und wurde daraufhin als pdf-Datei exportiert.

3 Ergebnisse

3.1 Stichprobe

In Abbildung 4 sind das Studiendesign und die Entwicklung der Stichprobe dargestellt. Insgesamt waren 124 Studierende zu den drei Modulen angemeldet. Von diesen gaben vier Studierende kein Einverständnis für die Teilnahme an der Studie. Drei Studierende waren beim Eingangstest nicht anwesend. Während der zehn E-Fallseminare kam es zur Kontamination der Studiengruppen. Konkret bedeutete „Kontamination“ hier, dass insgesamt zwölf Teilnehmer mindestens einmal im falschen Raum saßen und nicht die vorgesehene Intervention erhielten. Diese Kontaminationen hätten die Ergebnisse verfälscht, da Items getestet wurden, die eigentlich hätten gelesen werden müssen und andersherum. Die Daten dieser Studierenden wurden nachträglich von der Analyse ausgeschlossen. Zwei Studierende nahmen nicht am Ausgangstest teil, was zu einer Stichprobe von 103 Studierenden zum Ausgangstest führte. Beim Retentionstest fehlten 16 weitere Studierende, sodass sich für die gesamte Studie eine Stichprobe von 87 Studierenden ergab. Somit betrug die Rücklaufquote für Studierende mit kompletten longitudinalen Daten 70,2%.

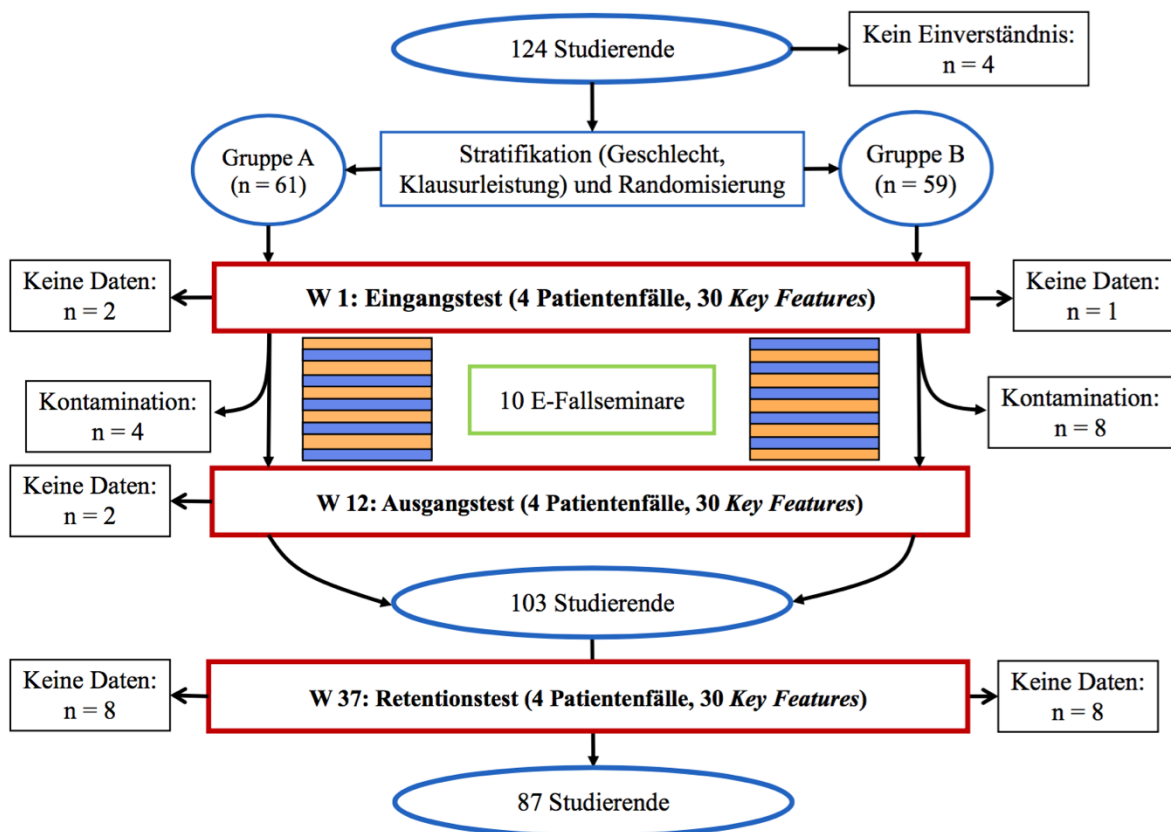


Abbildung 4: Entwicklung der Stichprobe, W=Woche

3.2 Deskriptive Ergebnisse

3.2.1 Beschreibung der Stichprobe

Tabelle 4: Beschreibung der Stichprobe insgesamt und randomisiert auf die Gruppen A und B, n = Anzahl der Studierenden

Stichprobe, n=87		Insgesamt	Gruppe A	Gruppe B	p-Wert
Geschlecht in n (%)	weiblich	51 (58,6%)	25 (55,6%)	26 (61,9%)	0,548
	männlich	36 (41,1%)	20 (44,4%)	16 (38,1%)	
Alter in Jahren		25,0 ± 2,9	25,4 ± 3,3	24,6 ± 2,5	0,202
Klausurleistung des Vorsemesters in %		86,4 ± 7,4	85,7 ± 8,8	87,2 ± 5,5	0,372

Die Studiengruppen unterschieden sich nicht signifikant voneinander.

3.2.2 Anwesenheit

Die Anwesenheit der Studierenden war hoch. Bei insgesamt 10 E-Fallseminaren fehlte kein Studierender mehr als zwei Mal. Da für die Berechnungen die Stichprobengröße von 87 Studierenden verwendet wurde, waren während des Eingangs-, Ausgangs- und Retentionstests 100% der Studierenden anwesend.

3.2.3 Gruppenweise Betrachtung der Ergebnisse

Beide Gruppen wurden zuerst unabhängig voneinander betrachtet. Bei beiden Gruppen bestand ein signifikanter Unterschied zwischen den Summenscores der Interventions- und Kontrollitems im Ausgangstest. In den 15 Interventionsitems der Gruppe A wurden von Studierenden in Gruppe A im Durchschnitt $9,6 \pm 4,5$ Punkte und von Studierenden in Gruppe B $7,6 \pm 3,6$ Punkte erreicht ($p = 0,027$). In den Interventionsitems der Gruppe B wurden von Studierenden in Gruppe B $8,1 \pm 3,6$ Punkte und von Studierenden in Gruppe A $6,4 \pm 3,8$ Punkte erreicht ($p = 0,035$).

Im Retentionstest gab es keinen signifikanten Unterschied zwischen Interventions- und Kontrollitems, wenn die Gruppen unabhängig voneinander betrachtet wurden. Studierende der Gruppe A erreichten von den maximal 15 Punkten der Interventionsitems ihrer eigenen Gruppe $9,1 \pm 3,8$ Punkte (Gruppe B: $7,8 \pm 3,7$; $p = 0,097$). In den Interventionsitems der Gruppe B wurden von Studierenden der Gruppe A $6,9 \pm 3,7$ Punkte und von Studierenden der Gruppe B $7,6 \pm 3,9$ Punkte erreicht ($p = 0,369$).

In den Tabellen 5 und 6 ist aufgezeigt, wie viele Studierende die Interventionsitems der Gruppe A beziehungsweise der Gruppe B im Eingangs-, Ausgangs- und Retentionstest richtig beantworteten. Insgesamt ist zu sehen, dass die Studierenden, die die Interventionsitems in den E-Fallseminaren mit Hilfe von *Key-Feature-Fragen* beantworteten, besser im Ausgangs- und Retentionstest abschnitten als Studierende, die die gleichen Items mit Hilfe von Lesefällen in den E-Fallseminaren bearbeiteten. Am Beispiel des Interventionsitems der Gruppe A „Erkennen einer Obstruktion in der Lungenfunktion“ (gelb markiert) ist erkennbar, dass, nachdem das Item Lungenfunktion im Laufe des Semesters zweimal getestet worden war, 68,9% der Studierenden die Lungenfunktion im Ausgangstest richtig interpretierten. Im Gegensatz dazu konnten nach zweimaligem Lesen der gleichen Inhalte während des Semesters nur 35,7% der Studierenden die Lungenfunktion richtig interpretieren ($p = 0,002$). Dasselbe Item hatten im Retentionstest 51,1% der Studierenden der Gruppe A (getestet) und 28,6% der Studierenden der Gruppe B (gelesen) richtig beantwortet.

Tabelle 5: Prozentsatz der Studierenden (n= Anzahl) mit einer richtigen Antwort in den Interventionsitems der Gruppe A im Eingangs-, Ausgangs- und Retentionstest

Interventionsitems Gruppe A	Eingangstest			Ausgangstest			Retentionstest		
	Gruppe A (n=45)	Gruppe B (n=42)	p-Wert	Gruppe A (n=45)	Gruppe B (n=42)	p-Wert	Gruppe A (n=45)	Gruppe B (n=42)	p-Wert
Erkennen eines nephrotischen Syndroms anhand der Labor- und Urinbefunde	17,8 (8)	14,3 (6)	0,658	75,6 (34)	59,5 (25)	0,110	62,2 (28)	57,1 (24)	0,629
V. a. Systemischer Lupus erythematodes bei typischer Klinik und wegweisendem AK-Befund	0,0 (0)	9,5 (4)	0,034	71,1 (32)	59,5 (25)	0,256	75,6 (34)	64,3 (27)	0,251
Nierenbiopsie zur Diagnosesicherung	4,4 (2)	33,3 (14)	0,001	71,1 (32)	59,5 (25)	0,256	71,1 (32)	69,0 (29)	0,834
Behandlung der Lupus-Nephritis durch einen Steroidstoß	31,1 (14)	33,3 (14)	0,825	71,1 (32)	64,3 (27)	0,496	71,1 (32)	73,8 (31)	0,778
Verdachtsdiagnose COPD bei chronischem Husten und Auswurf bei einem starken Raucher mit expiratorischem Giemen	37,8 (17)	52,4 (22)	0,171	66,7 (30)	76,2 (32)	0,327	75,6 (34)	71,4 (30)	0,663
Erkennen einer Obstruktion in der Lungenfunktion	22,2 (10)	11,9 (5)	0,203	68,9 (31)	35,7 (15)	0,002	51,1 (23)	28,6 (12)	0,032
Erkennen einer links-apikalen Lobärpneumonie im Röntgenbild	20,0 (9)	7,1 (3)	0,082	42,2 (19)	14,3 (6)	0,004	48,9 (22)	21,4 (9)	0,008
CRB-65-Index zur Entscheidung über die Empfehlung zur stationären Aufnahme	0,0 (0)	2,4 (1)	0,298	55,6 (25)	45,2 (19)	0,336	53,3 (24)	42,9 (18)	0,328
BGA bei V. a. beginnende CO ₂ -Narkose	35,6 (16)	31,0 (13)	0,649	71,1 (32)	64,3 (27)	0,496	60,0 (27)	57,1 (24)	0,787
Non-invasive Ventilation bei nachgewiesener Hyperkapnie	20,0 (9)	7,1 (3)	0,082	53,3 (24)	28,6 (12)	0,019	42,2 (19)	21,4 (9)	0,038
laborchemische Diagnose einer manifesten Hyperthyreose nach Kontrastmittelgabe	64,4 (29)	59,5 (25)	0,636	86,7 (39)	71,4 (30)	0,080	82,2 (37)	81,0 (34)	0,879
Absetzen von Amiodaron bei manifester Hyperthyreose	2,2 (1)	9,5 (4)	0,144	57,8 (26)	40,5 (17)	0,107	51,1 (23)	40,5 (17)	0,320
klinischer V. a. Lungenfibrose bei Dyspnoe und inspiratorischem Velcro-Knistern	11,1 (5)	2,4 (1)	0,108	53,3 (24)	38,1 (16)	0,154	44,4 (20)	42,9 (18)	0,881
Amiodaron als Verursacher einer Lungenfibrose	17,8 (8)	16,7 (7)	0,891	57,8 (26)	54,8 (23)	0,777	75,6 (34)	71,4 (30)	0,663
Indikationsstellung zur Langzeit-Sauerstofftherapie aufgrund eines pO ₂ < 55 mmHg in der arteriellen BGA	46,7 (21)	45,2 (19)	0,894	57,8 (26)	50,0 (21)	0,467	48,9 (22)	35,7 (15)	0,214

Tabelle 6: Prozentsatz der Studierenden (n= Anzahl) mit einer richtigen Antwort in den Interventionsitems der Gruppe B im Eingangs-, Ausgangs- und Retentionstest

Interventionsitems Gruppe B	Eingangstest			Ausgangstest			Retentionstest		
	Gruppe A (n=45)	Gruppe B (n=42)	p-Wert	Gruppe A (n=45)	Gruppe B (n=42)	p-Wert	Gruppe A (n=45)	Gruppe B (n=42)	p-Wert
Ableiten der Verdachtsdiagnose „Akute Lungenembolie“ bei typischer Klinik und Risikoprofil	51,1 (23)	42,9 (18)	0,441	48,9 (22)	59,5 (25)	0,320	64,4 (29)	57,1 (24)	0,486
Wells-Score zur Entscheidung für oder gegen eine D-Dimer-Bestimmung	4,4 (2)	9,5 (4)	0,350	60,0 (27)	76,2 (32)	0,106	51,1 (23)	76,2 (32)	0,015
sofortiges Thorax-CT bei hoher Wahrscheinlichkeit und klinischer Stabilität	8,9 (4)	11,9 (5)	0,644	55,6 (25)	42,9 (18)	0,236	44,4 (20)	26,2 (11)	0,076
Echokardiographie oder Troponinbestimmung zur Risikostratifizierung	2,2 (1)	0,0 (0)	0,331	22,2 (10)	45,2 (19)	0,023	24,4 (11)	42,9 (18)	0,069
Indikation zur Fibrinolyse bei hämodynamischer Instabilität	42,2 (19)	45,2 (19)	0,777	51,1 (23)	59,5 (25)	0,430	37,8 (17)	54,8 (23)	0,112
Verdacht auf sekundäre Hypertonie bei Manifestation >60 J., diskrepanten Organschäden und Therapierefraktärität	0,0 (0)	2,4 (1)	0,298	20,0 (9)	40,5 (17)	0,037	24,4 (11)	35,7 (15)	0,251
Verdacht auf diastolische Dysfunktion bei guter EF, NYHA II und erhöhtem BNP	2,2 (1)	0,0 (0)	0,331	20,0 (9)	21,4 (9)	0,869	17,8 (8)	23,8 (10)	0,488
ACE-Hemmer bei V. a. ACE-Hemmer-Husten absetzen	64,4 (29)	61,9 (26)	0,806	82,2 (37)	88,1 (37)	0,443	86,7 (39)	73,8 (31)	0,131
Krankenhaus-Einweisung bei Hyponatriämie < 120 mmol/l	2,2 (1)	2,4 (1)	0,961	6,7 (3)	31,0 (13)	0,003	26,7 (12)	35,7 (15)	0,362
Thiazid-Diuretika als mögliche Auslöser einer Hyponatriämie	40,0 (18)	31,0 (13)	0,379	60,0 (27)	78,6 (33)	0,061	60,0 (27)	59,5 (25)	0,964
zentrale pontine Myelinolyse als Komplikation einer zu raschen Natrium-Substitution	0,0 (0)	2,4 (1)	0,298	17,8 (8)	40,5 (17)	0,019	15,6 (7)	31,0 (13)	0,088
Schellong-Test bei V. a. orthostatisch bedingte Synkope	22,2 (10)	31,0 (13)	0,356	46,7 (21)	54,8 (23)	0,450	55,6 (25)	61,9 (26)	0,548
Schädel-CT zur Diagnose eines Hirninfarktes	62,2 (28)	59,5 (25)	0,797	66,7 (30)	66,7 (28)	1,000	62,2 (28)	64,3 (27)	0,842
Erkennen einer Tachyarrhythmia absoluta im EKG	37,8 (17)	23,8 (10)	0,159	33,3 (15)	38,1 (16)	0,643	46,7 (21)	64,3 (27)	0,099
CHA2DS2-VASc-Score zur Indikationsstellung zur dauerhaften Antikoagulation	11,1 (5)	11,9 (5)	0,908	53,3 (24)	71,1 (30)	0,082	71,1 (32)	54,8 (23)	0,114

3.2.4 Itemkennwerte und Cronbachs α

Die Itemkennwerte (Itemschwierigkeit und Itemtrennschärfe) wurden für jedes Item innerhalb eines E-Fallseminars berechnet (siehe Anhang 6). In Tabelle 7 sind die pro E-Fallseminar gemittelten Itemkennwerte und das jeweilige Cronbachs α angegeben. Die interne Konsistenz der E-Fallseminare, ausgedrückt als Cronbachs α , reichte von 0,56 bis 0,92. In sieben von 13 E-Fallseminaren betrug dieses untere Grenzmaß der Reliabilität $> 0,8$. Für drei E-Fallseminare lag das Cronbachs α zwischen 0,75 und 0,78. Die aus 30 Items bestehenden studienbezogenen formativen Prüfungen (Eingangs-, Ausgangs- und Retentionstest) wiesen besonders hohe interne Konsistenzen auf.

Tabelle 7: Mittelwerte der Itemschwierigkeiten und Itemtrennschärfen für die E-Fallseminare und Darstellung des Cronbachs α

E-Fallseminar	Itemschwierigkeiten			Itemtrennschärfen			Cronbachs α
	Mittelwert \pm Standard- abweichung	Minimum	Maximum	Mittelwert \pm Standard- abweichung	Minimum	Maximum	
Eingangstest	0,22 \pm 0,19	0,01	0,61	0,193 \pm 0,122	-0,100	0,442	0,67
2	0,66 \pm 0,28	0,03	0,97	0,348 \pm 0,120	0,097	0,519	0,78
3	0,39 \pm 0,25	0,04	0,90	0,185 \pm 0,199	-0,021	0,495	0,56
4	0,64 \pm 0,28	0,12	0,95	0,363 \pm 0,134	0,092	0,602	0,78
5	0,46 \pm 0,23	0,11	0,86	0,215 \pm 0,173	-0,141	0,462	0,61
6	0,60 \pm 0,23	0,07	0,87	0,410 \pm 0,119	0,157	0,607	0,83
7	0,60 \pm 0,18	0,19	0,94	0,314 \pm 0,127	0,053	0,533	0,75
8	0,56 \pm 0,23	0,11	0,85	0,439 \pm 0,138	0,122	0,647	0,85
9	0,39 \pm 0,23	0,02	0,82	0,351 \pm 0,188	-0,135	0,608	0,81
10	0,49 \pm 0,17	0,14	0,80	0,580 \pm 0,143	0,310	0,812	0,92
11	0,55 \pm 0,20	0,12	0,87	0,425 \pm 0,163	0,015	0,661	0,84
Ausgangstest	0,53 \pm 0,17	0,20	0,84	0,473 \pm 0,092	0,239	0,622	0,91
Retentionstest	0,50 \pm 0,16	0,19	0,80	0,464 \pm 0,124	0,107	0,657	0,90

3.2.5 Fokusgruppen

Es entstanden drei Fokusgruppen. Von den insgesamt 16 Teilnehmern waren acht Studierende weiblich und acht Studierende männlich. In der folgenden Tabelle 8 sind die Fokusgruppen beschrieben.

Tabelle 8: Beschreibung der Fokusgruppen

Fokusgruppe	Datum, Beginn	Dauer in Stunden	Teilnehmeranzahl
Gruppe 1	05.02.2014, 10 Uhr	1:12:41	6
Gruppe 2	06.02.2014, 14 Uhr	1:05:11	6
Gruppe 3	06.02.2014, 16 Uhr	0:52:02	4

3.3 Beantwortung der Studienfragen

3.3.1 Studentischer Lernerfolg in Abhängigkeit vom Präsentationsformat

Um die Studienfrage 1 („Erzeugt die Bearbeitung von wiederholten *Key-Feature*-Fragen einen größeren Lernerfolg hinsichtlich differenzialdiagnostischer und therapeutischer Fähigkeiten als das Lesen von wiederholten inhaltlich gleichen Lesefällen?“) zu beantworten, wurden die Interventions- und Kontrollitems unabhängig von den Studiengruppen betrachtet.

Insgesamt wurden im Durchschnitt im Eingangstest $22,6 \pm 11,3\%$, im Ausgangstest $53,0 \pm 24,4\%$ und im Retentionstest $52,4 \pm 23,4\%$ der maximal möglichen 30 Punkte erreicht. In Abbildung 5 sind die im Durchschnitt erreichten Prozentzahlen im Eingangs-, Ausgangs- und Retentionstest dargestellt. Im Eingangstest wurden gruppenübergreifend $22,2 \pm 12,9\%$ der maximal möglichen Punkte für die Interventionsitems und $22,9 \pm 12,9\%$ der maximal möglichen Punkte für die Kontrollitems erreicht. Im Ausgangstest wurden in den Interventionsitems im Mittel $59,3 \pm 27,7\%$ ($8,9 \pm 4,2$ von 15 Punkten) erreicht und in den Kontrollitems $46,7 \pm 24,8\%$ ($7,01 \pm 3,72$ von 15 Punkten). In den Interventionsitems wurden im Mittel $1,9 \pm 3,0$ Punkte mehr erreicht als in den Kontrollitems ($p < 0,001$). Auch im Retentionstest, sechs Monate nach dem Abschlusstest, konnte ein signifikanter Unterschied zwischen Interventionsitems und Kontrollitems nachgewiesen werden. In den Interventionsitems wurden durchschnittlich $1,1 \pm 2,8$ Punkte mehr erreicht als in den Kontroll-

items ($p < 0,001$). Von den Studierenden wurden $56,0 \pm 25,8\%$ ($8,4 \pm 3,9$ von 15 Punkten) in den Interventionsitems erreicht, während in den Kontrollitems nur $48,8 \pm 24,7\%$ ($7,3 \pm 3,7$ von 15 Punkten) erzielt wurden.

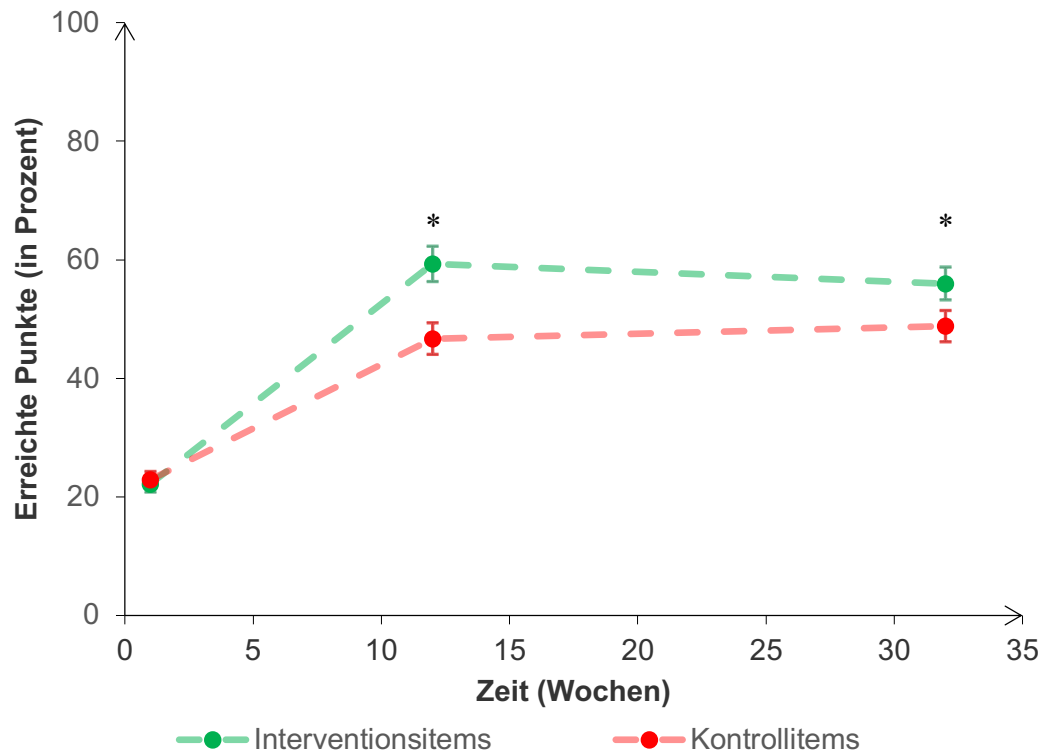


Abbildung 5: Durchschnittlich erreichte Punktzahlen (in Prozent) im Eingangs-, Ausgangs- und Retentionstest. Die Fehlerbalken stellen die Standardfehler dar. * $p < 0,001$ im gepaarten t-Test zur Untersuchung der Leistungsunterschiede zwischen Interventions- und Kontrollitems.

Somit wurden für Items, die mittels wiederholter formativer *Key-Feature*-Prüfungen trainiert worden waren, auch sechs Monate später noch signifikant bessere Leistungen beobachtet als für Items, die mittels wiederholter Exposition gegenüber Lesefällen trainiert worden waren.

3.3.2 Einfluss der Lernzeit auf den primären Endpunkt

Für die Studienfrage 2 („Welchen Einfluss hat die Lernzeit auf den Zusammenhang zwischen Präsentationsformat (Fragenfälle versus Lesefälle) und Lernerfolg?“) wurde eine lineare Regressionsanalyse durchgeführt.

Die durchschnittliche Login-Zeit bei den E-Fallseminaren lag bei $17:46 \pm 04:44$ Minuten. Hierbei fiel auf, dass die Studierenden im Mittel 8:17 Minuten mehr Zeit für E-

Fallseminare mit *Key-Feature*-Fragen als für E-Fallseminare mit Lesefällen benötigten ($21:47 \pm 05:27$ Minuten bzw. $13:30 \pm 05:24$ Minuten; $p < 0,001$).

Im Ausgangstest fand sich keine Korrelation zwischen dem Zeitüberschuss bei der Bearbeitung von Fragenfällen im Vergleich zu Lesefällen und der Differenz der erreichten Punktzahlen zwischen Interventionsitems und Kontrollitems ($r = 0,038$; $p = 0,724$). Im Retentionstest war eine Korrelation zwischen dem Zeitüberschuss bei der Bearbeitung von Fragenfällen im Vergleich zu Lesefällen und der Differenz der erreichten Punktzahlen zwischen Interventionsitems und Kontrollitems ($r = 0,273$; $p = 0,011$) festzustellen.

In der adjustierten Regressionsanalyse blieb der Haupteffekt (Zusammenhang zwischen Präsentationsformat und Prüfungsleistung) nach Adjustierung für das Geschlecht und die Leistung in den Modulklausuren des Wintersemesters 2013/2014 signifikant (Achsenabschnitt im Ausgangstest: $-8,40$, 95% Konfidenzintervall: $-13,71$ bis $-3,09$; $p=0,002$; Achsenabschnitt im Retentionstest: $-7,03$, 95% Konfidenzintervall: $-12,03$ bis $-2,03$; $p=0,006$).

3.3.3 Zeitliche Entwicklung der Testgütekriterien

Bezüglich der dritten Studienfrage („Wie entwickeln sich die Itemschwierigkeiten und -trennschärfen sowie die interne Konsistenz von Fragenfällen im Laufe eines Semesters?“) wurden die Itemkennwerte der Interventionsitems in den zehn E-Fallseminaren zwischen dem Eingangs- und dem Ausgangstest betrachtet. Tabelle 7 ist zu entnehmen, dass die gemittelten Itemschwierigkeiten der zehn E-Fallseminare in einem Bereich von $0,39 \pm 0,23$ bis $0,66 \pm 0,28$ lagen, wobei acht von zehn gemittelten Itemschwierigkeiten Werte um 0,5 aufwiesen. Außerdem ist festzustellen, dass im Laufe der Zeit die Itemtrennschärfen (und daher auch das Cronbachs α der gesamten Prüfungen) tendenziell zunahm. In Abbildung 6 sind die gemittelten Itemschwierigkeiten der Interventionsitems im zeitlichen Verlauf dargestellt. Jedes Interventionsitem wurde zweimal („Test 1“ und „Test 2“) im Zeitraum zwischen dem Eingangs- und dem Ausgangstest getestet. Der erste Test und der zweite Test jedes Items fanden nicht in den gleichen, sondern in jeweils unterschiedlichen E-Fallseminaren statt. Zum Beispiel wurde „Test 1“ des Items „Nierenbiopsie zur Diagnosesicherung“ im achten E-Fallseminar und „Test 2“ des Items „Nierenbiopsie zur Diagnosesicherung“ im zehnten E-Fallseminar durchgeführt, wohingegen „Test 1“ des Items „Absetzen von Amiodaron bei manifester Hyperthyreose“ im vierten E-Fallseminar und „Test 2“ des Items „Absetzen von Amiodaron bei manifester Hyperthyreose“ im achten E-Fallseminar stattfand. In Abbildung 6 werden die ersten Tests aller Items unter „Test 1“ und die zweiten Tests aller Items unter „Test 2“ zusammengefasst. Somit beschreibt der

Datenpunkt zum Zeitpunkt „Test 1“ den Mittelwert der Itemschwierigkeiten des jeweiligen ersten Tests aller Interventionsitems. Der Datenpunkt zum Zeitpunkt „Test 2“ der Abbildung 6 gibt den Mittelwert der Itemschwierigkeiten des jeweiligen zweiten Tests aller Interventionsitems an.

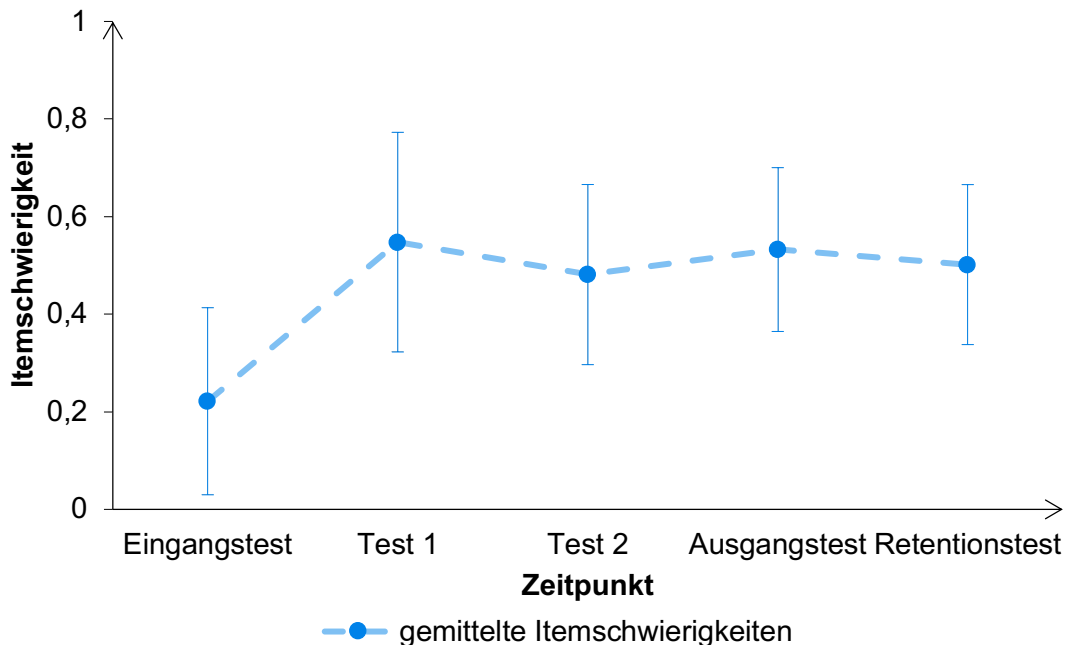


Abbildung 6: Zeitlicher Verlauf der gemittelten Itemschwierigkeiten der Interventionsitems. Die Fehlerbalken stellen die Standardabweichungen dar.

Es ist zu erkennen, dass die gemittelten Itemschwierigkeiten im „Test 1“ im Vergleich zum Eingangstest stark anstiegen. Der Wert sank im „Test 2“ und stieg erneut im Ausgangstest. Im Retentionstest fiel der Wert etwas ab. Insgesamt fällt hier auf, dass die gemittelten Itemschwierigkeiten um einen Wert von 0,5 lagen.

3.3.4 Studentisches Antwortverhalten in Abhängigkeit von der Leistung in vorangegangenen E-Fallseminaren

Um die vierte Studienfrage zu beantworten („Lassen sich distinkte Muster des Leistungserhaltes bzw. -verlustes bei wiederholter Bearbeitung von *Key-Feature*-Fragen identifizieren?“), wurde ein Baumdiagramm erstellt. Das Baumdiagramm (siehe Abbildung 7) beschreibt, wie häufig eine richtige bzw. falsche Antwort gegeben wurde, wenn das gleiche Item im vorherigen Test richtig bzw. falsch beantwortet wurde.

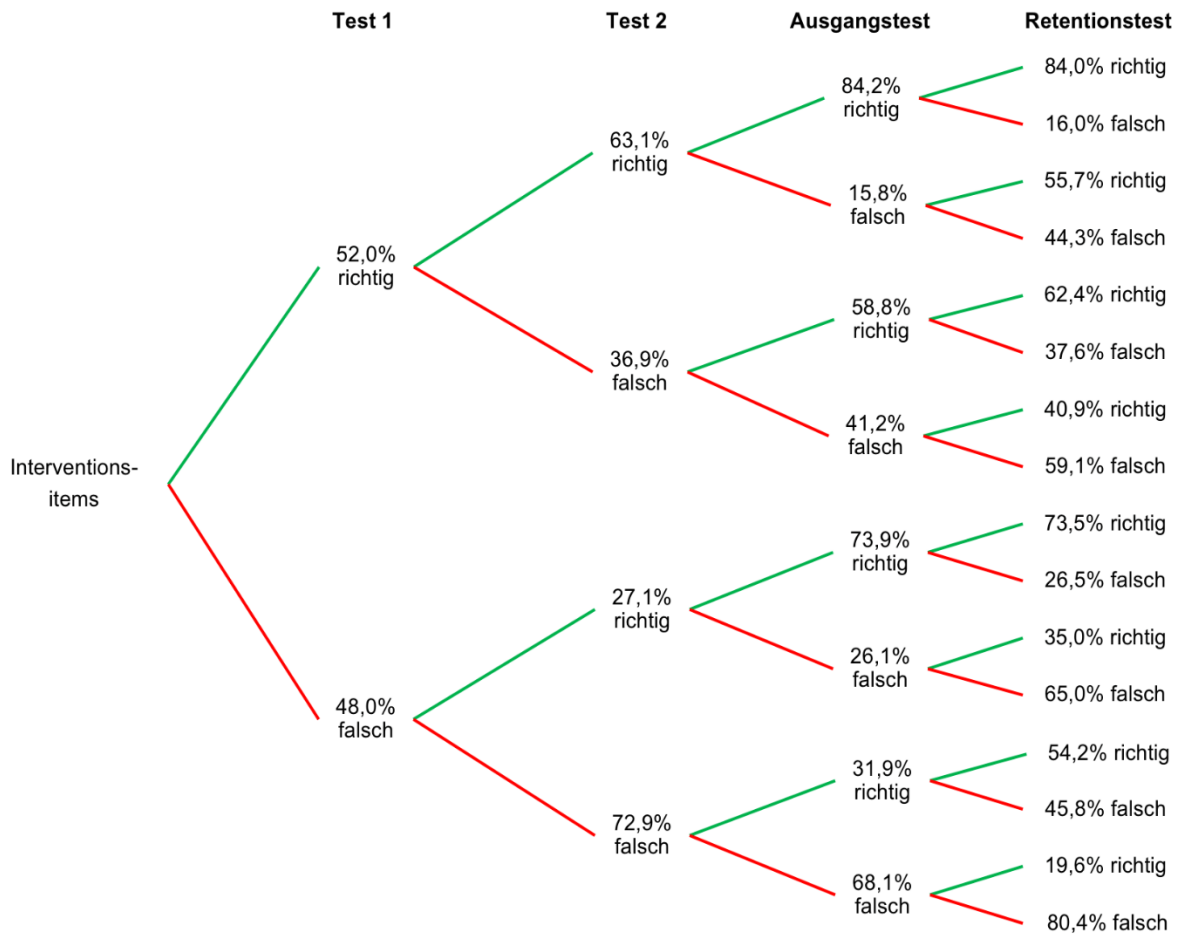


Abbildung 7: Baumdiagramm der Häufigkeiten in Prozent richtig bzw. falsch beantworteter Interventionsitems. Test 1= erstes Auftauchen des Items im E-Fallseminar, Test 2 = zweites Auftauchen des Items im E-Fallseminar.

Es ist zu erkennen, dass es nach der richtigen Beantwortung einer Frage tendenziell eher zu einem Erhalt des abgefragten Wissens kam, obwohl die jeweiligen Inhalte nach dem ersten Auftauchen des Items im E-Fallseminar nicht mehr gelehrt wurden. Items, die Studierende falsch beantworteten, wurden in jeweils nächsten Test in 20-60% der Fälle richtig beantwortet. Wenn ein Item von Studierenden mehrmals falsch beantwortet wurde, wurde es teilweise im Ausgangstest und Retentionstest richtig beantwortet. Seltener wurde ein Item falsch beantwortet, nachdem es zuvor richtig gelöst wurde.

Insgesamt ist festzustellen, dass wenn ein Item richtig beantwortet wurde, das gleiche Item im darauffolgenden Test in 72,0% der Fälle auch richtig beantwortet wurde. In diesem Zusammenhang ist besonders wichtig, dass falsch beantwortete Items in der konsekutiven Prüfung in 33,3% der Fälle richtig beantwortet wurden.

3.3.5 Ergebnisse der qualitativen Begleitstudie

Zur Beantwortung der fünften Studienfrage („Wie beurteilen Studierende die elektronischen Fallseminare?“) wurden die Fokusgruppengespräche mit den Studierenden qualitativ ausgewertet. Aus den studentischen Antworten auf die Leitfragen entwickelte sich eine Mind Map (siehe Anhang 7) mit fünf Hauptthemen: Präsentationsformate, Lernen, Implementierung, Inhalt und Optimierung. Im Folgenden wird auf jedes dieser Oberthemen näher eingegangen. Dabei werden insbesondere die häufig genannten Aspekte vorgestellt.

- **Präsentationsformate**

Die Präsentationsformate „Fragenfall“ und „Lesefall“ wurden diskutiert. Eines der Hauptargumente, welches für das Präsentationsformat „Fragenfall“ sprach, war das aktive Nachdenken, welches während der E-Fallseminare von den Studierenden wahrgenommen wurde. Nach Ansicht der Studierenden festigte das Nachdenken für das Finden einer Antwort das Wissen mehr als das Lesen des gleichen Inhalts.

Zitat: „Also bei mir blieb das mit den Fragen auf jeden Fall mehr hängen, weil man sich aktiv was überlegt und drüber nachdenkt, was da jetzt richtig sein könnte und dann auch vielleicht enttäuscht ist, wenn's dann nicht so richtig ist und das bleibt einfach mehr hängen.“

(weiblich, insgesamt elf ähnliche Aussagen und 28 Zustimmungen)

Im Vergleich von Fragenfällen und Lesefällen fiel den Studierenden auf, dass die Fragenfälle motivierten. Bei den Lesefällen sprachen die Studierenden von einem Konzentrations- und Motivationsverlust.

Zitat: „Nur zu lesen, fand ich irgendwie immer so ein bisschen langweilig.“

(weiblich, insgesamt sechs ähnliche Aussagen und sieben Zustimmungen)

Auffallend war, dass die Studierenden einen direkten Zusammenhang zwischen den Präsentationsformaten und ihrem eigenen Lernerfolg beobachteten. Die E-Fallseminare und insbesondere die Fragenfälle erforderten Konzentration und regten Emotionen an.

Zitat: „Also die E-Fallseminare hatten total viel mit Motivation und Lernerfolg und auch Misserfolg zu tun. Es ist auf jeden Fall für mich emotionaler als diese Kreuzfragen.“

(weiblich, insgesamt neun ähnliche Aussagen und elf Zustimmungen)

Laut den Aussagen der Studierenden mussten sie sich bei den *Key-Feature-Fragen* im Vergleich zu Einfachwahl-Fragen aktiv mit dem Lernstoff auseinandersetzen und zu einer eigenen Entscheidung kommen, da es keine vorgegebenen Antwortmöglichkeiten gab.

Zitat: „Ich fand’s vor Allem gut, dass man sozusagen aktiv zum Nachdenken gebracht wurde. Wenn man fünf Antwortmöglichkeiten hat, dann muss man ja eigentlich nur das passive Wissen anzapfen.“

(männlich, insgesamt fünf ähnliche Aussagen und 14 Zustimmungen)

- **Lernen**

Das Thema Lernen gliederte sich in der Mind Map in drei Hauptäste (Lernprozess, Lernerfolg und Lernverhalten). Nach Ansicht der Studierenden rief das Bearbeiten der Fragenfälle und das Wiederholen von Inhalten eine Sicherheit im Vorgehen zum Beantworten weiterer Fragen hervor.

Zitat: „Also mir hat es so ne Routine gebracht.“

(weiblich, insgesamt fünf ähnliche Aussagen und vier Zustimmungen)

Die Studierenden hatten den Eindruck, dass die Fragenfälle eher zu einem Wissenserhalt und –zuwachs als die Einfachwahl-Fragen führten. Nach den Meinungen der Studierenden gab es zwei Gründe für das schlechtere Merken von Inhalten bei der Bearbeitung von Einfachwahl-Fragen im Vergleich zu Fragenfällen. Einerseits wurde der strukturelle Aufbau von Einfachwahl-Fragen mit vorgegebenen Antwortmöglichkeiten bemängelt und andererseits das Prüfen von Faktenwissen kritisiert.

Zitat: „So konnte ich mir das eigentlich viel besser merken als wenn man jetzt bei den Multiple-Choice-Fragen sich fünf Fakten durchliest und das bleibt mir dann nicht so gut im Gedächtnis, wie wenn man halt selber einfach versucht diesen Lösungsweg zu finden.“

(weiblich, insgesamt vier ähnliche Aussagen und elf Zustimmungen)

Ein weiterer von den Studierenden positiv wahrgenommener Effekt der Fragenfälle war die ständige Wissenskontrolle durch das nachfolgende Feedback. Wenn Studierende annehmen gewisse Inhalte zu beherrschen, konnten sie besonders durch die Fragenfälle sich selbst prüfen, ob die Inhalte wirklich verstanden waren. Nach Ansicht der Studierenden erfolgte durch das Feedback eine Korrektur, welche als sehr hilfreich für den Lernerfolg wahrgenommen wurde.

Zitat: „Man hat immer direkt Feedback gekriegt, ob man das gerade kann oder nicht und dann war einem das viel mehr bewusst.“

(weiblich, insgesamt vier ähnliche Aussagen und acht Zustimmungen)

Studierende differenzierten zwischen den Auswirkungen von Richtig- oder Falschantworten. Nach der Auffassung der Studierenden hatten sie bei Richtigantworten ein kleines Erfolgserlebnis, bei Falschantworten wurden Wissenslücken aufgedeckt. Eine wichtige Rolle spielte laut den Studierenden in diesem Zusammenhang das Feedback, da die Studierenden die Möglichkeit sahen, direkt auf ihre Fehler eine Erklärung zu erhalten, warum die Frage falsch beantwortet wurde.

Zitat: „Ich habe dann auch daraus gelernt, wenn ich vor allen Dingen oft was falsch gemacht habe und mir dann in diesen Begründungen das noch mal durchlesen konnte und mir das dann aber schlüssig geworden ist, was das für ein Fehler von mir war oder warum das jetzt richtig ist.“

(weiblich, insgesamt vier ähnliche Aussagen und drei Zustimmungen)

Die Studierenden vertraten den Standpunkt, dass ihr Lernerfolg unter anderem durch das Hervorrufen von Emotionen zustande kam. Insbesondere Emotionen wie Ärger waren laut Aussagen der Studierenden ein wichtiger Faktor für das Merken von Inhalten.

Zitat: „Wenn man zweimal das Gleiche falsch gemacht hat, wie zum Beispiel mit diesem Amiodaron, dann war ich so sauer, dann hab ich mir das garantiert gemerkt.“

(männlich, insgesamt sechs ähnliche Aussagen und zwölf Zustimmungen)

Insbesondere die Fragenfälle sorgten für anschließende Diskussionen der Studierenden nach den E-Fallseminaren. Es wurden über Behandlungsstrategien von Erkrankungen diskutiert und unklare Inhalte wurden einander erklärt.

Zitat: „Man hatte halt ne Diskussionsgrundlage für danach.“

(männlich, insgesamt vier ähnliche Aussagen und neun Zustimmungen)

Nach den Meinungen der Studierenden entwickelte sich durch Fragenfälle ein fächerübergreifendes Wissen, welches an die Anforderungen des klinischen Alltags erinnerte.

Zitat: „Ja also mir ist auch besonders in Erinnerung geblieben, dass mir die Key Features vor allen Dingen gefallen haben, weil man dann so ein zusammenhängendes Denken entwickeln musste, auch fächerübergreifend.“

(weiblich, insgesamt fünf ähnliche Aussagen und vier Zustimmungen)

Die Studierenden gaben an, dass das Wiederholen von Inhalten eine positive Auswirkung auf den Lernerfolg hatte.

Zitat: „Ja, aber ich wüsste zum Beispiel nicht, wenn die jetzt nicht zweimal oder

dreimal Amiodaron gefragt hätten, ob ich's dann jetzt noch wissen würde.“

(männlich, insgesamt fünf ähnliche Aussagen und zehn Zustimmungen)

Das individuelle Lernverhalten wurde nach Ansicht der Studierenden weniger durch die E-Fallseminare verändert.

Zitat: „Ich muss sagen, speziell für die E-Fallseminare habe ich mich überhaupt nicht vorbereitet eigentlich und hatte trotzdem ne Lernkurve.“

(männlich, insgesamt drei ähnliche Aussagen und vier Zustimmungen)

- **Implementierung**

Der Aspekt „Implementierung“ umfasste alle Aussagen der Studierenden zum Modulbezug und zur Zeiteinteilung. Es wurde über die Zeiteinteilung für die Bearbeitung der Fragenfälle und die Zeit zum Lesen der Lesefälle diskutiert. Hierbei fiel auf, dass die Mehrheit der Gesprächsteilnehmer mehr Zeit für die Fragenfälle aufwandte.

Zitat: „Ich hab für die Fragenfälle, glaub ich mehr Zeit gebraucht als für die Lesefälle.“

(männlich, insgesamt sechs ähnliche Aussagen und zwei Zustimmungen)

Ein weiterer Aspekt war das Thema Modulbezug. Nach den Meinungen der Studierenden fiel positiv auf, dass die Inhalte fächerübergreifend und die Fälle gut ausgearbeitet waren. Besonders hervorgehoben wurde das Modul 3.1, in dem laut den Studierenden die E-Fallseminare einen besonders großen Lernerfolg erzeugten.

Zitat: Ich fand auch den besten Lerneffekt hatte ich, als wir die Kardiologiefragen semesterbegleitend halt zu KPL, also zu Modul 3.1, hatten.

(weiblich, insgesamt drei ähnliche Aussagen und fünf Zustimmungen)

Von den Studierenden wurde gelobt, dass die Inhalte der E-Fallseminare während des Moduls 3.1 den Lehrinhalten von Modul 3.1 entsprachen. In den Modulen 3.2 und 3.3 funktionierte laut den studentischen Aussagen diese zeitliche Abstimmung nicht mehr. Die Studierenden beschrieben, dass sie gedanklich schon beim nächsten Modul waren und dementsprechend die E-Fallseminare passend zum jeweiligen Modul erwarteten. Außerdem wünschten die Studierenden sich, sie hätten vor den Modulklausuren Themen bearbeitet, die klausurrelevant für das betreffende Modul und nicht für das vorherige Modul waren.

Zitat: „Es war ja in dem Modul 3.1 so, dass die Fragen meistens auch passend zu dem Thema waren, dass das wirklich zusammenhing und in den andern beiden Modulen war das ja eher nicht mehr der Fall dann.“

(männlich, insgesamt acht ähnliche Aussagen und zwölf Zustimmungen)

- **Inhalt**

Unter diesem Stichpunkt wurden die Lerninhalte und deren Relevanz zusammengefasst. Der Hauptpunkt „Inhalt“ wurde in drei Schwerpunkte unterteilt: Feedback, Motivation und Praxisbezug. Das Feedback wurde besonders hervorgehoben. Hier entwickelten sich zwei wichtige Aspekte. Erstens gaben die Studierenden an, dass sie bei den Fragenfällen das Feedback häufiger lasen. Zweitens habe nach Ansicht der Studierenden das Feedback den Effekt, dass der Wissensstand erweitert wurde. Nach Angaben der Studierenden geschah dies, indem die Studierenden ein tieferes Verständnis der Zusammenhänge entwickelten und eine Rückmeldung über das Können durch Aufschluss über die eigenen Fehler erhielten.

*Zitat: „Ich hab das glaub ich auch bei den Fragen deutlich ausführlicher gelesen, weil man dann halt irgendwie Bezug dann hatte zu der Antwort, die man gegeben hat.“
(männlich, insgesamt zwei ähnliche Aussagen und sechs Zustimmungen)*

Die Studierenden vertraten den Standpunkt, dass die Themen eine hohe praktische Relevanz hatten und dass sehr grundlegende Schritte von Diagnostik und Therapie vermittelt wurden, welche für den späteren Berufsalltag als nützlich empfunden wurden.

*Zitat: „Sprich, es war halt einfach gehalten und nicht son super komplizierter Fall, aber das ist ja auch erstmal das Wichtigste, denk ich, dass wir die Grundlagen können. Dann kann man dann nachher als Arzt immer noch weiter drauf aufbauen, aber dafür war das sehr gut geeignet, glaub ich.“
(männlich, insgesamt sieben ähnliche Aussagen und elf Zustimmungen)*

Nach Ansicht der Studierenden sorgten die als sehr realistisch wahrgenommenen Patientenfälle und die Aufforderung zum Treffen eigener Entscheidungen für eine hohe Motivation richtige Antworten in den Fragenfällen zu geben, da hierdurch das individuelle Zu-rechtkommen im Berufsalltag zu erahnen war.

*Zitat: „Jetzt muss ich mal was entscheiden, wie hinterher halt eben auch, dass man echt mal das Gefühl hatte, es geht jetzt mal in Richtung Arzt zu sein und nicht immer nur irgendwas auswendig zu lernen.“
(männlich, insgesamt acht ähnliche Aussagen und 20 Zustimmungen)*

- **Optimierung**

Unter dem Hauptpunkt „Optimierung“ wurden die Vorschläge der Studierenden zur Verbesserung des Lehrformats berücksichtigt. In diesem Abschnitt werden neben den oft genannten auch vereinzelt vorgeschlagene Ideen wiedergegeben, da nur so die Vielfalt an Ideen dargestellt werden kann. Der erste Aspekt betrachtet die Integration von E-Fallseminaren in das Medizinstudium. Die Studierenden schlugen vor, neue Fragenfälle für andere Module zu entwerfen und diese fest in das Curriculum der Universitätsmedizin Göttingen aufzunehmen. Ein Studierender regte an, das Projekt multidisziplinär zu etablieren.

Zitat: „Es wäre voll gut, wenns das für jedes Fach gäbe. Auf jeden Fall.“

(weiblich, insgesamt drei ähnliche Aussagen und sieben Zustimmungen)

Einige Studierende schlugen vor, die E-Fallseminare auf freiwilliger Basis abzuhalten (*insgesamt fünf ähnliche Aussagen und vier Zustimmungen*). Um eine fakultative Teilnahme noch attraktiver zu gestalten, könnten laut den Studierenden Büchergutscheine oder das mögliche Sammeln von Punkten für die Modulabschlussklausur eine Lösung sein (*1 Aussage*). Andere Studierende präferierten hingegen eine Bearbeitungspflicht, da sie sonst die E-Fallseminare nicht bearbeiten würden (*insgesamt vier ähnliche Aussagen*). In diesem Zusammenhang wurde von den Studierenden vorgeschlagen, *Key-Feature-Fälle* in zukünftige Modulabschlussklausuren zu integrieren (*insgesamt drei ähnliche Aussagen und zwei Zustimmungen*). Betont wurde von den Studierenden dabei, dass bei Einführung in das Curriculum die E-Fallseminare nur Fragenfälle beinhalten sollten, da die Fragenfälle zum Lernerfolg beitrugen und die Lesefälle keinen zusätzlichen Nutzen zu den ohnehin viel gelesenen Lehrbüchern brachten.

Zitat: „Immer Fragenfälle und nicht Lesefälle.“

(weiblich, insgesamt vier ähnliche Aussagen und acht Zustimmungen)

Ein Wunsch der Studierenden war es, E-Fallseminare von zu Hause aus zu bearbeiten. Nach Meinungen der Studierenden würden sich die Vorteile anbieten, frei einen Zeitpunkt für die Bearbeitung zu wählen und nicht verstandene Inhalte nachschlagen zu können.

Zitat: „Ich fänds gut, wenn mans von zu Hause aus machen könnte.“

(weiblich, insgesamt sechs ähnliche Aussagen und drei Zustimmungen)

Ein weiterer Aspekt war die Kritik am technischen Präsentationsformat und die darauffolgenden Ideen zur Weiterentwicklung der Technik. Die Studierenden wünschten sich mehr Flexibilität, Abwechslung und mehr Funktionen.

Zitat: „Ich glaub, die bisherige Plattform ist einfach zu starr und zu unflexibel.“

(männlich, insgesamt vier ähnliche Aussagen und fünf Zustimmungen)

Um die Vielfalt des Campus-Prüfungssystems und der Fragenfälle zu steigern, empfahlen die Studierenden mehr Medien in die Fälle zu integrieren. Die Studierenden schlugen vor, dies in Form von kurzen Videosequenzen, Auskultationsbefunden, EKGs, Originallaborausdrucken, Sonographien, Röntgenbildern und Fotos von eindrucksvollen Befunden umzusetzen.

Zitat: „Das ist natürlich viel mehr Arbeit, das dann zu erstellen, aber das wäre auch noch was, glaub ich, was motivieren könnte, wenn da mehr multimedialer Inhalt vielleicht noch drin ist.“

(männlich, insgesamt zwei ähnliche Aussagen und sechs Zustimmungen)

Nach Ansicht der Studierenden könnte durch eine auswählbare Hilfestellung in Form eines Oberarztcommentares oder durch interdisziplinäre Fälle der Lernerfolg noch optimiert werden *(insgesamt eine ähnliche Aussage und 7 Zustimmungen)*.

Ein Wunsch einiger Studierenden war es, das Campus-Prüfungssystem individueller zu gestalten. Vorgeschlagen wurde die Entwicklung eines Computerspiels mit der Möglichkeit individuell auf den Patientenfall zu reagieren und die eigenen Entscheidungen im Hinblick auf anschließende unterschiedliche Krankheitsverläufe mit Besserung oder Verschlechterung des Patientenzustandes, zu reflektieren.

Zitat: „Also ich glaub mein Traum wäre wirklich, diese ganzen Informationen mit Tönen und Videos reinzupacken und das weiter mit dieser Struktur zu haben, dass man halt immer Fragen beantworten muss und dann das Feedback dazu direkt kriegt und dass man im Prinzip den ganzen wirklich essentiellen Stoff zumindest komplett da drin hätte und das zu Hause einfach wie nen Computerspiel quasi durchspielen könnte, ich glaub viel effektiver könnte ich persönlich zumindest nicht lernen.“

(männlich, insgesamt vier ähnliche Aussagen und fünf Zustimmungen)

Die Studierenden regten an, dass nach der Bearbeitung der E-Fallseminare der einzelne Studierende eine persönliche Auswertung per E-Mail *(insgesamt eine ähnliche Aussage und zwei Zustimmungen)* oder eine persönliche Statistik *(insgesamt eine ähnliche Aussage)* erhalten solle. Die Studierenden wünschten sich für die Zukunft eine größere inhaltliche Vielfalt der wiederholten Fälle zu einem Thema.

Zitat: „Ich hatte so das Gefühl, dass es vielleicht ganz schön gewesen wäre, noch zwei, drei Fälle mehr zu haben oder zum Ende hin die Fälle zu ändern.“

(männlich, insgesamt drei ähnliche Aussagen und acht Zustimmungen)

3.3.6 Theorien zu den Mechanismen des *Testing Effects*

Die Ergebnisse der sechsten Studienfrage „Inwieweit wurden die Mechanismen des *Testing Effects* in Fokusgruppengesprächen bekräftigt?“ sind in den folgenden Tabellen 9 bis 11 dargestellt.

Tabelle 9: Aussagen der Studierenden entsprechend der *Elaborate Retrieval Theory*

Theorie		Aussagen der Studierenden
Elaborate Retrieval Theory	Associative Memory Theory	Zitat: „Ja also mir ist auch besonders in Erinnerung geblieben, dass mir die Key Features vor allen Dingen gefallen haben, weil man dann so ein zusammenhängendes Denken entwickeln musste, auch fächerübergreifend.“ (weiblich, insgesamt fünf ähnliche Aussagen und vier Zustimmungen)
	Degree of Semantic Elaboration	Zitat: „Also bei mir blieb das mit den Fragen auf jeden Fall mehr hängen, weil man sich halt aktiv was überlegt und drüber nachdenkt, was da jetzt richtig sein könnte...“ (weiblich, insgesamt acht ähnliche Aussagen und 15 Zustimmung)
		Zitat: „...wenn man das selber eintragen musste, dann musste man ja auch wirklich selber drüber nachdenken und das auch entscheiden, was man macht.“ (weiblich, insgesamt fünf ähnliche Aussagen und vier Zustimmungen)
		Zitat: „Ja bei den Fällen, die man halt nur gelesen hat, da hat man viel auch überlesen. Da hat man sich vielleicht schon auch noch ein bisschen Gedanken gemacht, aber man war natürlich nicht so gefordert wie bei den Key-Feature-Fragen.“ (weiblich, insgesamt vier ähnliche Aussagen und acht Zustimmungen)
		Zitat: „Das ist halt nicht diese passive Aufnahme oder versuchte Aufnahme von Wissen, sondern man muss halt echt drüber nachdenken und sich auch seine Gedanken dazu machen.“ (männlich, insgesamt vier ähnliche Aussagen und eine Zustimmung)

Tabelle 10: Aussagen der Studierenden entsprechend der *Transfer Appropriate Processing Theory*

Theorie	Aussagen der Studierenden
<p><i>Transfer Appropriate Processing Theory</i></p>	<p>Zitat: „...wenn sich Fälle geglichen haben nacheinander, dann hatte man irgendwie doch ne gewisse Sicherheit im Vorgehen. Nicht, dass immer alles richtig war, aber es hat schon was geholfen, wenn man schonmal nen ähnlichen Ablauf gelesen bzw. beantwortet hat. Und auf jeden Fall mehr bei den Key-Feature-Fragen als bei Fällen, die man einfach gelesen hat.“ (weiblich, insgesamt fünf ähnliche Aussagen und zehn Zustimmungen)</p>

Tabelle 11: Aussagen der Studierenden entsprechend der *Unspecific Goal Perspective*

Theorie	Aussagen der Studierenden
<p><i>Unspecific Goal Perspective</i></p>	<p>Zitat: „Ja das fand ich irgendwie auch ganz gut, dass halt so allgemeinere Dinge gefragt wurden. Bei den MC-Fragen ist es ja immer so, dass so krass spezielle Sachen immer nur abgefragt werden, die man vielleicht wahrscheinlich nie mehr braucht. Und dieses Mal wars halt so, dass man wirklich son Fall, der könnte immer passieren und passiert wahrscheinlich auch häufig und das fand ich schon sinnvoll.“ (weiblich, insgesamt fünf ähnliche Aussagen und sechs Zustimmungen)</p> <p>Zitat: „Also das ist ja das, was ich eben auch schon gesagt hab, dass es halt wirklich die Basics waren, aber ich mein das sind ja auch die Sachen, die halt wirklich sitzen müssen und ich glaube, dass die wenigsten Patienten, die man später dann sieht wirklich so das Vollbild immer einer Krankheit zeigen und wirklich genau die Symptome, die im Lehrbuch stehen.“ (männlich, insgesamt fünf ähnliche Aussagen und vier Zustimmungen)</p> <p>Zitat: „...und dann halt eben einfach diese Grundlagen, weil das find ich mehr bringt als wenn man jetzt irgendwelche Spezialsachen fragt, weil die kann man hinterher immer noch genug lernen, wenn man dann Assistenzarzt ist. Wichtig ist, dass die Basics sitzen. Aber es waren ja auch, also jedenfalls kam mir es so vor, überwiegend Basics.“ (männlich, insgesamt eine ähnliche Aussage und sechs Zustimmungen)</p>

Der Prozess *Associative Memory Theory* wurde in sechs Aussagen und vier Zustimmungen angesprochen. Insgesamt 25 Aussagen und 28 Zustimmungen bezogen sich auf den Prozess *Degree of Semantic Elaboration*. Die Theorie *Transfer Appropriate Processing Theory* wurde in insgesamt sechs Aussagen und zehn Zustimmungen angesprochen. Zur Theorie *Unspecific Goal Perspective* passten 14 Aussagen und 16 Zustimmungen.

4 Diskussion

4.1 Die wesentlichen Ergebnisse

In der vorliegenden Studie wurde erstmals die Effektivität wiederholten Prüfens mit *Key Features* bei Medizinstudierenden für die Festigung differentialdiagnostischer und therapeutischer Kompetenzen nachgewiesen. Die Bearbeitung wiederholter formativer *Key-Feature-Fragen* erzeugte sowohl kurz- als auch langfristig signifikant bessere studentische Leistungen hinsichtlich komplexer kognitiver Fertigkeiten als das Lesen wiederholter Lesefälle. Auf den ersten Blick wirkte das Gesamtergebnis im Retentionstest mit durchschnittlich 50% erreichten Punkten enttäuschend. Jedoch ist dieser Wert mit den Ergebnissen vorheriger Studien gut vergleichbar bzw. liegt sogar darüber (Larsen et al. 2009, Larsen et al. 2013). Der nachgewiesene Effekt von *Key-Feature-Fragen* blieb auch nach Adjustieren für die zusätzliche Bearbeitungszeit erhalten.

Die Ergebnisse der vorliegenden Studie könnten darauf beruhen, dass alle Prüfungen formativ gestellt waren. Aufgrund der Tatsache, dass die Studierenden keine Konsequenz in Form von z. B. Leistungspunkten zum Bestehen der Modulklausur erhielten, ist davon auszugehen, dass die Studierenden nicht für die E-Fallseminare lernten. Hierdurch wurde zwar einerseits ein Störfaktor durch einen evtl. drohenden Punktverlust (mit konsekutiver Änderung des Lernverhaltens) vermieden; andererseits hat sich dies möglicherweise negativ auf die Sorgfalt der Studierenden bei der Bearbeitung der E-Fallseminare und der abschließenden formativen Prüfungen ausgewirkt. Unter dieser Prämisse ist anzunehmen, dass die in dieser Studie beobachteten Effekte nicht durch artifizielle Lernanreize hervorgerufen oder überlagert wurden (siehe hierzu Raupach et al. 2013). Bemerkenswert war insbesondere die relativ geringe Abnahme der Leistung der Studierenden im Retentionstest im Vergleich zum Ausgangstest. Insgesamt gibt es nur wenige Studien im Rahmen medizinischer Ausbildung, die einen Zeitraum von sechs Monaten betrachtet haben. In Studien, die einen Zeitraum von sechs Monaten berücksichtigt haben, ist jedoch ein deutlicher Rückgang der Leistung über diesen Zeitraum zu beobachten (Larsen et al. 2012; Larsen et al. 2009, Larsen et al. 2013; Schmidmaier et al. 2011). In der Studie von Larsen et al. (2013) ereignete sich eine Reduktion der durchschnittlich erreichten Punkte mittels Testen um circa 40% nach sechs Monaten. In der vorliegenden Studie zeigte sich im Gegensatz dazu eine Abnahme von nur circa 3% der durchschnittlich erreichten Punkte in den Interventionsitems nach sechs Monaten. Wie in anderen Studien auch, entsprach der Ausgangstest dem

Retentionstest (Dobson und Linderholm 2015; Larsen et al. 2009; Schmidmaier et al. 2011). Folglich wäre eine mögliche Erklärung für die geringere Leistungsabnahme, dass sich die Studierenden an die Fälle, Fragen und Antworten erinnerten. Diese Erklärung ist jedoch nicht sehr plausibel, da innerhalb der sechs Monate zwischen Ausgangstest und Retentionstest keine Möglichkeit zur erneuten Bearbeitung der Fälle bestand. In den bereits zitierten Studien, in denen sich Abschluss- und Retentionstest ebenfalls glichen, zeigte sich bei Schmidmaier et al. (2011) eine Reduktion von 30% der durchschnittlichen Leistung in den Testitems und damit ein Verlust der Überlegenheit des Testens gegenüber dem Lesen auf eine ungefähr gleiche Leistung von Testen und Lesen nach sechs Monaten. Bei Larsen et al. (2009) reduzierte sich die Leistung der Studienteilnehmer in den Testitems um circa 12% im Retentionstest (nach sechs Monaten). Die Studierenden der vorliegenden Studie mussten sich außerdem an beides – die Interventions- und Kontrollitems – in gleichem Maße erinnern. Das heißt, auch wenn dieser Effekt Einfluss auf die Ergebnisse der vorliegenden Arbeit genommen hätte, wäre er nicht ausreichend groß genug gewesen, um die Überlegenheit des Testens gegenüber dem Lesen zu verbergen. Eine weitere Erklärung für das gute Behalten der Inhalte könnten zwischenzeitliche Famulaturen sein, in denen die Studierenden ihr Wissen anwendeten.

Die *Key-Feature*-Fälle umfassten eine große Bandbreite an grundlegenden und klinisch relevanten Erkrankungen wie „Lungenembolie“, „Pneumonie“ oder „Nephrotisches Syndrom“. Insbesondere für ressourcenintensive Lehrformen könnten *Key Features* eine reizvolle Alternative sein. Sobald die eher aufwändige Erstellung von *Key-Feature*-Fragen erfolgt ist, können *Key-Feature*-Prüfungen eine zeitlich effiziente Lehrform darstellen. Nicht nur in den quantitativen Ergebnissen, sondern auch in den qualitativen Ergebnissen, kristallisierten sich die Vorteile der Fragenfälle verglichen mit den Lesefällen heraus. Die Studierenden bemerkten, dass sie sich in den Fragenfällen aufgrund der aktiven Auseinandersetzung mit dem Lernstoff besser Inhalte merken konnten. In dieser Studie wurden alle vier Theorien zu den zugrundeliegenden Mechanismen des *Testing Effects* angesprochen. Außerdem wurden in den Fokusgruppen Optimierungsvorschläge gesammelt.

4.2 Studiendesign

Vorangegangene Studien, die sich mit der Effektivität von testgestütztem Lernen in der medizinischen Ausbildung beschäftigten, unterschieden sich von dieser Studie bezüglich des Studiendesigns, der Stichprobengröße, der Teilnehmer, des Ausbildungsstands, der Wissensniveaus, Fragetypen, Anzahl der Tests und des Nachbeobachtungs-Zeitraums. In

der Literatur finden sich sowohl Beobachtungsstudien (Logan et al. 2011) als auch randomisierte Studien (Kromann et al. 2009). Das Studiendesign wurde nach den Anregungen der aktuellen Literatur entworfen. Larsen et al. (2008) erstellten einen Überblick über die bisherige Literatur zum testbasierten Lernen in der medizinischen Ausbildung. Hieraus ergab sich die Empfehlung, in häufigen und regelmäßigen Abständen zu testen, um ein besseres Behalten von Inhalten zu gewährleisten. Weiterhin sollten Fragetypen (wie *Key-Feature-Fragen*) gestellt werden, die das Abrufen und die Produktion von Inhalten (*Production Tests*) für die Lösung von Problemen gewährleisten. (Larsen et al. 2008) Viele Studien führten einen Retentionstest nach einem Zeitraum von einer Woche durch (Dobson und Linderholm 2015, Baghdady et al. 2014; Larsen et al. 2008; Logan et al. 2011; Roediger und Karpicke 2006b). Hierbei besteht das Problem, dass in der Realität Inhalte länger als eine Woche behalten werden müssen. Somit wurde angeraten, den Retentionstest nach mehreren Monaten durchzuführen, um ein Behalten der Inhalte über eine längere Zeit zu untersuchen. Darüber hinaus empfahlen Larsen et al. (2008) ein Feedback mit Informationen zu den richtigen Antworten bereitzustellen. Außerdem wurde nahegelegt, Studien zu konzipieren, die in die medizinische Ausbildung und somit in reale Lehrsituationen eingebaut sind. (Larsen et al. 2008). Diese Empfehlungen wurden – soweit möglich – beim Design der vorliegenden Studie umgesetzt. Durch die Randomisierung sollte vermieden werden, dass die Studierenden in den beiden Gruppen ungleiche Voraussetzungen hatten. Dies war zwar für die primäre Analyse unerheblich (da diese unabhängig von den Studiengruppen war); für die weiteren Analysen konnten Unterschiede zwischen den Gruppen weder damit begründet werden, dass in der einen Gruppe mehr Frauen oder Männer waren, noch damit, dass eine Gruppe möglicherweise aus leistungsstärkeren Studierenden bestand. Durch den Einsatz des Cross-Over Designs diente jeder Studierende seiner eigenen Kontrolle.

In unserer Studie umfasste die Stichprobe 87 Medizinstudierende aus dem 3. klinischen Semester, während in anderen Studien beispielsweise 47 Medizinstudierende aus dem 1. Jahr (Larsen et al. 2013), 112 Zahnmedizinstudierende (Baghdady et al. 2014) oder 49 Assistenzärzte im Fachbereich Pädiatrie (DelSignore et al. 2016) eingeschlossen waren. Die Stichproben von Studien bezogen sich auf unterschiedliche Ausbildungsstände, sie reichten von Studienanfängern bis zu arbeitenden Ärzten (Larsen et al. 2009). Unterschiedliche Wissensniveaus wurden getestet; Schmidmaier et al. (2011) prüften Faktenwissen in Form von elektronischen Karteikarten, während Larsen et al. (2012) die klinische Anwendung von medizinischem Wissen an standardisierten Simulationspatienten untersuchten.

Auch die Fragetypen unterschieden sich in den Studien. Die Bandbreite der Fragetypen reichte von Einfachwahl-Fragen (Dobson und Linderholm 2015) über schriftliche Kurzantworten (Yeung et al. 2016) bis zu Aufsätzen (Larsen et al. 2013) und fünfminütigen Szenarien mit Checklisten (Kromann et al. 2009). Ebenso divers ist die Häufigkeit der Testwiederholungen vor dem Retentionstest. Sie variierte von einer einmaligen Testwiederholung (Schmidmaier et al. 2011) bis zu drei (Larsen et al. 2009) oder vier (Larsen et al. 2013) Testwiederholungen. Retentionstests folgten je nach Studie zwischen einer Woche (Dobson und Linderholm 2015, Baghdady et al. 2014, Logan et al. 2011) nach dem letzten Testen bis zu sechs Monaten (Larsen et al. 2013, Schmidmaier et al. 2011, DelSignore et al. 2016) nach dem letzten Testen.

4.3 Gemeinsamkeiten zu vorherigen Studien

Auch wenn es offensichtlich einige Unterschiede zwischen dieser Studie und vorherigen, thematisch ähnlichen Studien gab und somit die Vergleichbarkeit unserer Ergebnisse mit denen anderer Studien eingeschränkt war, fanden sich auch Gemeinsamkeiten. In den folgenden zwei Studien Larsen et al. (2009) und Larsen et al. (2013) wurde ebenfalls ein Nachbeobachtungszeitraum von sechs Monaten gewählt; in beiden Studien erwies sich wiederholtes Testen gegenüber wiederholtem Lernen als überlegen. Larsen et al. (2009) untersuchten testbasiertes Lernen in der Neurologie. An dieser Studie im randomisierten *Crossover*-Design nahmen 40 Assistenzärzte der Fachrichtung Neurologie teil. Diesen 40 Assistenzärzten wurden Lehrinhalte zu der Erkrankung Myasthenia gravis und zum Status epilepticus in einem Seminar vermittelt. Danach wurden sie jeweils einer von zwei Gruppen zugeteilt. Die eine Gruppe wurde zur Myasthenia gravis getestet und lernte Inhalte mittels eines Arbeitsblatts zum Status epilepticus. Die andere Gruppe wurde zum Status epilepticus getestet und studierte ein Arbeitsblatt zur Myasthenia gravis. Insgesamt nahmen die Assistenzärzte innerhalb von vier Wochen entweder an drei Tests oder an drei Terminen, an denen sie die Inhalte durchlasen, teil. Sechs Monate später ergab sich in einem Retentionstest, dass die Teilnehmer, die dreimal getestet wurden, besser abschnitten als die Teilnehmer, die die Inhalte gelesen hatten (durchschnittliche prozentuale Leistung im Retentionstest nach wiederholtem Testen: 39%; durchschnittliche prozentuale Leistung im Retentionstest nach wiederholtem Lesen: 26%) (Larsen et al. 2009). In einer anderen Studie derselben Arbeitsgruppe wurden 47 Medizinstudierende im ersten Studienjahr in den medizinischen Themen „Krampfanfall“, „Optikusneuritis“, „Myasthenia gravis“ und „Migräne“ unterrichtet. Daraufhin wurden die Studierenden in vier Gruppen eingeteilt. Es

gab vier verschiedene Präsentationsformate (Test; Test mit Selbsterklärung; Arbeitsblatt zum Lesen; Arbeitsblatt mit der Möglichkeit, eigene Formulierungen zu verfassen). Jedes der vier Themen wurde mit jeweils einem der vier Präsentationsformate gepaart. Demzufolge bearbeitete jeder Studierende alle Themen bzw. Präsentationsformate – jedes mit einer anderen Methode. Sechs Monate nach der ersten Veranstaltung wurden alle Studierenden über alle vier Krankheitsbilder getestet. Insgesamt fiel auf, dass in dem sechs Monate später durchgeführten Retentionstest durchschnittlich mehr Punkte erreicht wurden, wenn Inhalte in der Zwischenzeit wiederholt getestet und nicht nur gelesen wurden (36% gegen 20%) (Larsen et al. 2013). Ähnliche Ergebnisse wurden bei einer Studie erzielt, in der sowohl bei den Interventionen (formative Prüfungen) als auch im Retentionstest Schauspielpatienten eingesetzt wurden. (Larsen et al. 2012) Die Ergebnisse der Studie zeigten, dass Tests mit Schauspielpatienten (61% der erreichbaren Punkte) ein gleich gutes Ergebnis wie das schriftliche Testen mit Kurzantwortfragen hervorrufen können (61% der erreichbaren Punkte). Beide Präsentationsformate waren dem reinen Lesen von Inhalten überlegen (48% der erreichbaren Punkte) (Larsen et al. 2012). Die Ergebnisse bestätigten die Ergebnisse von Kromann et al. (2009), dass testbasiertes Lernen auch für die Festigung praktischer Fertigkeiten Anwendung finden kann. Problematisch ist jedoch, dass der Einsatz von Schauspielpatienten zeit- und ressourcenintensiv ist (Larsen et al. 2012).

4.4 Computergestütztes Lernen im Medizinstudium

David A. Cook et al. diskutierten 2008 in einer Metaanalyse, was für eine Relevanz computergestütztes Lernen in der Ausbildung von Gesundheitsberufen hat. Sie stellten heraus, dass computergestütztes Lernen im Vergleich zu keiner Instruktion große Lerneffekte erzeugt und eine ähnliche Effektivität wie die traditionellen Lehrmethoden aufweist (Cook et al. 2008). Es müsse in weiteren Studien untersucht werden, wie elektronisches Lernen effizient einsetzbar ist (Cook et al. 2008; Cook 2009). In Anbetracht der Wirtschaftlichkeit der Nutzung von elektronischen Lehrformaten (Greenhalgh 2001) und der Erkenntnis aus der vorliegenden Arbeit, dass die Bearbeitung von *Key Features* zur Festigung von komplexen kognitiven Fertigkeiten in der medizinischen Ausbildung von Bedeutung ist, ergeben sich aus den hier präsentierten Daten Empfehlungen für die Integration entsprechender Formate in die humanmedizinischen Curricula. Nach der vorliegenden Studie können insbesondere elektronische *Key-Feature*-Prüfungen effiziente Lernmöglichkeiten generieren, die mit anderen Mitteln nicht für alle Studierende verfügbar wären.

4.5 Stärken der Studie

Die Studierenden erhielten in der vorliegenden Studie ein sofortiges Feedback auf ihre Antworten. Diskutiert wird in der Literatur, zu welchem Zeitpunkt Feedback am sinnvollsten ist (Kulik und Kulik 1988). Rückblickend beschrieben Kulik und Kulik 1988, dass in Studien mit Tests im Rahmen realistischer Lehrsituationen Feedback sofort gegeben werden sollte. Die Ergebnisse anderer Studien zeigten, dass späteres Feedback effektiver ist als sofortiges Feedback (Butler et al. 2007; Butler und Roediger 2008). Ausschlaggebend ist hierbei, dass sich diese Ergebnisse auf Einfachwahl-Fragen bezogen und in weiteren Studien herausgefunden werden muss, ob diese Ergebnisse auf weitere Fragetypen wie *Key Features* übertragbar sind. In der vorliegenden Studie ergaben sich aus den studentischen Anmerkungen in den Fokusgruppengesprächen Hinweise darauf, dass ein sofortiges Feedback nach *Key-Feature*-Fragen sinnvoll sein kann.

Die Ergebnisse zum Leistungserhalt bzw. -verlust bei wiederholter Bearbeitung von *Key-Feature*-Fragen waren vergleichbar mit einer Studie von Smith et al. (2009), in der die Ergebnisse ebenfalls in einem Baumdiagramm dargestellt wurden. Die Ergebnisse von Smith et al. (2009) deuteten wie in der vorliegenden Studie an, dass einige Studierende, die einen Inhalt erst nicht verstanden und dementsprechend das Item falsch beantworteten, durch die Interventionen lernten und beim nächsten Mal das Item richtig beantworteten. Auch wenn eine Frage mehrmals hintereinander falsch beantwortet wurde, lernten zumindest einige Studierende durch das Testen und erzielten im Retentionstest richtige Antworten. Sowohl die Ergebnisse bei Smith et al. (2009) als auch die Ergebnisse der vorliegenden Arbeit zeigten außerdem, dass, wenn Studierende zu Beginn ein Item richtig beantworteten, sie tendenziell das Item in den nachfolgenden Tests auch richtig beantworteten. Letzteres gibt Hinweise darauf, dass, nachdem eine Frage richtig beantwortet worden war, wiederholtes Testen das Behalten der abgefragten Inhalte stärkte. In einer Studie von McConnell et al. (2015) konnte ein positiver Lerneffekt dargestellt werden, wenn Aufgaben sowohl in den Interventionen als auch im Retentionstests durchgehend identisch gestellt wurden. Trotz der nennenswerten Unterschiede zwischen den Fragen in den E-Fallseminaren und den Fragen im Retentionstest bei gleichen Lernzielen konnte in der vorliegenden Studie ein Effekt des wiederholten Testens nachgewiesen werden.

Kopp et al. (2006) und van Bruggen et al. (2012) verwiesen darauf, dass besonders die Erstellung von *Key Features* und deren Fallvignetten gutes Training von Autoren und Korrekturprozesse benötigt. In der vorliegenden Studie wurden die Fragenfälle von jeweils mehreren Autoren, welche sich teilweise intensiv mit Medizindidaktik und Ausbildungs-

forschung befassen, konzipiert und korrigiert. Des Weiteren empfahlen Page et al. (1995) eine Pilotierung von *Key Features*, welche in der vorliegenden Studie durch die Pilotstudie generiert wurde. Wie von Rotthoff et al. (2006) nahegelegt, wurde der Fragetyp *Long-Menu* in Kombination mit *Key Features* gewählt, um die Befähigung zum Treffen von klinischen Entscheidungen zu prüfen. Außerdem sorgten unterschiedliche *Long-Menu*-Listen für eine bessere Übersicht (Rotthoff et al. 2006).

Die Teilnehmerquote dieser Studie war akzeptabel. Durch die stratifizierte Randomisierung konnte ein Selektionsbias minimiert werden. Die gemittelten Itemschwierigkeiten wiesen auf eine durchschnittliche Schwierigkeit der Items hin (empfohlener Wertebereich der Itemschwierigkeit: 0,4 bis 0,8 (Möltner et al. 2006)). Die gemittelten Trennschärfen ließen darauf schließen, dass es möglich war, anhand der Items leistungsstarke Studierende von schwächeren Studierenden zu trennen (Werte von über 0,3 sind anzustreben (Möltner et al. 2006)). Ausgezeichnet war die interne Konsistenz des Ausgangstests und des Retentionstests.

Eine besondere Stärke dieser Studie war der *Mixed Methods Approach*. *Mixed Methods Approach* heißt, dass eine Studie sowohl einen quantitativen als auch einen qualitativen Anteil hat (Cook und Reichardt 1979). In der vorliegenden Studie wurde zusätzlich eine qualitative Datenanalyse durchgeführt mit den Zielen das Projekt zu evaluieren, eine Ergänzung zur Erhebung quantitativer Daten zu gewährleisten und um quantitative Ergebnisse mit Hilfe eines Einblicks in die Gedanken und Meinungen der Studierenden zu überprüfen (Henseling et al. 2006). Neben Selbsteinschätzungen und offengelegten Gedanken wurden Denkprozesse erläutert und neue Erkenntnisse diskutiert. Außerdem wurden durch die Anmerkungen und Verbesserungsvorschläge Weiterentwicklungsmöglichkeiten für die Arbeitsgruppe und weitere Projekte geschaffen.

4.6 Limitationen der Studie

Nichtsdestotrotz ist die Generalisierung der Ergebnisse dieser Studie durch die monozentrische Durchführung limitiert. Eine multizentrische Studie würde eine größere Stichprobe und höhere Aussagekraft bedeuten. Dennoch muss erwogen werden, ob eine Studie in dieser Form in anderen Zentren möglich wäre. Seit 2013 besteht an der Universitätsmedizin Göttingen die Möglichkeit elektronisch zu prüfen, so dass die Voraussetzungen für Prüfungsformate mit höherem taxonomischen Niveau geschaffen sind. Für eine multizentrische Studie müsste also gewährleistet sein, dass die Räumlichkeiten mit entsprechend vielen Computern in den Zentren vorhanden wären. Außerdem müsste herausgefunden

werden, ob die Lehre im Fach Innere Medizin mit derjenigen an der Universitätsmedizin Göttingen vergleichbar ist. Wenn nicht, müssten die Fälle entsprechend dem Curriculum umgeschrieben werden, was wiederum die Vergleichbarkeit zwischen verschiedenen Standorten einschränken würde.

Die vorliegende Studie gab Aufschluss über die Effektivität von testbasiertem Lernen in der Realität. Auf denkbare Störfaktoren wird im Folgenden eingegangen. *Key-Feature-Fälle* könnten möglicherweise das Selbststudium außerhalb der E-Fallseminare mehr angeregt haben als die Lesefälle, was zur Vergrößerung des Effekts geführt haben könnte. Ein anderer potenzieller Störfaktor ist die Bearbeitungszeit für die Präsentationsformate. Im Unterschied zu Studien, die im nicht-medizinischem Kontext stattfanden (Butler und Roediger 2007; McDaniel et al. 2007b), wurde keine Zeit vorgegeben, in der die Fälle bearbeitet werden sollten, sondern die Studierenden konnten selbst entscheiden, wann für sie das E-Fallseminar vorüber war. Dennoch wurden die Anmeldezeiten verzeichnet und in die Studie miteingeschlossen. In der lernpsychologischen Literatur wurde wiederholt diskutiert, dass der *Testing Effect* einfach nur dadurch zustande kommt, dass die Studierenden in der Prüfungssituation mehr Zeit verbringen als in der Lesesituation (*Total Time Hypothesis*, siehe Roediger und Karpicke 2006a). Auch in der vorliegenden Studie zeigte sich, dass die Studierenden mehr Zeit mit den Fragenfällen verbrachten als mit den Lesefällen. Es wurde also untersucht, ob ein Zusammenhang zwischen einer zusätzlichen Bearbeitungszeit und einer höheren Punktzahl in den *Key-Feature-Prüfungen* bestand. Der Störfaktor „Bearbeitungszeit“ wurde in einer linearen Regression in die Analyse einbezogen. Der beobachtete Effekt der Bearbeitung der *Key-Feature-Prüfungen* (im Vergleich zur Bearbeitung der Lesefälle) auf den Lernerfolg war nach Adjustierung für Geschlecht und vorherige Klausurleistungen dennoch weiterhin signifikant. Also kann die Hypothese abgelehnt werden, dass die zusätzliche Bearbeitungszeit in den Fragenfällen den beobachteten Leistungsunterschied zwischen Interventions- und Kontrollitems erklärt.

Schließlich wurde mit der vorliegenden Studie nicht untersucht, ob testbasiertes Lernen mit *Key-Feature-Fragen* einen Einfluss auf die Leistung von Studierenden im klinischen Alltag hat. Obgleich die weiter oben beschriebene Studie von Larsen et al. (2012) einen Hinweis darauf gibt, dass wiederholtes Testen zu besseren Ergebnissen in abschließenden Prüfungen mit Schauspielpatienten führt, muss weiter in diese Richtung geforscht werden, um eine kausale Beziehung zwischen wiederholtem Testen und einer besseren Patientenversorgung zu bestätigen.

In Bezug auf die Fokusgruppen kann sich die Rekrutierung der Teilnehmer äußerst schwierig gestalten (Carlsen und Glenton 2011). In der vorliegenden Studie wurden die Studierenden über den Semesterverteiler angeschrieben, und Teilnehmern wurde als Entschädigung für ihren Zeitaufwand ein Gutschein im Wert von 20 Euro zugesagt. Entsprechend ist zu vermuten, dass sich nur die besonders motivierten und generell engagierten Studierenden zurückmeldeten. Empfohlen ist die Durchführung eines Probe-Fokusgruppengespräch, in dem der Moderator das Moderieren üben kann und getestet werden kann, ob die Fragen in sich konsistent sind. Des Weiteren kann der Einsatz von Moderationsassistenten sinnvoll sein, um eine bestmögliche Aufnahme durch das Unterbinden von Nebengeräuschen und von Dazwischenreden zu ermöglichen. Gegebenenfalls kann ein Komoderator nebenher Ergebnisse visualisieren (Krueger und Casey 2015). In der vorliegenden Studie gab es weder ein Probe-Fokusgruppengespräch, noch einen Moderationsassistenten oder einen Komoderator. Aussagekräftiger als eine Tonaufzeichnung können Videoaufnahmen der Fokusgruppen sein, damit auch paraverbale Informationen durch die Körpersprache verzeichnet werden können. Außerdem könnte besser beobachtet werden, welcher Teilnehmer welche Sätze sagt. In den Tonaufnahmen dieser Studie konnte lediglich zwischen „weiblich“ und „männlich“ unterschieden werden, eine klare Spracherkennung war nicht möglich. Bei erneuten Fokusgruppengesprächen sollte dies berücksichtigt werden.

4.7 Fokusgruppengespräche

Die Studierenden hatten den Eindruck, dass sie durch das wiederholte Prüfen mit Fragenfällen sich Inhalte besser merkten als mit Lesefällen. In der Wahrnehmung der Studierenden entwickelte sich durch die Fragenfälle ein Wissen, welches über einen Fachbereich hinausreichte und eine empfundene Sicherheit für Arbeitsabläufe als Arzt im Arbeitsalltag. Im Vergleich zu den Lesefällen, schienen die *Key-Feature-Fragen* die bessere Alternative zu sein. Die Studierenden lobten, dass sie bei der Bearbeitung von *Key-Feature-Fragen* sich aktiv mit den Lerninhalten auseinandersetzten und im Vergleich dazu bei der Bearbeitung von Einfachwahl-Fragen die richtige Antwort nur wiedererkannten. Insgesamt deuteten die Studierenden hiermit die These an, dass *Production Tests* zu einem besseren Behalten von Inhalten führen als *Recognition Tests* (Roediger und Karpicke 2006a; Wenger et al. 1980). Die Themen der E-Fallseminare wurden als sehr relevant wahrgenommen, was folglich für eine hohe Motivation sorgte. Das Feedback wurde von den Studierenden sehr positiv aufgenommen und trug in den Augen der Studierenden zum Lernerfolg bei. Laut der Studierenden beeinflusste das Feedback auch beim wiederholten Testen positiv den

Lernerfolg. Die Studierenden schlugen vor, die E-Fallseminare für weitere Module zu gestalten. Solch ein Vorgehen würde bedeuten, dass die Lehrkoordinatoren der anderen Fachrichtungen entsprechende Fälle schreiben müssten. Als technische Weiterentwicklung wurden Ideen diskutiert wie Videosequenzen und Auskultationsbefunde in die Fragenfälle einzubauen oder die E-Fallseminare als Computerspiel zu programmieren.

Es bestand zum Teil eine Diskrepanz in den studentischen Aussagen. Einerseits wünschten sich die Studierenden ein fächerübergreifendes Lernen. Andererseits kritisierten sie, dass teilweise E-Fallseminare thematisch nicht passend zum Modul stattfanden und wünschten sich die Thematisierung klausurrelevanter Inhalte in den E-Fallseminaren vor den Modulklausuren. Letzteres kann ein Beispiel für das Axiom ‚*assessment drives learning*‘ darstellen. Einige Studierenden schlugen vor, dass Fragenfälle in die Modulabschlussklausuren eingeführt werden. Fraglich ist, welche Folgen die Umstellung von formativen Prüfungen auf summative Prüfungen in den E-Fallseminaren hätte.

4.7.1 Struktur der Fokusgruppen

Gewöhnlich bestehen Fokusgruppen aus homogenen Gruppen, welche sich dadurch kennzeichnen, dass die Gruppenteilnehmer einen sehr ähnlichen Hintergrund (Medizinstudium, 3. klinisches Semester) bezüglich der Thematik der Fokusgruppe aufweisen (Henseling et al. 2006). Ein Vorteil besteht darin, dass die Teilnehmer aufgrund von gemeinsamen „Anknüpfungspunkten“ leichter miteinander ins Gespräch kommen (Henseling et al. 2006). Dennoch wäre es sinnvoll, wenn sich die Teilnehmer in mindestens einem Merkmal (in dieser Studie: das Geschlecht) unterscheiden, um eine größere Vielfalt von Meinungen zu erzielen. Die meisten Fokusgruppen bestehen aus Teilnehmern, die sich untereinander nicht kennen. Vorteilhaft ist hierbei, dass Meinungen durch Beziehungen zueinander nicht verfälscht werden und eine Qualitätssicherung durch die Gruppe gegeben ist. In dieser Studie waren die Fokusgruppen „Realgruppen“, was bedeutet, dass die Gruppenteilnehmer sich zumindest teilweise untereinander kannten (Henseling et al. 2006). Das ließ sich in diesem Fall offensichtlich nicht anders organisieren, da die Stichprobe ein Semester im 4. Studienjahr umfasste. Insgesamt wird empfohlen Fokusgruppen mit jeweils fünf bis acht Gruppenteilnehmern durchzuführen, um eine gute Gruppendynamik zu erhalten (Krueger und Casey 2015). Zu große Fokusgruppen tendieren dazu, dass die Gruppe schwer zu kontrollieren ist und durch zu kurze Pausen im Gespräch nicht alle Gruppenteilnehmer ihre Meinungen mitteilen können. Kleinere Fokusgruppen ermöglichen, dass alle Teilnehmer zu Wort kommen. Jedoch ist in dem Fall die Bandbreite an Erfahrungen eingeschränkt.

In der Regel werden drei oder vier Fokusgruppen gebildet, um eine theoretische Sättigung zu erreichen. (Krueger und Casey 2015). Der Begriff „Theoretische Sättigung“ wurde 1967 von Glaser und Strauss geprägt und beschreibt in diesem Fall, dass die Fokusgruppen durchgeführt wurden, bis keine neuen Erkenntnisse mehr aufkamen und sich keine neuen Einzelheiten zum Vertiefen der Themen ergaben, sodass aufgrund von Wiederholungen die Fokusgruppengespräche beendet werden konnten (Glaser und Strauss 1967). In den durchgeführten Fokusgruppen mit insgesamt 16 Studierenden, schien die theoretische Sättigung erreicht, da in der dritten Fokusgruppe nahezu keine neuen Aspekte mehr aufkamen. Jedoch ist fraglich, ob diese 16 Studierenden repräsentativ für die insgesamt 124 Studierenden sind. Immerhin repräsentierten die 16 Studierenden ein Achtel des gesamten dritten klinischen Semesters. Im Vergleich dazu wurden in einer Studie von Schiekirka et al. (2012) nur 17 Studierende von insgesamt vier klinischen Semestern für drei Fokusgruppengespräche rekrutiert. Auch hier wurde die theoretische Sättigung erreicht (Schiekirka et al. 2012) und sowohl die Anzahl der Fokusgruppenteilnehmer als auch die Anzahl der Fokusgruppen lagen im Bereich von bisher durchgeführten Studien (Carlsen and Glenton 2011). Wichtig ist einen respektvollen und klar kommunizierenden Moderator für die Fokusgruppengespräche einzusetzen, der das Prinzip der Leitfragen verstanden hat. So bleibt der Fokus des Gespräches auf den Fragen, die gestellt werden (Krueger und Casey 2015). In der vorliegenden Studie führte eine Moderatorin die Gespräche, die nicht Medizin studierte und so nicht auf ablenkende Themen wie z. B. die Klausurrelevanz der Inhalte eingehen konnte. Sie ließ, wie empfohlen, die Gruppe das Gespräch unter sich führen und gab durch Fragen einen Leitfaden vor (Henseling et al. 2006).

4.8 Theorien zum *Testing Effect*

Die Fokusgruppengespräche wurden im Hinblick auf die Theorien zu den zugrundeliegenden Mechanismen des *Testing Effects* ausgewertet. Der Prozess *Associative Memory Theory*, welcher sich der *Elaborate Retrieval Theory* zuordnen lässt, wurde in den Fokusgruppengesprächen angesprochen, da sich nach Ansicht der Studierenden durch Fragenfälle ein Wissen entwickelte, welches über das geprüfte Fach hinausreichte. Einige der Stellungnahmen der Studierenden waren mit dem Prozess *Degree of Semantic Elaboration* vereinbar. Die Studierenden berichteten über ihre Erfahrung, dass das Lösen der Fragenfälle sie gedanklich forderte und sie das Gefühl hatten, dass sie dadurch Inhalte mit den Fragenfällen mehr im Gedächtnis behielten als bei den Lesefällen. Kang et al. (2007) zeigten, dass ein aufwendiges Abrufen von Informationen (*effortful retrieval*) ihre Retention besonders

stark fördert. *Production Tests* waren diesbezüglich effektiver als *Recognition Tests*. Auch Endres und Renkl (2015) bestätigten diese Ergebnisse. Die sich wiederholenden *Key-Feature*-Fragen waren inhaltlich nicht gleich, aber ähnlich. Nach Ansicht der Studierenden entwickelte sich bei ähnlichen *Key-Feature*-Fällen eine Sicherheit für ihr Vorgehen. Das könnte auf die *Transfer Appropriate Processing Theory* hindeuten, indem das Testen kognitive Prozesse durch Transfer zwischen den Fällen stärkte. Einige Zitate der Studierenden konnten im Sinne der Theorie *Unspecific Goal Perspektive* gewertet werden. Von den Studierenden wurde gelobt, dass mit den *Key-Feature*-Fragen Grundlagenwissen geprüft wurde und keine spezifischen Einzelheiten, wie es bei Einfachwahl-Fragen üblich ist. Nach den Meinungen der Studierenden sind Grundlagen aufgrund ihrer Prävalenz in der Berufspraxis relevanter und wichtiger als Einzelheiten, die sich sehr schwer merken lassen. Einzelheiten könnten laut den Studierenden später gelernt werden.

4.9 Ausblick und Schlussfolgerung

Key Features sollten in das Curriculum des Medizinstudiums integriert werden, um Absolventen mit adäquaten CR-Kompetenzen auszustatten. Auch wenn in dieser Studie keine Daten zu einem späteren Zeitpunkt im klinischen Berufsalltag erhoben wurden, legen die Ergebnisse der Fokusgruppengespräche nahe, dass auf diese Weise der Einstieg in den Berufsalltag erleichtert und möglicherweise sogar eine bessere Patientenversorgung gewährleistet werden könnte. Hierbei ist es von Bedeutung, dass medizinische Ausbilder wissen müssen, wie wiederholtes Testen effektiv und effizient genutzt werden kann.

Seit Jahren stellen Einfachwahl-Fragen den wesentlichen Fragetyp im Medizinstudium dar. In mehreren Studien konnte gezeigt werden, dass hinsichtlich des Lernerfolges *Production Tests* (z. B. *Key Features*) den *Recognition Tests* (z. B. Einfachwahl-Fragen) deutlich überlegen sind. Zusätzlich konnte analysiert werden, dass insbesondere *Key Features* prozedurales Wissen prüfen können (Hrynychak et al. 2014), welches essentiell für Medizinstudierende ist, um im klinischen Alltag zu bestehen. Konkret heißt das, dass die Fragenfälle auf alle Fachrichtungen ausgeweitet und – wo immer dies sinnvoll erscheint – Einfachwahl-Fragen ersetzen sollten. Hierfür sollte in weiteren Studien untersucht werden, ob der beobachtete Effekt auch auf andere Inhalte und auf andere Universitäten übertragbar ist. *Key Features* sollten an anderen Universitäten und möglichst in Staatsexamina integriert werden. Das Institut für Medizinische und Pharmazeutische Prüfungsfragen hatte im November 2017 angekündigt, dieses Fragenformat künftig zusätzlich zu dem bisher genutzten Format der Einfachwahlfrage verwenden zu wollen (Jünger 2017). Im Frühjahr 2018 er-

folgte der erstmalige Einsatz von *Key-Feature-Fragen* im 2. Abschnitt der Ärztlichen Prüfung (Vogel 2018). Im Herbst 2018 wurde das Fragenformat weiterhin eingesetzt. Bisher erfolgten die Prüfungen nur schriftlich. Es ist geplant, dass die Prüfungen künftig elektronisch unterstützt werden (<https://www.impp.de/informationen/kompetenzorientierte-gegenstandskataloge/masterplan-medizinstudium-2020-kopie.html>). Wie Kopp et al. (2006) vorschlugen, könnte es sinnvoll sein, eine nationale Datenbank von *Key-Feature-Fällen* zu erstellen, um die Prozesse zur Erstellung von *Key Features* zu erleichtern und ökonomischer zu gestalten. Mit elektronischen Prüfungen kann ressourcenarm geprüft – und unter Berücksichtigung des *Testing Effects* auch trainiert – werden. Zudem ist im Rahmen elektronischer Lehrformate eine Standardisierung der präsentierten Inhalte besonders einfach möglich. Digitale Lehrformate eignen sich in erster Linie für (komplexe) kognitive Lernziele und weniger für das Training praktischer Fertigkeiten. Entsprechend müssen der Unterricht am Krankenbett sowie spezifische Kommunikationstrainings ihre herausragende Stellung im Studium der Humanmedizin selbstverständlich behalten.

Eine wesentliche Weiterentwicklungsmöglichkeit der Fragenfälle wäre die Integration von Videosequenzen und Auskultationsbefunden in die Fragenfälle. Die Studierenden schlugen zudem vor, die E-Fallseminare als Computerspiel zu programmieren. Es müsste in weiteren Studien untersucht werden, ob diese Optimierungsvorschläge den Lernerfolg weiter steigern. Des Weiteren sollten Retentionstests im Abstand von ein bis zwei Jahren stattfinden, um zu testen, ob auf einen längeren Zeitraum die Fragenfälle immer noch den Lesefällen überlegen sind. Zusätzlich sollte untersucht werden, wie sich Absolventen in der klinischen Praxis zurechtfinden, die vorher *Key Features* bearbeiteten. Eine große Herausforderung bei der Planung und Implementierung solcher Studien ist die Kontrolle von Störfaktoren (u.a. Exposition gegenüber Zielinhalten in nicht-standardisierten Lehrformaten).

Die Aussagen der Studierenden in den Fokusgruppengesprächen waren mit den drei Theorien zu den dem *Testing Effect* zugrundeliegenden Mechanismen vereinbar (*Elaborate Retrieval Theory*, *Transfer Appropriate Processing Theory* und *Unspecific Goal Perspective*). Die Ergebnisse sowie die aktuelle Literatur implizieren, dass Tests eine erhöhte mentale Anstrengung von Studierenden fordern sollten. Weiterhin sollten in der Lernphase ähnliche Fragetypen wie im finalen Test gewählt werden und breiteren Themen der Vorzug gegenüber Detailwissen gegeben werden.

5 Zusammenfassung

Der Erwerb der Kompetenz *Clinical Reasoning* ist ein wichtiger Bestandteil der medizinischen Ausbildung. Bislang ist unklar, wie diese Kompetenz bestmöglich gelehrt, gelernt und geprüft werden kann. In der vorliegenden Studie wurde erstmals der in der lernpsychologischen Forschung beschriebene direkte *Testing Effect* hinsichtlich des Trainings solcher Kompetenz untersucht. Nach der dem *Testing Effect* zugrundeliegenden *Retrieval-Hypothese* wirkt sich der Zugriff auf einen Gedächtnisinhalt positiv auf die Retention dieses Inhaltes aus. Hinsichtlich der langfristigen Retention war wiederholtes Prüfen in Vorstudien effektiver als ein wiederholtes Studium (zumeist Lesen) entsprechender Lehrtexte. Die bisherige Forschung fand allerdings entweder in einem nicht-medizinischen Kontext statt oder war auf reines Faktenwissen oder praktische Fertigkeiten fokussiert.

Im Kontext von schriftlichen Prüfungen war der *Testing Effect* besonders ausgeprägt, wenn die Lösung von Fragen mit großer mentaler Anstrengung verbunden war. Bisher dominiert der Fragetyp Einfachwahl die Prüfungen der medizinischen Ausbildung einschließlich der medizinischen Staatsexamina in Deutschland. Aktuelle Daten legen nahe, dass prozedurales Wissen (u.a. die Befähigung zur Abwägung differentialdiagnostischen und –therapeutischen Vorgehens) besonders gut mittels sogenannter *Key-Feature-Fragen* geprüft werden kann.

In der vorliegenden Studie wurde die Frage beantwortet, ob wiederholtes Prüfen mit fallbasierten *Key-Feature-Fragen* hinsichtlich differentialdiagnostischer und therapeutischer Fähigkeiten einen größeren langfristigen Lernerfolg erzeugt als das wiederholte Studium von inhaltlich identischen Fallbeschreibungen. In wöchentlichen E-Fallseminaren wurden Studierende im Wechsel mit den Präsentationsformaten Fragefall und Lesefall konfrontiert; die *Key-Feature-Fragen* waren dabei formativer Natur, d. h., sie waren unbenotet. Zur Objektivierung des Lernerfolgs fand am Anfang des Semesters ein Eingangstest, nach zwölf Wochen ein Ausgangstest und nach sechs Monaten ein Retentionstest statt. Der primäre Endpunkt der Studie war der Unterschied der studentischen Leistungen im Retentionstest für Inhalte, die zuvor in zwei E-Fallseminaren mit *Key-Feature-Fragen* geprüft wurden, im Vergleich zu Inhalten, die zuvor in zwei E-Fallseminaren in Form von „Lesefällen“ (Lern-texten) präsentiert wurden. Zusätzlich zur quantitativen Analyse der Prüfungsergebnisse wurden in Fokusgruppengesprächen die Meinungen der Studierenden zu den E-Fallseminaren eingeholt; zudem wurden die Gespräche hinsichtlich der Vereinbarkeit mit gängigen Theorien (*Elaborate Retrieval Theory*, *Transfer Appropriate Processing Theory*,

Unspecific Goal Perspective) zu den zugrundeliegenden Mechanismen des *Testing Effects* ausgewertet.

Von 87 Studierenden lagen zu allen drei Messzeitpunkten vollständige Daten vor (effektive Rücklaufquote 70,2%). In der primären Endpunktanalyse zeigte sich ein signifikanter Unterschied in den studentischen Leistungen für mehrfach geprüfte Inhalte (Fragenfälle: $56,0 \pm 25,8\%$ der erreichbaren Punkte; Lesefälle: $48,8 \pm 24,7\%$ der erreichbaren Punkte; $p < 0,001$). Dieser Effekt war in einer linearen Regression nach Adjustierung für das Geschlecht und das Leistungsniveau der Studienteilnehmer stabil. Eine detaillierte Analyse des individuellen Antwortverhaltens zeigte, dass Studierende nach inkorrekt beantworteter Frage in einem E-Fallseminar beim erneuten Auftreten einer inhaltlich ähnlichen Frage in einem späteren E-Fallseminar tendenziell eher die richtige Antwort kannten. Zahlreiche studentische Aussagen in den Fokusgruppengesprächen waren mit den in der aktuellen Literatur diskutierten Theorien vereinbar. Aus den Ergebnissen dieser Studie ergeben sich Empfehlungen für die Integration von *Key-Feature*-Prüfungen in humanmedizinische Curricula.

6 Literaturverzeichnis

- Baghdady M, Carnahan H, Lam EW, Woods NN (2014): Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Med Educ* 48, 181-188
- Barrows HS (1996): Problem-based learning in medicine and beyond: A brief overview. *Nw Dir Teach Learn* 68, 3-12
- Bjork RA: Retrieval as a Memory Modifier: An interpretation of negative recency and related phenomena. In: Solso RL (Hrsg.): *Information Processing and Cognition: The Loyola Symposium*. Lawrence Erlbaum Associates, Hillsdale, NJ 1975, 123-144
- Black P, Wiliam D (1998): Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice* 5, 7-74
- Bloch R, Hofer D, Krebs R, Schläppi P, Weiss S, Westkämper R, Jakob L: *Kompetent prüfen; Handbuch zur Planung, Durchführung und Auswertung von Facharztprüfungen*. Institut für Ausbildung Weiterbildung und Fortbildung IAWF, Bern/Wien 1999
- Bortz J, Döring N: *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Auflage; Springer Medizin Verlag, Heidelberg 2006
- Brüstle P: *Kurzanleitung Prüfen mit MC-Fragen*. Universität Freiburg. Freiburg, 2011
- Bundesministerium für Gesundheit: *Approbationsordnung für Ärzte*. 2002
- Butler AC, Roediger HL 3rd (2007): Testing improves long-term retention in a simulated classroom setting. *Eur J Cogn Psychol* 19, 514-527
- Butler AC, Roediger HL 3rd (2008): Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem Cognit* 36, 604-616
- Butler AC, Karpicke JD, Roediger HL 3rd (2007): The effect of type and timing of feedback on learning from multiple-choice tests. *J exp Psychol Appl* 13, 273-281
- Cantillon P (2003): Teaching large groups. *BMJ* 326, 437-440
- Carlsen B, Glenton C (2011): What about N? A methodological study of sample-size reporting in focus group studies. *BMC Med Res Methodol* 11, 26
- Carpenter SK (2009): Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *J Exp Psychol Learn Mem Cogn* 35, 1563-1569
- Carrier M, Pashler H (1992): The influence of retrieval on retention. *Mem Cognit* 20, 633-642
- Clarke PN, Yaros PS (1988): Research blenders: commentary and response. *Transitions to new methodologies in nursing sciences*. *Nurs Sci Q* 1, 147-151
- Cohen J (1992): A power primer. *Psychol Bull* 112, 155-159

Cook DA (2009): The failure of e-learning research to inform educational practice, and what we can do about it. *Med Teach* 31, 158-162

Cook DA, Reichardt CS: *Qualitative and quantitative methods in evaluation research*. Sage Publications, Beverly Hills 1979

Cook DA, Levison AJ, Garside S, Dupras DM, Erwin PJ, Montori VM (2008): Internet-based learning in the health professions: a meta-analysis. *JAMA* 300, 1181-1196

Croskerry P (2009): A universal model of diagnostic reasoning. *Acad Med* 84, 1022-1028

Dawson NV (1993): Physician judgment in clinical settings: methodological influences and cognitive performance. *Clin Chem* 39, 1468-1480

DelSignore LA, Wolbrink TA, Zurakowski D, Burns JP (2016): Test-enhanced E-Learning Strategies in Postgraduate Medical Education: A Randomized Cohort Study. *J Med Internet Res*. 18, e299

Dempster FN: Distributing and managing the conditions of encoding and practice. In: Bjork EL, Bjork RA: *Memory*. Academic Press, San Diego 1996, 317-344

Dobson JL, Linderholm T (2015): Self-testing promotes superior retention of anatomy and physiology information. *Adv Health Sci Educ Theory Pract* 20, 149-161

Dolmans DH, Gijssels WH, Moust JH, de Grave WS, Wolfhagen IH, Van der Vleuten CP (2002): Trends in research on the tutor in problem-based learning: conclusions and implications for educational practice and research. *Med Teach* 24, 173-180

Dolmans DH, de Grave W, Wolfhagen IH, Van der Vleuten CP (2005): Problem-based learning: future challenges for educational practice and research. *Med Educ* 39, 732-741

Dunphy BC, Cantwell R, Bourke S, Fleming M, Smith B, Joseph KS, Dunphy SL (2010): Cognitive elements in clinical decision-making - toward a cognitive model for medical education and understanding clinical reasoning. *Adv Health Sci Educ Theory Pract* 15, 229-250

Elstein AS, Schwartz A (2002): Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ* 324, 729-732

Endres T, Renkl A (2015): Mechanisms behind the testing effect: an empirical investigation of retrieval practice in meaningful learning. *Front Psychol* 6, 1054

Ericsson KA (2004): Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 79, 70-81

Fischer MR, Kopp V, Holzer M, Ruderich F, Jünger J (2005): A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Med Teach* 27, 450-455

Frank JR: The CanMEDS 2005 physician competency framework. Better standards. Better physicians. Better care. The Royal College of Physicians and Surgeons of Canada, Ottawa 2005

Glaser BG, Strauss AL: The discovery of grounded theory: strategies for qualitative reasearch. Aldine Publishing, Chicago, IL 1967

Glover JA (1989): The "Testing" Phenomenon: Not Gone but Nearly Forgotten. J Educ Psychol 81, 392-399

Gordon J (2003): ABC of learning and teaching in medicine: One to one teaching and feedback. BMJ 326, 543-545

Greenglass ER: Proactive Coping. In: Frydenberg E (Hrsg.): Beyond coping: Meeting goals, vision, and challenges. Oxford University Press, London 2002, 37-62

Greenhalgh T (2001): Computer assisted learning in undergraduate medical education. BMJ 322, 40-44

Hasher L, Goldstein D, Toppino T (1977): Frequency and the Conference of Referential Validity. J Verbal Learning Verbal Behav 16, 107-112

Henseling C, Hahn T, Nolting K: Die Fokusgruppen-Methode als Instrument in der Umwelt- und Nachhaltigkeitsforschung. Institut für Zukunftsstudien und Technologiebewertung, Werkstattbericht 82, Berlin 2006

Herzig S, Biehl L, Stelberg H, Hick C, Schmeißer N, Koerfer A (2006): Wann ist ein Arzt ein guter Arzt? Eine inhaltsanalytische Untersuchung von Aussagen medizinischer Professionsvertreter. Dtsch Med Wochenschr, 131, 2883-2888

Hogan RM, Kintsch W (1971): Differential effects of study and test trials on long-term recognition and recall. J Verbal Learning Verbal Behav 10, 562-567

Hogarth RM: Educating Intuition. University of Chicago Press, Chicago, IL 2001

Hrynychak P, Glover Takahashi SG, 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

Jaques D (2003): Teaching small groups. BMJ 326, 492-494

Johnson RB, Onwuegbuzie AJ (2004): Mixed Methods Research: A Research Paradigm Whose Time Has Come. Educ Res 33, 14-26

Jünger J: Pläne und Ausblick zur Zusammenarbeit mit dem IMPP. Vortrag im Rahmen der Delegiertenkonferenz der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, Frankfurt am Main, 11.11.2017

- Kang SH, McDermott KB, Roediger HL 3rd (2007): Test format and corrective feedback modify the effect of testing on long-term retention. *Eur J Cogn Psychol* 19, 528-558
- Karpicke JD, Roediger HL 3rd (2007): Repeated retrieval during learning is the key to long-term retention. *J Mem Lang* 57, 151-162
- Karpicke JD, Roediger HL 3rd (2008): The critical importance of retrieval for learning. *Science* 319, 966-968
- Kassirer JP (2010): Teaching clinical reasoning: case-based and coached. *Acad Med* 85, 1118-1124
- Kern DE, Thomas PA, Howard DM, Bass EB: Curriculum development for medical education: A six-step approach. John Hopkins University Press, Baltimore, MD 1998
- Kopp V, Möltner A, Fischer MR (2006): Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden. *GMS Z Med Ausbild* 23, 3
- Krathwohl DR (2002): A revision of bloom's taxonomy: An overview. *Theory Pract* 41, 212-218
- Krebs R: Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung. Institut für Medizinische Lehre ILM, Abteilung für Ausbildungs- und Examensforschung AAE, Bern 2004
- Kromann CB, Jensen ML, Ringsted C (2009): The effect of testing on skills learning. *Med Educ* 43, 21-27
- Krueger RA, Casey MA: Focus groups: A practical guide for applied research. 5. Auflage; Sage Publications, Los Angeles, CA 2015
- Kulik JA, Kulik CC (1988): Timing of feedback and verbal learning. *Rev Educ Res* 58, 79-97
- Kunina-Habenicht O, Hautz WE, Knigge M, Spies C, Ahlers O (2015): Assessing clinical reasoning (ASCLIRE): Instrument development and validation. *Adv Health Sci Educ Theory Pract* 20, 1205-1224
- Larsen DP, Butler AC, Roediger HL 3rd (2008): Test-enhanced learning in medical education. *Med Educ* 42, 959-966
- Larsen DP, Butler AC, Roediger HL 3rd (2009): Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Med Educ* 43, 1174-1181
- Larsen DP, Butler AC, Lawson AL, Roediger HL 3rd (2012): The importance of seeing the patient: test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. *Adv in Health Sci Educ Theory Pract* 18, 409-425
- Larsen DP, Butler AC, Roediger HL 3rd (2013): Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Med Educ* 47, 674-682

- Logan JM, Thompson AJ, Marshak DW (2011): Testing to enhance retention in human anatomy. *Anat Sci Educ* 4, 243-248
- Mamede S, Schmidt HG, Rikers RM, Penaforte JC, Coelho-Filho JM (2007): Breaking down automaticity: case ambiguity and the shift to reflective approaches in clinical reasoning. *Med Educ* 41, 1185-1192
- Mayring P: *Qualitative Inhaltsanalyse - Grundlagen und Techniken*. 11. Auflage; Beltz Verlag, Weinheim 2010
- McConnell MM, St-Onge C, Young ME (2015): The benefits of testing for learning on later performance. *Adv Health Sci Educ Theory Pract* 20, 305-320
- McDaniel MA, Anderson JL, Derbish MH, Morrisette N (2007a): Testing the testing effect in the classroom. *Eur J Cogn Psychol* 19, 494-513
- McDaniel MA, Roediger HL 3rd, McDermott KB (2007b): Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev* 14, 200-206
- MFT Medizinischer Fakultätentag der Bundesrepublik Deutschland eV: *Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM)*. Berlin 2015
- Miller GE (1990): The assessment of clinical skills/competence/performance. *Acad Med* 65, 63-67
- Möltner A, Schellberg D, Jünger J (2006): Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Z Med Ausbild* 23, Doc53
- Morgan DL (1996): Focus groups. *Annu Rev of Sociol* 22, 129-152
- Morris CD, Bransford JD, Franks JJ (1977): Levels of processing versus transfer appropriate processing. *J Verbal Learning Verbal Behav* 16, 519-533
- Newble DI, Jaeger K (1983): The effect of assessments and examinations on the learning of medical students. *Med Educ* 17, 165-171
- Norman GR (2005): Research in clinical reasoning: past history and current trends. *Med Educ* 39, 418-427
- Norman GR, Schmidt HG (1992): The psychological basis of problem-based learning: A review of the evidence. *Acad Med* 67, 557-565
- Paas F, Kirschner F: The goal-free effect. In: Seel NM (Hrsg.) *Encyclopedia of the Sciences of Learning*. Springer, New York 2012, 1375-1377
- 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
- Page G, Bordage G, Allen T (1995): Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med* 70, 194-201

Raupach T, Muenscher C, Anders S, Steinbach R, Pukrop T, Hege I, Tullius M (2009): Web-based collaborative training of clinical reasoning: a randomized trial. *Med Teach* 31, 431-437

Raupach T, Muenscher C, Pukrop T, Anders S, Harendza S (2010): Significant increase in factual knowledge with web-assisted problem-based learning as part of an undergraduate cardio-respiratory curriculum. *Adv Health Sci Educ Theory Pract* 15, 349-356

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

Reusser K (2005): Problemorientiertes Lernen - Tiefenstruktur, Gestaltungsformen, Wirkung. *Beiträge zur Lehrerbildung* 23, 159-182

Roediger HL 3rd, Marsh EJ (2005): The Positive and Negative Consequences of Multiple-Choice Testing. *J Exp Psychol Learn Mem Cogn* 31, 1155-1159

Roediger HL 3rd, Karpicke JD (2006a): The Power of Testing Memory - Basic Research and Implications for Educational Practice. *Perspect Psychol Sci* 1, 181-210

Roediger HL 3rd, Karpicke JD (2006b): Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci* 17, 249-255

Rotthoff T, Baehring T, Dicken HD, Fahron U, Richter B, Fischer MR, Scherbaum WA (2006): Comparison between Long-Menu and Open-Ended Questions in computerized medical assessments. A randomized controlled trial. *BMC Med Educ* 6, 50

Schiekirka S, Reinhardt D, Heim S, Fabry G, Pukrop T, Anders S, Raupach T (2012): Student perceptions of evaluation in undergraduate medical education: A qualitative study from one medical school. *BMC Med Educ* 12, 45

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

Schmidt HG, Mamede S (2015): How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Med Educ* 49, 961-973

Schulze J, Drolshagen S (2006): Format und Durchführung schriftlicher Prüfungen. *GMS Z Med Ausbild* 23, Doc44

Schurter T: Anleitung zur Herstellung von Kurzantwortfragen (KAF). Institut für Medizinische Lehre ILM, Abteilung für Ausbildungs- und Examensforschung AAE, Bern 2009a

Schurter T: Anleitung zur Korrektur von Kurzantwortfragen (KAF). Institut für Medizinische Lehre ILM, Abteilung für Ausbildungs- und Examensforschung AAE, Bern 2009b

Schuwirth LWT (2009): Is assessment of clinical reasoning still the Holy Grail? *Med Educ* 43, 298-300

Schuwirth LWT, van der Vleuten CP (2003): ABC of learning and teaching in medicine: Written assessment. *BMJ* 326, 643-645

Schuwirth LWT, van der Vleuten CP, Donkers HH (1996a): A closer look at cueing effects in multiple-choice questions. *Med Educ* 30, 44-49

Schuwirth LWT, van der Vleuten CP, Stoffers HE, Peperkamp AG (1996b): Computerized long-menu questions as an alternative to open-ended questions in computerized assessment. *Med Educ* 30, 50-55

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

Spencer J (2003): Learning and teaching in the clinical environment. *BMJ* 326, 591-594

Stanovich KE: *The robot's rebellion: Finding meaning in the age of Darwin*. University of Chicago Press, Chicago, IL 2004

Thompson CP, Wenger SK, Bartling CA (1978): How recall facilitates subsequent recall: A reappraisal. *J Exp Psychol Hum Learn* 4, 210-221

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

van der Vleuten CP, Newble DI (1995): How can we test clinical reasoning? *Lancet* 345, 1032-1034

Vogel T (2018): Ärztliche Prüfung: Das Stex im Wandel. *Dtsch Arztebl Int* 14, 7-9

Vohs KD, Baumeister RF, Loewenstein G: *Do emotions help or hurt decision making? A Hedgesfoxian Perspective*. Russell Sage Foundation, New York, NY 2007

Wahlgren CF, Edelbring S, Fors U, Hindbeck H, Ståhle M (2006): Evaluation of an interactive case simulation system in dermatology and venereology for medical students. *BMC Med Educ* 6, 40

Weinert FE: Vergleichende Leistungsmessung in Schulen - eine umstrittene Selbstverständlichkeit. In: Weinert FE: *Leistungsmessungen in Schulen*; Beltz Verlag, Weinheim 2001, 17-31

Wenger SK, Thompson CP, Bartling CA (1980): Recall facilitates subsequent recognition. *J Exp Psychol Hum Learn* 6, 135-144

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

Wimmers PF, Schmidt HG, Splinter TA (2006): Influence of clerkship experiences on clinical competence. *Med Educ* 40, 450-458

Wirth J, Künsting J, Leutner D (2009): The impact of goal specificity and goal type on learning outcome and cognitive load. *Comput Human Behav* 25, 299-305

Yeung E, Kulasagarem K, Woods N, Dubrowski A, Hodges B, Carnahan H (2016): Validity of a new assessment rubric for a short-answer test of clinical reasoning. *BMC Med Educ* 16, 192

Internetquellen

<https://www.impp.de/informationen/kompetenzorientierte-gegenstandskataloge/masterplan-medizinstudium-2020-kopie.html>; Zugriff am 24.09.2019

<http://www.medizinische-fakultaet-hd.uni-heidelberg.de/CAMPUS-Lehr-und-Pruefungssystem.110009.0.html>; Zugriff am 10.01.2019

<https://www.ucan-assess.org/cms/de/tools/item-and-exam-management/>;
Zugriff am 10.01.2019

https://www.umg.eu/fileadmin/user_upload/Go__ttinger_Lernzielkatalog_Stand_19-07-2017.pdf; Zugriff am 09.01.2019

<https://www.umg.eu/studium-lehre/humanmedizin/>; Zugriff am 09.01.2019

<https://www.umg.eu/studium-lehre/humanmedizin/klinik/module/>; Zugriff am 09.01.2019

7 Anhang

Anhang 1: Zeitplan zur Durchführung der Studie	86
Anhang 2: Bildschirmfotos mit Erklärung zu der Bearbeitung eines Fragenfalls	88
Anhang 3: Bildschirmfotos mit Erklärung zu einem Lesefall	96
Anhang 4: Fotos eines DIPS-Raumes	99
Anhang 5: Fokusgruppen: Leitfragen und Zusatzfragen.....	100
Anhang 6: Itemschwierigkeit und Trennschärfe jedes Items der jeweiligen E-Fallseminare	102
Anhang 7: Mind Map zur Strukturierung und Visualisierung der Fokusgruppengespräche	115

Anhang 1: Zeitplan zur Durchführung der Studie

Semester- woche	Datum	Oberthema des E-Fallseminars	Fälle (Symptome und Hauptdiagnose)
1	23.10.2013	Eingangstest mit Befragung	mit 15 Interventions-Items und 15 Kontroll-Items
2	30.10.2013	Koronare Herzkrankheit	Fall 01: Thoraxschmerz, NSTEMI
			Fall 02: Thoraxschmerz, stabile Angina pectoris
			Fall 04: Dyspnoe, NSTEMI
			Fall 07: Thoraxschmerz, instabile Angina
3	06.11.2013	Herzinsuffizienz	Fall 06: Dyspnoe, Dilatative Kardiomyopathie
			Fall 09: Thoraxschmerz, diastolische Dysfunktion
			Fall 10: Dyspnoe, Chemo-Kardiomyopathie
			Fall 13: Thoraxschmerz, STEMI mit Bradykardie
4	13.11.2013	Vitien & Koronare Herzkrankheit	Fall 02a: Thoraxschmerz, stabile Angina pectoris
			Fall 03: Thoraxschmerz, STEMI
			Fall 20: Bewusstlosigkeit, Aortenstenose
			Fall 24: Dyspnoe, Mitralinsuffizienz
5	20.11.2013	Arrhythmien & Herzinsuffizienz	Fall 08: Dyspnoe, Myokarditis
			Fall 09a: Thoraxschmerz, diastolische Dysfunktion
			Fall 14: Palpitationen, supraventrikuläre Tachykardie
			Fall 18: Bewusstlosigkeit, Synkope
6	26.11.2013	Koronare Herzerkrankung, Vitien, Lungenerkrankung	Fall 04a: Dyspnoe, NSTEMI
			Fall 11: Dyspnoe, Pneumonie
			Fall 12: Dyspnoe, COPD
			Fall 22: Thoraxschmerz, Aortenstenose

Semester- woche	Datum	Oberthema	Fälle
7	02.12.2013	Herzinsuffizienz, Arrhythmien, Gefäßerkrankungen, entzündliche Herzerkrankungen	Fall 18a: Bewusstlosigkeit, Synkope
			Fall 19: Palpitationen, Vorhofflimmern
			Fall 21: Dyspnoe & Fieber, Endokarditis
			Fall 23: Dyspnoe, akute Lungenembolie
8	17.12.2013	Nephrotisches Syndrom, Lungener- krankungen, Vitien	Fall N04a: Unterschenkelödeme, nephrotisches Syndrom bei Systemischem Lupus erythematodes
			Fall 12a: Dyspnoe, COPD
			Fall 15: Dyspnoe, Lungenfibrose
			Fall 20a: Dyspnoe, Aortenstenose
9	19.12.2013	Homöostase, Gefäßerkrankungen, Lungenembolie, entzündliche Herzerkrankungen, Arrhythmien	Fall N01a: Somnolenz, diabetische Ketoazidose
			Fall N02: Abgeschlagenheit, Hyponatriämie
			Fall 05: Thoraxschmerz
			Fall 23a: Dyspnoe, akute Lungenembolie
10	09.01.2014	Nierenversagen, nephrotisches Syn- drom, Lungenerkrankungen	Fall N04a: Unterschenkelödeme, nephrotisches Syndrom bei Systemischem Lupus erythematodes
			Fall N06: Dyspnoe, chronische Niereninsuffizienz
			Fall 11a: Dyspnoe, Pneumonie
			Fall 15a: Dyspnoe, Lungenfibrose
11	14.01.2014	Anämie, Homöostase, Gefäßerkran- kungen, Lungenembolie	Fall H01: Abgeschlagenheit, Anämie
			Fall N02a: Abgeschlagenheit, Hyponatriämie
			Fall 19a: Palpitationen, Vorhofflimmern
			Fall 21a: Dyspnoe & Fieber, Endokarditis
12	24.01.2014	Ausgangstest mit Befragung	mit 15 Interventions-Items und 15 Kontroll-Items
13	14.07.2014	Retentionstest mit Befragung	im Rahmen einer HNO-Übungsklausur mit Einfachwahl-Fragen

Anhang 2: Bildschirmfotos mit Erklärung zu der Bearbeitung eines Fragenfalls

The screenshot shows a software interface for a medical exam. On the left, a sidebar contains a tree view of the exam structure. It lists four parts, each with six questions. The third part, 'Teil (3/4) (abhängig)', is currently selected and highlighted in blue. Below the tree is a 'Prüfung abschließen' button. The main area on the right displays the text of the selected question case. The text describes a 60-year-old patient, Alfred Häuser, who was brought to the hospital by his wife on Sunday evening. They were caught in a storm, and the patient has since developed cough and shortness of breath. He is a diabetic and has a heart condition. The patient is currently febrile (38.1°C) and has crackles in the lungs. The blood pressure is 100/65 mmHg. Below the text is a button labeled 'Weiter'. At the bottom right of the window, there are navigation buttons for 'A+', 'A-', and a yellow sticky note icon. The window title bar indicates the user is 'Prüfung: Andresen, Jill Catherine' and the version is '1.3'.

Teil (1/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)
Teil (2/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)
Teil (3/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)
Teil (4/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)
Prüfung abschließen

Der 60jährige Alfred Häuser wird von seiner Frau am Sonntagabend zu Ihnen in die Notaufnahme gebracht. Am Mittwoch waren sie gemeinsam spazieren und sind auf freiem Feld in ein entsetzliches Gewitter geraten; erst zwei Stunden später kamen sie völlig durchnässt zu Hause an. Seit Freitag leidet Herr Häuser nun unter Husten und heute ist auch noch Luftnot dazugekommen. Herr Häuser ist Diabetiker und hat zudem eine Herzinsuffizienz, die medikamentös behandelt wird. Er wirkt nicht besonders krank und interagiert vollkommen adäquat mit Ihnen. Bei der körperlichen Untersuchung stellen Sie Fieber (38,1°C) und klingende Rasselgeräusche über der Lunge fest. Der Blutdruck des Patienten beträgt 100/65 mmHg.

Weiter

Erklärung: Ein Fragenfall begann mit einer Fallvignette. Um zu der ersten Frage zu gelangen, klickten die Studierenden auf „Weiter“.

The screenshot shows a software interface for a medical exam. On the left, a sidebar lists the exam structure: 'Teil (1/4) (abhängig)' with questions 1/6 to 6/6; 'Teil (2/4) (abhängig)' with questions 1/6 to 6/6; 'Teil (3/4) (abhängig)' with question 1/6 highlighted in blue; and 'Teil (4/4) (abhängig)' with questions 1/6 to 6/6. At the bottom of the sidebar is a 'Prüfung abschließen' button. The main area on the right displays a question under the heading 'Innere Medizin'. The question text is 'Welche Verdachtsdiagnose stellen Sie bei Herrn Häuser?' with '(1 Punkt)' below it. Below the text is a dropdown menu. Underneath the dropdown is a 'Kommentar:' label followed by a text input field. At the bottom of the question area is a grey button labeled 'Auswahl bestätigen'. The top of the window shows 'Datei' and 'Prüfung: Andresen, Jil Catherine Version: 1.3 13:15'. The bottom right corner has '2/24', 'A+', 'A-', and a yellow flag icon.

Erklärung: In das freie Feld unterhalb der Frage, konnte nun die Antwort eingegeben werden. Hier öffnete sich eine Dropdown-Liste mit Antwortmöglichkeiten nach Eingabe von mindestens drei Buchstaben.

The screenshot shows a software interface for a medical exam. On the left is a navigation tree with sections like 'Teil (1/4)', 'Teil (2/4)', 'Teil (3/4)', and 'Teil (4/4)', each containing 'Frage' (questions). The main area displays a question under the heading 'Innere Medizin' (Internal Medicine). The question asks for a suspected diagnosis for a patient named Herr Häuser. A dropdown menu is open, showing 'Pneumonie' (Pneumonia) as the selected answer. Below the dropdown is a 'Kommentar:' (Comment) field. At the bottom of the question area is a button labeled 'Auswahl bestätigen' (Confirm selection). The interface also includes a status bar at the bottom with 'A+', 'A-', and a yellow icon.

Prüfung: Andresen, Jill Catherine Version: 1.3 13:17

Teil (1/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Teil (2/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Teil (3/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Teil (4/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Prüfung abschließen

Vignette anzeigen...

Innere Medizin

Welche Verdachtsdiagnose stellen Sie bei Herrn Häuser?
(1 Punkt)

Pneumonie

Kommentar:

Auswahl bestätigen

2/24

A+ A-

Erklärung: Die Studierenden wählten ihre jeweilig gewünschte Antwort aus der Dropdown-Liste und klickten auf „Auswahl bestätigen“.

The screenshot shows a software interface for a medical exam. On the left is a navigation tree with sections like 'Teil (1/4) (abhängig)', 'Teil (2/4) (abhängig)', 'Teil (3/4) (abhängig)', and 'Teil (4/4) (abhängig)'. The 'Teil (3/4) (abhängig)' section is expanded, showing 'Frage (1/6)', 'Frage (2/6)', 'Frage (3/6)', 'Frage (4/6)', 'Frage (5/6)', and 'Frage (6/6)'. The 'Frage (2/6)' is selected and highlighted in blue.

The main area displays the question:

Innere Medizin
Die Anamnese ist typisch für eine Pneumonie.

Ausführliche Erläuterungen zu Frage 1
Die Anamnese zeigt Husten, Sputum, Fieber und Abgehörtes Auskultationszeichen spricht die Diagnose für eine Pneumonie. Die Lungensymptome sind oben beschrieben. Die Anamnese zeigt Fieber bei typischer Infektionsgenese und kommt nur beim Lungensystem einer primären Virus-Respirationsstörung zu erwarten.
Als richtig wurden die folgenden Antwortoptionen genannt:
Pneumonie

Welche Untersuchung führen Sie durch, um diese Verdachtsdiagnose zu bestätigen?
(1 Punkt)

Below the question is a dropdown menu, a 'Kommentar:' text input field, and a large 'Auswahl bestätigen' button.

At the bottom right of the window, there are navigation buttons 'A+', 'A-', and a yellow sticky note icon. The status bar shows '1/24'.

Erklärung: Eine Lösung erhielten die Studierende in diesem Beispiel mit „Die Anamnese ist typisch für eine Pneumonie.“ Um ausführliche Erklärungen für die Antwort zu erhalten, bestand die Möglichkeit das Feedback „Ausführliche Erläuterungen zu Frage 1“ anzuklicken. Wenn Studierende das Feedback nicht anklicken wollten, konnten diese mit der nächsten Frage fortfahren.

The screenshot shows a software interface for a medical exam. On the left, a sidebar lists various parts of the exam, including 'Teil (1/4)', 'Teil (2/4)', 'Teil (3/4)', and 'Teil (4/4)', each with sub-items for 'Frage (1/6)' through 'Frage (6/6)'. The main area displays a question under the heading 'Innere Medizin'. The question text is: 'Die Anamnese ist typisch für eine Pneumonie. Welche Un (1 Punkt)'. Below the question, there are input fields and a 'Komme' button. A feedback window titled 'Ausführliche Erläuterungen zu Frage 1' is open, containing the following text: 'Die Klinik aus Husten, Dyspnoe, Fieber und klingenden Rasselgeräuschen spricht am ehesten für eine Pneumonie. Ein Lungenödem ist bei oben beschriebener Klinik unwahrscheinlich; erstens wäre hier Fieber kein typisches Begleitsymptom, und zweitens wären beim Lungenödem eher grobblasige feuchte Rasselgeräusche zu erwarten. Als richtig wurden die folgenden Antwortoptionen gewertet: - Pneumonie - Lungenentzündung'. The feedback window has a 'Schließen' button. At the bottom of the main area, there is an 'Auswahl bestätigen' button. The interface also shows a status bar at the bottom right with '3/24' and navigation buttons 'A+', 'A-', and a yellow icon.

Erklärung: Bei Anklicken des Feedbacks, vergrößerte sich das Feedback und der Studierende konnte die ausführlichen Erläuterungen lesen. Um zur nächsten Frage zu gelangen, musste das Fenster wieder per Klick geschlossen werden.

Prüfung: Andresen, Jill Catherine | Version: 1.3 | 13:26

Teil (1/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Teil (2/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Teil (3/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Teil (4/4)
(abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Prüfung abschließen

Vignette anzeigen...

Innere Medizin

Die Anamnese ist typisch für eine Pneumonie.

Ausführliche Erläuterungen zu Frage 1
Die Abbildung zeigt typische Symptome, Fieber und Abhängigkeit. Ausgeprägtere Symptome sind zu erwarten.
Die Abbildung zeigt typische Symptome, Fieber und Abhängigkeit. Ausgeprägtere Symptome sind zu erwarten.
Die Abbildung zeigt typische Symptome, Fieber und Abhängigkeit. Ausgeprägtere Symptome sind zu erwarten.
Die Abbildung zeigt typische Symptome, Fieber und Abhängigkeit. Ausgeprägtere Symptome sind zu erwarten.

Welche Untersuchung führen Sie durch, um diese Verdachtsdiagnose zu bestätigen?
(1 Punkt)

Weiß nicht

Kommentar: Blutentnahme

Auswahl bestätigen

3/24

A+ A-

Erklärung: Wenn Studierende die gewünschte Antwort zu einer Frage nicht in der Dropdown-Liste fanden, bestand die Möglichkeit, die Frage mit „Weiß nicht“ zu beantworten. Im Kommentarfeld darunter konnte nun die vermutete Antwort eingetragen werden und durch das Klicken von „Auswahl bestätigen“ gesichert werden.

Das beschriebene Prozedere führten die Studierenden bei jeder der fünf Fragen im Fragenfall durch. Des Öfteren umfassten Fragen EKG's, Laborbefunde oder Röntgenbilder.

Prüfung: Andresen, Jil Catherine Version: 1.3 13:28

Teil (1/4) (abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Teil (2/4) (abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Teil (3/4) (abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Teil (4/4) (abhängig)
Frage (1/6)
Frage (2/6)
Frage (3/6)
Frage (4/6)
Frage (5/6)
Frage (6/6)

Prüfung abschließen

Vignette anzeigen...


Innere Medizin

Zur Sicherung der Verdachtsdiagnose muss ein Röntgenbild des Thorax angefertigt werden - idealerweise in zwei Ebenen.

Ausführliche Erläuterungen zu Frage 2

haben die Adresse und die Applikation sowie zur Diagnose einer Pneumonie eine Erläuterung des Thorax angefertigt werden. Um sicher zu gehen, sollte der Thorax in zwei Ebenen dargestellt werden. Die Frontalansicht wird durch ein schalenförmiges Transparenzvermögen (Aussicht im Brustbereich) über das Thorax, wie die beiden Seiten sind, unterstützen können. Ein Bild kann ein Schlüsselstein sein.

Welche



<http://www.stanford.edu/dept/radiology/radiologysite>

(1 Punkt) Schließen

Kommentar:

Auswahl bestätigen

4/24

A+ A-

Erklärung: Ein Beispiel eines Röntgenbildes.

Prüfung: Andresen, Jill Catherine Version: 1.3 13:30

Teil (1/4) (abhängig)
 Frage (1/6)
 Frage (2/6)
 Frage (3/6)
 Frage (4/6)
 Frage (5/6)
 Frage (6/6)

Teil (2/4) (abhängig)
 Frage (1/6)
 Frage (2/6)
 Frage (3/6)
 Frage (4/6)
 Frage (5/6)
 Frage (6/6)

Teil (3/4) (abhängig)
 Frage (1/6)
 Frage (2/6)
 Frage (3/6)
 Frage (4/6)
 Frage (5/6)
 Frage (6/6)

Teil (4/4) (abhängig)
 Frage (1/6)
 Frage (2/6)
 Frage (3/6)
 Frage (4/6)
 Frage (5/6)
 Frage (6/6)

Prüfung abschließen

Vignette anzeigen...

Innere Medizin

Bevor Sie die antibiotische Therapie umstellen, sollten Sie Blutkulturen abnehmen, um einen Erregernachweis führen zu können. Vom Antibiogramm hängt dann die Entscheidung über die weitere antibiotische Therapie ab.

Ausführliche Erläuterungen zu Frage 5

Alle Punkte zu einer der folgenden Antwortoptionen gewährt:

- Blutkultur
- Antibiogramm
- Kultur von Sputum
- Kultur von Pleuralflüssigkeit
- Kultur von Urin
- Kultur von Stuhl
- Kultur von Nasensekret
- Kultur von Urin
- Kultur von Pleuralflüssigkeit
- Kultur von Sputum

MC-Frage

Diese Frage wurde im WS 2010/11 von 95% der Studierenden richtig beantwortet; ihre Trennschärfe betrug 0,25.

Ein Patient gibt an, bei einem Campingurlaub vor 10 Tagen in einen Bach gefallen und mit nasser Kleidung darauf 6 Stunden gewandert zu sein. Seit drei Tagen könne er schlechter durchatmen und habe bei tiefem Einatmen Schmerzen unter dem rechten Rippenbogen, gestern Nacht habe er stark geschwitzt, seit heute habe er Husten ohne Auswurf und deutliche Dyspnoe. Was ist zur Diagnosestellung einer Pneumonie vor Beginn einer Antibiose **am ehesten** hinreichend?
 (Bitte kreuzen Sie eine Antwort an!)

(1 Punkt)

- Sputumkultur
- CT Thorax
- Bronchoskopie mit bronchoalveolärer Lavage
- Auskultation, Palpation, Röntgen Thorax
- Pleurapunktion mit Zytologie und Kultur

Auswahl bestätigen

7/24

A+ A-

Erklärung: Am Ende des jeweiligen Fragenfalls beantworteten die Studierenden eine Einfachwahl-Frage (hier zum besseren Verständnis für die Studierenden als MC-Frage beschrieben). Nach Anklicken von „Auswahl bestätigen“, folgte der nächste Fragenfall.

Anhang 3: Bildschirmfotos mit Erklärung zu einem Lesefall

Datel Prüfung: Andresen, Jil Catherine Version: 1.3 13:39

Teil (1/1)
(unabhängig)

- Frage (1/4)
- Frage (2/4)
- Frage (3/4)
- Frage (4/4)

Prüfung abschließen

Innere Medizin

Der 60jährige Alfred Häuser wird von seiner Frau am Sonntagabend in die Notaufnahme gebracht. Der diensthabenden Ärztin berichtet er, dass er am Mittwoch mit seiner Frau gemeinsam spazieren war und dass sie auf freiem Feld in ein entsetzliches Gewitter geraten seien, erst zwei Stunden später seien sie völlig durchnässt zu Hause angekommen. Seit Freitag leide er nun unter Husten und heute sei auch noch Luftnot dazugekommen.

Herr Häuser ist Diabetiker und hat zudem eine Herzinsuffizienz, die medikamentös behandelt wird. Er wirkt nicht besonders krank und interagiert vollkommen adäquat mit der diensthabenden Ärztin auf der Notaufnahme. Bei der körperlichen Untersuchung stellt sie Fieber (38,1°C) und klingende Rasselgeräusche über der Lunge fest. Der Blutdruck des Patienten beträgt 100/65 mmHg. Zur Sicherung der offensichtlichen Verdachtsdiagnose einer Pneumonie lässt die Ärztin ein Röntgenbild des Thorax in zwei Ebenen anfertigen (siehe Bild - aus technischen Gründen unter der MC-Frage).

Die Bilder zeigen eine Lobärpneumonie des linken Oberlappens. Zur Beurteilung der Notwendigkeit einer stationären Aufnahme bestimmt die Ärztin den CRB-65-Index. Die Atemfrequenz beträgt aktuell zwar nur 20/min, aber dafür ist der systolische Blutdruck von Herrn Häuser langsam abgefallen und liegt jetzt bei 90 mmHg. Somit besteht eine Indikation zur stationären Behandlung. Die Ärztin leitet eine antibiotische Therapie der ambulant erworbenen Pneumonie ein und verlegt den Patienten auf die Normalstation.

Zwei Tage später wird der ärztliche Hausdienst von den zuständigen Pflegekräften kontaktiert, weil das Fieber von Herrn Häuser auf 39,2°C gestiegen ist. Rasch verordnet er ein Medikament zur Fiebersenkung und nimmt Blutkulturen ab.

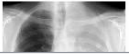

Hintergrundinformationen zum Fall:

Alle Rechte vorbehalten. Nachdruck, Vervielfältigung und Verbreitung, auch auszugsweise, ist ohne schriftliche Genehmigung der Elsevier GmbH, München, Deutschland. Die Weitergabe oder Nutzung ist nur für den persönlichen Gebrauch des Lesers gestattet. Nachdruck, Vervielfältigung und Verbreitung, auch auszugsweise, ist ohne schriftliche Genehmigung der Elsevier GmbH, München, Deutschland. Die Weitergabe oder Nutzung ist nur für den persönlichen Gebrauch des Lesers gestattet.

MC-Frage

Diese Frage wurde im WS 2010/11 von 95% der Studierenden richtig beantwortet; ihre Trennschärfe betrug 0,25.

Ein Patient gibt an, bei einem Campingurlaub vor 10 Tagen in einen Bach gefallen und mit nasser Kleidung darauf 6 Stunden gewandert zu sein. Seit drei Tagen könne er schlechter durchatmen und habe bei tiefem Einatmen Schmerzen unter dem rechten Rippenbogen, gestern Nacht habe er stark geschwitzt, seit heute habe er Husten ohne Auswurf und deutliche Dyspnoe. Was ist zur Diagnosestellung einer Pneumonie vor Beginn einer Antibiose **am ehesten** hinreichend?

0/4
A+ A-

Erklärung: Beim Lesefall lasen die Studierenden den Lesetext, welcher inhaltlich identisch zum Fragenfall war. Um Hintergrundinformationen zu erhalten, konnte hier das Feld „Hintergrundinformationen zum Fall“ angeklickt werden.

Hintergrundinformationen zum Fall:

Sowohl die Anamnese als auch die klinischen Symptome und Befunde (Husten, Dyspnoe, Fieber und klingende Rasselgeräusche) sprechen am ehesten für das Vorliegen einer Pneumonie. Neben der Anamnese und der Auskultation sollte zur Diagnose einer Pneumonie eine Röntgenaufnahme des Thorax angefertigt werden. Um nichts zu übersehen, sollte der Thorax in zwei Ebenen dargestellt werden. Die Frontalansicht wird dabei im posterior-anterioren Strahlengang aufgenommen (Aufnahme im Stehen, Strahlenquelle hinter dem Thorax); nur bei liegenden Patienten wird ein antero-posteriorer (a.p.) Strahlengang mit Strahlenquelle vor dem Thorax akzeptiert. Laboruntersuchungen können die Verdachtsdiagnose einer Pneumonie stützen (z.B. wenn das CRP und die Leukozytenzahl erhöht sind). Erhöhte Infektparameter sagen aber meist noch nichts über die Lokalisation der Infektion aus.

In den Röntgen-Aufnahmen wird eine Lobärpneumonie des linken Oberlappens dargestellt; die Lokalisation ist besonders gut im Seitbild zu erkennen, denn der linke Oberlappen ist ventral gelegen. Der typische radiologische Befund einer Lobärpneumonie ist das flächige Infiltrat (großflächige Transparenzminderung), unter Umständen mit luftgefüllten und daher dunkel dargestellten Bronchien. Dieses „positive Bronchopneumogramm“ fehlt bei der Bronchopneumonie, die außerdem durch eine segmentale (nicht flächige) Transparenzminderung gekennzeichnet ist. Bei einer interstitiellen Pneumonie wären eher fleckige bzw. netzartige Transparenzminderungen zu erwarten.

Die Empfehlung zur stationären Aufnahme hängt vom CRB-65-Index ab. Dieser Score wurde entwickelt, um das Sterblichkeitsrisiko einer ambulant erworbenen Pneumonie besser vorhersagen zu können. Im folgenden sind die vier Parameter aufgelistet, die in den Score eingehen:

C = Confusion (Verwirrtheit)
R = Respiratory rate (Atemfrequenz > 30/min)
B = Blood pressure (Blutdruck \leq 90mmHg oder < 60mmHg diastolisch)
Alter > 65 Jahre

Für jedes dieser Kriterien wird ein Punkt vergeben. Bei einem Punktescore von ≥ 1 Punkt ist eine stationäre Aufnahme indiziert.

Die Kriterien „confusion“, „respiratory rate“ und „age“ sind bei Herrn Häuser bisher negativ. Auch der Blutdruck ist initial mit 100 mmHg systolisch noch nicht pathologisch. Da der Blutdruck im Verlauf jedoch auf 90 mmHg systolisch fällt, ergibt sich ein Summenscore von 1 Punkt, womit die stationäre Aufnahme empfohlen wird.

Offenbar war die initial verordnete antibiotische Therapie nicht ausreichend – hierfür spricht die Tatsache, dass Herr Häuser im Verlauf wieder höheres Fieber bekam. Bevor die antibiotische Therapie umgestellt wird, sollten in einem solchen Fall Blutkulturen abgenommen werden, um einen Erregernachweis führen zu können. Vom Antibiogramm hängt dann die Entscheidung über die weitere antibiotische Therapie ab.

Schließen

Erklärung: Daraufhin öffnete sich ein Fenster mit den Hintergrundinformationen. Diese umfassten die Informationen der fünf Einzel-Feedbacks „Ausführliche Erläuterungen“ des jeweiligen Fragenfalls. Um zurück zum Lesefall zu gelangen, musste das Fenster per Klick geschlossen werden.

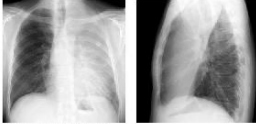
Prüfung: Andresen, Jil Catherine Version: 1.3 13:44

Teil (1/1)
(unabhängig)
Frage (1/4)
Frage (2/4)
Frage (3/4)
Frage (4/4)
Prüfung abschließen

Hintergrundinformationen zum Fall:
 Ein 45-jähriger Mann, Nichtraucher, berichtet über ein plötzliches, linksseitiges Brustschmerz, das bei tiefem Einatmen verstärkt wird. Er berichtet über eine leichte Erregung und Schweißausbrüche. Die Schmerzen sind nicht durch Bewegung oder Nahrungsaufnahme verstärkt. Er hat keine Fieber, keine Dyspnoe und keine andere Symptome. Die Schmerzen sind seit Beginn der Erkrankung nicht abgeklungen. Er hat keine Vorerkrankungen, keine Allergien und keine Einnahme von Medikamenten. Die Schmerzen sind seit Beginn der Erkrankung nicht abgeklungen. Er hat keine Vorerkrankungen, keine Allergien und keine Einnahme von Medikamenten. Die Schmerzen sind seit Beginn der Erkrankung nicht abgeklungen. Er hat keine Vorerkrankungen, keine Allergien und keine Einnahme von Medikamenten.

MC-Frage
 Diese Frage wurde im WS 2010/11 von 95% der Studierenden richtig beantwortet; ihre Trennschärfe betrug 0.25.

Ein Patient gibt an, bei einem Campingurlaub vor 10 Tagen in einen Bach gefallen und mit nasser Kleidung darauf 6 Stunden gewandert zu sein. Seit drei Tagen könne er schlechter durchatmen und habe bei tiefem Einatmen Schmerzen unter dem rechten Rippenbogen, gestern Nacht habe er stark geschwitzt, seit heute habe er Husten ohne Auswurf und deutliche Dyspnoe. Was ist zur Diagnosestellung einer Pneumonie vor Beginn einer Antibiose **am ehesten** hinreichend?



(Bitte kreuzen Sie **eine** Antwort an!)
(1 Punkt)

Auskultation, Palpation, Röntgen Thorax

CT Thorax

Bronchoskopie mit bronchoalveolärer Lavage

Sputumkultur

Pleurapunktion mit Zytologie und Kultur

Auswahl bestätigen

0/4

A+ A-

Erklärung: Zum Ende des Lesefalls beantworteten die Studierenden eine Einfachwahl-Frage (hier zum besseren Verständnis für die Studierenden als MC-Frage beschrieben). Nach dem Anklicken der gewünschten Antwort und anschließend „Auswahl bestätigen“, folgte der nächste Lesefall.

Anhang 4: Fotos eines DIPS-Raumes



Anhang 5: Fokusgruppen: Leitfragen und Zusatzfragen**1) Was ist Ihnen von den E-Fallseminaren besonders in Erinnerung geblieben? Was war besonders hervorstechend?**

- a. Hat sich Ihre Meinung über die E-Fallseminare im Verlauf des Semesters irgendwie geändert?
- b. Welche Rolle haben die MC-Fragen für Sie gespielt?

2) Womit haben Sie während der E-Fallseminare die meiste Zeit verbracht?

- a. Wonach haben Sie entschieden, wie lange Sie in einem E-Fallseminar sitzengeblieben sind?
- b. Wann bzw. unter welchen Umständen haben Sie während der E-Fallseminare das Gefühl, Ihre Zeit sinnvoll zu nutzen?

3) Was war in Ihren Augen der Zweck der E-Fallseminare?

- a. Was haben Sie aus den E-Fallseminaren mitgenommen?
- b. Welche Aspekte an den E-Fallseminaren waren besonders hilfreich/sinnvoll?
- c. Inwieweit hatten Sie das Gefühl, dass Sie durch die E-Fallseminare beim Lernen unterstützt wurden?
- d. Woran können Sie selbst erkennen, ob Sie in einem E-Fallseminar etwas gelernt haben?
- e. Was haben Sie in den E-Fallseminaren gelernt, was Sie in anderen Lehrveranstaltungen nicht gelernt haben?
- f. Inwiefern haben die E-Fallseminare Ihnen dabei geholfen, Inhalte aus verschiedenen Modulen miteinander zu verknüpfen?

4) Wie relevant für die Praxis erschienen Ihnen die Inhalte der E-Fallseminare?

- a. Wie realistisch erschienen Ihnen die Fälle im E-Fallseminar? Wodurch wären sie evtl. noch realistischer geworden?
- b. Wie haben Sie in den E-Fallseminaren zwischen wichtigen und weniger wichtigen Inhalten unterschieden?
- c. Welchen Effekt hatte es für Sie, dass Fälle wiederholt aufgetaucht sind?
- d. Hat es für Sie persönlich eine Rolle gespielt, ob die E-Fallseminare während des Kardiologie-, Nephrologie- oder Onkologiemoduls stattgefunden haben?
- e. Wie gut konnten Sie sich während der E-Fallseminare auf die Inhalte konzentrieren?

- f. Wie ernsthaft haben Sie die Fragenfälle bearbeitet? (Im Ausgangstest anders als während des Semesters?)

5) Inwiefern hat die Teilnahme an den E-Fallseminaren Ihr Lernverhalten beeinflusst?

- a. (Wie) haben Sie sich auf die E-Fallseminare vorbereitet?
- b. (Wie) haben Sie die E-Fallseminare nachbereitet?
- c. Hatten die E-Fallseminare irgendeine motivierende Wirkung auf Sie? D. h. haben Sie nach einem E-Fallseminar irgendetwas gemacht, was Sie sonst nicht gemacht hätten?
- d. Haben die E-Fallseminare Sie manchmal bzgl. Ihres Wissens verunsichert (in dem Sinne, dass Sie gemerkt haben, etwas nicht mehr zu wissen, was Sie eigentlich schon gelernt hatten)? Wenn ja, welchen Effekt hatte das?

6) (falls noch nicht ausreichend beantwortet) Im wöchentlichen Wechsel haben Sie entweder Fragenfälle oder Lesefälle bearbeitet. Was ist für Sie der wesentliche Unterschied zwischen den beiden Formaten?

- a. Haben Sie eines der beiden Formate (Lese- oder Fragenfälle) als interessanter empfunden? Wenn ja, warum?
- b. Welche Aspekte/Teile der E-Fallseminare haben Ihnen mehr Spaß gemacht, welche weniger? Warum?
- c. Wie hat Ihnen der Aufbau der Fragenfälle und der Lesefälle gefallen (Fragenformat / Feedback)?
- d. Inwiefern sind Sie unterschiedlich an die Fragen- und Lesefälle herangegangen?
- e. Wie intensiv haben Sie das Feedback bzw. die Hintergrundinfos genutzt? Was hat Ihnen das gebracht?
- f. Haben Sie die Inhalte, die Sie bei den Fragenfällen behandelt haben oder die, die Sie in den Lesefällen behandelt haben, besser behalten? Haben Sie eine Idee warum?

7) Wie könnte das Lehrformat „E-Fallseminar“ in Ihren Augen weiter verbessert werden?

- a. Wie könnte sich Ihr Lernerfolg durch die E-Fallseminare weiter steigern lassen?
- b. Wie müssten E-Fallseminare gestaltet sein, damit Sie freiwillig hingehen würden?

Anhang 6: Itemschwierigkeit und Trennschärfe jedes Items der jeweiligen E-Fallseminare

Eingangstest	Item	Itemschwierigkeit	Trennschärfe
Fall 1	KF 1: Ableiten der Verdachtsdiagnose „Akute Lungenembolie“ bei typischer Klinik und Risikoprofil	0,46	0,191
	KF 2: Wells-Score zur Entscheidung für oder gegen eine D-Dimer-Bestimmung	0,08	0,105
	KF 3: sofortiges Thorax-CT bei hoher Wahrscheinlichkeit und klinischer Stabilität	0,09	0,340
	KF 4: Echokardiographie oder Troponinbestimmung zur Risikostratifizierung	0,03	0,164
	KF 5: Indikation zur Fibrinolyse bei hämodynamischer Instabilität	0,41	0,248
	KF 6: Erkennen eines nephrotischen Syndroms anhand der Labor- und Urinbefunde	0,15	0,372
	KF 7: V. a. Systemischer Lupus erythematodes bei typischer Klinik und wegweisendem AK-Befund	0,04	0,139
	KF 8: Nierenbiopsie zur Diagnosesicherung	0,17	0,190
	KF 9: Behandlung der Lupus-Nephritis durch einen Steroidstoß	0,37	0,371
Fall 2	KF 1: Verdachtsdiagnose COPD bei chronischem Husten und Auswurf bei einem Raucher mit expiratorischem Giemen	0,47	0,086
	KF 2: Erkennen einer Obstruktion in der Lungenfunktion	0,15	0,099
	KF 3: Erkennen einer links-apikalen Lobärpneumonie im Röntgenbild	0,13	0,236
	KF 4: CRB-65-Index zur Entscheidung über die Empfehlung zur stationären Aufnahme	0,01	0,038
	KF 5: BGA bei V. a. beginnende CO ₂ -Narkose	0,30	0,248
	KF 6: Non-invasive Ventilation bei nachgewiesener Hyperkapnie	0,13	0,125
Fall 3	KF 1: Verdacht auf sekundäre Hypertonie bei Manifestation > 60 J., diskrepanten Organschäden und Therapierefraktärität	0,02	0,015
	KF 2: Verdacht auf diastolische Dysfunktion bei guter EF, NYHA II und erhöhtem BNP	0,02	0,054
	KF 3: ACE-Hemmer bei V. a. ACE-Hemmer-Husten absetzen	0,60	0,331
	KF 4: Krankenhaus-Einweisung bei Hyponatriämie < 120 mmol/l	0,04	-0,100
	KF 5: Thiazid-Diuretika als mögliche Auslöser einer Hyponatriämie	0,35	0,350
	KF 6: zentrale pontine Myelinolyse als Komplikation einer zu raschen Natrium-Substitution	0,01	0,010
Fall 4	KF 1: Schellong-Test bei V. a. orthostatisch bedingte Synkope	0,22	0,173
	KF 2: Schädel-CT zur Diagnose eines Hirninfarktes	0,61	0,442
	KF 3: Erkennen einer Tachyarrhythmia absoluta im EKG	0,29	0,263
	KF 4: CHA ₂ DS ₂ -VASc-Score zur Indikationsstellung zur dauerhaften Antikoagulation	0,11	0,249
	KF 5: laborchemische Diagnose einer manifesten Hyperthyreose nach Kontrastmittelgabe	0,60	0,195
	KF 6: Absetzen von Amiodaron bei manifester Hyperthyreose	0,08	0,222
	KF 7: klinischer Verdacht auf Lungenfibrose bei Dyspnoe und inspiratorischem Velcro-Knistern	0,08	0,242
	KF 8: Amiodaron als Verursacher einer Lungenfibrose	0,17	0,233
	KF 9: Indikationsstellung zur Langzeit-Sauerstofftherapie aufgrund eines pO ₂ < 55 mmHg in der arteriellen BGA	0,45	0,152

E-Fallseminar 2	Item	Itemschwierigkeit	Trennschärfe
Fall 1	KF 1: Verdachtsdiagnose „Akuter Infarkt“ aufgrund typischer Symptomatik und positiven Risikoprofils	0,97	0,323
	KF 2: Analgetische Therapie mit Morphin als Teil der Akuttherapie (ASS, Clopidogrel, Heparin, Nitro, Betablocker, Sauerstoff)	0,93	0,519
	KF 3: Bestimmung des Troponin T bei normalem EKG und weiterhin bestehendem Verdacht	0,90	0,374
	KF 4: Entscheidung zur sofortigen Koronarangiographie bei positivem Troponin T	0,84	0,460
	KF 5: Indikationsstellung zur PTCA/Stentimplantation bei filiformer Stenose	0,84	0,475
Fall 2	KF 1: Ergometrie zum Ischämienachweis	0,66	0,371
	KF 2: Szintigraphie zum Ischämienachweis bei fehlender Ausbelastung in der Ergometrie und Z. n. Schrittmacherimplantation	0,33	0,255
	KF 3: Indikation zur Koronarangiographie bei positivem Ischämienachweis in der Szintigraphie	0,82	0,409
	KF 4: Indikation zur operativen Myokardrevaskularisation bei 3-Gefäß-KHK mit Hauptstammstenose	0,54	0,344
	KF 5: Rehabilitation nach operativer Myokardrevaskularisation	0,07	0,231
Fall 7	KF 1: Erkennen der Diagnose „instabile Angina pectoris“	0,97	0,097
	KF 2: Indikation zur Ergometrie bei instabiler Angina pectoris ohne Troponin-Erhöhung	0,56	0,300
	KF 3: Indikation zur Koronarangiographie bei pathologischer Ergometrie	0,64	0,472
	KF 4: Indikation zur PTCA + Stentimplantation bei isolierter RCA-Stenose	0,82	0,437
	KF 5: Tabakentwöhnung als wichtigste prognostische Maßnahme	0,77	0,518
Fall 4	KF 1: Interpretation eines leicht erhöhten Troponins im Sinne einer hypertensiven Entgleisung	0,23	0,221
	KF 2: Indikation zur Koronarangiographie bei steigendem Troponin und Beschwerdepersistenz	0,92	0,341
	KF 3: Verdacht auf Vorderwand-Aneurysma bei persistierenden ST-Hebungen	0,03	0,129
	KF 4: Echokardiographie zur Bestätigung der Verdachtsdiagnose	0,82	0,242
	KF 5: V. a. Stentverschluß bei Beschwerden nach verfrühtem Absetzen der dualen Thrombozytenaggregationshemmung	0,57	0,435

E-Fallseminar 3	Item	Itemschwierigkeit	Trennschärfe
Fall 6	KF 1: Verdachtsdiagnose „Globaldekompensation" bei typischer Klinik und passendem klinischen Untersuchungsbefund	0,77	-0,021
	KF 2: Transthorakale Echokardiographie zur Bestimmung der linksventrikulären Ejektionsfraktion	0,90	0,381
	KF 3: Indikation zur diuretischen Therapie bei Dekompensation	0,37	0,309
	KF 4: Kaliummangel als Ursache einer ventrikulären Arrhythmie	0,31	-0,032
	KF 5: Empfehlung zur täglichen Gewichtskontrolle zur frühzeitigen Erkennung einer drohenden Dekompensation	0,58	0,177
Fall 9	KF 1: Verdachtsdiagnose „sekundäre arterielle Hypertonie" bei typischer Anamnese	0,21	0,181
	KF 2: Verdachtsdiagnose „Obstruktives Schlafapnoesyndrom" bei typischer Anamnese	0,50	0,495
	KF 3: Poly(somno)graphie bei V. a. Obstruktives Schlafapnoesyndrom	0,63	0,490
	KF 4: Verdacht auf diastolische Dysfunktion bei normaler LVEF, NYHA II und erhöhtem BMP	0,17	0,457
	KF 5: Verdachtsdiagnose „Tako-Tsubo-Kardiomyopathie" bei unauffälligen Koronarien und apical ballooning	0,27	0,309
Fall 10	KF 1: Transthorakale Echokardiographie bei Verdacht auf Herzinsuffizienz	0,40	0,040
	KF 2: Identifikation der Chemotherapie als Auslöser der Herzinsuffizienz	0,15	0,145
	KF 3: koronarangiographischer Ausschluß einer KHK als alternativer Ursache einer Herzinsuffizienz	0,33	-0,048
	KF 4: Kontraindikation von Spironolacton bei Niereninsuffizienz mit Kreatinin > 2,5 mg/dl	0,10	-0,006
	KF 5: ACE-Hemmer absetzen, wenn V. a. ACE-I-Husten besteht	0,79	0,380
Fall 13	KF 1: Erkennen eines Inferolateralwandinfarktes im EKG	0,67	0,260
	KF 2: Indikation zum passageren Pacing bei Bradykardie während der Koronarangiographie	0,13	0,296
	KF 3: Auskultation bei V. a. pulmonale Stauung	0,04	0,021
	KF 4: Indikation zur nicht-invasiven Beatmung bei Lungenödem	0,19	-0,220
	KF 5: Verdachtsdiagnose „Papillarmuskelabriss" bei neuem Systolikum nach Infarkt	0,23	0,082

E-Fallseminar 4	Item	Itemschwierigkeit	Trennschärfe
Fall 24	KF 1: Erkennen eines STEMI der Inferolateralwand im EKG	0,31	0,372
	KF 2: Verdachtsdiagnose „Mitralklappeninsuffizienz" bei passendem Auskultationsbefund	0,83	0,474
	KF 3: Echokardiographie zur Quantifizierung der Mitralklappeninsuffizienz	0,95	0,481
	KF 4: Erkennen einer pulmonalen Hypertonie bei PAP syst. = 80 mmHg im Echo	0,31	0,433
	KF 5: Indikation zur Klappen-Operation bei symptomatischer Mitralklappeninsuffizienz (NYHA III)	0,31	0,396
Fall 20	KF 1: Verdachtsdiagnose „Aortenklappenstenose" bei typischem Untersuchungsbefund und Synkope	0,90	0,260
	KF 2: Echokardiographie zur Sicherung der Verdachtsdiagnose „Aortenklappenstenose"	0,95	0,602
	KF 3: Indikation zum Klappenersatz bei symptomatischer hochgradiger Aortenklappenstenose	0,93	0,314
	KF 4: laborchemische Diagnose einer manifesten Hyperthyreose	0,78	0,292
	KF 5: Absetzen von Amiodaron bei manifester Hyperthyreose	0,12	0,178
Fall 2a	KF 1: korrekte Benennung der stabilen Angina pectoris	0,66	0,344
	KF 2: Indikation zur Ergometrie bei V. a. Koronarinsuffizienz	0,55	0,192
	KF 3: Indikation zur Koronarangiographie bei positivem Ischämienachweis	0,81	0,560
	KF 4: Indikation zur operativen Myokardrevaskularisation bei 3-Gefäß-KHK mit Hauptstammstenose	0,62	0,454
	KF 5: Rehabilitation nach operativer Myokardrevaskularisation	0,31	0,092
Fall 3	KF 1: sofortiges EKG bei instabiler Angina pectoris	0,88	0,459
	KF 2: sofortige Koronarangiographie bei STEMI	0,88	0,293
	KF 3: Indikation zur PTCA + Stentimplantation bei RIVA-Verschuß	0,91	0,493
	KF 4: Defibrillation bei Kammerflimmern	0,64	0,393
	KF 5: Verdachtsdiagnose „Pericarditis epistenocardica" bei neuer ST-Hebung nach stattgehabtem Infarkt	0,16	0,174

E-Fallseminar 5	Item	Itemschwierigkeit	Trennschärfe
Fall 8	KF 1: Erkennen einer Tachyarrhythmia absoluta im EKG	0,43	0,177
	KF 2: Verdachtsdiagnose „Myokarditis“ bei passender Anamnese und eingeschränkter EF	0,14	0,310
	KF 3: Frequenzsenkung als Therapieziel bei Tachyarrhythmia absoluta	0,23	0,141
	KF 4: Betablocker zur Frequenzsenkung	0,64	0,067
	KF 5: Indikation zur primärprophylaktischen ICD-Implantation bei persistierend niedriger EF	0,46	-0,037
Fall 14	KF 1: Verdachtsdiagnose „supraventrikuläre Tachykardie“ (SVT) anhand der EKG-Beschreibung	0,55	0,301
	KF 2: Valsalva-Maneuver zur Terminierung eines SVT-Anfalls	0,16	-0,141
	KF 3: Adenosin zur Terminierung eines SVT-Anfalls	0,55	0,381
	KF 4: Elektrophysiologische Untersuchung zur Sicherung der Diagnose und ggf. Therapie	0,39	0,292
	KF 5: Indikation zur Schrittmacher-Implantation bei kompletter AV-Blockierung	0,54	0,462
Fall 9	KF 1: Verdachtsdiagnose „sekundäre arterielle Hypertonie“ bei typischer Anamnese	0,11	0,226
	KF 2: Verdachtsdiagnose „Obstruktives Schlafapnoesyndrom“ bei typischer Anamnese	0,64	0,342
	KF 3: CPAP zur Obstruktives Schlafapnoesyndrom-Therapie	0,23	0,083
	KF 4: Verdachtsdiagnose „Diastolische Dysfunktion“ bei normaler LVEF, NYHA II und erhöhtem BNP	0,23	0,443
	KF 5: Indikation zur Koronarangiographie bei typischer Angina pectoris und neu aufgetretenem Linksschenkelblock	0,79	0,400
Fall 18	KF 1: Diagnose „Synkope“ bei typischer Anamnese	0,86	0,374
	KF 2: Frage nach Krampfäquivalenten zur Abgrenzung von der Epilepsie	0,38	0,368
	KF 3: Transthorakale Echokardiographie zum Ausschluß einer strukturellen Herzerkrankung	0,86	0,041
	KF 4: Schellong-Test bei V. a. orthostatisch bedingte Synkope	0,63	0,065
	KF 5: Kipptisch-Untersuchung bei V. a. Reflex-Synkope	0,34	0,005

E-Fallseminar 6	Item	Itemschwierigkeit	Trennschärfe
Fall 11	KF 1: Verdachtsdiagnose „Pneumonie“ bei typischen Symptomen	0,75	0,555
	KF 2: Röntgendiagnostik zur Sicherung der Verdachtsdiagnose	0,78	0,341
	KF 3: Erkennen einer Lobärpneumonie des linken Oberlappens im Röntgenbild	0,69	0,335
	KF 4: CRB-65-Index zur Entscheidung über die Empfehlung zur stationären Behandlung	0,58	0,567
	KF 5: Abnahme von Blutkulturen bei erneutem Fieber > 38,5°C	0,65	0,290
Fall 4a	KF 1: Interpretation eines leicht erhöhten Troponins im Sinne einer hypertensiven Entgleisung	0,42	0,374
	KF 2: Indikation zur Koronarangiographie bei steigendem Troponin und Beschwerdepersistenz	0,87	0,592
	KF 3: Verdacht auf Vorderwandaneurysma bei persistierenden ST-Hebungen	0,11	0,407
	KF 4: Transthorakale Echokardiographie zur Bestätigung der Verdachtsdiagnose	0,80	0,382
	KF 5: V. a. Stentverschluß bei erneuten Beschwerden nach verfrühtem Absetzen der dualen Thrombozytenaggregationshemmung	0,67	0,564
Fall 12	KF 1: Verdachtsdiagnose „COPD“ bei chronischem Husten und Auswurf und typischem Untersuchungsbefund	0,84	0,415
	KF 2: Erkennen einer Obstruktion in der Lungenfunktion	0,64	0,443
	KF 3: BGA bei V. a. beginnende CO ₂ -Narkose	0,73	0,430
	KF 4: Non-invasive Ventilation bei nachgewiesener Hyperkapnie	0,45	0,376
	KF 5: Indikation zur Bronchoskopie bei Hämoptysen	0,42	0,157
Fall 22	KF 1: Stellen der Diagnose „stabile Angina pectoris“	0,73	0,473
	KF 2: Ergometrie zum Ischämienachweis	0,69	0,335
	KF 3: Verdachtsdiagnose „LV-Hypertrophie“ aufgrund des EKG's	0,29	0,316
	KF 4: Auskultation vor Ergometrie	0,07	0,239
	KF 5: Empfehlung zum Klappenersatz bei symptomatischer hochgradiger Aortenklappenstenose	0,87	0,607

E-Fallseminar 7	Item	Itemschwierigkeit	Trennschärfe
Fall 23	KF 1: Verdachtsdiagnose „Lungenembolie“ bei typischer Klinik und Risikoprofil	0,58	0,311
	KF 2: Wells-Score zur Entscheidung für oder gegen eine D-Dimer-Bestimmung	0,73	0,442
	KF 3: sofortiges Thorax-CT bei hoher klinischer Wahrscheinlichkeit	0,42	0,156
	KF 4: Echokardiographie oder Troponinbestimmung zur Risikostratifizierung	0,38	0,376
	KF 5: Indikation zur Fibrinolyse bei hämodynamischer Instabilität	0,71	0,353
Fall 21	KF 1: Verdachtsdiagnose „Endokarditis“ bei Fieber und neuem Herzgeräusch	0,46	0,271
	KF 2: Abnahme von Blutkulturen zur Sicherung der Diagnose	0,44	0,240
	KF 3: Kreatinin-Kontrolle unter Aminoglykosid-Gabe	0,54	0,174
	KF 4: Indikation zum operativen Klappenersatz bei Therapie-Refraktärität	0,65	0,288
	KF 5: Schädel-CT bei V. a. perioperativen Hirninfarkt	0,58	0,425
Fall 18a	KF 1: Diagnose „Synkope“ bei typischer Anamnese	0,94	0,450
	KF 2: Frage nach Krampfäquivalenten zur Abgrenzung von der Epilepsie	0,81	0,486
	KF 3: Transthorakale Echokardiographie zum Ausschluß einer strukturellen Herzerkrankung	0,77	0,205
	KF 4: Schellong-Test bei V. a. orthostatisch bedingte Synkope	0,81	0,208
	KF 5: Kipptisch-Untersuchung bei V. a. Reflex-Synkope	0,81	0,194
Fall 19	KF 1: EKG zur Diagnose von Vorhofflimmern bei passendem klinischen Untersuchungsbefund	0,48	0,533
	KF 2: Antikoagulation als erste Maßnahme bei normofrequenter Arrhythmia absoluta	0,19	0,053
	KF 3: Transösophageale Echokardiographie zum Thrombenausschluß vor Kardioversion	0,48	0,382
	KF 4: CHA2DS2-VASc-Score zur Indikationsstellung zur dauerhaften Antikoagulation	0,63	0,259
	KF 5: Schrittmacher-Implantation bei symptomatischem AV-Block III°	0,56	0,479

E-Fallseminar 8	Item	Itemschwierigkeit	Trennschärfe
Fall 15	KF 1: Verdachtsdiagnose „Lungenfibrose“ bei Dyspnoe und typischem Auskultationsbefund	0,18	0,122
	KF 2: Amiodaron als Auslöser einer Lungenfibrose	0,37	0,397
	KF 3: Erkennen einer Restriktion in der Lungenfunktionsdiagnostik	0,11	0,193
	KF 4: Bronchoalveoläre Lavage zur Differenzierung zwischen medikamenten-induzierter Fibrose und Sarkoidose	0,29	0,320
	KF 5: Indikation zur Langzeit-Sauerstofftherapie aufgrund der BGA	0,29	0,397
Fall 12a	KF 1: Verdachtsdiagnose „COPD“ bei chronischem Husten und Auswurf und typischem Untersuchungsbefund	0,81	0,458
	KF 2: Erkennen einer Obstruktion in der Lungenfunktion	0,66	0,272
	KF 3: BGA bei V. a. beginnende CO ₂ -Narkose	0,65	0,580
	KF 4: Non-invasive Ventilation bei nachgewiesener Hyperkapnie	0,53	0,386
	KF 5: Indikation zur Bronchoskopie bei Hämoptysen	0,45	0,535
Fall 20a	KF 1: Verdachtsdiagnose „Aortenklappenstenose“ bei typischem Untersuchungsbefund und Synkope	0,73	0,590
	KF 2: Echokardiographie zur Sicherung der Verdachtsdiagnose „Aortenklappenstenose“	0,73	0,572
	KF 3: Indikation zum Klappenersatz bei symptomatischer hochgradiger Aortenklappenstenose	0,85	0,515
	KF 4: laborchemische Diagnose einer manifesten Hyperthyreose	0,68	0,647
	KF 5: Absetzen von Amiodaron bei manifester Hyperthyreose	0,24	0,279
Fall N04	KF 1: Diagnose „Nephrotisches Syndrom“ bei typischer Klinik und Laborkonstellation	0,71	0,447
	KF 2: Diagnose „Systemischer Lupus erythematodes“ bei typischer Klinik und Laborkonstellation (Antikörper)	0,81	0,488
	KF 3: Indikation zur Nierenbiopsie zur Diagnose und Therapieplanung bei nephrotischem Syndrom	0,69	0,542
	KF 4: Steroidstoß als initiale Therapie des nephrotischen Syndroms bei Systemischem Lupus erythematodes	0,82	0,549
	KF 5: Phlebothrombose als mögliche Komplikation des nephrotischen Syndroms	0,68	0,483

E-Fallseminar 9	Item	Itemschwierigkeit	Trennschärfe
Fall N02	KF 1: Krankenhauseinweisung bei Hyponatriämie < 120 mmol/l	0,69	0,493
	KF 2: Thiazid-Diuretika als mögliche Auslöser einer Hyponatriämie	0,71	0,258
	KF 3: Wirkort von Thiaziden: distaler Tubulus	0,57	0,445
	KF 4: Absetzen von Thiaziden zur Therapie der Hyponatriämie	0,37	0,168
	KF 5: Zentrale pontine Myelinolyse als Komplikation einer zu raschen Natrium-Substitution	0,08	0,185
Fall N01a	KF 1: Diagnose „Diabetische Ketoazidose“ anhand einer BGA	0,82	0,492
	KF 2: Therapie der Hyperkaliämie durch Korrektur der Azidose	0,08	0,204
	KF 3: Flüssigkeitsgabe zur Korrektur der Azidose	0,02	-0,135
	KF 4: Insulingabe zur Korrektur der Azidose	0,06	0,035
	KF 5: Urinstatus bei V. a. Harnwegsinfekt als Ursache der Entgleisung	0,31	0,282
Fall 23a	KF 1: Verdachtsdiagnose „Lungenembolie“ bei typischer Klinik und Risikoprofil	0,45	0,333
	KF 2: Wells-Score zur Entscheidung für oder gegen eine D-Dimer-Bestimmung	0,61	0,552
	KF 3: sofortiges Thorax-CT bei hoher klinischer Wahrscheinlichkeit	0,35	0,248
	KF 4: Echokardiographie oder Troponinbestimmung zur Risikostratifizierung	0,31	0,565
	KF 5: Indikation zur Fibrinolyse bei hämodynamischer Instabilität	0,41	0,608
Fall 5a	KF 1: Krankenhauseinweisung bei vernichtendem Thoraxschmerz	0,25	0,446
	KF 2: Blutdrucksenkung bei hypertensiver Entgleisung	0,31	0,445
	KF 3: Verdachtsdiagnose „Aortendissektion“ bei wanderndem Schmerz und D-Dimer-Erhöhung	0,31	0,409
	KF 4: Thorax-CT zur Sicherung der Diagnose „Aortendissektion“	0,41	0,527
	KF 5: Indikation zur sofortigen Operation bei Typ A-Dissektion	0,71	0,459

E-Fallseminar 10	Item	Itemschwierigkeit	Trennschärfe
Fall N06	KF 1: Diagnose Lungenödem bei Dyspnoe und feuchten Rasselgeräuschen	0,43	0,483
	KF 2: Transthorakale Echokardiographie zur Beurteilung der systolischen LV-Funktion	0,63	0,577
	KF 3: Trinkmengenbeschränkung zur Verhinderung einer Dekompensation	0,57	0,625
	KF 4: Laborparameter, die durch eine Dialyse reguliert werden	0,80	0,743
	KF 5: Substitutionspflichtige Substanzen unter Dialyse	0,57	0,451
Fall N04a	KF 1: Diagnose „Nephrotisches Syndrom“ bei typischer Klinik und Laborkonstellation	0,63	0,674
	KF 2: Diagnose „Systemischer Lupus erythematodes“ bei typischer Klinik und Laborkonstellation (Antikörper)	0,61	0,755
	KF 3: Indikation zur Nierenbiopsie zur Diagnose und Therapieplanung bei nephrotischem Syndrom	0,61	0,812
	KF 4: Steroidstoß als initiale Therapie dese nephrotischen Syndroms bei Systemischem Lupus erythematodes	0,64	0,783
	KF 5: Phlebothrombose als mögliche Komplikation des nephrotischen Syndroms	0,39	0,618
Fall 15a	KF 1: Verdachtsdiagnose „Lungenfibrose“ bei Dyspnoe und typischem Auskultationsbefund	0,34	0,486
	KF 2: Amiodaron als Auslöser einer Lungenfibrose	0,59	0,672
	KF 3: Erkennen einer Restriktion in der Lungenfunktionsdiagnostik	0,14	0,310
	KF 4: Bronchoalveoläre Lavage zur Differenzierung zwischen medikamenten-induzierter Fibrose und Sarkoidose	0,30	0,342
	KF 5: Indikation zur Langzeit-Sauerstofftherapie aufgrund der BGA	0,27	0,410
Fall 11a	KF 1: Verdachtsdiagnose „Pneumonie“ bei typischen Symptomen	0,55	0,654
	KF 2: Röntgendiagnostik zur Sicherung der Verdachtsdiagnose	0,70	0,672
	KF 3: Erkennen einer Mittellappenpneumonie im Röntgenbild	0,27	0,438
	KF 4: CRB-65-Index zur Entscheidung über die Empfehlung zur stationären Behandlung	0,43	0,609
	KF 5: V. a. parapneumonischen Erguss bei typischer Klinik	0,38	0,480

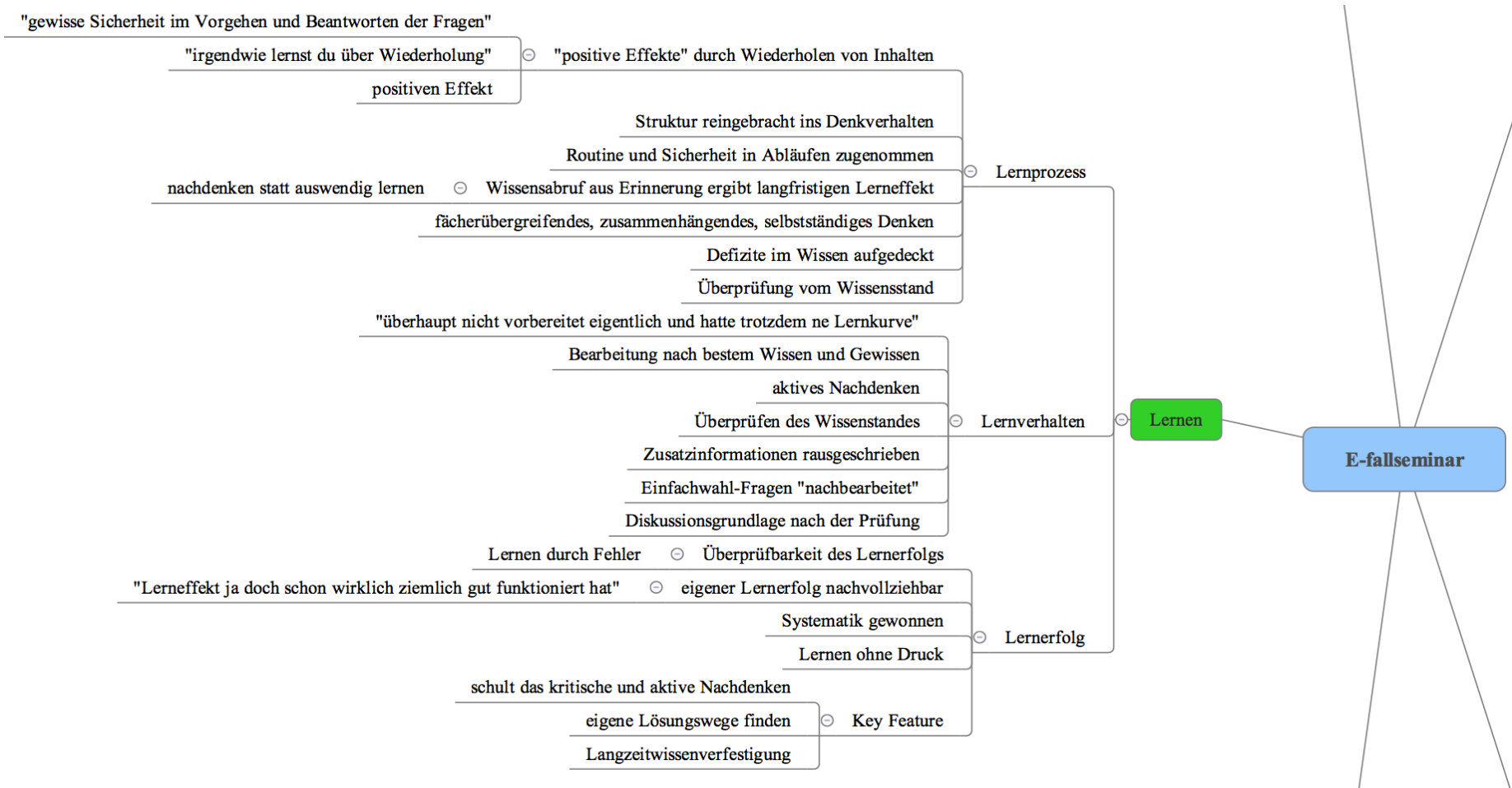
E-Fallseminar 11	Item	Itemschwierigkeit	Trennschärfe
Fall H01	KF 1: Bestimmung der Erythrozytenindizes zur Einordnung einer Anämie	0,52	0,29
	KF 2: Benennen einer Anämie als hypochrom und mikrozytär	0,81	0,63
	KF 3: V. a. Eisenmangelanämie bei passender Anamnese und Laborkonstellation	0,87	0,57
	KF 4: Ferritinbestimmung zur Bestätigung der Verdachtsdiagnose "Eisenmangel"	0,81	0,63
	KF 5: Substitution von zweiwertigem Eisen	0,81	0,66
Fall 19a	KF 1: EKG zur Diagnose von Vorhofflimmern bei passendem klinischen Untersuchungsbefund	0,71	0,53
	KF 2: Antikoagulation als erste Maßnahme bei normofrequenter Arrhythmia absoluta	0,12	0,02
	KF 3: direkte Kardioversion bei Vorhofflimmer-Dauer < 48 Stunden	0,48	0,28
	KF 4: CHA2DS2-VASc-Score zur Indikationsstellung zur dauerhaften Antikoagulation	0,50	0,51
	KF 5: Indikation zur Phenprocoumon-Gabe bei Niereninsuffizienz	0,60	0,23
Fall 21a	KF 1: Verdachtsdiagnose „Aortenklappenendokarditis“ bei Fieber und neuem Herzgeräusch	0,46	0,24
	KF 2: Abnahme von Blutkulturen zur Sicherung der Diagnose	0,58	0,53
	KF 3: Kreatinin-Kontrolle unter Aminoglykosid-Gabe	0,42	0,28
	KF 4: Indikation zum operativen Klappenersatz bei Therapie-Refraktärität	0,37	0,44
	KF 5: Schädel-CT bei V. a. perioperativen Hirninfarkt	0,56	0,52
Fall N02a	KF 1: Krankenhauseinweisung bei Hyponatriämie < 120 mmol/l	0,44	0,47
	KF 2: Thiazid-Diuretika als mögliche Auslöser einer Hyponatriämie	0,62	0,56
	KF 3: Wirkort von Thiaziden: distaler Tubulus	0,60	0,42
	KF 4: Absetzen von Thiaziden zur Therapie der Hyponatriämie	0,54	0,36
	KF 5: Zentrale pontine Myelinolyse als Komplikation einer zu raschen Natrium-Substitution	0,13	0,34

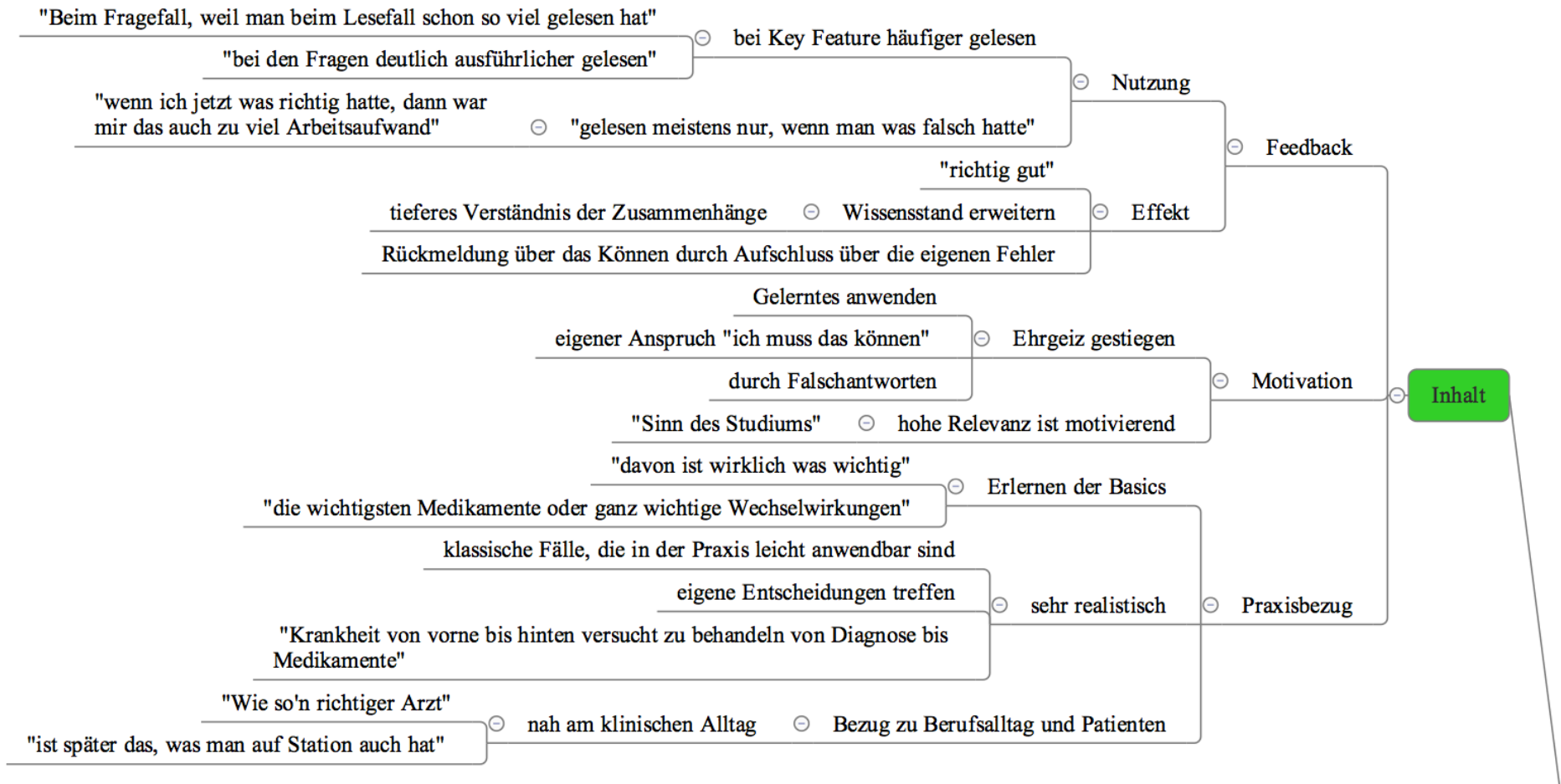
Ausgangstest	Item	Itemschwierigkeit	Trennschärfe
Fall 1	KF 1: Ableiten der Verdachtsdiagnose „Akute Lungenembolie“ bei typischer Klinik und Risikoprofil	0,54	0,414
	KF 2: Wells-Score zur Entscheidung für oder gegen eine D-Dimer-Bestimmung	0,68	0,496
	KF 3: sofortiges Thorax-CT bei hoher Wahrscheinlichkeit und klinischer Stabilität	0,47	0,365
	KF 4: Echokardiographie oder Troponinbestimmung zur Risikostratifizierung	0,34	0,466
	KF 5: Indikation zur Fibrinolyse bei hämodynamischer Instabilität	0,53	0,549
	KF 6: Erkennen eines nephrotischen Syndroms anhand der Labor- und Urinbefunde	0,66	0,622
	KF 7: V. a. Systemischer Lupus erythematodes bei typischer Klinik und wegweisendem AK-Befund	0,66	0,554
	KF 8: Nierenbiopsie zur Diagnosesicherung	0,65	0,564
	KF 9: Behandlung der Lupus-Nephritis durch einen Steroidstoß	0,69	0,591
Fall 2	KF 1: V. a. COPD bei chronischem Husten und Auswurf bei einem starken Raucher mit expiratorischem Giemen	0,75	0,469
	KF 2: Erkennen einer Obstruktion in der Lungenfunktion	0,53	0,239
	KF 3: Erkennen einer links-apikalen Lobärpneumonie im Röntgenbild	0,32	0,458
	KF 4: CRB-65-Index zur Entscheidung über die Empfehlung zur stationären Aufnahme	0,51	0,545
	KF 5: BGA bei V. a. beginnende CO ₂ -Narkose	0,67	0,529
	KF 6: Non-invasive Ventilation bei nachgewiesener Hyperkapnie	0,41	0,403
Fall 3	KF 1: V. a. sekundäre Hypertonie bei Manifestation > 60 J., diskrepanten Organschäden und Therapierefraktärität	0,32	0,325
	KF 2: V. a. auf diastolische Dysfunktion bei guter EF, NYHA II und erhöhtem BNP	0,20	0,496
	KF 3: ACE-Hemmer bei V. a. ACE-Hemmer-Husten absetzen	0,84	0,583
	KF 4: Krankenhaus-Einweisung bei Hyponatriämie < 120 mmol/l	0,21	0,271
	KF 5: Thiazid-Diuretika als mögliche Auslöser einer Hyponatriämie	0,69	0,512
	KF 6: zentrale pontine Myelinolyse als Komplikation einer zu raschen Natrium-Substitution	0,28	0,469
Fall 4	KF 1: Schellong-Test bei V. a. orthostatisch bedingte Synkope	0,52	0,472
	KF 2: Schädel-CT zur Diagnose eines Hirninfarktes	0,66	0,453
	KF 3: Erkennen einer Tachyarrhythmia absoluta im EKG	0,34	0,360
	KF 4: CHA ₂ DS ₂ -VASc-Score zur Indikationsstellung zur dauerhaften Antikoagulation	0,67	0,534
	KF 5: laborchemische Diagnose einer manifesten Hyperthyreose nach Kontrastmittelgabe	0,77	0,516
	KF 6: Absetzen von Amiodaron bei manifester Hyperthyreose	0,48	0,497
	KF 7: klinischer Verdacht auf Lungenfibrose bei Dyspnoe und inspiratorischem Velcro-Knistern	0,46	0,545
	KF 8: Amiodaron als Verursacher einer Lungenfibrose	0,58	0,510
	KF 9: Indikationsstellung zur Langzeit-Sauerstofftherapie aufgrund eines pO ₂ < 55 mmHg in der arteriellen BGA	0,54	0,374

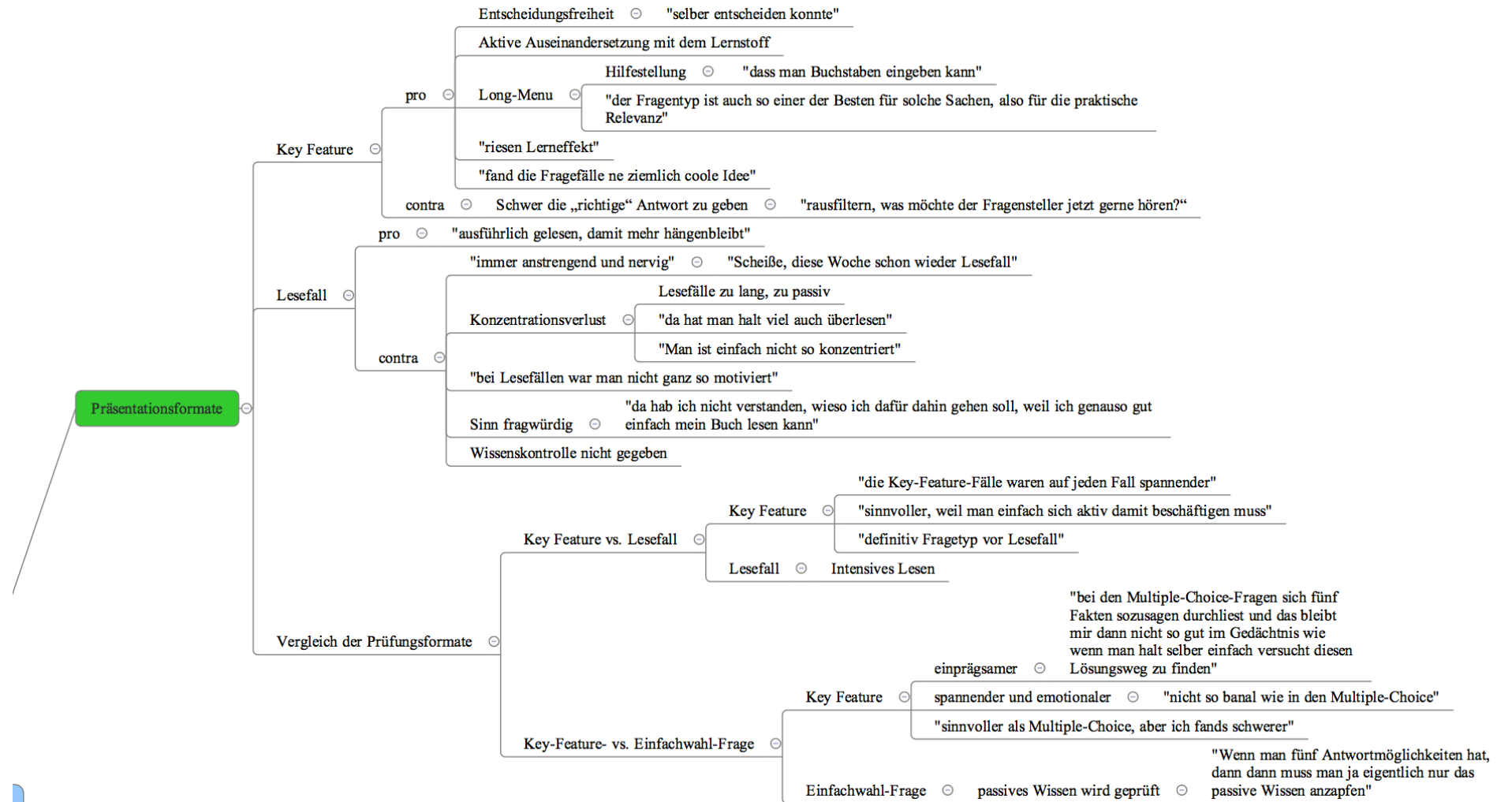
Retentionstest	Item	Itemschwierigkeit	Trennschärfe
Fall 1	KF 1: Ableiten der Verdachtsdiagnose „Akute Lungenembolie“ bei typischer Klinik und Risikoprofil	0,58	0,531
	KF 2: Wells-Score zur Entscheidung für oder gegen eine D-Dimer-Bestimmung	0,60	0,476
	KF 3: sofortiges Thorax-CT bei hoher Wahrscheinlichkeit und klinischer Stabilität	0,36	0,357
	KF 4: Echokardiographie oder Troponinbestimmung zur Risikostratifizierung	0,33	0,325
	KF 5: Indikation zur Fibrinolyse bei hämodynamischer Instabilität	0,44	0,411
	KF 6: Erkennen eines nephrotischen Syndroms anhand der Labor- und Urinbefunde	0,55	0,657
	KF 7: V. a. Systemischer Lupus erythematodes bei typischer Klinik und wegweisendem Antikörper-Befund	0,66	0,599
	KF 8: Nierenbiopsie zur Diagnosesicherung	0,66	0,543
	KF 9: Behandlung der Lupus-Nephritis durch einen Steroidstoß	0,70	0,601
Fall 2	KF 1: Verdachtsdiagnose COPD bei chronischem Husten und Auswurf bei einem Raucher mit expiratorischem Giemen	0,71	0,546
	KF 2: Erkennen einer Obstruktion in der Lungenfunktion	0,43	0,271
	KF 3: Erkennen einer links-apikalen Lobärpneumonie im Röntgenbild	0,34	0,400
	KF 4: CRB-65-Index zur Entscheidung über die Empfehlung zur stationären Aufnahme	0,46	0,552
	KF 5: BGA bei V. a. beginnende CO ₂ -Narkose	0,55	0,546
	KF 6: Non-invasive Ventilation bei nachgewiesener Hyperkapnie	0,31	0,186
Fall 3	KF 1: Verdacht auf sekundäre Hypertonie bei Manifestation >60 J., diskrepanten Organschäden und Therapierefraktärität	0,28	0,411
	KF 2: Verdacht auf diastolische Dysfunktion bei guter EF, NYHA II und erhöhtem BNP	0,19	0,453
	KF 3: ACE-Hemmer bei V. a. ACE-Hemmer-Husten absetzen	0,78	0,437
	KF 4: Krankenhaus-Einweisung bei Hyponatriämie < 120 mmol/l	0,32	0,455
	KF 5: Thiazid-Diuretika als mögliche Auslöser einer Hyponatriämie	0,55	0,475
	KF 6: zentrale pontine Myelinolyse als Komplikation einer zu raschen Natrium-Substitution	0,22	0,447
Fall 4	KF 1: Schellong-Test bei V. a. orthostatisch bedingte Synkope	0,53	0,412
	KF 2: Schädel-CT zur Diagnose eines Hirninfarktes	0,61	0,532
	KF 3: Erkennen einer Tachyarrhythmia absoluta im EKG	0,51	0,496
	KF 4: CHA ₂ DS ₂ -VASc-Score zur Indikationsstellung zur dauerhaften Antikoagulation	0,59	0,641
	KF 5: laborchemische Diagnose einer manifesten Hyperthyreose nach Kontrastmittelgabe	0,80	0,558
	KF 6: Absetzen von Amiodaron bei manifester Hyperthyreose	0,43	0,567
	KF 7: klinischer Verdacht auf Lungenfibrose bei Dyspnoe und inspiratorischem Velcro-Knistern	0,43	0,540
	KF 8: Amiodaron als Verursacher einer Lungenfibrose	0,70	0,384
	KF 9: Indikationsstellung zur Langzeit-Sauerstofftherapie aufgrund eines pO ₂ < 55 mmHg in der arteriellen BGA	0,42	0,107

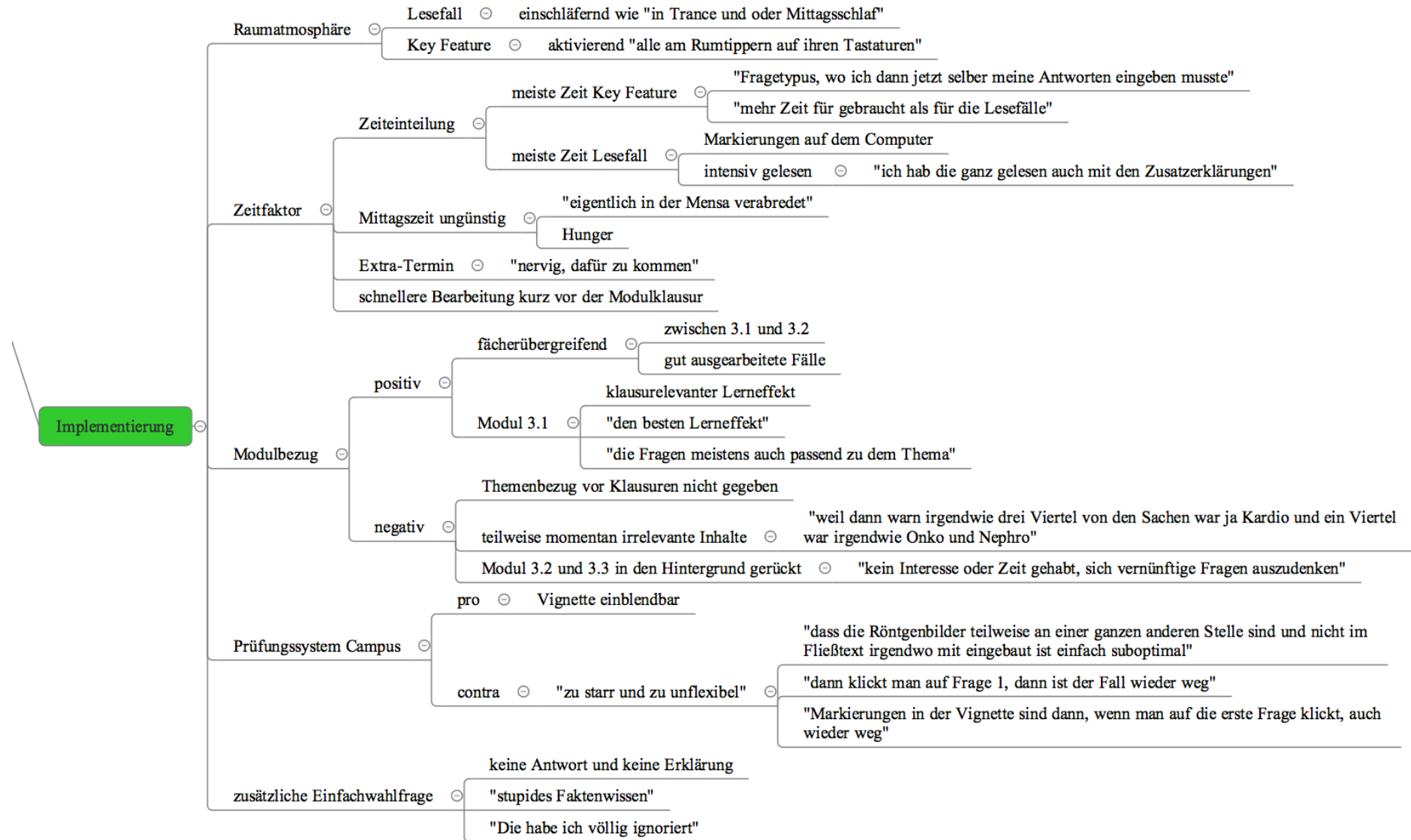
Anhang 7: Mind Map zur Strukturierung und Visualisierung der Fokusgruppengespräche

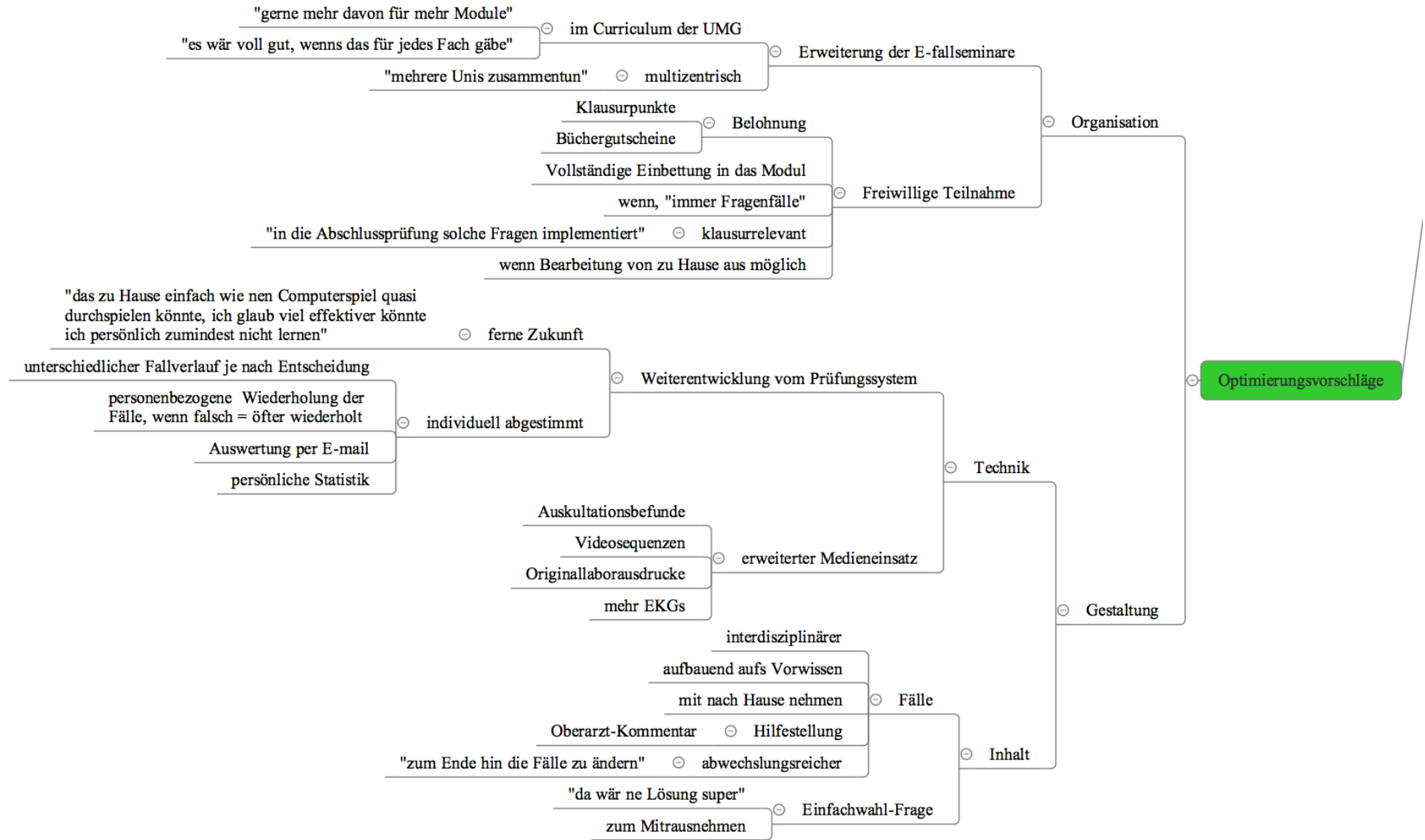
Bedauerlicherweise ist die erstellte Mind Map zu groß, um sie im Gesamten abzubilden. Deshalb ist sie hier auf fünf verschiedenen Fotos zu sehen. Es ergaben sich fünf Hauptkategorien: Lernen, Inhalt, Präsentationsformate, Implementierung und Optimierungsvorschläge. Die Hauptkategorien (grün) sind auf den Fotos im Uhrzeigersinn dargestellt.











Danksagung

Mein herzlichster und größter Dank gilt Herrn Prof. Dr. Raupach für das Ermöglichen dieser Arbeit im Bereich der Medizindidaktik und Ausbildungsforschung der Georg-August-Universität Göttingen. Ich möchte mich insbesondere für die außerordentlich gute Betreuung, das große Engagement und die inhaltlich konstruktive Kritik bedanken. Die wissenschaftliche Zusammenarbeit mit Herrn Prof. Dr. Raupach war äußerst lehrreich und hat große Freude bereitet. Zusätzlich war er bei der Vorbereitung der Vorträge didaktisch, methodisch und inhaltlich eine große Hilfe und sorgte für ein sehr angenehmes Arbeitsklima in seiner Arbeitsgruppe.

Ausdrücklich bedanken möchte ich mich des Weiteren bei Herrn Prof. Dr. Koziolok, Herrn Dr. Jung und Dr. Lisa Strobel für die fachlichen Beiträge und für die Erstellung der Fragen- und Lesefälle. Außerdem danke ich Herrn Dr. Jamie Brown für die Hilfe bei der statistischen Analyse und Prof. Dr. Anders für die Anregungen zum Studiendesign.

Ebenso möchte ich mich beim Geschäftsbereich G3-71, insbesondere bei Herrn Münscher und Herrn Hermann, für die technische Umsetzung und bei Katharina Meyer für die Mithilfe bei der Digitalisierung und Beaufsichtigung der E-Fallseminare ganz herzlich bedanken.

Ein besonderer Dank gilt der Arbeitsgruppe des Bereichs Medizindidaktik und Ausbildungsforschung unter Leitung von Prof. Dr. Raupach für die produktiven Teamtreffen.

Abschließend bedanke ich mich bei den Studierenden für die Teilnahme an den E-Fallseminaren und den Fokusgruppengesprächen.

Curriculum Vitae

Am 15.12.1990 wurde ich in Hannover geboren. Hier wuchs ich zusammen mit meinen Eltern Ursula Ingrid Andresen und Kay Andresen und meiner Schwester Leonie Josephine Andresen auf. Von 1996 bis 2000 besuchte ich die Grundschule Kestnerstraße und von 2000 bis 2002 die Orientierungsstufe Luerstraße. Im Jahre 2002 wurde ich im Kaiser-Wilhelm-Ratsgymnasium unterrichtet, bis ich 2006 für ein Jahr im englischen Internat Malvern College zur Schule ging. Danach besuchte ich für zwei Jahre das Wirtschaftsgymnasium in Hannover. Währenddessen wurde ich als Schulsprecherin gewählt und arbeitete zusätzlich ehrenamtlich in der Schülerversammlung und im Schulvorstand. An dieser Schule erreichte ich 2009 meine allgemeine Hochschulreife. Zum Wintersemester 2009/2010 begann ich an der Georg-August-Universität Humanmedizin zu studieren. Meine Pflegepraktika absolvierte ich zwei Monate an der Medizinischen Hochschule Hannover und einen Monat an der Charité Universitätsmedizin Campus Mitte. 2011 wurde ich zur Semestersprecherin gewählt und arbeitete in diesem Amt für ein Jahr. Im Wintersemester 2012/2013 studierte ich für ein Semester an der Université Paris 11. Famulaturen absolvierte ich unter anderem an der Harvard Medical School im Brigham and Women's Hospital, im Universitätsklinikum Antoine Bécélère in Paris, am Charité Universitätsklinikum Campus Mitte und Campus Virchow, in der Asklepiosklinik St. Georg in Hamburg und in Galle in Sri Lanka. Im Oktober 2013 begann ich meine Promotion unter der Leitung von Prof. Dr. Tobias Raupach. In diesem Rahmen hielt ich 2014 zwei Vorträge auf der Jahrestagung der Gesellschaft für Medizinische Ausbildung. Ein Paper zu der vorliegenden Studie wurde im Jahr 2016 veröffentlicht. Im April 2015 schrieb ich das zweite Staatsexamen. Das erste Tertial des Praktischen Jahres verbrachte ich zur einen Hälfte in der Inneren Medizin an der Universitätsmedizin Göttingen und zur anderen Hälfte am Kantonsspital Luzern in der Schweiz. Das zweite Tertial absolvierte ich in der Augenklinik der Ludwig-Maximilians-Universität München und das dritte Tertial an der Universitätsmedizin Göttingen in der Unfallchirurgie und in der Thorax-, Herz- und Gefäßchirurgie. Im Mai 2016 folgte das dritte Staatsexamen. Zum heutigen Zeitpunkt arbeite ich als Assistenzärztin in der Universitätsaugenklinik Bonn.